Characterization of Pseudo CT From MRI With Deep Learning : Application to Brain Tumors Treated With Radiotherapy

Emilie Alvarez Andres

▶ To cite this version:

Emilie Alvarez Andres. Characterization of Pseudo CT From MRI With Deep Learning : Application to Brain Tumors Treated With Radiotherapy. Cancer. Université Paris-Saclay, 2021. English. NNT : 2021UPASL070 . tel-03816279

HAL Id: tel-03816279
https://theses.hal.science/tel-03816279

Submitted on 16 Oct 2022

université
PARIS-SACLAY

Characterization of pseudo CT from MRI with deep learning – Application to brain tumors treated with radiotherapy

Caractérisation de pseudo CT générés à partir d'images IRM à l'aide de méthodes deep learning - Application aux tumeurs cérébrales traitées par radiothérapie

Thèse de doctorat de l'université Paris-Saclay

Ecole doctorale n°582 Cancérologie : biologie – médecine – santé (CBMS)
Spécialité de doctorat : Recherche clinique, innovation technologique, santé publique
Unité de recherche : Université Paris-Saclay, Institut Gustave Roussy, Inserm, Radiothérapie Moléculaire et Innovation Thérapeutique, 94800, Villejuif, France
Référent : Faculté de Médecine

Thèse présentée et soutenue à Villejuif, le 15/10/2021, par

Emilie ALVAREZ ANDRES

Composition du Jury

Pr. Elizabeth MOYAL PU-PH, Université Paul Sabatier, Toulouse, France	Présidente
Pr. Nick REYNAERT Professeur / Physicien médical, Institut Jules Bordet, Bruxelles, Belgique	Rapporteur & Examinateur
David SARRUT DR, CNRS, Lyon, France	Rapporteur & Examinateur
Vincent LEPETIT DR, ENPC ParisTech, Paris, France	Examinateur
Philippe MEYER Docteur / Physicien médical, ICANS, Strasbourg, France	Examinateur
Pr. Hervé SAINT-JALMES PU, Université de Rennes 1, Rennes, France	Examinateur

Direction de la thèse

Charlotte ROBERT MCU, Université Paris-Saclay, Institut Gustave Roussy, Inserm U1030, Villejuif, France	Directrice de thèse
Pr. Nikos PARAGIOS PU, TheraPanacea, Paris, France	Co-Directeur de thèse

Thèse de doctorat

NNT : 2021UPASL070

Remerciements

Tout au long de cette thèse, j'ai eu l'occasion de rencontrer de nombreuses personnes, qui m'ont toutes apporté un « petit quelque chose », et qui ont fait que j'ai pu en arriver là aujourd'hui.

Je tiens en premier lieu à remercier ma maître de thèse, Charlotte Robert. Merci de m'avoir fait confiance il y a presque 4 ans… Et merci du fond du cœur pour tout ce que tu m'as appris durant cette thèse. Tu as toujours su me donner les bons conseils et me tirer vers le haut pour que je me surpasse. Je me souviendrai pendant pas mal de temps de cette nuit de soumission à l'ESTRO 2021, ou 6h avant la deadline, on s'est dit qu'on n'allait pas soumettre un, mais deux abstracts ! Tu as pris le premier, moi le deuxième, et on s'est laissé 1h pour avancer. Une soirée assez sportive donc… Encore un grand merci à toi !

Je tiens en deuxième lieu à remercier Nikos Paragios pour ses connaissances techniques ainsi que son expertise en intelligence artificielle dont j'ai pu bénéficier tout au long de cette thèse.

Merci à Eric Deutsch pour son accueil au sein de l'U1030 et tous les précieux conseils qu'il m'a donnés concernant la thèse, mais aussi plus largement sur mes projets futurs et les directions que je devais suivre pour me lancer dans une carrière académique.

Ensuite, je tiens à remercier tous mes collègues de l'U1030 avec qui j'ai pu échanger et apprendre au quotidien. Merci en particulier à Angela, Nathan et Stéphane qui partageaient mon bureau, et avec qui j'ai vécu des supers moments de complicité.
Je remercie aussi François, le nouvel hôte de mon (ancien) bureau, pour sa vivacité et sa grande implication dans son stage. Je suis sûre que tu vas faire une super thèse à GR !
Je remercie Cristina et Ibrahima pour tout leur soutien apporté lors de cette thèse, notamment en ce qui concerne la partie dosimétrie et plus particulièrement calcul de dose.
Un grand merci à Alexandre, pour son aide toujours bénéfique en termes d'informatique.
Merci aussi à Marvin, Enzo, Théo, Amaury et Théophraste pour leurs conseils en IA et leur humour… assez épicé !

Et justement en parlant d'humour épicé, je tiens à remercier Roger ! Merci à toi pour tes nombreux conseils concernant la dosimétrie ou encore les statistiques ainsi que pour ta bonne humeur à toute épreuve.

Je remercie aussi Marion, qui a toujours su se montrer bienveillante et disponible pour l'équipe !

J'ai aussi pu bénéficier de l'aide de collègues qui travaillent dans les départements de Radiothérapie et d'Imagerie de GR.

En tout premier lieu, Anne Gasnier, avec qui j'ai pu collaborer lors de la seconde étude. Merci à toi Anne, pour tout ce que tu m'as apporté en termes de dosimétrie, technique de traitement… Tu as toujours été très pédagogue dans tes explications ! Merci aussi à toi pour cette préparation « technique » à mon entretien pour le postdoc, qui m'a permis de mieux m'approprier le contexte du sujet.

Je remercie aussi Nathalie, Vithun, Julie, Ilhem, Isabelle, Pauline, Florent pour leur intérêt pour cette thèse, leurs précieux conseils ou encore leur aide concernant de nombreux aspects cliniques, comme les courbes de calibration ou encore les algorithmes de calcul de dose.

J'ai aussi une pensée toute particulière pour Bruno, Agathe et Florian en dosimétrie, avec qui j'ai eu l'occasion de collaborer sur la deuxième étude et qui ont toujours fait preuve de bienveillance à mon égard. Merci à vous d'avoir fait partie de cette aventure !

J'en profite aussi pour remercier tout particulièrement Simon Corbin, qui a œuvré à la mise en place de cette collaboration, et sans qui l'étude des différents impacts dosimétriques de la deuxième étude (algorithme de calcul de dose, technique de traitement) n'aurait pas été possible !

Merci plus largement à Karima, Ludovic, Jane, Marjolaine, Rodrigo, Quentin sur qui j'ai pu compter tout au long de la thèse pour m'apporter des connaissances techniques en termes de dosimétrie mais aussi une certaine notion de la réalité clinique.

Je remercie aussi Samy Ammari pour sa grande disponibilité, son expertise en IRM et son enthousiasme dans tous les projets auquel il a été associé.

Enfin un grand merci à Frédéric Dhermain et Pierre Blanchard pour leurs connaissances pointues et conseils avisés en termes de radiothérapie des tumeurs cérébrales et ORL/génito-urinaires respectivement.

Je remercie Lucas Fidon, qui a développé le réseau de neurones à la base de cette thèse, et a toujours su se montrer présent et pédagogue à chaque difficulté rencontrée concernant la partie technique de la génération des peudo CT. Bon courage à toi pour tes derniers mois de thèse !

Un grand merci à Norbert, Rafael, Aurélien avec qui j'ai fait mes débuts à TheraPanacea, et qui ont été d'une aide précieuse lorsque j'ai débuté la programmation.
Je remercie aussi Benjamin, Christophe, Sibylle, Thierry, Eugénie, Sonia, Catherine Nemoz, pour leur bonne humeur, leurs petites blagues et leur soutien sans faille durant cette thèse !

Dans le même temps, je souhaite remercier Maria, avec qui j'ai pu travailler lors de la première étude. Thank you for all your advice and help during this thesis… and I really wish you the best for the future!

Un grand merci aussi à Ludovic De Marzi du CPO, pour son implication dans la troisième étude de cette thèse qui porte sur la protonthérapie, ses conseils avisés et pour le co-encadrement de deux stagiaires !

J'aimerais à présent remercier des personnes qui ne plus affiliées à GR ou à TheraPanacea, et avec qui j'ai pu tisser des liens forts.
Jade, merci pour ton précieux soutien et ton optimisme légendaire. Je te souhaite sincèrement de réussir la fac de médecine, même si je ne me fais pas trop de soucis de ce côté-là !
Merci à Arthur, Maélie, Louis et Mehdi pour les supers stages que vous avez faits à GR, pour votre joie de vivre et…toutes nos sorties dans le 5$^{\text{ème}}$!
Un grand merci à Sylvain, Marie, Kim, Anthony et Saturnin avec qui j'ai fait mes débuts à GR. Merci pour votre accueil chaleureux, votre disponibilité et tous vos conseils bienveillants. J'espère que l'on aura l'occasion de se revoir.
Je remercie aussi Michel et Léa que j'ai tout récemment rencontrés à GR et avec qui j'ai pu partager de nombreux moments de complicité. Je vous souhaite beaucoup de succès dans la suite de vos études, c'est amplement mérité !
J'ai aussi une pensée pour Caroline et Cindy, qui se sont toujours montrées bienveillantes et m'ont soutenue durant cette thèse, et ce, dans la joie et la bonne humeur !

Un grand merci à Henri que j'ai pu rencontrer à TheraPanacea, et qui est à présent devenu un ami. Merci pour ton précieux soutien et tes encouragements tout au long de la thèse, en particulier pendant la période de rédaction !
Mes remerciements s'adressent à présent à Rutger. You have been always present, happy to share advice from your own PhD experience. You also helped me a lot in my postdoc research. So, I sincerely would like to thank you. And don't worry, I am sure you will be a super dad!

Enfin, et de manière un peu plus personnelle, je tiens à remercier toute ma famille ainsi que mes proches qui m'ont toujours soutenue et ont toujours cru en moi.
J'ai une pensée pour ma mère qui s'est toujours grandement intéressée à la thèse et qui a suivi l'évolution du manuscrit jour par jour. Un grand merci à toi pour ton optimisme dans chacun des projets que j'entreprends.
Je tiens aussi à remercier sincèrement mon père pour son aide constante et précieuse tout au long de mes études, sa générosité de cœur ainsi que son humour !
Enfin, je remercie ma sœur pour sa grande implication dans la thèse. Elle a partagé mes joies comme la publication de mon premier papier. Elle s'est transformée en DRH lorsqu'il a fallu préparer mon entretien pour le postdoc. Au mois d'aout, elle a enfilé une dernière fois sa tenue de Wonder Woman pour m'aider dans mon déménagement pour l'Allemagne... Un grand merci à elle donc !

Synthèse

Les traitements actuels des tumeurs cérébrales par radiothérapie incluent la double acquisition d'un scanner et d'une Imagerie par Résonance Magnétique (IRM). Le scanner est nécessaire à la délinéation de certains organes à risque (OAR) ainsi que pour la dosimétrie. En effet, ses intensités, les Unités Hounsfield (UH), sont directement reliées aux densités électroniques via la courbe de calibration implémentée dans le logiciel de planification de traitement. L'IRM, quant à elle, est préférée durant la segmentation des volumes cibles ainsi que des OAR de par son excellent contraste dans les régions contenant des tissus mous.

Cette double acquisition impose un recalage spatial, qui a pour but de placer les deux images dans le même référentiel. Cependant, Ulin et al. ont prouvé qu'un recalage rigide de l'IRM sur le CT pouvait induire des erreurs jusqu'à 3mm. Celles-ci sont actuellement prises en compte à l'aide d'une augmentation des marges de sécurité, et engendrent donc une irradiation potentielle des tissus sains environnants. Dans ce contexte, il apparait pertinent de générer un pseudo scanner (pCT) directement à partir de l'IRM, afin d'éviter cette étape de recalage. Cette thèse avait pour but de générer des pCT à partir d'IRM à l'aide de Deep Learning (DL) pour des patients atteints de tumeurs cérébrales traitées par radiothérapie.

La première étude a consisté à investiguer l'impact de paramètres clés de l'image ainsi que du réseau de neurones. Les données de 402 patients atteints de tumeurs cérébrales, représentant 182 couples scanner/IRM pondérée en T1 (T1), 180 couples scanner/IRM pondérée en T1 avec injection de produit de contraste (T1-Gd) et 40 scanners/T1/T1-Gd, ont été rétrospectivement collectées. Les IRM ont été acquises à l'aide de cinq machines différentes afin d'assurer la robustesse du modèle développé. Le réseau de neurones utilisé par défaut était une version modifiée du HighResNet 3D (Li et al.), se caractérisant notamment par ses connexions résiduelles et ses filtres de convolution dilatés. L'impact de la taille du jeu d'entrainement a tout d'abord été investigué, à l'aide de jeux de données d'entrainement de taille variable, i.e. 242, 121, 60, 30 et 15 patients. L'impact de la séquence IRM utilisée en entrée du réseau a été analysé en entrainant deux réseaux en parallèle basés sur des séquences T1 ou T1-Gd uniquement. Trois méthodes de standardisation des IRM ont fait l'objet d'analyses, à savoir une harmonisation basée sur des histogrammes, imposant une moyenne à zéro et une variance à 1 ou white stripe. L'impact du filtre N4 de correction d'inhomogénéités de champ a été quantifié. Enfin, le HighResNet 3D a été comparé au UNet 3D, afin d'en extraire l'architecture du réseau optimale. Les images ont été évaluées à l'aide d'une métrique basée sur les intensités, à savoir l'Erreur Absolue Moyenne (EAM). De plus, une étude dosimétrique a été réalisée sur une sous-cohorte du jeu de test, composée de 39 patients traités à l'aide d'une radiothérapie conformationnelle par modulation d'intensité. Le plan initial du CT a été transféré sur le pCT dans le but de recalculer la dose sans optimisation des paramètres, à l'aide d'un algorithme pencil beam. Les gamma index globaux sans seuil 1%/1mm, 2%/2mm et 3%/3mm ainsi que les différences sur les grandeurs D2%, D50%, D95% and D98% dérivées des Histogrammes Dose/Volume (HDV) pour le Planning Target Volume (PTV, Volume de Planification de Traitement) ont été calculés. Des performances dosimétriques équivalentes ont été prouvées pour tous les paramètres testés, excepté pour la taille du jeu d'entrainement. Concernant ce dernier, une amélioration notable de l'EAM avec l'augmentation du nombre de patients considéré dans le jeu d'entrainement a été observée. Une possibilité de travail futur consisterait à évaluer l'impact dosimétrique de chaque taille du jeu d'entrainement.

Dans le but final d'implémenter une dosimétrie sur pCT en clinique ainsi que de prouver la non-infériorité d'une planification de traitement de radiothérapie basée uniquement sur IRM vis-à-vis de celle conventionnellement basée sur scanner, les métriques dosimétriques les plus pertinentes pour une évaluation dosimétrique des pCT, c'est-à-dire non biaisées par le volume des volumes cibles, ont été déterminées lors d'une seconde étude. Les différents gamma index globaux impliqués étaient les 1%/1mm avec des seuils de dose égaux à 0%, 10%, 20%, 50%, 80%. Des critères locaux 1%/1mm avec seuils de 0% et 10% ont aussi été étudiés. Concernant les données, les jeux d'entrainement et de validation du réseau de l'étude 1 ont été réutilisés. 71 patients représentant une large diversité en termes de localisation tumorale, i.e. cervelet (4 patients), encéphale in toto (26 patients), lésion frontale (14 patients), lésion occipitale (3 patients), lésion pariétale (12 patients), lésion temporale (12 patients), et de techniques de traitement, à savoir arc dynamique conformationnel (17 patients), radiothérapie conformationnelle 3D (27 patients) et arcthérapie volumétrique modulée (27 patients), ont été utilisés en jeu de test. Le réseau de neurones et ses hyperparamètres associés étaient ceux utilisés dans l'étude 1. Là encore, les plans de traitement initiaux ont été transférés sur les pCT, sans optimisation. La dose a été recalculée à l'aide des algorithmes pencil beam (17 patients) ou collapsed cone (54 patients). Des analyses de corrélations à l'aide du test Spearman entre les taux de passage du test du gamma index et les volumes des PTV ont été réalisées afin de sélectionner deux critères pour la suite de l'étude présentant un score de corrélation faible. Les corrélations entre ces deux gamma index, l'EAM ainsi que les coefficients de Dice, ont ensuite été calculées pour établir le lien entre gamma index et métriques d'évaluation basées sur les intensités. Enfin, trois différentes qualités de pCT ont été générées en stoppant l'entrainement du réseau à différentes itérations (epochs), à savoir 3, 14 et 48, afin de simuler des scenarios d'erreurs extrêmes et quantifier leurs impacts dosimétriques. Les gamma index 1%/1mm global et local avec seuils de 50% et 10% respectivement ont été définis comme métriques dosimétriques optimales. Ils doivent donc être préférentiellement reportés pour qualifier la qualité de pCT cérébraux, ainsi que les différences de métriques basées sur les HDV pour le PTV et les organes à risque. Concernant l'impact des trois scenarios d'erreurs de pCT, les métriques définies préalablement ont prouvé leur caractère discriminant en termes de qualité d'image.

Enfin, une troisième étude, visant à évaluer la généralisabilité du modèle de génération de pCT préalablement développé sur des adultes à une cohorte pédiatrique traitée par protonthérapie, a été réalisée. Pour ce faire, 18 patients atteints de craniopharyngiomes ont été inclus. Pour chaque patient, un scanner et une IRM T1-Gd ont été acquis, avec un délai moyen égal à 2.0 jours +/- 3.0 jours. Le modèle préalablement obtenu en standardisant les IRM à l'aide d'une moyenne à zéro et d'une variance à 1 a ensuite été utilisé pour générer les pCT. Dans un premier temps, les images ont été évaluées à l'aide de l'EAM. Dans un second temps, une étude dosimétrique impliquant un transfert de plan et un re-calcul de dose à l'aide de l'algorithme pencil beam a été réalisée. Les critères d'évaluation dosimétrique appliqués étaient ceux recommandés par l'étude 2. Des performances dosimétriques acceptables ont été atteintes pour la plupart des patients, suggérant une transférabilité et une robustesse du modèle développé. Les travaux futurs consistent à évaluer 4 méthodes d'entrainement et de validation du réseau, i.e. basées sur des patients pédiatriques uniquement, sur des patients adultes uniquement, sur des cohortes mixtes adultes/pédiatrie, ou sur un pré-entrainement du réseau sur une cohorte de patients adultes suivi d'un ajustement des paramètres sur une cohorte pédiatrique. Les quatre modèles générés sont ensuite appliqués à la même cohorte de test pédiatrique. L'originalité de cette approche réside dans les données du réseau utilisées pour calculer la fonction coût, qui ne sont pas les Unités Hounsfield, mais les cartes de pouvoirs d'arrêt, afin de s'affranchir de l'influence de chaque scanner sur l'image, et donc de potentiellement générer des bases de données plus importantes et des modèles généralisables.

Ainsi, une caractérisation précise des paramètres influençant la qualité de pCT générés par DL a été réalisée dans cette thèse, ainsi qu'une compréhension complète des métriques utilisées pour les évaluer en dosimétrie. Cette dernière pourra jouer un rôle clé dans la quantification des critères d’évaluation, lors de la mise en place d’essais cliniques, tels que GliopCT.

Table of Contents

Introduction

Cancer is one of the leading causes of death worldwide, with 19.3 new million cases and a mortality of 10.0 million in 2020 (2). Main locations among the new patients are female breast (11.7%), lungs (11.4%), colorectum (9.8%) and prostate (7.3%). Brain and nervous system tumors represented 308 102 new cases in 2020, and a mortality equal to 251 329, and are the first cause of solid tumors for paediatric patients (3).

Conventional treatments for brain tumors require the acquisition of a Computed Tomography (CT) and Magnetic Resonance Imaging (MRI) sequences. The first is used during the delineation step, and due to its acquisition process, has the advantage to have its units directly related to Electron Densities (ED). The latter presents an excellent soft tissue contrast, enabling a highly accurate target volume delineation.

Yet, Ulin et al. (4) proved that the MRI to CT registration applied to place the MRI in the CT spatial frame, resulted in errors up to 2mm +/- 1mm. They are currently considered in the whole radiotherapy workflow through a margins increase, and thus induce a potential irradiation of healthy tissues.

As a result, generating pseudo Computed Tomography (pCT) from MRI appears appealing since the synthetic image would directly be in the MRI spatial frame, avoiding a registration application.

The aim of the thesis was to generate brain pCT from MRI with Deep Learning (DL), and to extensively quantify the associated intensity-based and dosimetric quality to assess optimal network and preprocessing-based parameters as well as evaluation metrics enhancing performances.

Chapter 1 presents a general overview of head tumors treated with radiotherapy, including the whole workflow description and the urgent need for a more accurate patient care.

Chapter 2 is an introduction to Artificial Intelligence (AI), and more specifically to DL. Overview of basic concepts and AI-based implementation into clinics are provided.

Chapter 3 introduces the concept of brain pCT generated from MRI. Commonly evaluation metrics are reported, as well as a literature review of main approaches developed for such a task.

Chapter 4 is a study performed to quantify the impact of key parameters on the pCT quality. These parameters were the training set size, the MR input sequence, the strategy used for MRI standardization and biased correction, as well as the network architecture. The computed error was evaluated with an intensity-related metric and dosimetry analysis.

Chapter 5 is an investigation of Planning Target Volume (PTV)-unbiased dosimetry metrics, as well as the assessment of their relationships with intensity-based metrics. In this study, dosimetry impacts of extreme errors scenarios are analysed via the generation of three different pCT qualities.

Chapter 6 is the generalizability evaluation of the Chapter 4 pCT model to paediatric patients treated with protontherapy. Obtained pCT are evaluated with an intensity-based metric and Chapter 5 recommended dosimetry metrics.

Chapter 7 is a global discussion about the remaining challenges of MRI-only radiotherapy workflow implementation into clinics.

1. Central nervous system tumors

1.1. Pathology

Central nervous system tumors refer to abnormal cells amounts which can be benign or malign. The latter comprises either primary tumors, i.e. deriving from brain or spinal cord cells, or secondary tumors (metastases).

A classification and grading system have been introduced and were recently updated by the World Health Organization (WHO) to describe a tumor based on its histological and molecular characteristics (5). Seventeen main categories were defined, namely: diffuse astrocytic and oligodendroglial tumors, other astrocytic tumors, ependymal tumors, other gliomas, choroid plexus tumors, neuronal and mixed neuronal-glial tumors, tumors of the pineal region, embryonal tumors, tumors of the cranial and paraspinal nerves, meningiomas, mesenchymal/non-meningothelial tumors, melanocytic tumors, lymphomas, histiocytic tumors, germ cell tumors, tumors of the sellar region, metastatic tumors.

Regarding primary tumors, 85% to 95% are located within the brain (6). In France, primary brain tumors represent approximately 5000 new cases each year. Typical brain tumor symptoms can involve headaches, confusion, convulsions, nausea, vomiting. Various causes were reported, including sex, radiotherapy dose exposure and genetics (7). The latter comprises diseases such as neurofibromatosis types 1 and 2, respectively linked to NF1 and NF2 genes, Von Hippel-Lindau syndrome associated to VHL gene or Li-Fraumeni syndrome referring to TP53 gene (8,9). Mutations in these genes known as tumor suppressors would fatally lead to the development of cancers.

Surgery is one of the pillars to totally or partially remove a tumor. For instance, low grade gliomas often require only surgery. This method is a balance between removing the maximum of tumor and protecting vital surrounding organs to avoid neurologic disorders. It has benefitted from technology advances, such as cortical mapping, which is based on electrostimulation to recognize functional areas (10) or the endoscopic endonasal approach, consisting in the tumor removal through natural cavities, such as the nose. It is particularly effective for craniopharyngiomas (11,12), meningiomas (13,14) and chordomas (15,16) and has the advantage to be less invasive than the standard transcranial approach.

Regarding chemotherapy, drugs are used to achieve three different purposes: either the chemotherapy is adjuvant, after a surgery for instance to ensure the tumor vanishing, or it is neoadjuvant, i.e. prescribed before the surgery in case the tumor is too large for resection, or it is palliative, to improve patients life conditions without a total tumor removal (17). Standard drugs used in clinics are temozolomide, lomustine, procarbazine, vincristine and carmustine, and depend on the brain tumor histology. Doses are usually based on the Body Surface Area calculated in m^2, and range from $0.4mg/m^2$ to $200mg/m^2$ per day.

Lastly, radiotherapy using high particles energy beams to destroy tumor cells can be adopted, possibly in combination with the previous approaches. Prescribed doses for low and high grades tumors respectively range from 45Gy to 54Gy and from 54Gy to 60Gy, according to Société Française de Radiothérapie Oncologique guidelines of 2007 (18). It depends on the histology, patient age and patient general condition.

For glioblastomas treated with concomitant chemotherapy based on Temozolomide ($75mg/m^2$ per day, 7 days per week, during the whole radiotherapy process), the standard treatment consists in delivering 60Gy to the tumor, with 2Gy/session and 5 sessions per week. Adjuvant chemotherapy with Temozolomide, for 5 days every 28 days, is finally administrated (19). This Stupp protocol is designed for people less than 70 years old, in a good condition (WHO score ≤ 2) (20). Older people, still in good condition, may benefit from hypo-fractionated radiotherapy, i.e. delivering 40.5Gy in 15 fractions (20).

Regarding grade III glioma, delivered doses for patients with bad prognosis characteristics, such as wild type isocitrate dehydrogenase status (21), are similar to the doses delivered for glioblastomas (22). Good conditions patients and older than 70 years old may benefit from a dose decrease to 40.5Gy, delivered in 15 fractions.

Concerning grade II gliomas, a total dose varying from 45Gy to 50.4Gy should be delivered in fractions of 1.8Gy (20).

Medulloblastomas are usually treated with doses comprised between 30Gy to 36Gy (with a maximum of 54Gy in the posterior fossa), delivered in 15 to 18 fractions. Worse prognosis patients may also benefit from chemotherapy (etoposide and carboplatin) before and after radiotherapy (23).

Lastly, primary CSN lymphomas are treated with 23Gy to 40Gy, delivered in fractions of 1.8 to 2Gy, 5 fractions per week (24).

Life expectancy depends on various parameters, such as histology, age, tumor location, surgery, genes mutations, etc. In UK, 20% and 5% patients with grade III and IV patients survive 5 years or more (25). In US, five-year survival rates for grade III gliomas are respectively equal to 29% and 15% for patients ages comprised between [45; 54] and [55; 64] (26). Regarding glioblastomas, rates decrease to 9% and 6%, according to the American Cancer Society (26).

1.2. Radiotherapy workflow

1.2.1. Images acquisition

1.2.1.1. CT

1.2.1.1.1. History

Scanners were first developed in 1972, by Godfrey Newbold Hounsfield, based on mathematical theory introduced by Allan McLeod Cormack. So far, four different CT generations have been developed.

In 1970, an X-ray tube producing a pencil beam was placed in front of a single detector. Due to the narrow beam, the source had to translate through the whole slice before applying 1° rotation. The whole process was repeated, leading to long acquisition times (~5 minutes per rotation).

Second generation of CT was composed of a source producing a fan beam with an angle comprised between 5° and 10° and arrays containing about 30 detectors. The beam coverage was still limited, leading to a translation/rotation approach, as described for the first CT generation. Acquisition times were significantly reduced to 20 seconds to 2 minutes per slice.

Note that due to these long times to acquire one image, first and second generations CT were only used for the head.

In 1976, third CT generation was developed, based on a rotate-rotate process. Indeed, a wide fan beam, whose angle ranged from 30° to 50°, was used to completely cover the patient. Hundreds curved detectors were placed in front of the source, and rotated similarly to it. Time to acquire a slice was equal to 1 second.

The fourth CT generation was a rotate-stationary process. The detectors formed a closed ring around the patients, and were more numerous than the third generation CT (between 600 and 4800 elements). Regarding the source, which was still a fan beam, two designs were described: either rotating within or outside the detectors ring. In the latter case, detectors positions were adapted, to avoid irradiating electronics behind the detectors (27). Acquisition times were still equal to 1 second.

Figure 1 presents the four CT generations, and was reported by Luke et al. (28).

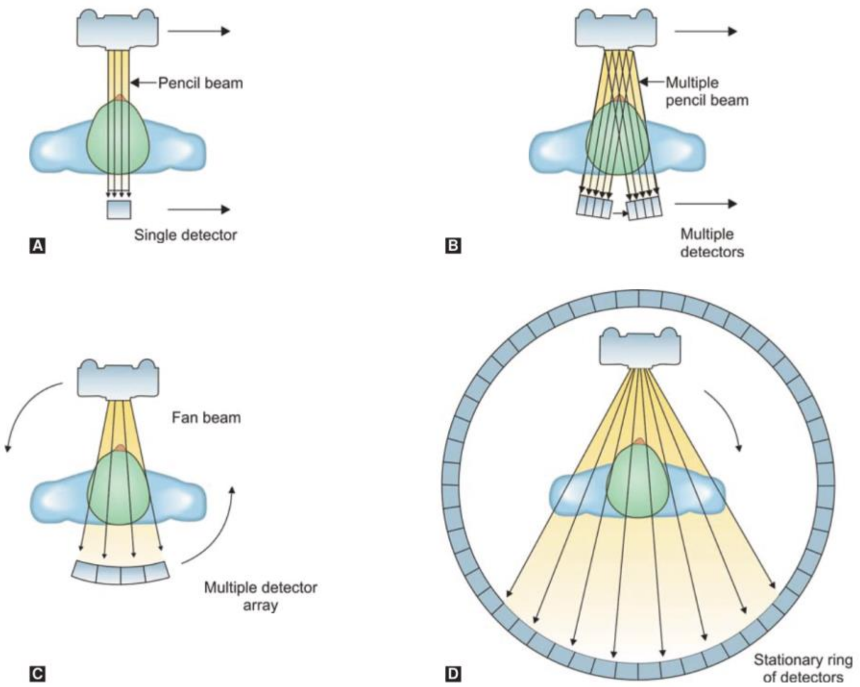

Figure 1: 1rst, 2nd, 3rd and 4th CT generations respectively presented in A, B, C and D. Reprinted from "Comparison of Spiral Computed Tomography and Cone-Beam Computed Tomography", Luke AM, Shetty KP, Satish SV, Kilaru K, 2013, J Indian Acad Oral Med Radiol., 25(3). CC BY-NC-SA (2013).

As properly stated by McCollough et al. (29), attenuation coefficients depend on various parameters, such as material properties, beam energy, etc, resulting in the potential assignment of equal coefficients for different materials. Dual energy CTs have been recently introduced to overcome this issue and rely on two acquisitions with different tube voltages, such as 80kVp and 140kVp. For instance, iodine, representing the contrast product used in CT, has a K edge of energy of 33.2keV. A tension of 80kVp enables to obtain photons with energies close to 33.2keV, leading to an increased attenuation (30) and a better visualization than with a single energy CT device.

1.2.1.1.2. X-Ray beam production

To produce the X-ray beam, a tungsten filament is heated to 2400K, and electrons are emitted via thermionic effect at the cathode. They are accelerated to the anode, often constituted of Tungsten, with a tube voltage, often ranging from 90kV to 140kV. They enter the tungsten target to a depth of approximately 0.5mm resulting into two types of interactions with atoms.

First, the accelerated electrons interact with the target nucleus electric field via the Coulomb force, leading to an energy loss as a X-ray, an incident particle deceleration and deviation (Bremsstrahlung). The emitted X-ray energy ranges from 0 to the incident electron energy, resulting in a continuous spectrum. The associated cross section is proportional to Z^2/m^2, with Z and m the target atomic number and the incident particle mass.

Second, electrons interact with the anode inner shell atoms electrons. For inelastic collisions, an energy transfer E occurs, with a cross section proportional to $1/E^2$. If E is inferior to 10eV, it results in excitations and heat transfers. Atoms ionizations occur for energies comprised between 10eV and 100eV. Energies above 100eV correspond to delta electrons, i.e. remote electrons trajectories for incoming particle. Processes such as ionizations lead to electron knocks out, whole filling with superficial shells electrons associated to energy release. The emitted X-ray energy is the difference between the two binding energies of the two enrolled shells. This phenomenon is named fluorescence.

Krestel et al. (31) defined the X-ray production efficiency η as:

$$\eta = aUZ \qquad \text{Equation 1}$$

With U, Z, a, respectively representing the tube kilovoltage, the target atomic number and a constant equal to $1.1\text{x}10^{-9}V^{-1}$. Michael et al. (32) reported a relevant example of a tube kilovoltage of 100kV associated with a Tungsten target leading to an efficiency lower than 0.1%, the rest of the energy being transformed into heat.

1.2.1.1.3. Detection

Three main generations of detectors have been developed so far.

Gas detectors are composed of an ensemble of chambers containing gas under pressure. The incident photon, which has a probability of 60% to 87% to be absorbed by the detector, interacts with the gas via Photoelectric effect, resulting in a gas ionization. The electric field applied between the chambers leads to a collection of charges and the creation of a current (27,33). Note that the gas usually used is xenon, due to its stability when pressurized.

The second detectors category corresponds to solid-state, i.e. the combination of a scintillator and a photodiode. The former, usually CsI, BGO or $VdWO_4$, converts the X-ray into visible light. This process can be split into three main stages. First, the incident beam creates hole/electron couples in the valence and conduction bands respectively, interacting with the crystal via Photoelectric, or Compton interactions. Second, the hole/electron pairs travel in the crystal to reach specific sites to scintillate. Finally, the electron gets trapped and emits visible radiation light. It is then transformed into electric signals by the photodiode via Photoelectric effect. A final analog-to-digital converter has the role transform the electric signals into digits to enable computers interpretation.

More recently, full electronic detectors have been introduced, such as Stellar Detector (Siemens Healthineers, Erlangen, Germany) (34). Contrary to solid-state detectors, the photodiode and analog-to-digital converter are combined into one component, highly limiting the signal path. Noise is also drastically decreased, enabling images acquisition at lower mAs to obtain a similar noise level. For instance, an initial 300mAs-based image noise is achievable with a tube current of 250mAs, resulting in a reduction of 20%.

1.2.1.1.4. Image reconstruction and visualization

The Radon transform of a function f represents all the projections for angles ranging from 0 to π, as described by:

$$p(u,\theta) = \int_{-\infty}^{+\infty} f(x,y)dv \qquad \text{Equation 2}$$

With p the projection and f the signal representing attenuation.

The goal in reconstruction is to assess f from the multiple p projections.

Two main approaches have been described in the literature for image reconstruction, i.e. transforming a set of projections, corresponding to different angles, to an attenuation map.

First, analytic solutions, relying on continuous approach of the problem, have been introduced. The most popular is the Filtered Back Projection (FBP), corresponding to (35):

$$f(x,y) = \int_{0}^{\pi} \hat{p}_\theta(t)d\theta \qquad \text{Equation 3}$$

With $f(x,y)$ the signal linked to the attenuation, θ the acquisition angle, $\hat{p}_\theta$ the filtered projection p_θ corresponding to an angle θ.

First, it consists in calculating the Fourier transform of the input sinogram. The signal is then filtered to obtain the exact Radon transform, before calculating the Fourier inverse and back propagating the signal, to achieve the attenuation associated to each voxel. As a result, it appears it is a simple approach, easily and quickly implemented. However, several limitations have been highlighted, such as the noise quantity which is directly proportional to the delivered dose. Thus, low dose CT, presenting a higher patient safety, result in non-satisfying quality images (36).

Second type of algorithms are iterative, and correspond to a discrete problem formulation as:

$$p = R \times f \qquad \text{Equation 4}$$

With p the projection, R the Radon transform and f the signal from spatial frame representing attenuation.

They consist in iteratively projecting the ith-image to be reconstructed. The error with the ground truth projection is then calculated, before forwarding it to the spatial frame. This error is finally applied on the image, to obtain the i+1th image. These algorithms have the advantage to strongly reduce noise and artefacts, compared to FBP. However, their computational time is high, and this approach presents an over smoothing risk, potentially leading to the removal of fine objects, as highlighted by Stiller (37).

The beam attenuation deriving from the two previous reconstruction approaches is finally converted into Hounsfield Units (HU) via the formula:

$$N_{HU} = \frac{\mu - \mu_{eau}}{\mu_{eau}} \times 1000 \qquad \text{Equation 5}$$

With N_{HU} the HU number of a given voxel, μ the voxel measured attenuation, μ_{eau} the water attenuation.

HU for several tissues are provided in Table 1.

Table 1: Tissues and corresponding HU.

Tissue	**HU**
Air	-1000
Lung	-600 to -400
Fat	-100 to -50
Water	0

Muscle	10 to 40
White matter	20 to 30
Grey matter	37 to 45
Bone	700 (spongy) to 3000 (dense)

For an adapted visualization to the investigated tissues, CT are observed via a window width and a window level, respectively corresponding to the HU range and the HU window centre. The HU lower than the minimum and higher than the maximum window are respectively displayed in black and white. Two main windows are commonly used for head tumors. First, bone window displays bone related injuries such as calcification. Regarding brain window, differences in soft tissues, such as blood and brain, are clearly visible as well as bone fractures (38). Usual bone and brain couples of width and level windows are respectively equal to (1800, 400) and (80, 40). Figure 2 presents a head CT in the two previously described contrasts.

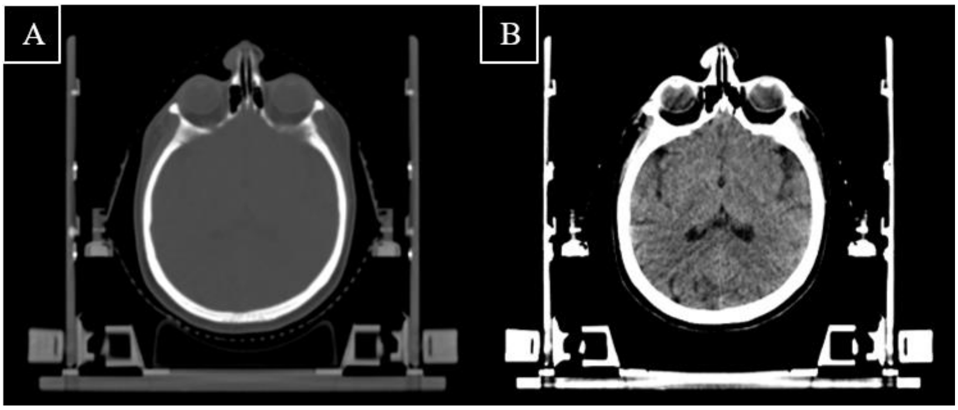

Figure 2: Head CT in bone (A) and brain (B) windowings.

1.2.1.1.5. Image quality

1.2.1.1.5.1. Spatial resolution

Resolution refers to the smallest distance required by the imaging system to distinguish two adjacent points. It can be quantified by Full Width Half Maximum of the Point Spread Function (PSF), representing how blurry a perfect object point becomes when passing through the imaging system. It is defined as (39):

$$PSF(x,y) = \mathcal{F}^{-1}\left(\frac{\mathcal{F}\big(I(x,y)\big)}{\mathcal{F}\big(O(x,y)\big)}\right) \qquad \text{Equation 6}$$

With $\mathcal{F}$, $\mathcal{F}^{-1}$, I, O the Fourier Transform, inverse Fourier Transform, the image and object functions respectively.

A second approach to quantify resolution relies on the Modulation Transfer Function (MTF), describing the imaging system ability to preserve small details. More formally, the MTF is the PSF Fourier Transform amplitude and is defined as (40):

$$MTF(u,v) = \frac{|OTF(u,v)|}{|OTF(0,0)|} \qquad \text{Equation 7}$$

With $OTF(u,v) = \mathcal{F}(PSF(x,y))$

To measure CT spatial resolution, a phantom using a point source, such as for instance thin wires or spheres, is the most popular technique (41).

Current CT devices reach performances equal to 0.5mm and comprised between 0.5mm and 0.625mm for the in-plane and transverse resolutions (42).

1.2.1.1.5.2. Noise

Noise corresponds to random fluctuations around a mean value. It can be quantified with the Signal to Noise Ratio (SNR), or the Contrast to Noise Ratio (CNR). The former represents the ideal signal divided by the noise. For periodic signals, the latter, also known as Michelson contrast, can be defined as:

$$CNR = \frac{S_1 - S_2}{S_1 + S_2} \qquad \text{Equation 8}$$

With S_1 and S_2 representing the signals from two distinct regions.

Different noise origins exist. First, X-rays exhibit quantum properties. Photons distribution follows a Poisson law, resulting in heterogeneous photons beams on detectors (43). This phenomenon is known as quantum noise and is proportional to the inverse of the photons number square root. Second category is the electronic noise, deriving from detection system, and more specifically when converting the analog signal form the photodiode to digital signal (44). Third category corresponds to the noise originating from reconstruction. In several studies, iterative methods have been proved to create less noisy images. Kuo et al. reported a noise reduction of up to 53.7% for the bladder when using iterative model reconstruction than FBP for abdomen CT reconstruction (45). In 2013, Shuman et al. compared model-based iterative reconstruction and FBP for liver lesions CT, respectively leading to CNR equal 34.4 +/- 29.1 and 6.3 +/- 6.0 (46).

1.2.1.1.5.3. Artefacts

Artefacts are a misrepresentation of the reality, potentially leading to biased diagnoses. Different types of artefacts exist, as precisely investigated by Barrett et al. (47).

First category is hardware-based, such as detector miscalibrations resulting in concentric dark or white circles and corresponding to rings artefacts (48). Solid state detectors are more sensitive to this artefact than gas detectors. A proper detector calibration associated with a maintenance software are key for ring artefacts avoidance.

Moreover, out-of-field artefacts can occur, when the body part to examine is not in the CT Field Of View (FOV), leading to missed projections and streaking artefacts. Dotson et al. recently reported promising decrease in out of field artefact severity mean from 2.6 to 0.2 for water and iodine density images respectively, highlighting the potential of the iodine density images from dual energy CT (49).

Second class concerns physics-related artefacts. One of them is beam hardening. It is associated with the polychromatic X-ray energy spectrum and corresponds to a higher attenuation of low energetic photons compared to high energetic ones. It leads to streaking and cupping artefacts. Regarding the former, different source/detector angular positions result in beam hardening differences leading to artefacts. The latter is well described by Barrett et al. (47) and can be simulated with a photons beam irradiating a cylinder phantom. Photons at the centre are more hardened than at the extremities. It results in an increase of their mean energy when passing through the detectors and thus an artificially smaller attenuation than in the cylinder edges. Physical filters are usually placed by the vendors to reduce this effect. For instance, Davis et al. reported that 6mm of Aluminium are enough for a 90kV standard beam to exhibit similar attenuation properties as a monoenergetic beam (50).

Partial volume effect occurs when different tissues attenuations are gathered in one voxel, leading to an assignment of the attenuation average. A possible approach to reduce this effect is to acquire thinner slices, remembering noise is inversely proportional to slice thicknesses. In 2016, Monnin et al. investigated the optimal slice thicknesses, and reported a size equal to 75% of the object width (51).

Third type is related to the patient. Mobile tumors owing to patient motion result in blurring, streaking or shading artefacts. For voluntary movements, contentions should be used for immobilization. Involuntary motions, such as breathing, are handled during 4D CT, which consists in a 3D CT acquired at different periods of the respiratory cycle. To do so, various different approaches exist, including breathing holding or Respiratory Adaptive Computed Tomography (REACT) technique designed to acquire only when the patient breathing is regular (52). This approach has recently been introduced into clinical workflow (53).

Streaking artefacts, deriving from metal, happen when high-density material is scanned for a dental filling or prostheses for instance. Note that it can also generate physics-related artefacts such as beam hardening which was previously described. Operators may first optimize the acquisition parameters to avoid the metallic area (especially for dental filling cases). In addition, techniques are available for metal artefact reduction such as for instance Iterative Metal Artefact Reduction method (iMAR, Siemens Oncology Care Systems, Erlangen, Germany) (54) or more recently on convolutional neural networks (55).

1.2.1.1.6. Parameters influencing image quality

Beam energy spectra have two main contributions. First, the accelerated electrons interacting with the anode nucleus result in X-rays emissions via Bremsstrahlung. It is represented via a continuous spectrum, with energies ranging from 0 to the maximum electrons energy, i.e. the tube kilovoltage.

Second contribution are characteristics X-rays emitted when an electron of anode inner shells atoms is ejected, resulting in a vacancy. It is then filled with a higher-level electron, resulting in the emission of a specific X-ray. Its energy is equal to the binding energies differences between the two shells involved. These specific X-rays are thus dependent on the anode material.

A last influence concerns the filtration, whose goal is to remove low energy photons useless for image reconstruction, and results in a mean energy beam increase. Two types of filtrations exist, namely inherent and additional. The former is constituted by any non-removable X-ray tube components attenuating the beam, such as the window or the cooling oil. The latter is a non-permanent additional beam filtration having the role to decrease the entrance skin exposure, commonly with several millimetres of aluminium.

All these three parameters, namely tube kilovoltage, anode material and filtration, influence the final beam spectrum delivered to the patient. An example is provided in Figure 3, with a tungsten anode and a tube kilovoltage of 90kVp (56)

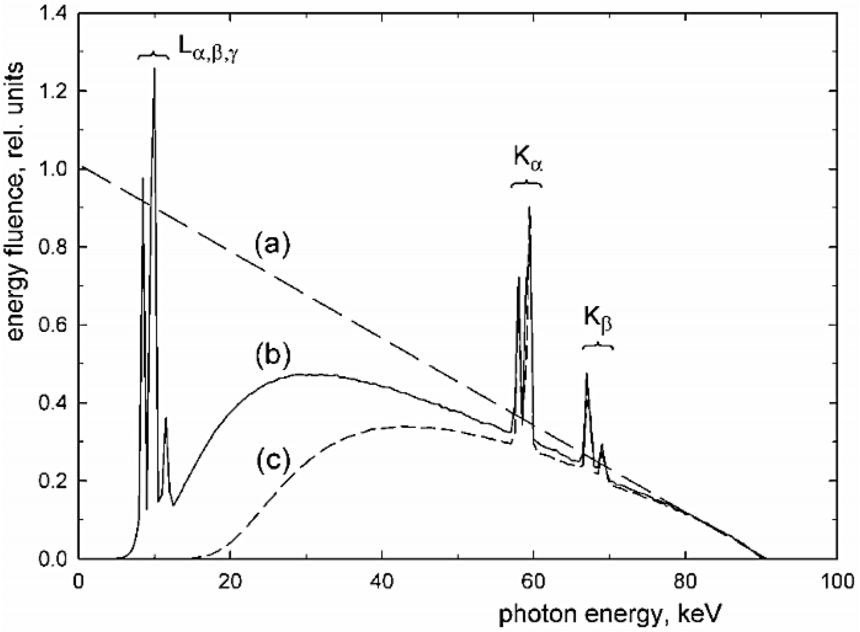

Figure 3: Beam energy spectra in case of no filtration (a), inherent filtration of 1mm Beryllium only (b) and combined with additional 2.5mm equivalent Aluminium (c). The tube anode is in Tungsten, and the potential difference is 90kVp. Reprinted from “Diagnostic Radiology Physics: A Handbook for Teachers and Students”, Dance D, Christofides S, Maidment A, McLean I., © IAEA, 2014.

As a result, the function used to convert the HU into ED, named the calibration curve is also affected by these parameters. An example of such a curve is provided in Figure 4.

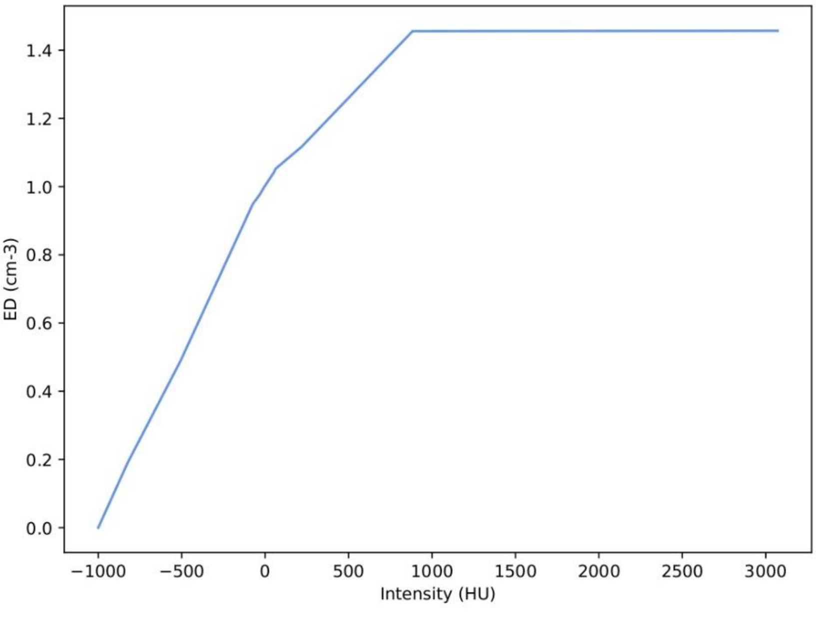

Figure 4: Calibration curve available for intracranial irradiation in the software iPlan RT 4.5 Dose (Brainlab, Munich, Germany). The associated CT device was a SOMATOM Sensation

Open (Siemens Healthineers, Erlangen, Germany) composed of a STRATON X-Ray tube used in this case with a kilovoltage of 120kVp.

1.2.1.2. MRI

1.2.1.2.1. Principles

MRI is non-radiating technology relying on the Nuclear Magnetic Resonance (NMR). This section focuses on the particular case of a flip angle equal to 90°. Under this assumption, the NMR process is presented in Figure 5. The upper and lower parts respectively represent the rotating spins and the main magnetization. Due to its large quantity in human body, hydrogen is commonly used. However, other nuclei, including phosphorus, fluorine or sodium, can be imaged.

Protons are composed of an intrinsic magnetic momentum, named spin. Initially, the protons spins have different directions and no magnetization exist (Figure 5.A). From the patient's point of view, it corresponds to the protons state before entering the MRI magnet.

Then, the patient is placed at the centre of a bore creating a uniform B_0 magnetic field along the z axis. It results in an alignment of all the protons spins either up or down, corresponding to two different energy levels: $-\frac{1}{2}\gamma\hbar B_0$ and $\frac{1}{2}\gamma\hbar B_0$, with γ the gyromagnetic ratio, ħ the reduced Planck constant and B_0 the magnetic field. The energy gap ΔE between the two levels is equal to:

$$\Delta E = \gamma\hbar B_0 \qquad \text{Equation 9}$$

To assess the spins distribution between these two energy levels, Maxwell-Boltzmann statistics is used, defined as:

$$\frac{N_{low}}{N_{high}} = e^{\frac{\Delta E}{kT}} \qquad \text{Equation 10}$$

With ΔE, k, T, N_{low}, N_{high}, the energy difference between the two levels, the Boltzmann constant, the temperature, the numbers of protons aligning spins up and down respectively.

For a magnetic field of 1.5T, hydrogen proton-based MRI and a temperature of 37°C, the previous ratio is equal to 1.00001, highlighting the low sensitivity of the NMR process.

It results in the creation of a main magnetization M, equal to M_0 and parallel to B_0 (Figure 5.B).

In addition, the Planck-Einstein formula is defined as follows:

$$\Delta E \; = hf \qquad \text{Equation 11}$$

Thus, the required wave frequency f to disturb the previously described equilibrium state is named Larmor frequency and is calculated as:

$$f \; = \frac{\gamma}{2\pi} B_0 \qquad \text{Equation 12}$$

As a result, certain protons spins go from the low energy level to the high energy level, i.e. spins down, resulting in the disappearance of the longitudinal magnetization M_z (Figure 5.C).

Moreover, the spins synchronize in phase, leading to a transversal magnetization M_{xy} (Figure 5.D). This magnetization generates a temporally variable magnetic field. It leads to a potential difference, according to the Faraday law, across the receiving coil.

When no radiofrequency is delivered, the spins stop the synchronization phase and return to the initial state. This step corresponds to the relaxation, and is described by the Bloch equation as:

$$\frac{dM}{dt} \; = \gamma M \wedge B_0 - \frac{M_{xy}}{T2} + \frac{M_0 - \; M_z}{T1} \qquad \text{Equation 13}$$

Regarding the solution for the transversal M_{xy} magnetization, it is described, in the moving frame, as:

$$M_{xy}(t) = \; M_0 e^{-\frac{t}{T_2}} \qquad \text{Equation 14}$$

Thus, the transversal magnetization first disappears (Figure 5.E) via a T2 damped sinusoidal function. It corresponds to the T2 relaxation, or spin-spin relaxation.

Second part of the relaxation focuses on M_z magnetization appearance via a T1 exponential calculated as:

$$M_z(t) = \; M_0 \left(1 - e^{-\frac{t}{T_1}}\right) \qquad \text{Equation 15}$$

It represents the T1 relaxation, or spin-lattice relaxation, and is accompanied by an energy release to the surrounding tissues (Figure 5.F).

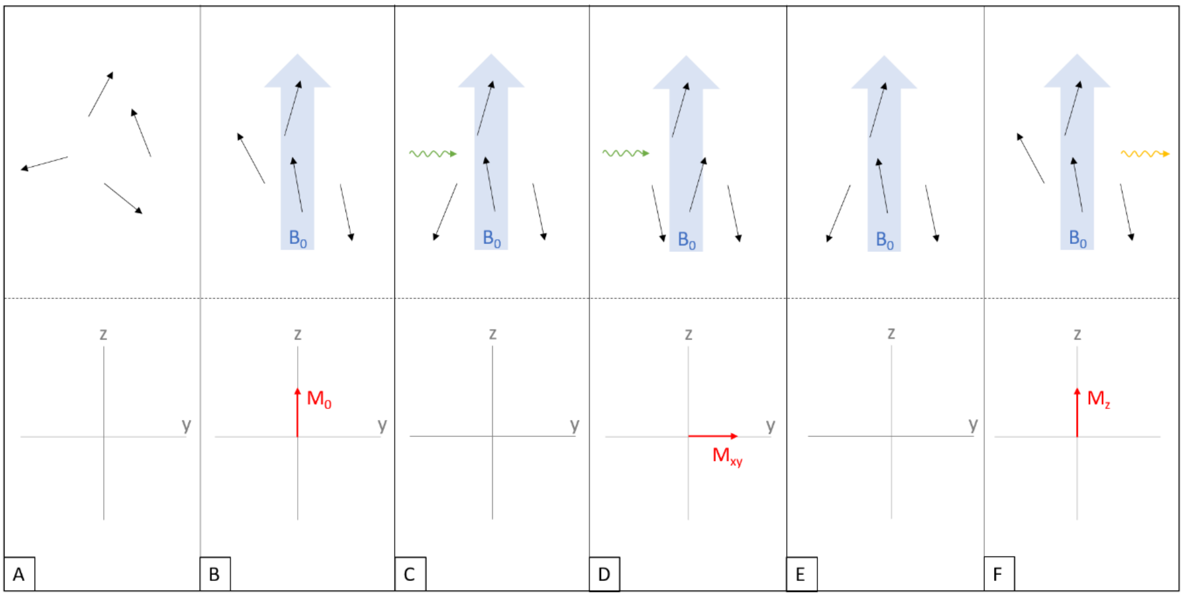

Figure 5: NMR process. First, spins, represented by black arrows, randomly rotate (A). When applying a B_0 magnetic field (blue arrow), the spins align either up or down, leading to a longitudinal magnetization M_0 (red arrow) appearance (B). The perturbation consists in applying a RF wave (green array) which first changes the energy state from low to high for some protons, implying the reduction of the magnetization along z (C). After a spins synchronization, a transversal magnetization M_{xy} appears in the xy plane (D). The last step, namely relaxation, is first composed of a disappearance of the transversal magnetization M_{xy} (E) and an appearance of the longitudinal magnetization M_z combined with an energy release (yellow array) as shown in F.

As a result, tissues are directly characterized by their relaxation times T1 and T2. Table 2 presents relaxation times for various tissues, as reported in the literature review recently performed by Zavala Bojorquez et al. (57).

Table 2: T1 and T2 relaxation times of various tissues categories in ms at 3T in case of healthy volunteers. For each tissue and time, the minimum and maximum values investigated by Zavala Bojorquez et al. (57) are reported.

Tissue	**T1**	**T2**
Fat	[253 +/- 42; 450 +/- 26]	[41; 371 +/- 8]
White matter	[728 +/- 433; 1735 +/- 205]	[65 +/- 6; 75 +/- 3]
Gray matter	[968 +/- 85; 1815 +/- 170]	[83 +/- 4; 109 +/- 11]

1.2.1.2.2. Image acquisition

To acquire the whole image and enable a signal localization, three different magnetic gradients are required.

In 2D, a first gradient along the z-axis has the role to select the slice to be imaged. Indeed, the RF pulse tips every proton whose Larmor frequency is equal to the emitted frequency wave. Thus, a gradient G_z along z is applied simultaneously to the RF wave, to ensure a linear B_0 variation as defined by:

$$B_0(z) = B_0 + G_z z \qquad \text{Equation 16}$$

Replacing in the Larmor frequency formula results in:

$$f(z) = f_0 + \frac{\gamma}{2\pi} G_z z \qquad \text{Equation 17}$$

With f_0 equal to $\frac{\gamma}{2\pi} B_0$.

Deriving according to z leads to:

$$\frac{\Delta f}{\Delta z} = \frac{\gamma}{2\pi} G_z \qquad \text{Equation 18}$$

It appears that the slice thickness Δz is selected through the frequency span Δf of the incoming RF pulse.

Second, a phase encoding gradient is applied, conventionally along the y-axis. During its application, protons frequencies increase or decrease, depending on their locations on the y-axis. When turned-off, protons return to their initial frequencies. However, dephasing occurs between two different given lines perpendicular to the gradient (along x for instance). This shift is proportional to the gradient strength and application duration. Thus, at the end of this step, protons from lines perpendicular to the gradient rotate at the same frequency, but with different phases.

Third, a frequency-encoding gradient is turned-on, usually along the x-axis, creating variations in frequencies along this direction. This step corresponds to the signal collection.

Steps 1, 2 and 3 are repeated, with increasing phase encoding time durations or strengths, until acquiring the signal for the whole selected slice.

In 3D, the whole acquisition process slightly varies. First, the gradient along the z-axis excites a whole volume representing a thick slice, instead of a thin slice. Second, an extra phase encoding gradient, with varying strengths or application times, is applied along the z-axis. Concerning phase encoding along the y direction and frequency encoding along x-axis, they remain unchanged. Owing to the addition of a new phase encoding gradient, large acquisition durations are observed, being equal to the Repetition Time (TR, section 1.2.1.2.4.) times the number of steps of phase encoding gradient 1 times the number of steps of phase encoding gradient 2 (58).

1.2.1.2.3. Image reconstruction

The collected signal S corresponds to the echoes obtained from different dephasing and frequency gradients, represented in the 2D k-space.

Its two axes are the spatial frequencies, k_x and k_y, respectively defined as:

$$k_x = \frac{\gamma}{2\pi} G_x t \qquad \text{Equation 19}$$

$$k_y = \frac{\gamma}{2\pi} G_y \tau \qquad \text{Equation 20}$$

With t, G_x, τ, G_y, the application time and strength of the signal collection gradient, the application time and strength of the dephasing gradient.

The k-space needs to be totally or partially filled in, to re-create a MRI slice. The k-space filling is a function of the MRI sequence. A first k-space completion, corresponding to gradient echo, is presented in Figure 6.A. The initial state is represented as the blue dot, at the centre of the space. Then, a dephasing gradient is applied combined with a negative read-out gradient, corresponding to the orange arrow. Finally, a positive read-out gradient is turned on, to collect signal (green arrow). A second k-space fill, obtained with a spin echo sequence, is presented in Figure 6.B. Dephasing and read-out gradients are first turned on (orange arrow). A 180° RF is then applied leading to the opposite vector direction. A read-out gradient is finally applied, and signal is collected (green arrow).

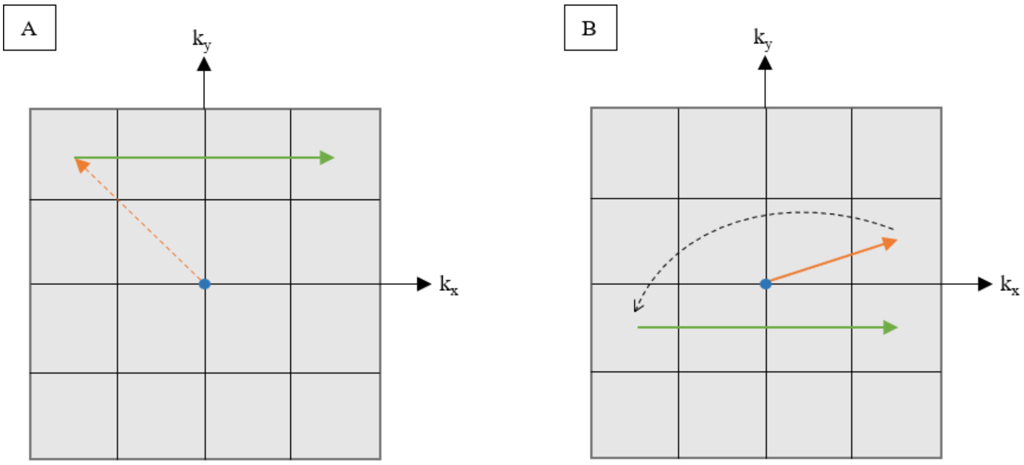

Figure 6: k-space completions with gradient echo (A, section 1.2.1.2.4.2.) and spin echo (B, section 1.2.1.2.4.1.) MRI sequences.

More generally, the low k-space centre frequencies are linked to the general shape of the head. Remaining high frequencies correspond to finer MRI details.

The collected signal is equal to:

$$S(t,\tau) = \iint M(x_k, y_l) e^{-i\gamma G_x t x} \, e^{-i\gamma G_y \tau y} dx\, dy \qquad \text{Equation 21}$$

With $M(x_k, y_l)$ the transversal magnetization of coordinates x_k and y_l in the image frame.

Thus, the transversal magnetization, directly representing the pixel-wise intensities, is obtained via an inverse Fourier transform of the collected signal.

1.2.1.2.4. Main MRI sequences

Two main sequences have been defined so far, namely gradient echo and spin echo. For each type, three key parameters are defined. First, the TR is the time between two RF pulses. The Echo Time (TE) represents the time between the application of the RF signal and the signal collection peak. Lastly, the flip angle represents the angle between the main magnetic field and the initial magnetization, before the application of the radiofrequency wave.

1.2.1.2.4.1. Spin echo

Figure 7.a presents a common spin echo pulse sequence (59). It is first composed of a radiofrequency excitation associated with a flip angle commonly equal to 90°. Simultaneously, the slice selection gradient is turned on. A negative re-phasing gradient is then applied, to offset the dephasing induced by the previous gradient. To induce dephasing between spins, a phase encoding gradient is turned on. Note that its strength is not constant during the MRI examination. This phase encoding gradient is combined with a positive read-out gradient to induce differences in spins frequencies. A 180° radiofrequency pulse is applied at TE/2 to reverse the spins. Thus, the smallest frequencies spins are actually the closest to the optimal common phase for signal collection, and reversely. An echo is finally obtained TE/2 later, and the read-out gradient is simultaneously turned on to acquire the signal.

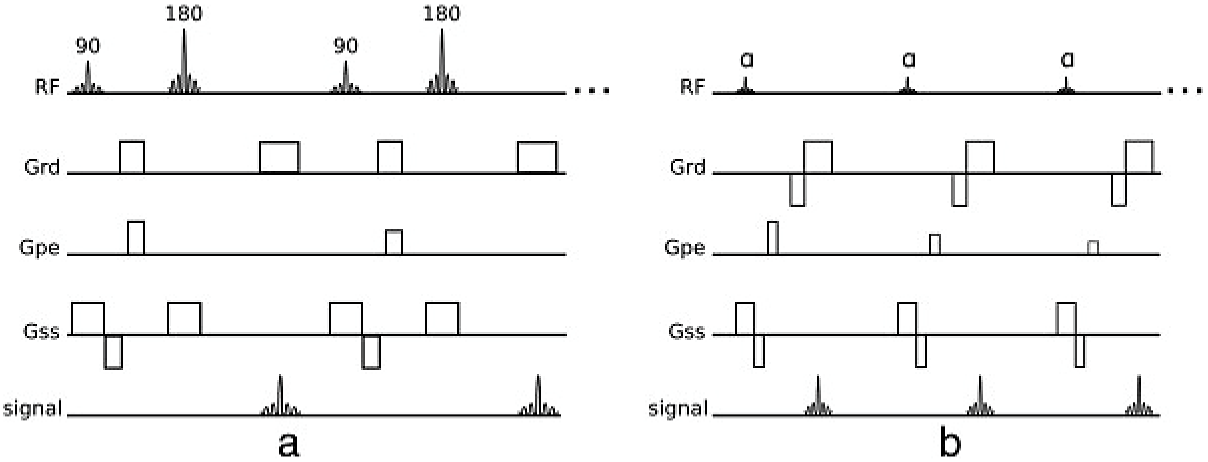

Figure 7: Spin echo (a) and gradient echo (b) pulse sequences. RF: radiofrequency pulse. Grd: read-out gradient. Gpe: phase-encoding gradient. Gss: slice selection gradient. Reprinted from "Cardiovascular magnetic resonance artefacts", Ferreira P, Gatehouse P, Firmin D, 2013, Journal of cardiovascular magnetic resonance: official journal of the Society for Cardiovascular Magnetic Resonance. CC BY 2.0 (2013).

One major advantage of spin echo sequences is the application of the refocusing 180° pulse, ensuring a pure T2 signal decay, without magnetic field imperfections or effects based on susceptibility differences considerations for instance (60).

1.2.1.2.4.2. Gradient echo

A standard gradient echo pulse sequence is presented in Figure 7.b (59). A radiofrequency pulse with a flip angle commonly inferior to 90° is first applied associated with a positive slice selection gradient. As in spin echo sequences, a re-phasing gradient is then turned on, before applying the phase encoding gradient, with a strength varying between each signal acquisition. Finally, a first read-out gradient is applied to pre-dephase the spins, resulting in spins dephasing speed up. A second read out gradient, equal to the first one but with an opposed polarity, is applied. It leads to an echo and the signal collection.

Regarding the optimal flip angle to apply, i.e. leading to the maximum signal for a given tissue characterized with a T1, Ernst et al. proposed to calculate it as (61):

$$\propto = \arccos\left(e^{-TR/T1}\right) \qquad \text{Equation 22}$$

Note that the Ernst angle optimizes the signal for a given tissue, but does not automatically provide the best contrast between two tissues.

Markl et al. investigated the key role of the flip angle in gradient echo sequences (60). As illustrated in Figure 8, intermediate flip angles ranging from 40° to 50° lead to T1 MRI (section 1.2.1.2.5.1.) with a strong signal. Larger flip angles, comprised between 60° and 80°, result in a high T1 contrast between tissues and a low SNR.

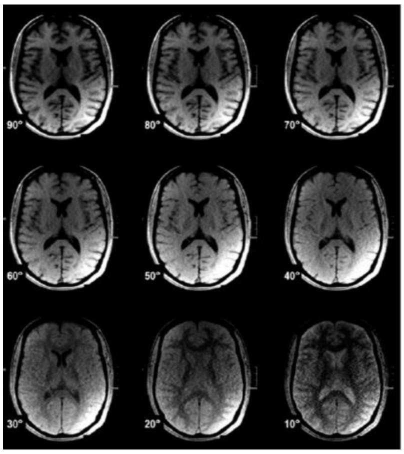

Figure 8: MRI contrast as a function of the flip angle. For every case, a gradient echo sequence was acquired with static parameters (TR = 500ms, TE = 4ms). Adapted by permission from John Wiley and Sons: Wiley, Journal of Magnetic Resonance Imaging, “Gradient echo imaging”, Markl M, Leupold J., Copyright (2012).

In gradient echo sequences, the apparent transversal relaxation time T2* is shorter than the T2 relaxation time occurring under perfect conditions, owing to the absence of refocusing pulses and the presence of main magnetic field inhomogeneities. According to Chavhan et al. (62), the latter can either be classified as macroscopic (e.g. metal implants, air/tissue interfaces) or microscopic (e.g. contrast product, iron accumulation). Tang et al. proposed to calculate the T2* as (63):

$$\frac{1}{T2*} = \frac{1}{T2} + \gamma \, \Delta B \qquad \text{Equation 23}$$

With γ and ΔB the magnetogyric ratio and the variations of the main magnetic field respectively. It leads to reduced spin-spin relaxations and potential signal loss.

1.2.1.2.5. MRI weighting

Three different types of weighted MRI can be acquired: T1, T2 or proton density images, but only T1 and T2 are detailed in this study, since they are the most common for the pCT synthesis task.

1.2.1.2.5.1. T1

T1 contrast aims to highlight differences between tissues T1 relaxations and relies on short TR and TE. Thus, tissues quickly recovering i.e. with a low T1, such as fat for instance, have a larger longitudinal magnetization than water before the application of the next RF pulse. It leads to a higher signal on the final image. In air, there is no protons, resulting in an absence of signal. Concerning short T2 tissues, such as for instance cortical bone, no signal can be acquired since the conventional MRI TE are longer than the T2, meaning that the signal has disappeared by the time of signal collection. Cakirer et al. investigated the brain diseases associated to hypersignal in T1 and classified them (64). A non-exhaustive list is presented below:

- Blood rupture vessel: primary tumors (e.g: pituitary adenoma, anaplastic astrocytoma, oligodendroglioma) or secondary tumors (e.g: from melanoma)
- Lesions composed of proteins (e.g: craniophryngioma)
- Lesions composed of fat (e.g: intraventricular lipoma, lipomatous meningioma)
- Calcified lesions (e.g: craniophryngioma, oligodendroglioma, choroid plexus papilloma, meningioma)
- Lesions composed of melanin (e.g: melanoma metastases)
- Other (e.g: Neurofribromatosis type I)

1.2.1.2.5.2. Enhanced T1 weighted MRI

MRI can be acquired with contrast agent to locally modify the tissue magnetic properties resulting in an improved lesion visualization. These agents are usually either superparamagnetic or paramagnetic (65).

The former is almost not used in clinics any longer, and is not detailed in this thesis.

Regarding the latter, it refers to chelates of ions with unpaired electrons, such as Fe^{3+}, Mn^{2+} or Gd^{3+}, which is actually the most widely used due to its stability. It has the role to reduce T1 and T2 relaxation times within the tissue they accumulate. However, in clinical imaging, T1 relaxation reductions are predominant. It results in a faster longitudinal magnetization relaxation. After applying the RF pulse, there is a larger amount of transverse magnetization in the xy plane, and thus a hypersignal, leading to a positive contrast on T1 images. The relaxation time decrease is defined by the formula (66):

$$\frac{1}{T_i} = \frac{1}{T_i^0} + r_i \times c \, ; i = 1 \; or \; 2 \qquad \text{Equation 24}$$

With T_i, T_i^0, r_i, c representing the tissue relaxation time after contrast agent administration, the initial tissue relaxation time, the contrast agent specific relaxivity and concentration. Typical r_1 and r_2 of Magnevist (Schering AG, Berlin, Germany), a contrast agent used for intra and extracranial lesions investigation, are respectively equal to $4.1 L.mmol^{-1}.s^{-1}$ and $4.6 L.mmol^{-1}.s^{-1}$ at 37°C and 1.5T (67).

Gadolinium belongs to the extracellular agents category. In case of Blood Brain Barrier (BBB) disruption, occurring when a tumor is present for instance, contrast product enters brain and accumulates in neurons and neuronal interstitium (68).

On an enhanced T1 weighted contrast MRI with gadolinium image (T1-Gd), normal, abnormal vessels are visible as well as BBB lesions.

1.2.1.2.5.3. FLAIR

Regarding T2, long TR and TE are applied. As a result, long T2 relaxation time tissues, such as cerebro-spinal fluid, have a stronger collected signal leading to bright contrast. On the other hand, quickly T2 recovering tissues, such as white matter, appear in dark. Clinically, T2 is used to visualize ventricular system, vasogenic edema and vessels.

T2 image contrast can be improved removing specific signal, such as water, corresponding to a T2 Fluid Attenuated Inversion Recovery MRI (FLAIR). It is composed of an inversion 180° RF pulse, leading to a longitudinal magnetization oriented along $-B_0$. Then, a long inversion time TI (e.g. TI ~ 2000ms) and corresponding to the delay for the water longitudinal magnetization to be equal to zero, is required. At this time, a 90° RF is applied, leading to images with suppressed signal for water. The T2 weighting is obtained applying long TR and TE. On a FLAIR image, cerebro-spinal fluid containing water appears in black, whereas the cortex is in light grey. FLAIR is clinically used to assess the vasogenic edema.

Figure 9 presents different MRI sequences, namely T1 (A), T1-Gd (B) and T2 FLAIR (C).

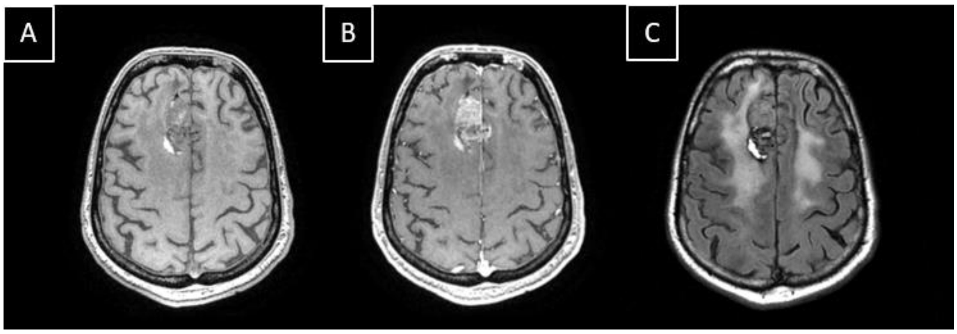

Figure 9: T1 (A), T1-Gd (B) and T2 FLAIR (C).

1.2.1.2.6. Artefacts in MRI

Various MRI artefacts have been described in the literature: truncation, motion, aliasing, chemical shift (69) and susceptibility-based (70).

First, truncation artefacts derive from the inverse Fourier transform used to reconstruct the final image. According to Fourier’s theorem, every continuous and periodic signal can be decomposed into an infinite sum of sinusoid signals. Yet, in MRI reconstruction, a finite sampling corresponding to discrete spatial frequencies, is used to approximate the final image (Figure 6), leading to k-space truncations at the boundaries. For small changing areas, no effect is visible. However, for high contrast interfaces, an ensemble of parallel lines appears (Figure 10.A). A method to reduce this effect consists in extrapolating the collected signal to introduce extra data and obtain smooth k-space boundaries (71).

Second, motion-related artefacts mostly appear as ghosting on the MRI, i.e. noisy translated replications of the original image (Figure 10.B). This artefact is due to the k-space periodic lines modifications, which are directly equal to the final number of ghosts (72). Since phase and frequency encodings have different duration times, several seconds and milliseconds respectively, this phenomenon is more prominent in the dephasing gradient direction. Motion artefacts can be overcome via the removal of specific moving tissue signals, such as fat for instance or more recently with DL network based on residual connections (73).

Aliasing occurs when the FOV is too small. Only the phase encoding direction can be affected, owing to the application of an analogical or numerical filter reducing aliasing in the frequency encoding direction. For instance, regarding phase encoding, extra body parts located on the right and left sides of the image also experience a dephasing gradient. For these regions, the phase shifts for the first encoding step are inferior to 0° and superior to 360° respectively, i.e. out of the range. The final affected shift would be equal to initial shift modulo 360. For instance, a 400° phase shift, located on the image left side, leads to a shift of 40°, placed on the image right side. It results in a superimposition of extra body parts on opposite image sides (Figure 10.C). Possible ways to reduce it are to enlarge the FOV, use specific software or apply a DL network (74).

Chemical shifts (Figure 10.D) are based on different chemical environment between air and fat protons. It results in slight differences between water and fat local magnetic fields and thus precession frequencies. This relative difference is equal to 3.5ppm, which was reported to be equal to 220Hz and 440Hz at 1.5T and 3T respectively (75). This leads to localization and/or amplitude artefacts. To reduce them, fat suppression sequences have been proved to overcome chemical shifts for cerebral lesions, especially for those located near optic nerves (76).

A material susceptibility refers to the material capability to be magnetized when a magnetic field is applied. If a weak magnetization occurs, the material is characterized as diamagnetic or paramagnetic. Most tissues are diamagnetic, such as air and fat compartments presenting low volume magnetic susceptibility being respectively equal to -9ppm and -7.8ppm in SI units. On the contrary, nickel, composing prostheses, exhibits a relative magnetic susceptibility superior to 10 000ppm. Thus, the material is highly magnetized when undergoing a magnetic field, and is classified as ferromagnetic. When two materials with different susceptibilities are close to each other, a magnetic field gradient is created and is superimposed on imaging gradients. It results in susceptibility artefacts, represented either as image distortions or blooming (Figure 10.E). Such artefacts can be reduced by increasing the frequency gradient strength or decreasing the dephasing time, i.e. the TE.
Figure 10 presents the previously described MRI artefacts.

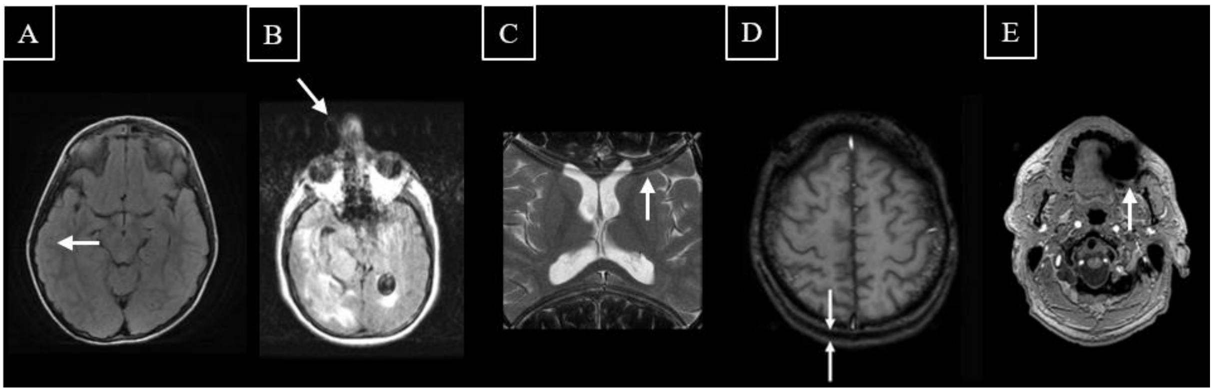

Figure 10: MRI artefacts, namely truncation (77) (A), motion (B), aliasing (78) (C), chemical shift (D) (79) and susceptibility magnetic-related (E). White arrows highlight some of the images distortions.

1.2.1.2.7. Standardizing MRI

MRI weighting, such as T1 or T2, depends on the site and the selected protocol (TR, TE and flip angle). In addition, applied scales during DICOM images records vary between MRI vendors. As a result, it appears necessary to standardize the MRI. Various methods have been introduced in the literature. A popular approach described by Nyul et al. (80) consists in a histogram-based standardization (HB). It relies on the matching of image percentiles to pre-defined template values derived from the MR images of the cohort, ideally corresponding to the training set. The intensity match is then obtained via a piece-wise linear transformation applied to the image intensities. A second approach is a normalization of the intensity distribution within the head (skull stripped or not) of each patient to zero mean and unit variance (ZMUV) (81). Lastly, White Stripe (WS) (82) is similar to the ZMUV approach, but based on the Normal Appearing White Matter (NAWM), as it is known to be homogeneous. Based on T1 images, the NAWM mean μ_{NAWM} is first selected on the histograms and corresponds to the highest peak (81). The associated standard deviation is defined by a stripe whose extremities are equal to +/- 5% of $\mu_{NAWM.}$ Recently, DL has also been proposed to standardize MRI across devices. For instance, DeepHarmony introduced in 2019 by Dewey et al. (83) relied on a 2.5D UNet for brain MRI contrast harmonization. Twelve subjects were enrolled, either classified as healthy (2 subjects) or presenting multiple sclerosis (10 subjects). Each subject was imaged with two different Philips Achieva (Philips Healthcare, Best, The Netherlands) and different protocols. Regarding the network, the selected loss function was the Mean Absolute Error (MAE, section 3.2.1.1.) and the optimizer was the Adaptive Momentum Estimation (Adam, section 2.3.4.2.). DeepHarmony presented volumes differences between devices 1 and 2 comprised between -1.89% +/- 19.53% (white matter lesions) to 0.35% +/- 0.88% (cerebral white matter), proving the high potential of such approach.

1.2.2. Images registration

Registration is a crucial step consisting in defining a function T to transform a source image onto a target image. Theoretically, registration algorithms seek for maximizing a similarity criterion or minimizing a distance metric between the transformed source image and the target image, being described as:

$$\arg\min_{T \in \tau} f\big(Y, T(X)\big) \qquad \text{Equation 25}$$

With Y the target image, X the source image, T the investigated transformation, τ the domain of all possible transformations, f the distance metric.

In head tumors clinical practices, the target and source images respectively refer to the CT and MRI.

Different types of transforms have been developed. First type of transform is rigid, and corresponds to six degrees of freedom, i.e. three representing translations and three for rotations, leading to distances conservation. The 3D transformation is presented below (84):

$$X' = T_{rigid}X \qquad \text{Equation 26}$$

Where X', X respectively represent the new and old image coordinates. T_{rigid} is defined as:

$$T_{rigid} = \begin{pmatrix} r_{11} & r_{12} & r_{13} & t^x \\ r_{21} & r_{22} & r_{23} & t^y \\ r_{31} & r_{32} & r_{33} & t^z \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

Whith t^x, t^y, t^z respectively refering to the translations along x, y and z. Regarding the rotation matrix, each element is defined as:

$$r_{il} = r_{ij}^{x} r_{jk}^{y} r_{kl}^{z}$$

With axis-specific rotations matrices, calculated as:

$$r^x = \begin{pmatrix} 1 & 0 & 0 \\ 0 & \cos\alpha^x & -\sin\alpha^x \\ 0 & \sin\alpha^x & \cos\alpha^x \end{pmatrix}; r^y = \begin{pmatrix} \cos\alpha^y & 0 & \sin\alpha^y \\ 0 & 1 & 0 \\ -\sin\alpha^y & 0 & \cos\alpha^y \end{pmatrix}; r^z = \begin{pmatrix} \cos\alpha^z & -\sin\alpha^z & 0 \\ \sin\alpha^z & \cos\alpha^z & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

With α^i, r^i the rotation angle around axis i and the rotation matrix around axis I respectively.

No significant changes occur for head radiotherapy patients due to the skull rigidity, except when comparing before/after surgery images. Thus, this type of transform is sufficient for intra-patient images registration.

Second type is non-rigid deformations. They can be split into parametric, i.e. the transformation can be parametrically represented, or non-parametric, typically represented by voxel-wise deformation fields. Regarding the parametric approach, affine registrations refer to rigid registration combined with correction factors. They correspond to scaling and shear, both adding three degrees of freedom to the transform T. Shear factors are particularly useful when dealing with CT since tilted gantries result in such distortion (85). Rueckert et al. (86) and Penny et al. (87) precisely described the equation:

$$T_{affine} = T_{shear} T_{scaling} T_{rigid} \qquad \text{Equation 27}$$

With T_{shear}, $T_{scaling}$, T_{rigid}, the shearing, scaling and rigid transformations respectively.

$$with\ T_{shear} = \begin{pmatrix} 1 & h_1 & h_2 & 0 \\ 0 & 1 & h_3 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} and\ T_{scaling} = \begin{pmatrix} q_1 & 0 & 0 & 0 \\ 0 & q_2 & 0 & 0 \\ 0 & 0 & q_3 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

Another parametric approach is the spline-based registration and belongs to the free form deformation category. It relies on the definition of matching points in the two involved images. Regarding non parametric methods, diffeomorphic registrations, based on bijective differentiable transforms whose inverses are also differentiable, have been widely used to deform brain MRI (88,89).

1.2.3. Segmentation

1.2.3.1. Basics

Second step of a radiotherapy treatment is the segmentation, i.e. locating and individually contouring target and Organs At Risk (OAR) volumes.

Regarding target volumes, they were defined by the International Commission of Radiation Units and Protection (ICRU), Reports 50, 62 and 83 (90–92), and were accurately detailed by Chavaudra et al. (93). The Gross Tumor Volume (GTV) corresponds to the macroscopic volume, visible on imaging such as MRI. If a surgery is required, it corresponds to the volume which is resected. The second volume, namely Clinical Target Volume (CTV), refers to the microscopic extension of the GTV. It is calculated considering the past disease history, knowledge about general tumor growth and predicted progression. Lastly, margins are added to the CTV to create the PTV, considering radiotherapy process geometric uncertainties related to the non-reproducible set up, beam delivery fluctuations, registration errors, etc. The margins extension from CTV to PTV may not be isotropic, owing to OAR proximities for instance. Lastly, note that for some tumors locations such as lungs, an intermediate volume corresponding to Internal Target Volume (ITV) is created to account for internal CTV movements, related to breathing, bladder filling variations, etc. This volume is comprised between CTV and PTV. Figure 11 illustrates the different volumes used in radiotherapy (94).

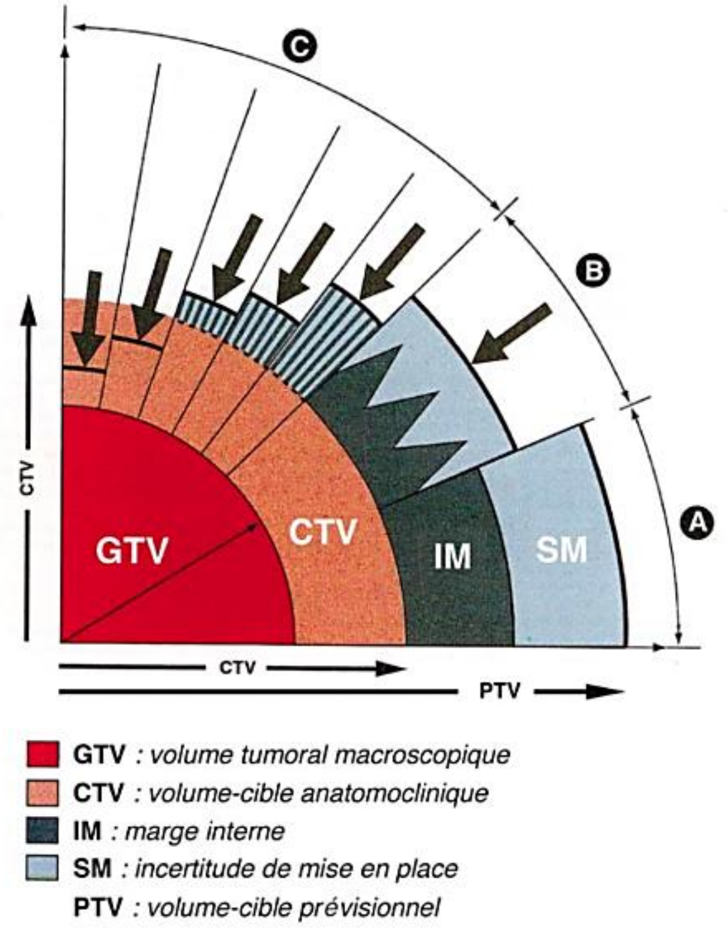

Figure 11: GTV (red), CTV (red+orange), ITV (red+orange+dark green) and PTV (red+orange+dark green+light blue) constituting the different target volumes. Three scenarios are illustrated. Scenario A involves GTV, CTV, ITV and PTV margins. Scenario B illustrates a reduced PTV to avoid healthy tissue irradiation. Scenario C presents a higher probability assigned to uncertainties risks linked to CTV and PTV margins. Reproduced from Kantor G, Maingon P, Mornex F, Mazeron JJ. Contours des volumes cibles en radiothérapie. Généralités. Cancer/Radiothérapie 2002; 6:56-60. Copyright © 2002 Elsevier Masson SAS. All rights reserved.

Delineation guidelines depend on the tumor type, age, tumor location, etc, and are commonly performed on MRI for target volumes. For instance, for glioblastomas, ESTRO-ACROP recommends to consider, for GTV, the resection cavity and hypersignal areas on T1-Gd (95). CTV are derived applying 20mm of margins, in high tumor spread probability directions. However, in some cases such as bone, this margin can be reduced to 0mm. Regarding peadiatrics receiving cerebro-spinal irradiation, SIOPE guidelines should be followed (96). They recommend considering the frontal lobe combined with cribriform plate for the $\text{CTV}_{\text{cranial}}$ delineation. A second CTV, corresponding to spine and including the intervertebral foramina, is also defined. Regarding PTV calculations, $\text{CTV}_{\text{cranial}}$ and $\text{CTV}_{\text{spinal}}$ are usually extended by margins respectively ranging from 3mm to 5mm and from 5mm to 8mm.

Concerning brain tumors OAR, CT is used to segment optic nerves, cochlea and lenses are usually contoured. Chiasma and brainstem are assessed with MRI.

Figure 12 presents an example of software clinically used for delineation (RayStation, RaySearch Laboratories, Stockholm, Sweden).

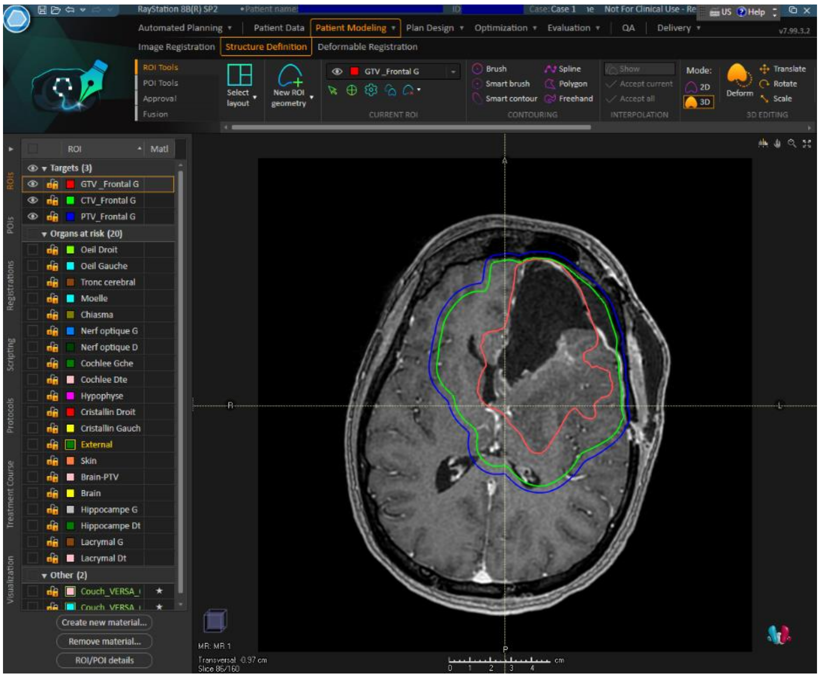

Figure 12: Segmentation panel from RayStation version 8B(R) (RaySearch Laboratories, Stockholm, Sweden). The red, green, blue colours respectively represent the GTV, CTV and PTV of a glioblastoma, assessed on the T1-Gd.

Common clinical practises include the delineation of external contour via automatic methods, based on foreground/background thresholding. For head tumors, the remaining contoured volumes are either manually or semi-automatically (e.g. smart brush) segmented. Thus, segmentation is a time-consuming task, raising the issue of inter-operators reproducibility.

Recently, new approaches have been developed to automatically and rapidly generate contours based on DL (97–100).

1.2.3.2. Evaluation metrics

Volumetric *Dice*-Sørensen Coefficient (VDSC) is commonly assessed. It is calculated as:

$$VDSC = 2 \times \frac{|A \cap B|}{|A| + |B|} \qquad \text{Equation 28}$$

With A and B are the two structures to be compared, |X| represents the number of elements in X.

VDSC directly reflects the overlap between the structures A and B. It ranges from 0, i.e. the two contours are completely different, to 1 implying a perfect contours match.

In addition, several metrics, relying on the comparison of tested and gold standard contours, have been introduced. These concepts evaluate a segmentation-based discrimination between two distributions, via the sensitivity, specificity, Positive Predictive Value (PPV) and Negative Predictive Value (NPV).

Sensitivity reflects the probability for a given voxel to belong to the evaluated segmentation when it belongs to the gold standard contour. It is defined as:

$$Sensitivity = \frac{True\ positive}{True\ positive + False\ negative} \qquad \text{Equation 29}$$

Specificity is a voxel probability not to belong to the evaluated contour when it does not belong to the gold standard one. It is calculated with:

$$Specificity = \frac{True\ negative}{True\ negative + False\ positive} \qquad \text{Equation 30}$$

PPV refers to the probability for a voxel to belong to the gold standard contour when it belongs to the evaluated contour.

The formula is as follow:

$$PPV = \frac{True\ positive}{True\ positive + False\ positive} \quad \text{Equation 31}$$

Lastly, NPV describes the probability for a voxel not to belong to the gold standard contour when it does not belong to the evaluated contour and is assessed with:

$$NPV = \frac{True\ negative}{True\ negative + False\ negative} \quad \text{Equation 32}$$

1.2.4. Dosimetry

1.2.4.1. Basics

Third step is the dosimetry and consists in calculating the dose and optimizing it to deliver the prescribed energy to the target volume while sparing OAR. It is performed on the CT, since HU are directly linked to the ED via the calibration curve.

Concerning dose constraints, ICRU report 62 recommends the dose to the PTV to range from 95% to 107% of the prescribed dose for 3D Conformal Radiation Therapy (3DCRT, section 1.2.4.5.1.) (101). In case of Intensity Modulated Radiation Therapy (IMRT, section 1.2.4.5.2.) for instance, ICRU Report 83 recommends maintaining a maximum absorbed dose of 5% difference for 85% of the target volume in low gradients areas, i.e. regions with dose changes inferior to 20% per cm (92). OAR dose constraints, in 2Gy equivalent fractionation, are provided in Table 3.

Table 3: OAR dose constraints in 2Gy equivalent fractionation.

Structures	Dose constraints
Optic nerves	$D_{max} < 54$ Gy
Chiasma	$D_{max} < 54$ Gy
Cochlea	$D_{mean} < 45$ Gy
Brainstem	$D_{max} < 54$ Gy $D_{max} < 60$ Gy* $D_{59Gy} < 10$ cc*
Lenses	$D_{max} < 6$ Gy

*: Applicable for tumors located near brainstem or invaded brainstem.

1.2.4.2. Dose calculation algorithms

1.2.4.2.1. Monte Carlo (MC)

Monte Carlo (MC) is a highly precise statistical method, which consists in probabilistic simulations, for a finite number of particles, of total tracks through per-defined media. Discrete photon/matter and electron /matter interactions potentially combined with dose depositions are randomly reproduced based on their cross sections. As a result, the particle trajectory is computed step by step, until reaching a residual energy.

MC approach mainly relies on physics modelling, cross sections tables, random numbers generation, and final statistics scoring (102).

Numerous physics models are available, with varying application domains. One of the most popular model is Livermore. Its data are extracted from three different databases: Evaluated Atomic Data Library, Evaluated Electrons Data Library and Evaluated Photons Data Library. It has been designed for particles energies ranging from 250eV to 1GeV and models electrons and gamma only. No representation of the positron is provided, representing a limitation for pair production interaction. Atomic numbers from 1 to 100 can be considered. Lastly, the representation of fluorescence and Auger mechanisms occurring in most of atom shells is available (103). Note that this model template is changeable, as well as other models, turning on/off mechanisms or modifying cut off energies for instance.

Interactions tables store information for photon and electron individual interactions. Regarding photon, specific cross sections for each interaction are provided, as well as the total cross section σ_{tot} calculated via (102):

$$\sigma_{tot} = \tau + \sigma + \varsigma + \kappa_n + \kappa_a + \sigma_{photo\,nucl} \qquad \text{Equation 33}$$

With τ, σ, ς, κ_n, κ_a, $\sigma_{photo\,nucl}$ the cross sections of photoelectric effect, Compton effect, Rayleigh scattering, pair production in the nucleus and atoms fields, photonuclear, all depending on parameters such as the photon energy or the atomic number. Concerning electrons interactions, collisions can occur, either classified as soft or hard. The former corresponds to excitations or outer shell ionizations. The latter refers to inner shells ionizations. Interactions between the incoming electron and the nucleus are known as Bremsstrahlung, and correspond to the emission of a X-ray combined with an incident electron trajectory deviation. Associated stopping powers are required for dose calculation, as well as extra information such as fluorescence and Auger shell-wise cross sections (102).

Note that in practice, and to ensure repeatability, pseudorandom numbers are generated, meaning relaunching the experience with the same initial seed N_0, leads to the same random numbers placed in the same order (104).

Final statistics refer to the dose computation, from each individual history. Since MC is based on a stochastic approach, large number of incident particles should be simulated to decrease errors, leading to long computation times. As a result, it is often used in second step, for a last accurate dose calculation.

MC has the advantage to be highly realistic and to offer a satisfying secondary particles transport. In addition, dose calculation in case of heterogeneities or small beams are available. However, a detailed geometry knowledge is required for unbiases dose calculations. In addition, achieving accurate results is time-consuming, due to the simulation of high particles number. A last limitation relies on statistics, since an uncertainty is associated to each computed result.

A possible approach to reduce it, consists in increasing the incident particles. However, it also results in a computation time increase. Variance reduction techniques have been introduced to achieve a limited uncertainty in delays in agreement with a clinical use. The Russian Roulette is a method to combine particles resulting in timesaving. This technique is often used for electrons in linac head, which are not of major interest for the ROI dose deposition. The particle has a probability p to survive to the Russian Roulette. If so, its weight is increased by a multiplicative factor equal to 1/p. The probability for the particle to be removed is equal to 1-p (105).

This method is often combined with an additional highly effective variance reduction technique, namely Directional Bremsstrahlung photon splitting, first introduced by Kawrakow et al. (106). It relies on the fact that tracking photons outside the ROI is time-consuming and useless. On the contrary, photons within the ROI should be as numerous as possible, to reduce uncertainty (107). The directional Bremsstrahlung photon splitting consists in splitting every resulting X-ray photon resulting from Bremsstrahlung interactions, into N photons assigned with a weight equal to 1/N. For each sub-photon, its direction is analysed. If it is towards the ROI, the photon is kept, and named as “non-fat photon”. Otherwise, the Russian roulette is applied. First, a random number is selected. If it is superior to a given threshold T, the photon is removed. If it is inferior to T, the photon is kept with a new weight multiplied by 1/T (108). To implement this approach, three parameters are required namely N, the ROI radius and the source to surface distance.

1.2.4.2.2. Convolution/superposition algorithms

1.2.4.2.3. General approach

This approach relies on separation between primary particles, i.e. photons, and secondary, namely photons and electrons.

First important concept is TERMA T and refers to the total energy released in the matter. In dose algorithms, T concerns primary photons only, and is proportional to the primary fluence. In a given point placed in $\vec{r}'$, it is calculated as:

$$T(\vec{r}', E) = \frac{\mu}{\rho}\, \Psi(\vec{r}', E) \qquad \text{Equation 34}$$

With μ, ρ, Ψ and E the total linear attenuation coefficient, the density, the primary energy fluence, the beam energy.

Second, a point kernel, referring to secondary particles, represents the deposited energy in $\vec{r}$ from a thin beam of mono energy E whose primary photon interacts in $\vec{r}'$, denoted as K($\vec{r}$, $\vec{r}'$, E) (109). They are commonly pre-calculated using MC. The final dose in $\vec{r}$ is obtained superposing, i.e. integrating, contributions from all $\vec{r}'$ positions and all energies beam, as shown below (109):

$$D(\vec{r}) = \int dE' \int d^3r'\, T(\vec{r}', E')\, K(\vec{r}, \vec{r}', E') \qquad \text{Equation 35}$$

When hypothesizing the medium homogeneity, the above expression can be described as the following convolution (109):

$$D(\vec{r}) = \int dE' \int d^3r'\, T(\vec{r}', E')\, K(|\vec{r} - \vec{r}'|, E') \qquad \text{Equation 36}$$

Figure 13 presents an example of dose calculation using the convolution approach (110).

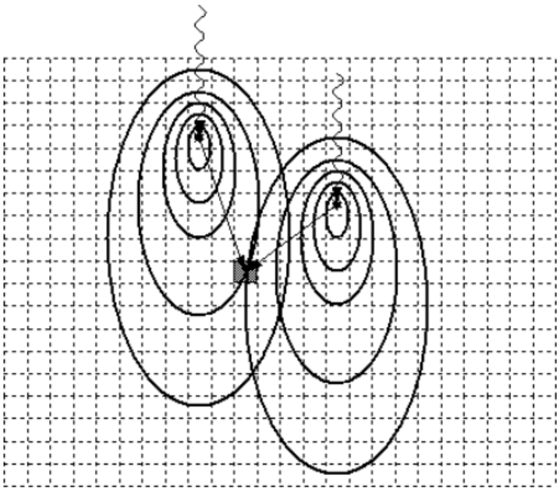

Figure 13: Absorbed dose in $\vec{r}$ (grey square) results from the convolution of the TERMA T in $\vec{r}'$ (black dots) and the corresponding kernel value in $\vec{r}$. The process is repeated for every interaction point $\vec{r}'$. Reprinted from "3D dose computation algorithms", Knöös T, 2017, J Phys Conf Ser. CC BY 3.0 (2017).

This approach has the advantage de consider electrons transport and is well-suited for modulated treatment deliveries. However, this convolution approach is relatively time-consuming. Indeed, it has been shown that for a phantom containing 10^6 voxels, corresponding to a size of $30 \times 30 \times 30\text{cm}^3$, combined with a resolution of 3mm, 10^{12} calculations were needed, leading to duration times of several minutes (111). It enhanced the need for approximations to simplify convolution/superposition equations and enable a use into clinics.

1.2.4.2.4. Collapsed Cone (CC)

The Collapsed Cone (CC) was first introduced in 1989 by Ahnesjo (112). It consists in considering point kernel distribution $K(\vec{r}, \vec{r}', E)$ as a finite number of cones, centred on $\vec{r}'$. For a given cone, the energy transportation occurs in the central axe, and is thus "collapsed". Figure 14 illustrates the process (113).

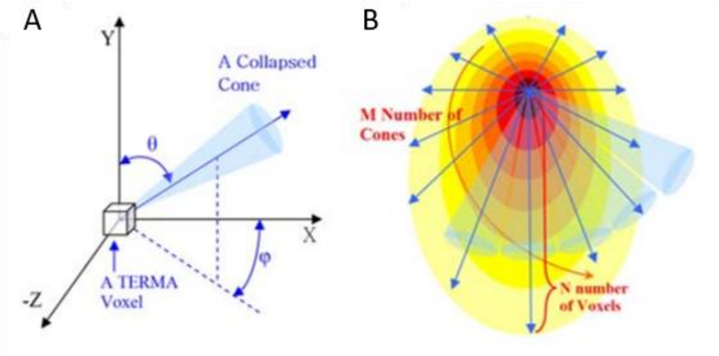

Figure 14: An example of cone (light blue) for a given interaction voxel (black square). It is defined via spherical coordinates. Energy collapsed along the cone central axis (blue arrow) (A). Global view of the discretized kernel (B). Adapted by permission from Springer Nature Customer Service Centre GmbH: Springer, J Korean Phys Soc, "Practical implementation of a collapsed cone convolution algorithm for a radiation treatment planning system", Cho W, Suh T-S, Park J-H, Xing L, Lee J-W, Copyright, 2012.

Required calculations for the dose assessment are now equal to $M \times N^3$, where M is the number of cones and N is the cube calculation size (114), resulting in faster dose calculations than convolution/superposition method. In addition, CC offers an approximate electrons transport and secondary photons transport in 3D heterogeneous media, as properly reported by Fogliata et al. (115). They enrolled 20 patients with left breast tumors. Different lungs air filling was used. The prescribed dose was equal to 50Gy, delivered in 25 fractions. Half of the patients were planned with two conventional tangential beams, while the remaining 10 patients were treated with a 3-field technique. Regarding free breathing, mean lung doses were respectively equal to 8.4Gy +/- 1.8Gy and 8.9Gy +/- 1.8Gy for CC and MC, the latter representing the ground truth. Thus, high agreement was proved between CC and MC approaches.

1.2.4.2.5. Pencil Beam

A pencil beam corresponds to the pre-integration of all the kernels along a narrow beam in the depth direction. As a result, dose of a given point located in $\vec{r}$, is considered to derive only from interactions occurring along this beam. To simulate larger beams, the pre-integration is duplicated, leading to a series of pre-integrated kernels. Such pencil beam is illustrated in Figure 15 (116).

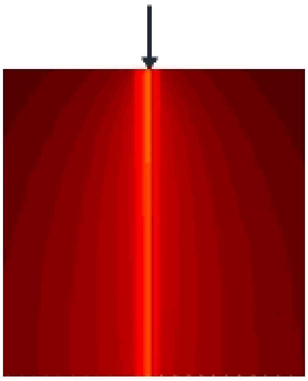

Figure 15: Pencil beam, with a primary photon reaching the studied volume (black arrow). Adapted from “What should we know about photon dose calculation algorithms used for radiotherapy? Their impact on dose distribution and medical decisions based on TCP/NTCP”, Chaikh A, Khamphan C, Kumar T, Garcia R, Balosso J, 2016, Int J Cancer Ther Oncol., 4(4). CC BY 3.0 (2016).

The dose assessment is then performed convoluting previously described pencil beam with the energy fluence, as described (110):

$$D(x,y,z) = \int\int \Psi(x',y') \cdot P(x-x',y-y',z)dxdy \qquad \text{Equation 37}$$

With Ψ, P, the energy fluence and the pencil beam with kernels integrated along z.

In case of heterogeneities, no modulation of the energy fluence maps is performed. Regarding kernel adaption to densities discrepancies, radiological scaling has been developed. It consists in basing the dose calculation on the radiological depth Z_{rad}, instead of the standard depth, which is defined as (117):

$$Z_{rad} = \int_0^Z \frac{\rho(t)}{\rho_{water}} dt \qquad \text{Equation 38}$$

With ρ, ρ_{water} the tissue and water ED, respectively.

However, this scaling has the disadvantage to consider only ED variations along the integration direction (in this thesis z), and not those from surrounding tissues, potentially leading to biases (118).

1.2.4.3. Optimization

Two approaches are proposed by the current Treatment Planning System (TPS) software. First, direct planning implies a definition of numerous parameters values, such as the number of beams, the energy, the collimators, to enable the dose calculation. The plan is finally either validated or modified to agree with the official guidelines. This approach is used for 3DCRT (section 1.2.4.5.1.).

Second, inverse planning has been proposed for delivery techniques such as IMRT. A large number of degrees of freedom are introduced from the decomposition of treatment fields into multiple individual unit field with a specific weight (119). Thus, inverse planning role is to enable a robust and easier dosimetry task. This approach mainly relies on the specification of dose constraints on both target and OAR volumes by the user and the parameters optimization to satisfy these criteria. Automatic optimization of technical parameters is performed either with physical cost functions, such as DVH, or biological cost functions, including Poisson statistics cell kill model (120). The latter offers the possibility to estimate the impact of radiation onto tissues, and thus being more realistic.

Different approaches have been developed for the objective function minimization, such as Simulated Annealing (SA) which is well-adapted for large dimensional data problems. It is inspired by solid-state physics theory (121). The object is first heated to a given temperature T, before stopping and observing the temperature t decrease (122). For each t<T, the energy is computed, and the atoms configuration is either accepted or rejected. In IMRT, SA iteratively generates a variation in the beam weights. If the cost function decreases, the variation is kept. If the cost function increases, the variation is also accepted with a given probability. Owing to its stochastic approach, SA is less prone to fall in local minimum, contrary to classical gradient descent. This approach has recently been applied to beam angles optimization by Dias et al. (123). Ten head and neck patients were included. They were beforehand characterized as critical cases, partly owing to the balance between satisfying PTV coverage and parotid glands sparing. The treatment beam was delivered at 5 different angles. Ground truth was calculated considering uniformly distributed beams. Concerning PTV1 receiving 70Gy, minimum D95% were equal to 65.4Gy and 65.2Gy for the ground truth and best SA model respectively. Regarding left parotid, Dmean reductions from ground truth model to best SA model were comprised between 0.1Gy and 3.2Gy, except for 3 patients. It potentially suggested that SA improved OAR sparing, while keeping unchanged target-related metrics.

1.2.4.4. Plan quality evaluation metrics

Several criteria can be used to evaluate the plan quality.

Dose Volume Histograms (DVH) are commonly computed. They present, for a given studied volume, the volume against the absorbed dose. Two types of DVH are currently available:

- Differential: The dose for a given volume is the dose received by this volume (Figure 16.A).
- Cumulative: The dose for a given volume is the minimal dose received by this volume (Figure 16.B). This DVH type is more commonly used in clinics.

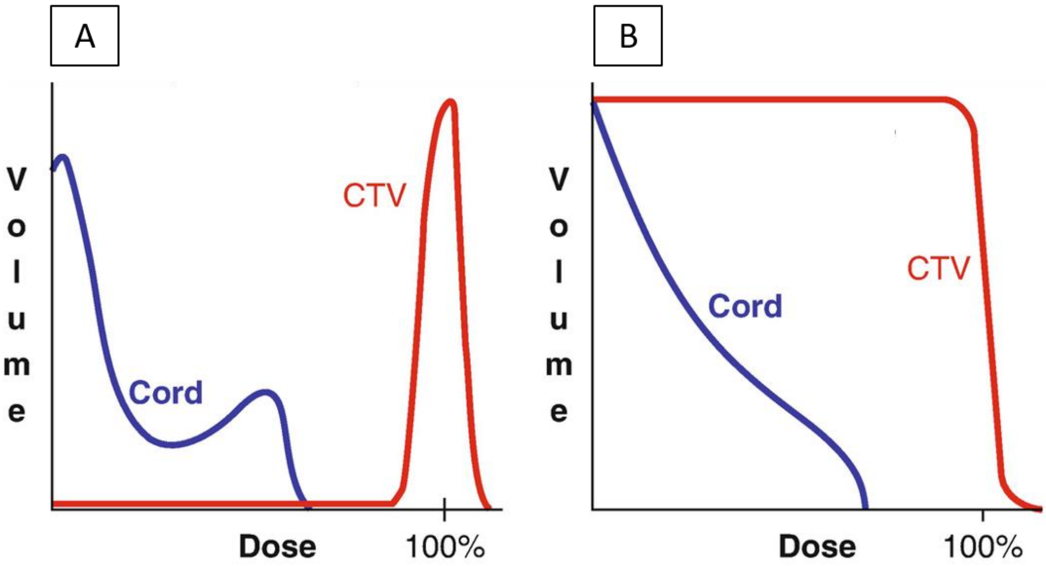

Figure 16: Differential (A) and cumulative (B) DVH. The red and blue lines respectively represent the CTV, i.e. a target volume, and the cord, an OAR. Adapted by permission from Springer Nature Customer Service Centre GmbH: Springer, Practical Radiation Oncology, "Plan Evaluation in 3D Conformal Radiotherapy", Pandit S., Copyright (2020).

Concerning cumulative DVH, $D_{X\%}$ represents the dose received by X% of the given structure.

As one can notice, no spatial information is provided constituting a non-negligible limitation to this metric.

Following ICRU 83 guidelines (92), three additional quality metrics, corresponding to conformity indexes, can be computed. First, the quality of coverage is calculated as follows:

$$Coverage = \frac{I_{min}}{RI} \qquad \text{Equation 39}$$

With I_{min} and RI representing the minimal isodose around the target and the reference isodose.

Coverage ideal value is equal to 1. An index inferior to 1 indicates a smaller volume than the PTV is irradiated. Values greater than 1 signify the irradiation of a larger volume than the PTV (124).

The tumor homogeneity index HI reflects the PTV dose distribution homogeneity. It is assessed as (125):

$$HI = \frac{D_{2\%} - D_{98\%}}{D_{50\%}} \qquad \text{Equation 40}$$

Acceptable HI are inferior or equal to 2. HI ranging from 2 to 2.5 and superior to 2.5 respectively represent small and important protocol transgressions (125,126).

For the specific case of stereotactic treatments, the conformity index refers to the amount of non PTV volume receiving radiations. It is computed as follow (125):

$$Conformity\ index = \frac{V_{RI}}{V_t} \qquad \text{Equation 41}$$

With V_{RI}, V_t the reference isodose volume and the target volume.

1.2.4.5. Delivery techniques

1.2.4.5.1. 3DCRT

3DCRT is based on a small number of fixed beams which deliver 3D uniform dose distributions adapted to the tumor shape. Beams parameters are optimized via direct planning. If a MultiLeaf Collimator (MLC) is present, it remains static during the irradiation.

3DCRT has been proved to outperform 2D, by Trignani et al. (127). More specifically, they investigated the parotid glands avoidance, in case of whole brain radiotherapy. The cohort was composed of seven patients with brain metastases, who received 30Gy in ten fractions. 2DCRT and 3DCRT techniques were used, with two lateral beams. For six patients out of seven, the average CTV minimum dose was equal to 49.47%, being highly inferior to 95%, and leading to a smaller target volume coverage than expected with 2DCRT planning.

1.2.4.5.2. IMRT

IMRT relies on a heterogeneous dose distribution based on the decomposition of each beam, into beamlets, whose weights are adjusted via inverse planning. This non-uniform intensities delivery is enabled via the MLC. The optimization is performed via inverse planning.

Two approaches exist. First, sliding window (also known as dynamic IMRT) consists in dynamically moving the MLC while delivering the beam. Its current clinical implementation is based on Convery et al. (128). For simplicity, two leaves, A and B, are considered, as well as a 2D intensity profile. The leaves movement direction is from left to right. Let t_{xB} represent the beginning of a point x irradiation, and t_{xA} the end. It has been shown that any point x fluence was proportional to t_{xB} - t_{xA} (129). Thus, reaching a given intensity is achievable manipulating leaves latencies. Figure 17 illustrates such a process (130).

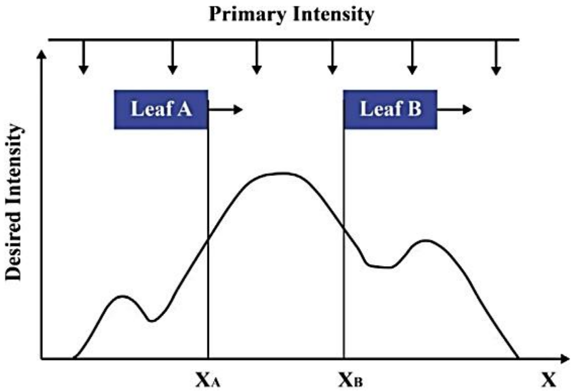

Figure 17: Moving leaves A and B along a patient axis x. Reprinted from “Intensity modulated radiation therapy: A review of current practice and future outlooks”, Rehman J ur, Zahra, Ahmad N, Khalid M, Noor ul Huda Khan Asghar HM, Gilani ZA, et al., 2018, J Radiat Res Appl Sci., 11(4). CC BY-NC-ND 4.0 (2018).

The second approach is known as step-and-shoot and corresponds to the sequential decomposition of the previous sliding window approach. Treatment beams are commonly decomposed into 2 to 20 segments (131). For each segment, the leaves are placed while the beam is off. It is then turned on, while the leaves remain static.

To compare IMRT approaches, Iqbal et al. (132) constituted a cohort composed of 13 prostate patients. Sliding window and step-and-shoot plans were generated for every patient, for 5, 7, 9 and 13 IMRT fields. All prescribed doses were equal to 50Gy. Means PTV coverages ranged from 0.95 to 0.96, with non-significant differences. Concerning bladder, maximum mean doses obtained for sliding window and step-and-shoot IMRT respectively were equal to 38.11% +/- 2.6% and 37.0% +/- 2.3% respectively. Regarding rectum, these values were equal to 42.92% +/- 10.7% and 42.44% +/- 9.1%. Homogeneity indices were equal to 0.12 +/- 0.02 and 0.13 +/- 0.02 for sliding window and step-and-shoot approaches. As a result, target performances were similar for both IMRT techniques. However, concerning OAR, lower doses and thus higher safety was observed for step-and-shoot technique.

The last step before transferring plan to the treatment device and irradiating the patient corresponds to the fluences sequencing. It aims at assessing the mechanically feasible fluence, from the ideal fluence determined during optimization. It is mostly performed with the MLC whose leaves positions must be optimized based on the initial fluence map, to reach the lowest treatment time. To do so, two types of constraints must be considered, namely dosimetry-based (e.g. transmission through the leaves and their extremities) and mechanics-based (e.g. maximum leaves speed, possible extreme positions), as reported by Marchesi (133). A possible evolution to face the high number of local minima deriving from the MLC modelling problem was proposed by Kelly et al. (134) in 2019. First, the leaves trajectories over time were represented with splines. Second, the leaf blocking irradiation function was not a sharp step, but a smooth function. More recently, Medeiros et al. (135) proposed a theoretical approach of the leaf sequencing problem consisting in the decomposition of an incoming dose matrix into a weighted sum of binary matrices, corresponding to the MLC positions leaves. Regarding the weights, they referred to the monitor units. In case of step-and-shoot IMRT, they proposed an optimization algorithm based on the reduction of three key parameters, namely the numbers of monitor units, MLC configurations (i.e. segment) and leaves travelling distances. It was first based on randomized solutions generation, followed by re-arrangement of segments, before optimizing the distance-related objective function. To evaluate the tool performance, 15x15 random virtual dose matrices were generated for decomposition. Values ranged from 0 to L, with L the maximum matrix comprised between 3 and 16. For each L, 1000 matrices were simulated. For L equal to 7 for instance, means of the numbers of segments and total distance travelled by the leaves were respectively equal to 12.69 and 91.262cm. For comparison, an additional algorithm proposed by Lust et al. (136) respectively resulted in rates equal to 13.08 and 100.54cm, proving the superiority of the approach proposed by Medeiros et al. (135).

1.2.4.5.3. Volumetric Modulated Arc Therapy (VMAT)

VMAT is one of the latest technologies which offers the highest degrees of freedom, via a continuous treatment delivery with one, two or three arcs generally. It is characterized by three main beam-on varying parameters, namely the MLC, the linac gantry speed and the dose rate (137). It enables a high number of entry points, reducing the dose to the OAR.

A major point concerns the gantry sampling and the corresponding MLC movements and positions, which has been extensively described by Otto et al (138). Suppose a single VMAT arc. Initially, low number of source positions are defined, namely a, b, c, d, e, f (Figure 18.A). After several optimization algorithm iterations, a first extra sample is equally distance-placed between a and b samples (Figure 18.B). The novel MLC position is obtained linearly interpolating a and b samples positions. New monitor units for the first, a and b samples are then defined as:

$$MU_1^{new} = \frac{MU_a^{old}}{3} + \frac{MU_b^{old}}{4} \qquad \text{Equation 42}$$

$$MU_a^{new} = \frac{2MU_a^{old}}{3} \qquad \text{Equation 43}$$

$$MU_b^{new} = \frac{3MU_b^{old}}{4} \qquad \text{Equation 44}$$

With MU_i^{new}, MU_i^{old} respectively the new and old monitor units for sample i.

Weights equal to 3 and 4 are used to overcome non-consistent weights between two consecutive samples. It occurs when a new sample is added, resulting in a non-uniform samples distribution. Additional samples are then added to reach the last defined sample, i.e. sample f (Figure 18.C). The whole process is then re-iterated, with for instance the addition of a sixth extra sample between sample a and sample 1 (Figure 18.D), until obtaining the sought samples number.

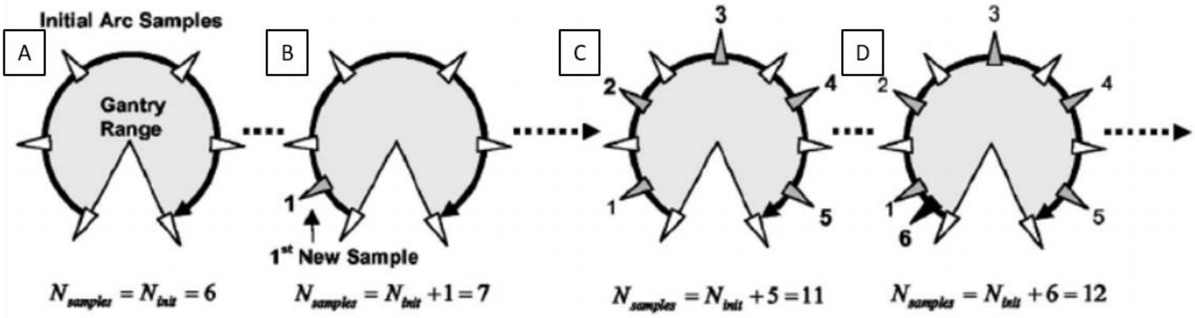

Figure 18: Gantry sampling process, based on a first sampling (A), the addition of the first extra sampling (B), the addition of four extra samples to reach the last sample, the beginning of a new similar process, with the addition of a sixth sample (D). Adapted by permission from John Wiley and Sons: Wiley, Med Phys., “Volumetric modulated arc therapy: IMRT in a single gantry arc”, Otto K., Copyright (2008).

Compared to IMRT, VMAT has the advantage to present smaller treatment sessions durations and a reduced number of monitor units delivered (139). According to Wagner et al. (140), stages III and IV glioma treatment sessions were proved to be five times higher for IMRT sliding window with 5 to 9 beams, than for 1 arc-based VMAT. Regarding monitor units, Shaffer et al.(141) selected 10 high grade glioma patients and performed a double planification, either with VMAT based on a single arc or with a 7-beams sliding window IMRT. A significant decrease in the monitor units to deliver 2Gy was shown, from 789 +/- 112 to 363 +/- 45 for IMRT and VMAT approaches respectively.

Figure 19 presents 3DCRT, IMRT and VMAT dose distributions for a patient with orbital lymphoma (142). As one can notice, a superior conformity to target volume is achieved with IMRT and VMAT. Higher dose levels are delivered to surrounding healthy tissues for 3DCRT, than with IMRT and VMAT. The latter appears to present the lowest irradiated region, reducing the secondary malignant tumors risk.

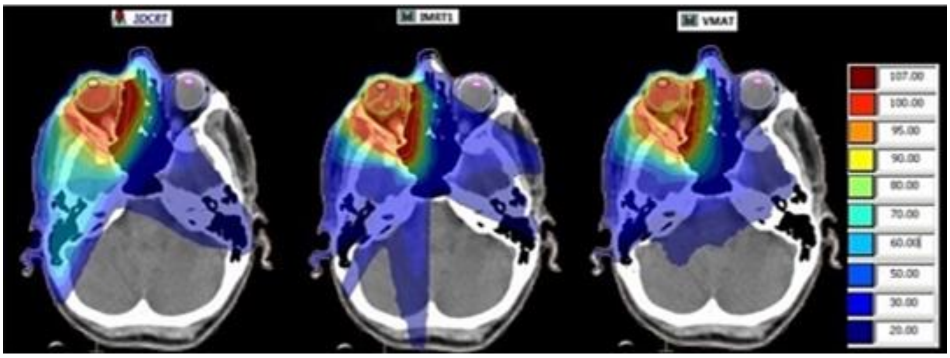

Figure 19: 3DCRT (left), IMRT (middle) and VMAT (right) dose distributions comparison. Reprinted from “EP-1567: Inverse planning versus forward planning for orbital lymphoma”, Rey EMA, Muñoz AR, Jiménez DN, Pardos RG, Truyols MC, 2017, Radiother Oncol. CC BY-NC-ND (2017).

Finally, note that VMAT, similarly to IMRT, requires a fluences sequencing step (section 1.2.4.5.2.).

1.2.4.5.4. DYNamic conformal ARC (DYNARC)

DYNARC approach is based on a rotating gantry and moving MLC. However, it induces PTV conformation and not dose modulation (143), as illustrated by Bokrantz et al. in Figure 20 (144). The optimization of arcs parameters is based on direct planning.

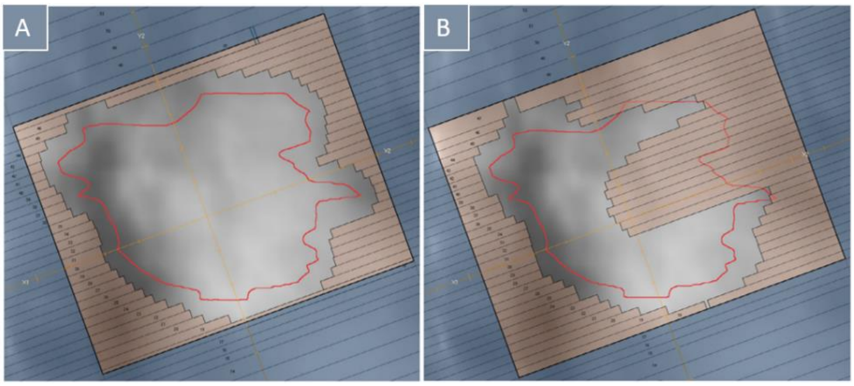

Figure 20: Example of MLC (brown), jaws (blue) and target volume (red) for DYNARC (A) and VMAT (B) cases. Adapted from "Dynamic conformal arcs for lung stereotactic body radiation therapy: A comparison with volumetric-modulated arc therapy", Bokrantz R, Wedenberg M, Sandwall P., 2020, J Appl Clin Med Phys. CC BY 4.0 (2020).

Various studies compared the different treatment deliveries, investigating the best strategies. Morales-Paliza et al. (145) compared DYNARC and IMRT plannings for 15 patients either with prostate or brain or neck or thoracic tumors. Focusing on prostate patients, IMRT was either composed of 7 fields (3 patients) or 5 fields (9 patients). DYNARC consisted in two arcs with coplanar beams. Maximum doses to PTV were equal to 105.4% +/- 2.6% and 104.3% +/- 2.5% of the prescribed dose (76Gy) for IMRT and DYNARC respectively. Mean monitor units used for IMRT were about five times higher than DYNARC, suggesting its high potential.

Concerning VMAT, Uto et al. (146) collected ten craniopharyngioma patients who were planned with DYNARC (2 coplanar and 2 non-coplanar arcs), coplanar VMAT (2 coplanar arcs) and non-coplanar VMAT (1 coplanar and 2 non-coplanar arcs). Prescribed dose was equal to 52.2Gy, delivered in 29 sessions. HI were equal to 0.114 +/- 0.010, 0.103 +/- 0.008 and 0.099 +/- 0.005 for DYNARC, coplanar arcs VMAT and non-coplanar arcs VMAT ($p \leq 0.005$), suggesting the superiority of the latter approach. A similar study was performed by Molinier et al. (147) enrolling 25 cranial lesions patients, planned with DYNARC (3 to 4 non-coplanar arcs per lesion), VMAT_1 (2 coplanar arcs), VMAT_2 (if 1 lesion, configuration similar to DYNARC; else, 4 non-coplanar arcs combined with one isocentre), VMAT_3 (2 coplanar arcs, 10° and 350° table rotations). Regarding patients with one lesion, DYNARC provided the lowest volumes receiving 10Gy, equal to 8.35cc +/- 5.61cc. For patients with more than one lesion, VMAT_2 presented a monitor units rate of 4746MU +/- 792MU, which was lower than the other techniques. Lastly, for patients with a lesion close to an OAR, VMAT_3 achieved the highest conformity index, being equal to 0.72 +/- 0.06. As a result, lesions number and location should be considered to assess the optimal delivery treatment.

1.2.4.5.5. Stereotactic Radiation Therapy (SRT)

Stereotactic Radiation Therapy (SRT) is a highly precise treatment relying on the large dose delivery into one (Stereotactic radiosurgery, SRS) or a few sessions (Hypofractionated Stereotactic Radiation Therapy, HSRT) to small extra or intracranial targets, such as for instance unresectable metastases. In the latter case, treatment sessions number varies from 2 to 5, with dose per session comprised between 5 to 9Gy (148). However, SRT treatment has the disadvantage to result in radionecrosis for 3% to 24% of the patients (149). Thus, it appears of upmost interest to define prognosis factors linked to SRT. To this aim, Gu et al. (150) collected 161 patients with brain metastases comprised between 1 and 7, treated either with SRT alone (99 patients) or SRT combined with a whole brain radiation therapy (62 patients). Biological effective dose E(D,N,t) to SRT were assessed, representing the biological effect of a dose per fraction D delivered N times to a tissue t with an associated α/β ratio. E(D,N,t) superior to 50Gy were reported to be good prognosis factors.

1.2.5. Treatment – Patient set up

1.2.5.1. Contentions

Contentions are immobilization devices ensuring reproducible patient set up. Their types depend on the tumor location and patient health condition.

For brain tumor treatment, a plastic mask is commonly used. Two types are commercially available. First, mesh plastic masks (Figure 21), being the most popular type, are contentions first requiring to be heated, before being placed on the patient head. During cooling, the mask is adjusted to perfectly fit the anatomy. Second type corresponds to Perspex® masks and are less commonly used in clinics. Paris bandage plaster stripes are applied on the patient face to create the mould used to derive the Perspex® mask.

Figure 21: Thermoplastic head mask (151). Adapted by permission from Elsevier, Radiother Oncol J Eur Soc Ther Radiol Oncol., “Repositioning accuracy of a commercially available thermoplastic mask system”, Fuss M, Salter BJ, Cheek D, Sadeghi A, Hevezi JM, Herman TS, Copyright, 2004.

Masks are placed on the patient for every treatment session, potentially being stressful for the patient. Thus, Arino et al. (152) investigated mask-related anxiety for 19 patients with either head and neck or brain or lymphoma tumors. They reported low levels, with averages ranging from 26 to 31, and a patients mask positive representation for 12 patients.

1.2.5.2. Image Guided Radiation Therapy (IGRT)

1.2.5.2.1. Concept

Radiotherapy images and planning are performed before beginning the treatment and correspond to one snapshot of the patient anatomy and position. Since numerous changes occur, including weight loss, tumor shrinking or differences in patient set up, the computed dose differs from the real absorbed dose. To consider this fact, margins are currently applied between CTV and PTV, potentially leading to healthy tissues irradiation (153). However, frequently imaging the patient could be a solution to monitor changes and ensure treatment reproducibility.

Image Guided Radiation Therapy (IGRT) consists in acquiring in-room images and consequently optimizing the treatment accuracy, via couch position adjustments or re-planning for instance. It assumes that improved accuracy and precision signify a possibility to reduce CTV to PTV margins.

IGRT process is three-step based, including the in-room image acquisition, the performance of a registration to the original CT, and the application of correction methods.

In the case of brain tumors, the goal mostly consists in evaluating the position “of the day” and adapting it to reproduce the initial radiotherapy CT configuration. To this aim, three translational shifts, in x, y and z respectively, and up to three rotational shifts, namely pitch, roll and yaw are applied. Note that, for most of standard couches, only yaw rotations are mechanically feasible.

1.2.5.2.2. Imaging devices

1.2.5.2.2.1. 2D kV X-ray

Two main designs have been developed.

First, the X-ray source and the flat-panel amorphous silicon detector can be mounted on two arms of the linac. It concerns the Varian On-Board Imager (OBI) system (Varian Medical

Systems, Palo Alto, CA) whose main possibilities are 2D X-ray, Cone Beam Computed Tomography (CBCT, see 2.5.2.2.2), tracking and fluoroscopy (154). Another commercially available solution is the Elekta X-ray Volumetric Images (XVI) system (Elekta AB, Crawley, UK), proposing 2D, 3D and motion imaging options (155). Two 2D orthogonal images, such as for instance in the anterior/posterior and right/left directions, are acquired.

Second type of systems is independent of the linac gantry, such as Brainlab ExacTrac (BrainLAB AG, Feldkirchen, Germany) for instance. The patient is set up via a first system, composed of two infrared cameras fixed to the ceiling and emitting infrared signal (156). It is reflected on spheres located on a frame previously placed on the patient's head. Signal is collected and analysed via the cameras. ExacTrac second components are the two X-ray tubes, located in the ground, and two detectors suspended from the ceiling. Sources and detector are fixed, and 3.62m apart (157). This configuration enables the delivery of two oblique X-ray beams with an angle of 45°.

The acquired images are bony structures-based registered to 2D Digitally Reconstructed Radiographs (DRR), obtained projecting the planning 3D CT onto a 2D plan.

Figure 22 illustrates the DRR, acquired X-ray and the output warped image of a cranial tumor patient (158).

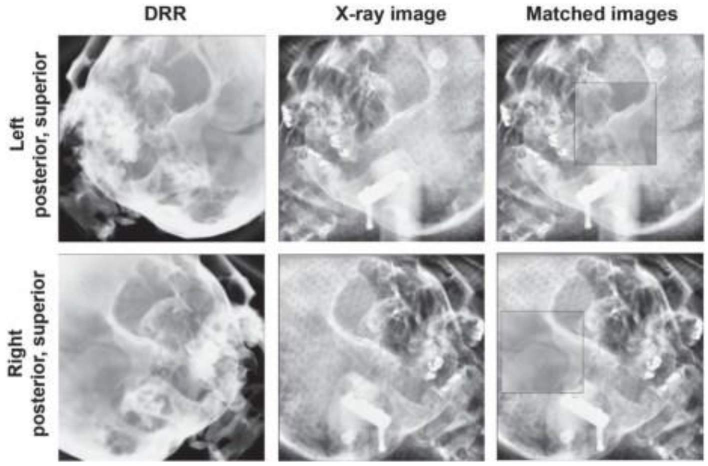

Figure 22: DRR, acquired X-ray and the output warped image for both X-ray tubes. Adapted from "ExacTrac X-ray 6 degree-of-freedom image-guidance for intracranial non-invasive stereotactic radiotherapy: comparison with kilo-voltage cone-beam CT", 93(3), Ma J, Chang Z, Wang Z, Jackie Wu Q, Kirkpatrick JP, Yin F-F, Radiother Oncol J Eur Soc Ther Radiol Oncol, 602-608, Copyright 2009, with permission from Elsevier.

1.2.5.2.2.2. 3D CBCT

Kilovoltage X-ray imaging systems mounted on the linac also offer the possibility to acquire volumetric images, namely CBCT, based on one rotation around the patient. This technology differs from a standard CT in terms of beam shape, being a cone, not a fan. Thus, it covers a larger part and results in acquisition length ranging from 14cm to 26cm and duration times from 1 to 2 minutes. Concerning average dose, it has been proved that it was comprised between 0.2 and 2cGy (159). To assess the image quality, Stützel et al. (160) compared various IGRT devices, including the Primatom (Siemens Oncology Care Systems, Erlangen, Germany) CT with a tube kilovoltage of 130kVp, and the Artiste prototype (Siemens Oncology Care Systems, Erlangen, Germany) providing the possibility to perform kV CBCT. Regarding the latter configuration, the source produced X-ray of 121kV and was located 100cm apart from the isocentre. Final slice thicknesses were respectively equal to 3mm and 4mm for the CT and CBCT. Concerning in-plane resolutions, a value of 0.75mm was obtained for the two different devices. Achieved SNR were equal to 144.50 and 66.73 for CT and CBCT respectively. These same devices led to 0.67% and 1.51% noise percentages. As a result, CBCT presents lower performances than CT, mostly attributed to its larger scatter proportion.

The acquired CBCT is registered to the planning CT, based on both bone and/or soft tissues. Figure 23 presents planning CT, CBCT and warped images for a cranial tumor patient (158).

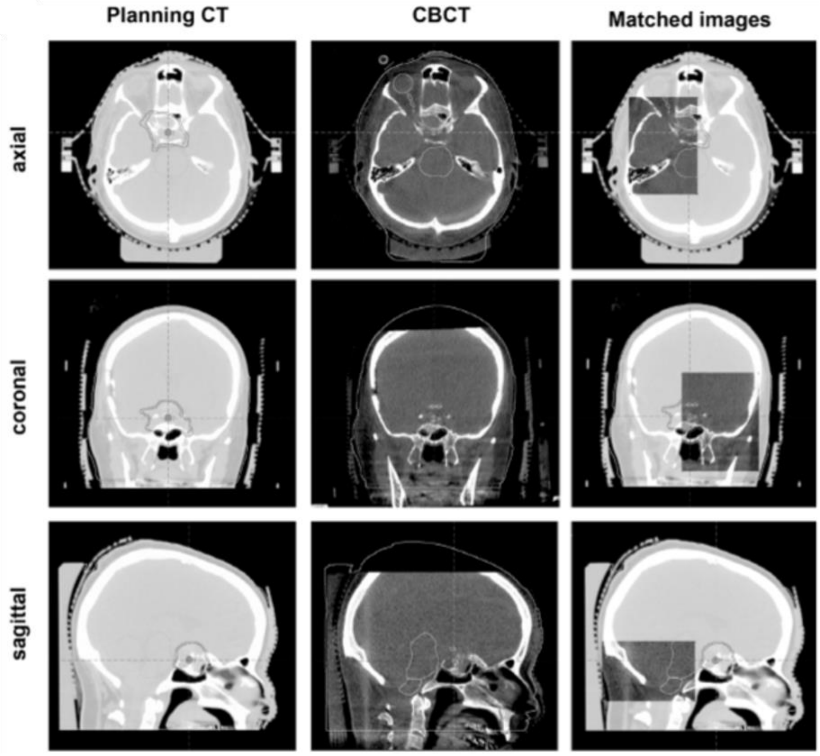

Figure 23: Planning CT, CBCT and registered images in the axial, coronal and sagittal views. Adapted from “ExacTrac X-ray 6 degree-of-freedom image-guidance for intracranial non-invasive stereotactic radiotherapy: comparison with kilo-voltage cone-beam CT”, 93(3), Ma J, Chang Z, Wang Z, Jackie Wu Q, Kirkpatrick JP, Yin F-F, Radiother Oncol J Eur Soc Ther Radiol Oncol, 602-608, Copyright 2009, with permission from Elsevier.

1.2.5.2.2.3. Portal imaging

Portal imaging consists in acquiring images with treatment beam. Owing to both the high megavoltages used in therapy and the predominance of Compton effect at this energies range, and thus the non-negligible scattered photons, image quality is very low. Historically, it was performed with films, and nowadays via Electronic Portal Imaging Devices (EPID). EPID are mounted on gantries of recent linacs. First EPID were either ionization chambers matrices-based, no longer commercially available, or video-based devices (161). The latter consists in copper associated to a gadolinium oxysulfide phosphor interacting with incident beams and producing a signal, which is then deviated with mirrors to reach a camera (162). Nowadays, EPID have evolved towards flat panels. Most of them are based on indirect conversion, i.e. copper associated to a scintillator (163). They respectively have the main roles to block scattered photons/secondary electrons and to convert signals into visible light. The latter is then absorbed by photodiodes located on a flat panel imager whose purpose is to convert light into charges. EPID have the advantage to be instantaneously displayed enabling a fast patient set up and beam re-positioning (164). A second major advantage is its digital format, facilitating software postprocessing and analysis.

1.3. Radiotherapy process deviations

1.3.1. Errors deriving from registration and patient set up

1.3.1.1. Concept

The error is defined as the distance between the true and measured values. Errors are classified into systematic or random. The former implies a constant shift between the measured and true values. Concerning the latter, shifts with the true value unpredictably vary (165).

Based on this observation, it appears of most interest to define appropriate PTV margins to ensure a satisfying dose CTV coverage. In 2000, Van Herk et al. (165) proposed:

$$PTV_{margin} = 2.5\Sigma + 1.64\sigma - 1.64\sigma_p \qquad \text{Equation 45}$$

With Σ, σ, σ_p the total standard deviation of systematic errors, the total standard deviations of random errors, the beam penumbra width.

The latter formula ensures to deliver at least 95% of the specified point dose to the PTV for 90% of the patients cohort.

Previously defined errors have non-negligible impacts on the final dose. Random errors result in dose distribution blurring (166). Assuming fluctuations arise only from patient set up variations, Leong modelled the 2D blurred dose via a convolution (167). The 3D adaption is computed as follows:

$$\mathfrak{D}(x', y', z') = \iiint D(x, y, z) g(x - x') h(y - y') i(z - z') dxdydz \qquad \text{Equation 46}$$

With D, g, h, i the dose distribution, the dose probability distributions functions characterizing translations in x, y, z directions respectively.

However, according to Van Herk et al. (168,169), a proper representation of random errors on dose distribution should be performed via a Gaussian penumbra, with a width proportional to the standard deviation of the random errors divided by the square root of the fractions number. Concerning systematic errors, they result in a shift between planned and absorbed doses.

In 2021, Abubakar et al. (170) released a MATLAB-based algorithm to calculate geometric errors in case of offline IGRT, with the final goal to reduce PTV margins. To do so, 25 head and neck tumor patients treated with IMRT were collected. Prescribed doses ranged from 60Gy to 70Gy, all delivered in 2Gy per fraction. CBCT were performed on fractions 1, 2 and 3. Each

CBCT was registered to the CT and mean set up errors were calculated. Then, two possibilities were presented: either the previous errors were used to correct the rest of the fractions (No Action Level protocol, NAL); or, the errors were used to correct only the remaining week fractions, i.e. fractions 4 and 5. In this case, corrections were then weekly updated, with the acquisition of CBCT on every first fraction of the remaining weeks (extended No Action Level protocol, eNAL). Concerning the algorithm, it took as input the shifts for all the patients and for all the fractions. It resulted in a mean and standard deviation for every patient, respectively corresponding to individual systematic and random errors. Population systematic errors corresponded to the standard deviation of the means. Population random errors were calculated as the root mean square of standard deviations. Finally, PTV margins were assessed with Van Herck formula (Equation 45). Maximum population systematic errors were 1.0mm, 0.8mm and 0.5mm for no correction, NAL and eNAL methods respectively. Regarding individual systematic errors, largest deviations occurred along the z direction, and were equal to 2.4mm, 2.2mm and 1.2mm for no correction, NAL and eNAL approaches respectively. The corresponding maximum individual random errors were 1.1mm, 1.6mm and 1.4mm, suggesting that offline IGRT was able to reduce systematic but not random errors. In addition, PTV margins along z were equal to 3.2mm, 3.0mm and 2.2mm without correction, NAL and eNAL methods respectively, highlighting a non-negligible reduction of the margins with offline IGRT.

1.3.1.2. Example 1 of a systematic errors process: segmentation

Segmentation induces systematic errors owing to the well-established variability between observers preventing a reproducible delineation. Observer variability has been quantified based on various metrics, such as delineated volumes standard deviations or ratio of the largest over smallest volumes (171,172). A quick literature review of this variability applied to brain tumors is presented in Table 4.

Table 4: Review of brain tumor delineation inter-observer variability.

Authors	**Tumor type**	**Delineated structures**	**Number of volumes to contour per clinician**	**Number of clinicians to contour**	**Evaluation metrics and associated results**
Weltens et al. (173)	Supratentorial brain tumor	GTV	5	9	Ratios largest/smallest

					volumes varied from 1.5 to 2.7 for a CT and MRI-based delineation
Coles et al. (174)	Medulloblastoma	CTV	1	17	CTV = 198mL +/- 21mL (posterior fossa) and 12mL +/- 6mL (tumor bed)
Stanley et al. (175)	Brain metastases	Brain metastases	14	8	Ratios largest/smallest volumes comprised between 1.26 and 4.47 after removing one outlier
Cattaneo et al. (176)	Glioblastoma (7 patients) and oligodendroglioma (2 patients)	CTV	7	5	Intersection volume = 67% +/- 15% of the mean volume in the case of registered CT and MRI
Sandström et al. (177)	Cavernous sinus meningioma (1 patient), pituitary adenoma (1 patient), vestibular	Targets and OAR	6	12	Contoured volumes were comprised between [5.29cm^3; 7.80cm^3] (meningioma),

	schwannoma (1 patient), large metastasis (1 patient), medium metastasis (1 patient), small metastasis (1 patient)				[1.67cm^3; 2.15cm^3] (adenoma), [3.56cm^3; 4.48cm^3] (schwannoma), [10.30cm^3; 14.55cm^3] (large metastasis), [1.27cm^3; 3.33cm^3] (medium metastasis), [1.42cm^3; 2.26cm^3] (small metastasis)

1.3.1.3. Example 2 of a systematic errors process: image registration

Image registration is a key task inducing non-negligible errors in the workflow, depending on the software used, the approach (rigid or non-rigid), the optimization algorithm, the optional use of immobilization devices, tumor location, etc.

This error has been quantified by Ulin et al. (4) in 2010 in case of a small left occipital lobe lesion. First, a couple of CT and MRI, from the same patient, were provided to 45 institutions. The lesion was only visible on two MRI slices. Institutions were requested to register via their own clinical software the MRI to the CT, to delineate the target volume on the MRI and to report its centre coordinates in the CT spatial frame. The ground truth target location was defined as the mean of all reported coordinates, considering only centres whose coordinates differences with the mean were inferior to two standard deviations. In total, 11 different registration software were used in association with various methods, such as manual or automatic. Illustration of quantitative deviations in the three views is presented in Figure 24. Mean distances between reported coordinates and true value were equal to 1.8mm +/- 2.2mm, proving that registration induced errors.

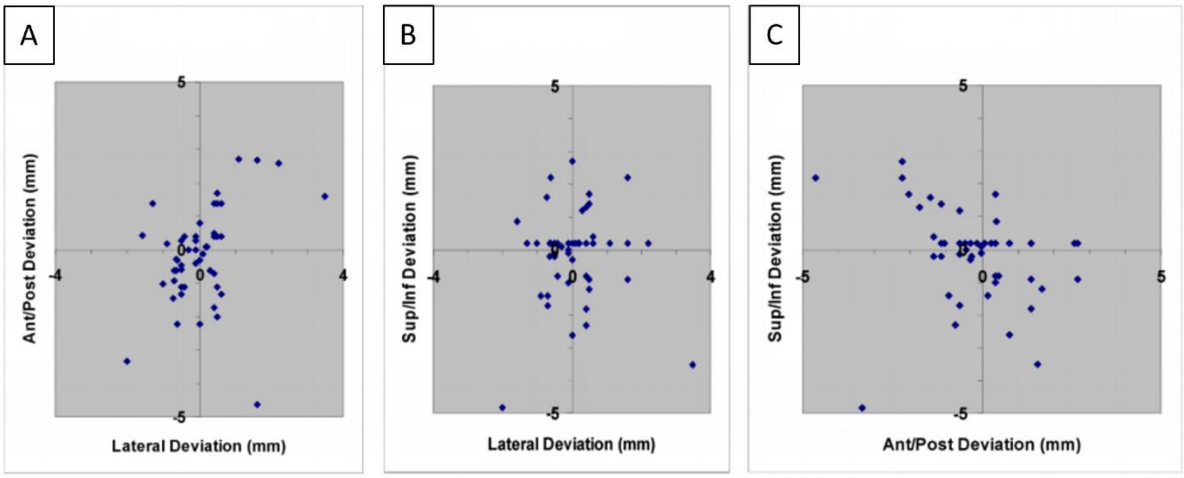

Figure 24: Distances between the real value at the centre of the scatter plot, and reported distances by institutions in the axial (A), coronal (B) and sagittal (C) views. Adapted from “Results of a multi-institutional benchmark test for cranial CT/MR image registration”, 77(5), Ulin K, Urie MM, Cherlow JM, Int J Radiat Oncol Biol Phys, 1584-1589, Copyright 2010, with permission from Elsevier.

1.3.1.4. Example 3 of a systematic and random errors process: patient set up

Global uncertainties related to head tumors patient set up were reported to range from 1 to 2 mm by the International Atomic Energy Agency (IAEA), report number 31 (178). This observation is in agreement with a recent study published by Kanakavelu et al (179) in which twenty brain IMRT radiotherapy patients were included and frequently imaged with MV 2D X-ray and MV CBCT. Deviations along x, y and z were recorded, respectively leading to averages of 0.57mm +/- 1.13mm, -0.95mm +/- 1.09mm and -0.01mm +/- 0.54 mm for the whole cohort. Since most of the radiotherapy contention masks used nowadays are non-invasive, misalignments and errors can occur potentially leading to target underdosage and OAR overdosage. Assessment of specific Raycast©-HP thermoplastic mask (Orfit Industries, Wijnegem, Belgium) performances has been performed by Fuss et al. (151) for 22 head lesions patients. One to four CT control CT were acquired per patient, enabling the assessment of the target isocentre displacements. Translations and rotations along x, y and z were respectively equal to 0.74mm +/ 0.53mm, 0.75mm +/- 0.60mm, 0.93mm +/- 0.78mm and 0.67° +/- 0.66°, 0.61° +/- 0.63°, 0.67° +/- 0.61°.

Quantifying systematic and random errors set up is not trivial. Tryggestad et al. (180) proposed a method to investigate four different immobilization configurations accuracies dedicated to intracranial lesions radiotherapy patients. To do so, 121 patients with intracranial lesions were collected. The detailed configurations are presented in Figure 25. kV CBCT were performed as imaging guidance. Interfraction systematic errors were patient-wise determined as the mean of all deviations occurring during treatments. Random interfraction errors were computed subtracting the deviation “of the day” to the previous systematic error. Systematic errors associated to the last configuration, including the mouthpiece, were the lowest, with deviations of 0.34mm +/- 0.70mm, 0.34mm +/- 1.20mm, 0.56mm +/-0.90mm along the medio-lateral, cranio-caudal, anterior-posterior directions. Based on the same directions, reported random errors were equal to 0.9mm, 1.0mm and 0.8mm. All the masks were concluded to be suitable for IGRT.

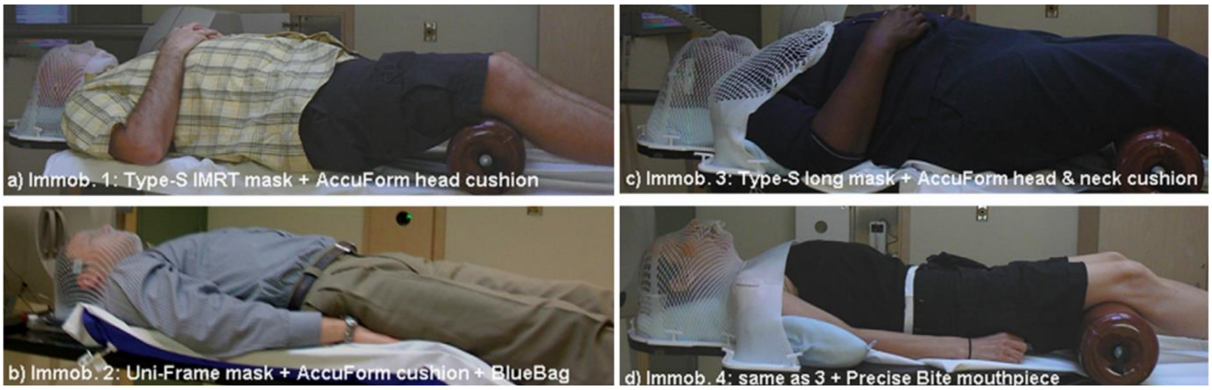

Figure 25: Immobilization configurations. Reprinted from “Inter- and intrafraction patient positioning uncertainties for intracranial radiotherapy: a study of four frameless, thermoplastic mask-based immobilization strategies using daily cone-beam CT”, 80(1), Tryggestad E, Christian M, Ford E, Kut C, Le Y, Sanguineti G, et al., Int J Radiat Oncol Biol Phys, 281-290, Copyright 2011, with permission from Elsevier.

1.3.2. Uncertainties in radiotherapy measurements

1.3.2.1. Concept

The uncertainty reflects the dispersion of a measurement. For instance, in radiotherapy, the absorbed dose is assumed to be represented as a Gaussian centred on the prescribed dose.

Two types of uncertainties have been defined: either type A or type B. The former is evaluated via statistical approaches implying the repetition of an experiment and the associated standard deviation computation, the latter via other approaches requiring critical skills.

The final uncertainty σ_{combined} is obtained combining previously described uncertainties as follows:

$$\sigma_{combined} = \sqrt{\sigma_A^2 + \sigma_B^2} \qquad \text{Equation 47}$$

With σ_A, σ_B, types A and B uncertainties.

1.3.2.2. Quantitative approach

Typical uncertainties magnitudes associated to the bilateral beam 3DCRT treatment delivery were provided by Mijnheer et al. (181). It included, for instance, the calibration of the ionization chamber (Type A = 0.9%, Type B = 1.1%), the assessment of the absorbed dose at given patient points (Type A = 1.0%, Type B = 1.3%) and the patient irradiation with two parallel MV beams (Type A = 1.5%, Type B = 0%). The obtained combined uncertainty was reported to be equal to 4.2%.

Similarly, Castro et al. (182) investigated the uncertainties associated to two chambers, namely a cylindrical PTW 30013 (PTW, Freiburg, Germany) and a plane-parallel Markus (PTW, Freiburg, Germany) used to measure the absorbed dose to water. Reproducibility, representing the variation of a given measurement when repeating the same experiment, is a well-established Type A uncertainty. Irradiations were performed with a ^{60}Co device, producing gamma rays with energies equal to 1.17MeV and 1.33MeV. Reproducibility was found to be equal to 0.03% for both detectors. Additional uncertainties linked to Type B, were reported for these chambers. For instance, leakage, potentially resulting in smaller measurements, was overestimated to ensure encompassing the true value. A value of 0.01% was assigned for both chambers. Long-term stability, evaluating detectors consistency over time, were equal to 0.29% and 0.48% for PTW 30013 and Markus chambers respectively. Regarding linearity, rates were equal to 0.03% for both chambers, based on a ^{90}Sr source. Final uncertainties, combining Types A and B, were 0.29% and 0.48% for PTW 30013 and Markus chambers, suggesting the superiority of PTV 30013.

1.4. MRI-Linac opportunities

1.4.1. Description

As previously mentioned, in the standard radiotherapy workflow, MR-imaging and treatment are performed with two independent devices, at different treatment times. Recently, hybrid devices combining a linac to a MRI have been developed, offering an optimal guidance and enhancing personalized patient care. Technically, two MRI-linac designs have been proposed, i.e. placing the irradiation beam perpendicular or parallel to the magnetic field (Figure 26 (183)).

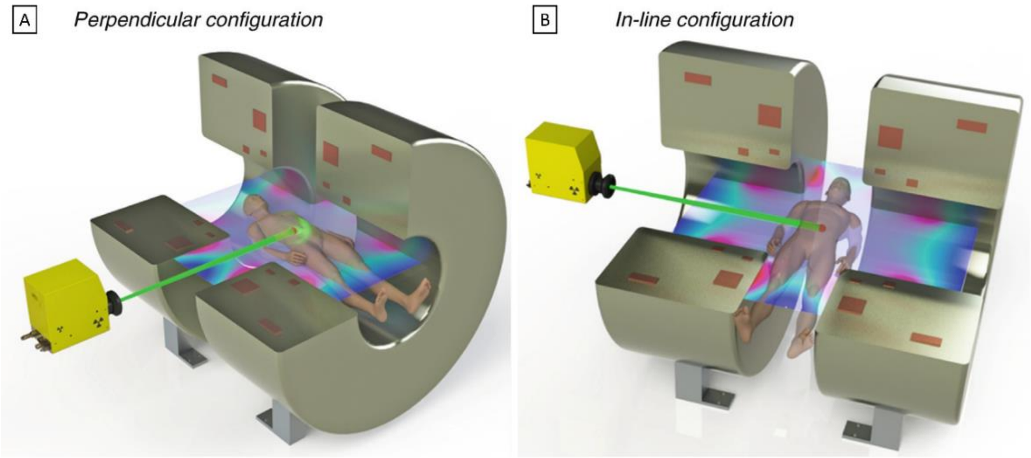

Figure 26: Perpendicular (27.A) and inline (27.B) MRI-Linac configurations. Adapted by permission from Springer Nature Customer Service Centre GmbH: Springer, "MRI Linac Systems", Whelan B, Oborn B, Liney G, Keall P, Copyright, 2019.

In 2017, the first human treatment MRI-Linac based occurred in Utrecht, The Netherlands with the Elekta Unity device (184). Four patients with spine metastases were enrolled in the study. Reported geometric accuracies derived from MRI were comprised between 0.2mm and 0.4mm proving the high device precision.

Nowadays, four different MRI-linac types have been developed, including the Elekta Unity, MRIdian, Australian MRI linac and Aurora RT. Technical details are provided in Table 5.

Table 5: MRI-linac characteristics.

Devices	**Commercial availability**	**Magnetic field strength**	**Photon energy beam**	**Technical configurations possibilites**
Elekta Unity (Elekta, Stockholm, Sweden)	Yes	1.5T Philips Ingenia (Philips Healthcare, Best, The Netherlands)	7 MV flattening filter free linac	Perpendicular
MRIdian (Viewray, Oakwood, United States)	Yes	0.35T	3 Co^{60} sources (γ) or 6 MV flattening filter free linac	Perpendicular

Australian MRI-Linac, (Liverpool, Australia)	No	1T MRI (Agilent Technologies, Oxford UK)	4 and 6MV flattening filter free (Varian Linatron-MP)	Both
Aurora RT (MagnetTx, Edmonton, Canada)	No	0.6T open bore magnet	6 MV	Inline

1.4.2. Advantages

MRI offers a higher soft tissues contrast than CT, and enables an access to the anatomy, but also to physiology (185). The latter provides numerous and meaningful information concerning the tumor, such as the quantity of hypoxic tissues, whose radio-resistance is two to three times higher than normal tissue (186). Similarly, Diffusion weighted MRI representing water molecules diffusion in the body, provides apparent diffusion coefficients whose increase has been proved to correlate with tumor response, since treatments induce cell deaths and improved water mobility (187).

This enhanced contrast allows for real time MRI. Combining it with MLC tracking enables the continuous tumor imaging while the beam is on. This technique is particularly adapted for mobile tumors located for instance in liver or lungs, ensuring an access to target intrafraction movements and ensuring a high treatment delivery accuracy. However, note that this approach requires latency times between the MRI acquisition and the MLC adaption to be low, to avoid wrong irradiation delivery. This latency was quantified by Borman et al. (188), who reported durations of 0.5 times the acquisition time for radial k-space sampling sequences.

Lastly, MRI-linac allows for Adaptive Radiation Therapy (ART), whose goal is to modify the treatment based on the patient variations related to the positioning, OAR or target volumes (189). Three types of strategies have been developed. First, offline approach relies on the acquisition of the MRI "of the day", and possibly re-planning before the next treatment session. This method is clinically acceptable as long as changes do not occur during fractions or between two fractions (190). Second approach is online, and consists in adapting the treatment before delivering the fraction "of the day". Elekta proposes for instance the *adapt to shape* option.

First, the online MRI is acquired. The planning CT is then deformably registered to the MRI "of the day". Contours are projected from the CT to the online MRI, and possibly corrected to match the anatomy of the day. Then, the crucial pCT generation step occurs. Finally, the plan is re-calculated or re-optimized (191). Thus, online ART combined with a MRI-linac is a promising pCT generation application. Last approach is real-time adaption, i.e. during a fraction. Its first clinical linac implementation feasibility occurred in 2018, as reported by Keall et al. (192). To do so, 8 prostate tumor patients were enrolled, with associated prescribed dose equal to 36.25Gy. Real time imaging relied on Kilovoltage Intrafraction Monitoring (KIM) and MLC tracking respectively to identify the target location and accordingly adapt the treatment beam. Reported KIM geometric accuracy were equal to -0.1mm +/- 0.4mm, 0.2mm +/- 0.2mm and -0.1mm +/- 0.6mm for left/right, superior/inferior and anterior/posterior directions. In addition, 100% and 95% of CTV received the prescribed dose when applying real-time ART and without, proving the high efficiency of real-time ART.

1.4.3. Challenges

The MRI-Linac devices raise various physical and technological issues, since the linac impacts the MRI and vice versa.

First, moving electrons, deriving from photons with matter interactions, in an electro-magnetic field result in a Lorentz force $\vec{\mathrm{F}}$, defined as:

$$\vec{\mathrm{F}} = e\vec{\mathrm{E}} + \ e\vec{\mathrm{v}} \wedge \vec{\mathrm{B}} \qquad \text{Equation 48}$$

With $\vec{\mathrm{E}}$, $\vec{\mathrm{B}}$ $\vec{\mathrm{v}}$, e, the electric field, magnetic field, electron velocity and electron charge.

Since the Lorentz force is perpendicular to the electrons displacement, it results in electrons re-entering the volume they exit at water/air interfaces, known as Electron Return Effect (193). Raaijmakers et al. (194) quantified the dose increase, based on both MC simulations and the irradiation of a phantom containing a Kodak X-Omat V film and placed in an electromagnetic field. They reported a rate of 40% on the surface where the beam exits the phantom. A possibility to reduce this extra dose is to place a bolus on the patient (195).

The integration of the electron return effect into clinical TPS has mainly be performed with MC. In 2019, Chen et al. (196) investigated the Monaco Research version (Elekta, Stockholm, Sweden) TPS ability, which is based on MC, to calculate the dose for a pre-clinical MRI-linac

(Elekta, Stockholm, Sweden). An inhomogeneous phantom was designed, assembling various ED tissues to mimic lung, soft tissues and bones. The selected ionization chamber was A26MR (Standard Imaging, Middleton, WI, United States). Two hundred monitor units were delivered with a $10cm^2$x$10cm^2$ field size. Difference between the measured dose and the calculated dose was equal to 0.98%, highlighting a high agreement between simulated and measured values. More recently, Shortall et al. (197) confirmed this observation especially close to air cavities evaluating Monaco (Elekta, Stockholm, Sweden) TPS. The Unity (Elekta, Stockholm, Sweden) was used to irradiate four phantoms, respectively containing air sphere of diameters equal to 0cm, 0.5cm, 3.5cm and 7.5cm. Ground truth measurements were performed with GafChromic EBT3 films. 3D global gamma indices with 20% dose threshold were calculated. It resulted in 3%/3mm pass rates equal to 98.3%, 95.7%, 98.3% and 95.6% for respectively phantoms with air sphere of diameter 0cm, 0.5cm, 3.5cm and 7.5cm, confirming the clinical use feasibility.

Another major MRI-linac limitation has been reported by Liney at al. (183) and concerns interactions occurring between the treatment beam and the MRI receiver antenna, leading to a beam attenuation, dose skin increase and electronic disequilibrium. To overcome this issue, they proposed a radiation transparent RF coil designed for 1T magnetic field (Magnetica Pty Ltd, Australia) and reported a high SNR of 42.6 +/- 0.9 (198).

Lastly, measuring the absorbed dose with dosimeters, such as air filled-ionization chambers, is challenging owing to the impact of the magnetic field on the air contained in the detectors (199). Indeed, as previously described, the Lorentz force deviates secondary electrons especially at air/water interfaces, resulting in dose changes. In 2013, Smit et al. (200) investigated this variation for the NE2571 Farmer-type ionization chambers (NE Technology Limited, Berkshire RG7 5PR, England). The MR-Linac was an Elekta prototype, with a 6MV beam. The presence of the 1.5T magnetic field strength was proved to increase the measured dose by a factor of 4.9% +/- 0.2%. One possible approach to overcome this limitation consists in applying correction factors. O’Brien et al. (201) recently reported corrections for various detectors placed parallel to the magnetic field in case of the Elekta Unity (Elekta, Stockholm, Sweden). For instance, PTW30013, PTW30012 and PTW30011 corrections were respectively equal to 0.994, 0.992 and 1.000 (PTW-Freiburg, Freiburg, Germany).

To conclude, current brain radiotherapy workflows are complex procedures requiring the dual acquisition of a CT and a MRI, used during segmentation/dosimetry and segmentation-only steps respectively. Errors, uncertainties or inter-observer variabilities arise from most of the processes, including registration algorithm, targets and OAR delineations, dose calculation and patient set up resulting in a large treatments variability. A possible option to overcome certain of the previously listed limitations consists in drastically reducing human intervention via tasks automation with state-of-the-art AI-based algorithms.

2. AI

2.1. General concept

AI aims to mimic human’s intelligence. The latter precise concept was first introduced in 1950, by Alan Turing with the imitation game (202). Three characters were defined: Person A, Person B and Computer. Person A was in a room apart, with two terminals. Each terminal corresponded either to Person B or Computer. Person A asked questions to find out who the terminal belonged to (203). The test was considered to be passed if the machine was able to fool 30% of test people. The machine was then considered as intelligent.

Nowadays, this technology is present in the daily life via multiple aspects, such as Natural Language Processing (NLP), which aims at developing sophisticated mechanisms to enable computers a human language comprehension (204). It is not a trivial task, owing to numerous traps, including the undertones, expressions, hyperboles, etc. One example of technologies relying on NLP are chatbots, consisting in robots designed to discuss with real human. Many firms have already developed their own technologies, such as Eno (Capital One), Blue Bot (KLM) and Anatole (SNCF), to quickly and accurately answer to customers. Yet, note that humans/chatbots conversations have been reported to be inferior, in terms of all various intelligence types, than humans/humans ones (205). A second application are self-driving cars, developed for various objectives, such as decreasing car accidents and enabling a smoother traffic (206). For example, Waymo developed by Google, is an autonomous car composed of Lidar sensor placed on the roof, additional sensors placed around the car, radars associated with a hardware design to enable a constant evaluation of the environment and a prediction of what will happen next. AI has been proved to be involved in various steps, such as the motion monitoring, traffic sign recognition, obstacle apprehension, etc (207).

2.2. Machine Learning (ML)

AI sub-category is Machine Learning (ML). Learning refers to the ability for a machine to improve itself with experience, as mentioned by Mitchell (208). ML can be mainly split into four classes, depending on the experience type.

First, supervised learning requires to provide the algorithm with couples of data, composed of the input x and the output y. The goal is to find a function transforming the input into the output and minimizing a loss function L, as described below (209):

$$\min_f \frac{1}{N} \sum_{(x,y)} L(f(x), y) \qquad \text{Equation 49}$$

With f, L, N the investigated function, the loss function, the total number of points used to build the model.

Supervised leaning is used for two tasks. First, classification predicts discrete output and aims at defining a threshold to separate 2 or more classes. It is for instance implemented in the anti-spam system of a mailbox. The input are features extracted from the email, and the output is either 0 or 1, respectively indicating a non-spam and spam emails (209). Second type is the regression and corresponds to continuous predictions. The simplest regression is linear and hypothesizes a linear relationship between x and y. Figure 27 illustrates a classification problem between blue squares and red triangles (A) and a regression task (B) (210).

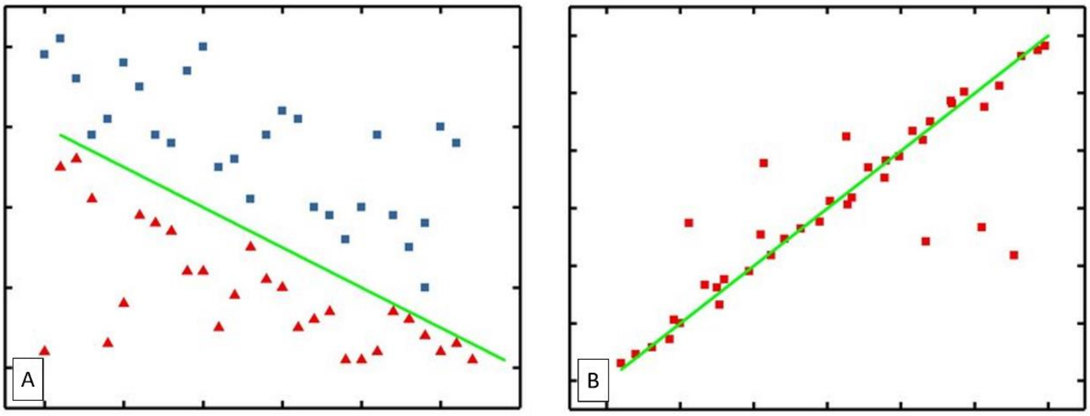

Figure 27: Classification (A) and regression tasks. Adapted from “Regression Analysis With Differential Privacy Preserving”, Fang X, Yu F, Yang G, Qu Y, 2019, *IEEE Access*. CC BY 4.0 (2019).

Since collecting labelled data is time-consuming, unsupervised learning has been developed and consists in providing the machine with data without label. The goal of the machine is to define robust features re-usable on an unseen cohort, to predict an output (211). Unsupervised learning offers the possibility to explore unexpected domains, contrary to supervised learning requiring to guide the machine with ground truth data (212). One type of unsupervised learning is clustering and consists in gathering data into groups, sharing similar properties. It relies on the reduction of inter-elements distances within a same group. On the contrary, distance between groups is augmented. This domain has been successfully applied to genomics, by Battistella et al. (213). 4615 genomics samples were retrieved from the public database The Cancer Genome Atlas, representing various tumors locations including head and neck, liver, breast, rectum, etc. The optimized signature, based on LP-stability and composed of 27 genes, was proved to be at least equivalent or even superior to the K-Means signature (30 genes) for 9 tested tumor types out of 10.

Semi-supervised approaches are hybrid methods relying on both labelled and unlabelled data collection, introduced for classification tasks. The required condition is the pertinence of the unlabelled data for the task. More formally, Chapelle et al. (214) suggested that p(y|x), representing the probability of predicting y knowing x, must benefit from p(x), corresponding to the information derived from unlabelled data. One possible approach for semi-supervised consists in first applying a clustering algorithm to the large unlabelled data amount. For each cluster, the centroid is selected and annotated. Thus, it results in the labelling of smaller dataset, corresponding to the number of clusters. A supervised approach, such as neural networks, are finally applied for classification model assessment.

A last ML category is Reinforcement Learning (RL). It is composed of an agent evolving in an environment. RL goal is to teach the agent with the action it has to perform in the environment via rewards. More precisely, it is an iterative process, first requiring the agent to detect the environment. Then, it performs an action. The latter is either positively or negatively rewarded (Figure 28 (215)). The agent accordingly upgrades a function (216), i.e. learns. RL have been used in various fields, such as robotics (217), chemistry (218) and finance (219).

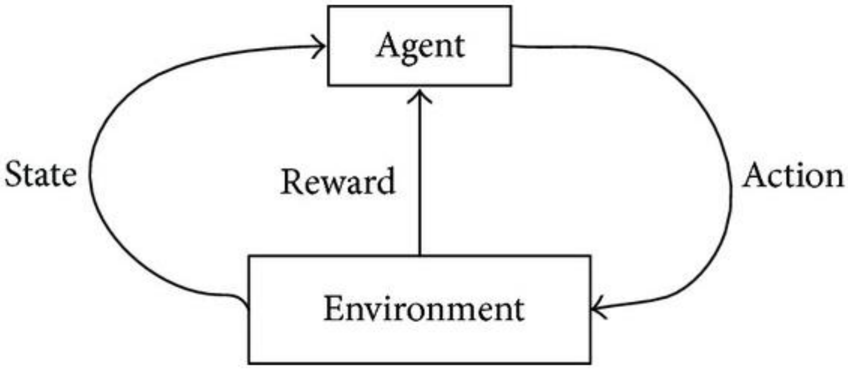

Figure 28: Whole RL process including the environment state observation by the agent, the action onto it and the reward. Reprinted from “A Sarsa(λ)-Based Control Model for Real-Time Traffic Light Coordination”, Zhou X, Zhu F, Liu Q, Fu Y, Huang W, 2014, Scientific World Journal. CC BY 3.0 (2014).

2.3. DL

2.3.1. Artificial neurons

DL is a ML sub-category in which the network automatically extracts features for the prediction. DL mostly relies on artificial neurons. Historically, this concept was first introduced by Rosenblatt (220), with the description of a single unit neuron, i.e. a perception, developed for a binary classification task. It was inspired by biological neurons, which are entities composed of several entries, a cell and a single output signal. It consists in an ensemble of weights and a bias, and can be described as:

$$y_n = \varphi\left(\sum_{p=1}^{i} x_p \times w_p + b_n\right) \qquad \text{Equation 50}$$

Where y_n is the output of neuron n, φ the activation function, x_p is the input from neuron p, w_p the corresponding weight applied to input p, and b_n the bias associated to neuron n.

An illustration of the perceptron is provided in Figure 29.

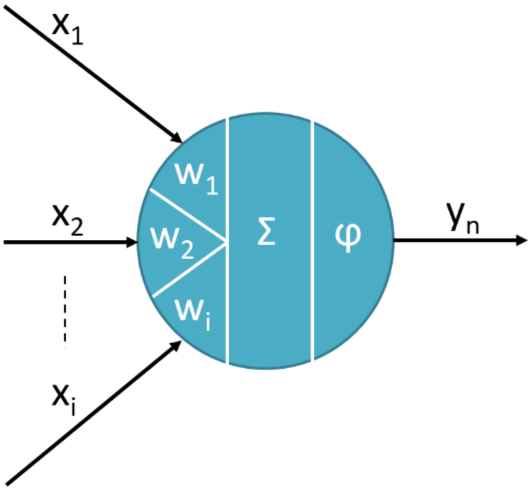

Figure 29: A perceptron, composed of i input, carrying the signal x_i. Different weights w_i are applied to each input, which are then summed. The network bias is added, before applying the activation function φ.

Table 6 presents common activation functions, and corresponding formulas.

Table 6: Examples of activation functions.

Activation functions	**Formula**
Linear	$\varphi(\mathrm{x}) = \mathrm{ax}$
Sigmoid	$\varphi(\mathrm{x}) = \frac{1}{1+e^{-\mathrm{x}}}$
Tanh	$\varphi(\mathrm{x}) = 2\ \mathrm{sigmoid}(2\mathrm{x}) - 1$
Rectified Linear Unit (ReLu)	$\varphi(\mathrm{x}) = \max(0, \mathrm{x})$

The most popular activation function is ReLu, since there is no gradient convergence to zero, and thus no gradient vanishing effect. Another advantage of the function is its low computation cost, compared to the exponential in sigmoid function for instance.

2.3.2. Networks design

Two types of configurations have been proposed to organize neurons.

First, feedforward networks being currently the most frequently used, are characterized by the unidirectional information path, i.e. from the input to the output. No back loops are integrated.

Neurons are assembled into several layers. Concerning the simplest architecture, namely the multilayer perceptron, all neurons of two consecutive layers are connected, as illustrated in Figure 30.

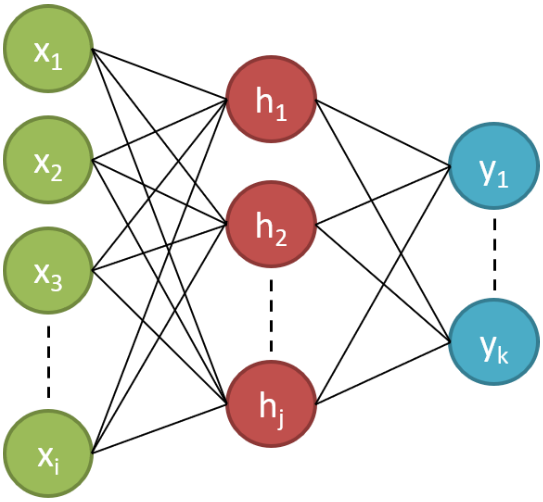

Figure 30: Multilayer perceptron, composed of i neurons on the input layer (green), j neurons on the hidden layer (red) and k output neurons (blue).

Concerning the ability of the network to predict the output, Hornik (221) reported about the Universal Approximation Theorem, stating that any continuous function on compact sub-ensembles can be approximated with a one-hidden layer feedforward multilayer perceptron. In other words, when dealing with multilayer perceptron, networks failures to achieve the desired output cannot be attributed to the architecture.

Second type of architecture is feedback networks where signal can go forward, i.e. from input to output, and backward, i.e. from output to input. Figure 31 presents an example of feedback network, with backward connections from the output Y(j) to two input nodes, Out1(i,j) and Out1(i,j+1) for instance, representing output signals from Y(j) to layer i (222).

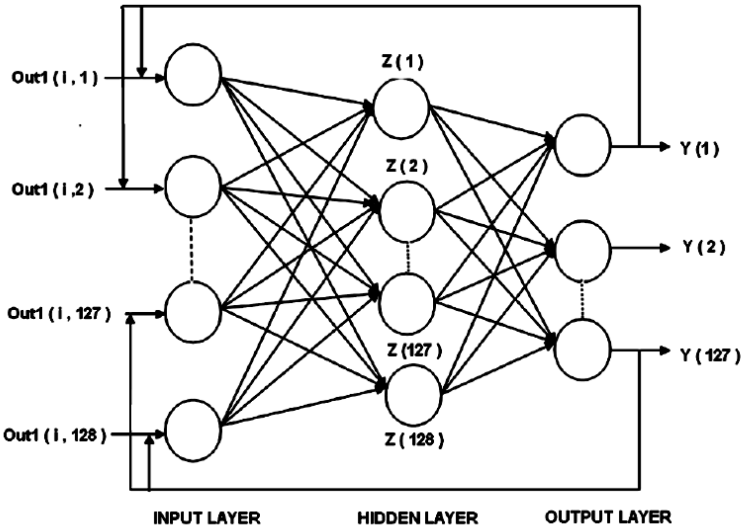

Figure 31: Feedback network, with four backward connections represented. Reprinted from "Image Hash using Neural Networks", 63(22), Desai V, Rao D, Int J Comput Appl., Copyright 2013, with permission from IJCA Edition.

These networks have three main advantages, as reported by Zamir et al. (223). First, it enables quicker predictions. Indeed, backward networks are based on an iterative process, meaning that output are first computed, before being fed back to the network, etc. On the contrary, for a forward network, the output is computed once, at the end of the prediction. Second advantage concerns the prediction configuration, which is performed hierarchically, becoming finer and finer with increasing iterations. To predict a French Bulldog for instance, the network at first iteration identifies the general category, namely animals, among other categories such as person or food. Then, it selects dogs among other pet classes, such as cats or horses. Finally, the correct bulldog output is predicted, among highly specific categories, Labradors Retrievers and Siberian Huskies. Last major advantage concerns curriculum learning which can be applied. It consists in selecting the data to present to the network and its order to iteratively boost complexity and decrease duration times (224). Thus, the feedback connections can be presented as the memory of previous actions.

2.3.3. General process

The whole process can be split into three sub-tasks, namely, training, validating and testing a model. Since supervised algorithms are more common for pCT generation task than unsupervised, the description has been adapted to this learning category.

2.3.3.1. Network training

The basic training idea relies on the calculation of a distance between the network prediction and the ground truth via a loss function. Network parameters, namely weights and biases, are then accordingly modified to achieve the loss function minimum.

Concerning the workflow, weights and biases are first initialized, before defining the input layer signals based on the training set. Information is then propagated forward, with the calculation of increasingly deep layers output with Equation 50. When obtaining the output signal, the loss function is calculated, comparing the prediction with the ground truth provided in the training set. Finally, the error is backpropagated from the output to the first hidden layer nodes to enable weights and biases update (section 2.3.4.).

Various loss functions have been introduced, depending on the DL task to solve. Concerning general classifications tasks, the categorical cross entropy CE has been widely used, and is calculated as follows:

$$CE = -\sum_{i=1}^{C} y_i^{gd} \log\left(y_i^{pred}\right) \qquad \text{Equation 51}$$

With C, y_i^{gd} and y_i^{pred} the number of classes, the ground truth and the network prediction for class i.

Concerning regression tasks, the Mean Squared Error (MSE) is commonly used, and defined as:

$$MSE = \frac{1}{N}\sum_{i=1}^{N}\left(y_i^{gd} - y_i^{pred}\right)^2 \qquad \text{Equation 52}$$

With N, y_i^{gd} and y_i^{pred} the total number of samples, the ground truth value and the predicted value.

In theory, MSE could be used in combination with a softmax activation to get probabilities for a classification task. However, in practice, distances between continuous ranges of predicted values and ground truth values result in lower loss function and thus a lower penalization than cross entropy.

2.3.3.2. Network validation

Validation aims at optimizing the hyperparameters, i.e. the parameters accessible by the operator such as the loss function or the optimal data preprocessing. An additional parameter requiring optimization is the number of epochs corresponding to the number of times the whole training dataset pass through the network. To avoid defining a threshold, early stopping has been introduced. Its role is to prevent overfitting referring to a model which is not generalisable to other sets than the training one. It is performed selecting the model corresponding to the lowest validation error. This task is not always trivial, as highlighted by Figure 32 which presents training and validation loss function curves under ideal (A) and real (B) conditions (225).

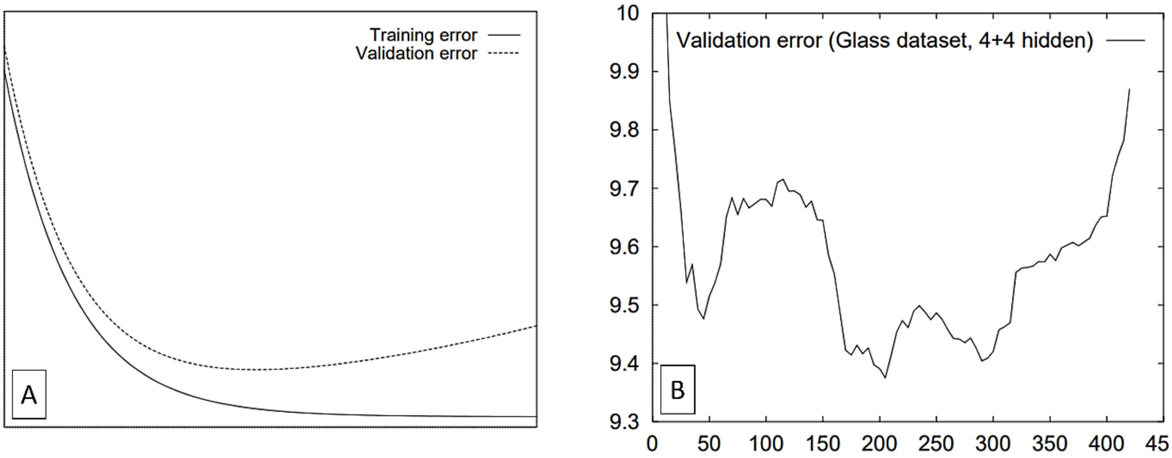

Figure 32: Error against time for training (continuous line) and validation (dotted line) in idealistic conditions (A) and for validation set in reality (B). Adapted from “Automatic early stopping using cross validation: quantifying the criteria”, 11(4), Prechelt L, Neural Netw., 761-767, Copyright 1998, with permission from Elsevier.

A pseudo code describing the process implemented for the studies below is provided, from an initial epoch N equal to 2:

```
Initialisation: A = 0; N=2; Reference_error = error(N=1); Stopping_criterion=S; Loss_function=L
• error(N) = L(Validation set, N) #Compute the error on validation set, for epoch N
• if error(N) < error(N-1):
    o Reference_error = error(N) #Reference error is the lowest achieved error
    o A=0 #Assign 0 to A, representing the number of epochs from reference error epoch to current epoch
• else:
    o A=A+1 #A is incremented
    o if A > S:
        ▪ Stop training #If A superior to a pre-defined threshold S, for instance 8, stop the training to avoid overfitting
• iterate the process with N=N+1 #Otherwise, continue training with the following epoch N+1
```

An additional parameter to be optimized is the batch size and refers to the number of elements fed to the network before calculating the loss, and updating the weights and biases. This parameter is a trade-off between decreasing duration computing time and enhancing convergence speed respectively occurring when the batch size increases and decreases (226).

2.3.3.3. Network testing

Finally, the model is tested on an unseen cohort, and performance is evaluated to assess the model quality. To avoid biases, the testing set should be fed to the network once, i.e. at the end of the training and validation optimization processes. Otherwise, the risk is to learn on the testing set.

2.3.4. Optimization

2.3.4.1. Standard gradient descents

Gradient descent aims at identifying the global minimum of a loss function via network parameters upgrades and optimal rates achievement. This is performed via successive gradients computations, i.e. loss function derivative estimations. The velocity to upgrade weights and biases is based on the learning rate. Caution must be observed when selecting this parameter. Indeed, an excessively low learning rates imply small displacements along the loss function, and the possibility to get trapped in a local minimum. On the contrary, large learning rates mean large displacements associated to the risk of missing the global minimum. Figure 33 presents a 1D illustration of the global gradient process (227).

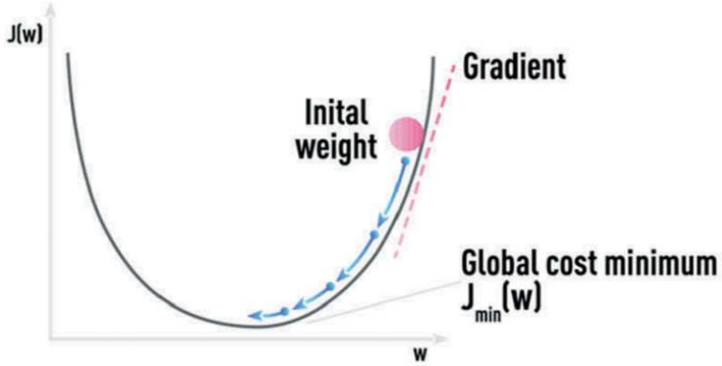

Figure 33: Gradient descent approach, with the loss function (black curve), gradient a given point (pink dotted line) and successive parameters updates (blue arrow). Adapted from "Synergy of physics-based reasoning and machine learning in biomedical applications: Towards unlimited deep learning with limited data", Gavrishchaka V, Senyukova O, Koepke M, 2019, Adv Phys X. CC BY (2019).

Three different types of gradient descents have been developed (228). First Vanilla, or batch gradient descent, calculates the error for each training element. However, the parameters update is performed once the whole training data is passed through the network, as follows:

$$\theta^t = \theta^{t-1} - \eta \nabla_\theta L(\theta^{t-1}; \tau) \qquad \text{Equation 53}$$

Where τ, η, θ^i, ∇ refers to the whole training set, the learning rate, the parameter to be updated (weight or bias) and the derivative of L per θ.

The main disadvantage of this approach is the long computation time, especially for large training datasets.

To overcome this limitation, stochastic gradient descent has been introduced, relying on an error calculation and parameters update performed for each training element, randomly selected in the training set, as described:

$$\theta^t = \theta^{t-1} - \eta \nabla_\theta L\big(\theta^{t-1}; (x_i; y_i)\big) \qquad \text{Equation 54}$$

With (x_i, y_i) a training couple composed of x_i the network input, and the corresponding ground truth y_i.

An intermediate approach is the mini batch gradient descent, consisting in updating the parameters for every N training elements contained in a mini-batch B with:

$$\theta^t = \theta^{t-1} - \eta \nabla_\theta L(\theta^{t-1}; B) \qquad \text{Equation 55}$$

The main advantage of such approach compared to stochastic gradient descent, is the variance decrease, since a more representative sample is used for optimization.

2.3.4.2. Adam

Adam was first introduced by Kingma et al. (229) in 2014. It aims at computing gradient first and second orders and relies on a learning rate specific to each parameter. First and second moments definitions are calculated as follows:

$$m_t = \beta_1 m_{t-1} + (1 - \beta_1) g(L, w_t) \qquad \text{Equation 56}$$

$$v_t = \beta_2 v_{t-1} + (1 - \beta_2) g(L, w_t)^2 \qquad \text{Equation 57}$$

With m_t, $g(L,w_t)$, v_t the gradient first moment at time t, the gradient of the loss function L regarding the weight w_t at time t, second gradient order at time t. β_1 and β_2 are two hyperparameters, commonly equal to 0.9 and 0.999.

The previously described equations are valid for t=1. Thus, concerning t=0, m_0 and v_0 are equal to 0, leading to a bias. To overcome it, the authors proposed corrected first and second order gradients, namely $\widehat{m}$ and $\hat{v}$, definition:

$$\widehat{m}_t = \frac{m_t}{1 - \beta_1^t} \qquad \text{Equation 58}$$

$$\hat{v}_t = \frac{v_t}{1 - \beta_2^t} \qquad \text{Equation 59}$$

The final parameters θ update is performed with:

$$\theta^t = \theta^{t-1} - \eta \frac{\widehat{m}_t}{\sqrt{\hat{v}_t} + \varepsilon} \qquad \text{Equation 60}$$

With ε a constant, whose possible value is 10^{-8}.

This optimizer has been successfully applied to numerous radiotherapy fields, such as target and OAR segmentation (230), beam orientation for planning optimization (231) and brain pCT generation (232).

2.3.5. Architectures examples

2.3.5.1. UNet

UNet was first introduced in 2015, by Ronneberger et al. (233) and was initially designed for medical image segmentation. Reported network parameters are equal to 7.76M. This network is a 2D fully convolutional neural network composed of two main parts, namely the encoder and decoder.

The former is composed of four blocks, each containing two convolutions. Convolutions consist in applying filters, composed of network weights, to a given image. In this case, the filter, i.e. kernel, size was equal to 3x3. The two convolutions are followed by a ReLu activation function, which was previously described. Last block element is a MaxPooling operation. It is a 2x2 filter selecting the maximum of the four wrapped elements, to decrease by two the features maps sizes. Thus, the encoder results in a decrease of the input image from 572x572 to 32x32. In the mid time, the kernels number increases implying a learning of gradually deep features.

Finally, the decoder has the role to reconstruct the high level 32x32 features maps. To do so, four blocks are consecutively applied. Each block is composed of two convolutions with a 3x3 filter, ReLu activation function and an up-convolution with a filter size of 2x2 to increase feature maps sizes. Two 3x3 and one last 1x1 convolutions are applied to obtain the final classification maps. Finally, note that for each block, skip connections enable features maps to be transferred from the encoder to the decoder ensuring a proper image reconstruction. Figure 34 presents the original network architecture (233).

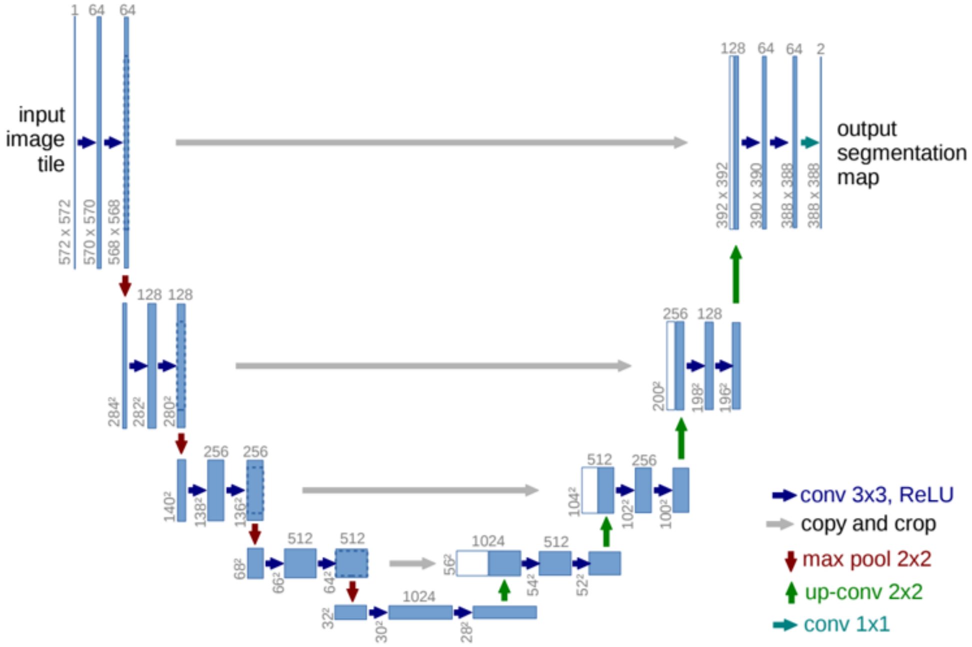

Figure 34: UNet architecture. Reprinted by permission from Springer Nature Customer Service Centre GmbH: Springer, Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015. Lecture Notes in Computer Science, "U-Net: Convolutional Networks for Biomedical Image Segmentation", Ronneberger O, Fischer P, Brox T, Copyright, 2015.

Despite its great success in medical field, UNet faces several limitations, such as for instance difficulties to handle images with various scales. To overcome this issue, Ibtehaz et al. (234) proposed to view the two consecutive 3x3 convolutions present in the architecture as a 5x5 convolution, and to perform 3x3 and 7x7 convolutions in parallel. Second limitation is the low high level information amount contained in the first network layers, resulting in non-informative skip connections and difficulties in construction. A Channel Attention Gate was reported by Khanh et al. (235) to overcome this issue aiming at collecting information from both encoder and decoder to assess meaningful characteristics.

2.3.5.2. Residual neural networks

During error backpropagation, the derivative of the loss function L with respect to the weight w associated to the layer l is proportional to the product of all layers norms deeper than l. Thus, in the case of these norms are inferior to 1 and a large layers number, the error derivative tends to zero. This phenomenon, referred as *vanishing gradient*, results in a significant decrease in the velocity to train the superficial networks layers. To overcome this issue, residual connections have been introduced and consist in summing the output layers block with the input as illustrated in Figure 35 (236).

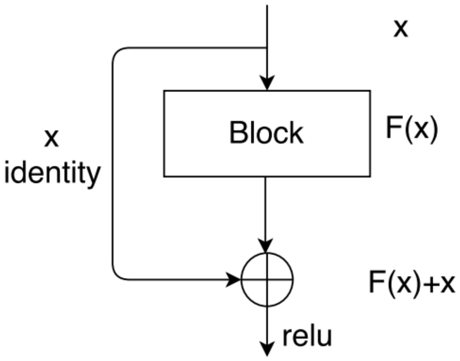

Figure 35: Residual connection summing a block input x to the output F(x). Reprinted from "Camera-Based Blind Spot Detection with a General Purpose Lightweight Neural Network", Zhao Y, Bai L, Lyu Y, Huang X, 2019, Electronics. CC BY (2019).

This concept has been re-used by Li et al. (237) to generate a 3D HighResNet, designed for a classification task. To enable pCT generation, the network was adapted, modifying the output channel number to 1, removing the softmax layer, and applying instance normalization instead of batch normalization layers (238). It is first composed of a sub-block of 3x3x3 convolution, instance normalization and ReLu activation. Then, there are nine residual blocks, each containing two sub-blocks of instance normalization, ReLu and 3x3x3 convolutions with a dilation factor of 1, 2 or 4. Finally, two last sub-blocks containing instance normalization, ReLu and either 3x3x3 convolution with a dilated factor of 1 or 1x1x1 convolution are placed at the end of the architecture. The resulting architecture is presented in Figure 36.

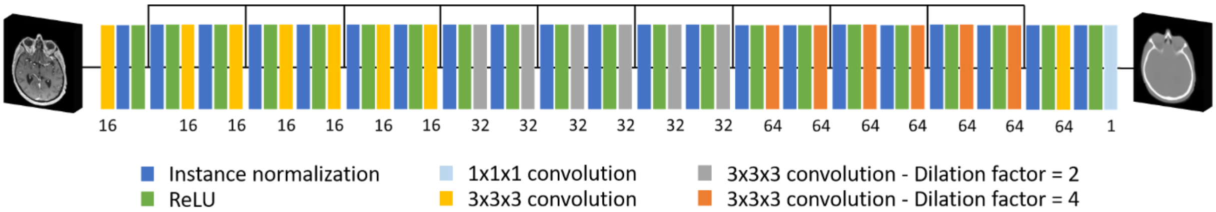

Figure 36: Modified 3D HighResNet.

2.3.5.3. Generative Adversarial Networks (GAN)

GAN are composed of two opposite models. First, the generator takes as input a random noise to compute highly realistic images. Concerning the discriminator, its role is to differentiate between real, i.e. from the training set, and fake images, i.e. from the generator. Its output is a probability for the image to be real. Training a GAN is a dynamic process, in which both the generator and discriminator aim at improving themselves. A balance should exist between the two models performances. Once the training and validation are performed, the discriminator is removed to keep a realistic images generator. Figure 37 presents a GAN illustration (239).

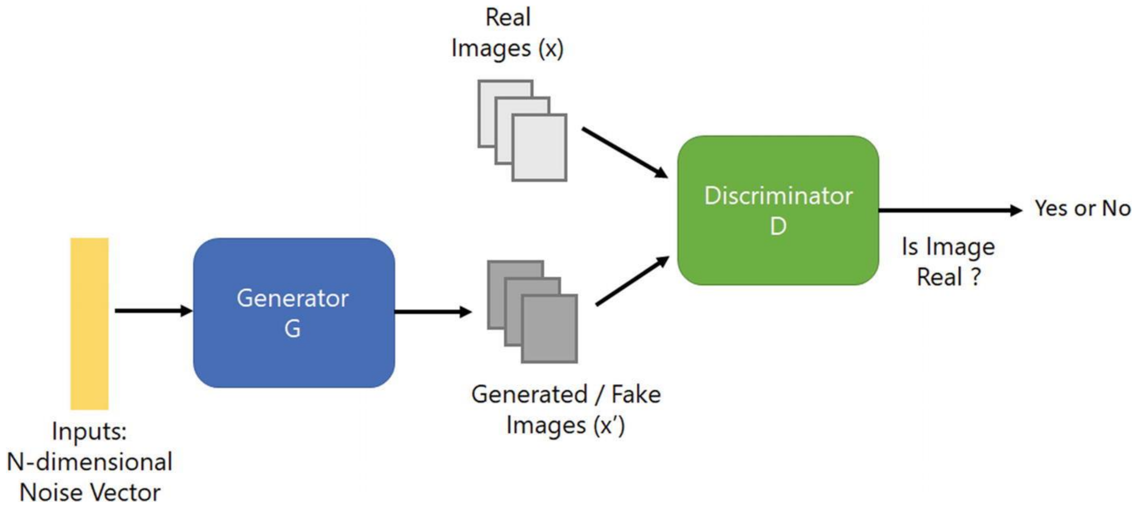

Figure 37: GAN composed of a generator (blue) computing fake images (dark grey) which are compared to ground truth images (light grey) by the discriminator (green) to predict the image type, i.e. real or fake. Reprinted by permission from Springer Nature Customer Service Centre GmbH: Springer, Deep Learning with Azure, “Generative Adversarial Networks”, Salvaris M, Dean D, Tok WH, Copyright, 2018.

Conditional GAN are one evolution of traditional GAN, and were first introduced by Mirza et al. (240). They consist in providing the generator and discriminator with additional information mostly to guide the generation of the fake image from noise ensuring higher performances. Conditional GAN have been widely used in radiotherapy, especially for pCT from MRI synthesis (232,241,242), prostate MRI fiducials detection (243) and linac and MLC treatment plan parameters prediction (244).

An additional GAN evolution is the cycle GAN designed for image-to-image transformation and not requiring paired data. It is composed of two generators and two discriminators. Generator 1 produces images from domain A to B via a function f. Generator 2 takes as input images from domain B to transform them into A with a function g. Associated Discriminators 1 and 2, as in a common GAN, have the role to assess the trueness of input images which are either generated or real. The crucial condition is the bijectivity of f and g. In other words, let image b in domain B represent the transformation of image a from domain A with f. Applying function g on b should lead to the initial image a. More theoretically, a cycle consistency loss is added to the common GAN loss function to ensure the one-to-one mapping (245). Main radiotherapy-related tasks were the generation of pCT from MRI (246–248), image quality improvements such as CT metallic artefact (249) and CBCT artefact (250) reductions.

Concerning architectures, GAN generators are commonly fully convolutional neural networks, i.e. architectures where fully connected layers have been replaced by 1x1 convolutions. Discriminators are simple convolutional neural networks according to Nie et al. (251). This observation is in agreement with GAN strategies in radiotherapy, since reported networks architectures for the generator are the original UNet (252,253) and modified UNet versions (254–256). Regarding the discriminator, consecutive convolutions layers (254) have been used as well as Patch GAN (253,255,256) consisting in predicting 2D maps where each voxel is assigned a probability to be real or fake.

2.4. AI in radiotherapy

AI potential applications in radiotherapy have been widely investigated these past few years. In 2020, Huynh et al. (257) proposed a global overview of AI integration in radiotherapy and the associated consequences on the clinical members staff. For instance, regarding radiation therapists, the treatment session dedicated to each patient will probably be shortened owing to automated set up and beam delivery. Dosimetrists were predicted to spend more time on difficult cases since the easiest ones would be automatically computed. Combinations between AI approaches and radiotherapy workflow tasks were also proposed, such as fully convolutional neural networks for segmentation and GAN for dose distribution prediction.

Regarding the underlying mathematical concepts, they have been introduced by Siddique et al. (258). Technical AI concepts useful for radiotherapy were provided, such as support vector machine and K-nearest neighbour approaches, both potentially playing a key role during segmentation or K-mean clustering to evaluate the patients who would benefit from ART.

Chan et al. (259) investigated the specific integration of AI in Quality Assurance (QA). One of its main roles is to ensure that prescribed and minimum doses are respectively delivered to target and OAR. Concerning machine-related QA, a possible reported AI application consisted in predicting a linac time stability with a neural network composed of one hidden layer with 6 neurons. More generally, ML was proved to have a great potential for both IMRT and VMAT plans QA with tasks such as gamma indices computation, dosimetry-derived errors classification, assessment of factors influencing dosimetry accuracy, etc.

2.4.1. Images synthesis and preprocessing

Many radiotherapy workflow steps can largely benefit from AI. First, images generation, such as synthetic CT (260,261) or synthetic MRI (262,263), has been successfully investigated with the promise to reduce the number of examinations per patient. A DL-based synthetic CT generation pipeline named MRI Planner (Spectronic Medical AB, Helsingborg, Sweden) is already CE-marked, and thus commercially available (section 3.3.3.). An additional application consists in preprocessing the acquired images for an improved quality, such as metal artefact reduction (264,265), still not implemented in clinics.

2.4.2. Brain registration

Regarding brain registration, various lines of research have been proposed tackling both monomodal (266,267) and multimodal transformations (268,269) with DL. A recent study published by Islam et al. (270) consisted in first artificially augmenting data via random rotations and translations. An initial symmetry-based alignment was optionally applied on fixed and moving images, before training a 3D registration convolutional network and developing a second network to classify the input images as fixed or moving. VDSC superior to 0.99 were obtained when evaluating CT to MRI registrations performances, suggesting the high efficiency of the proposed method. Figure 38 presents the registration output for six patients.

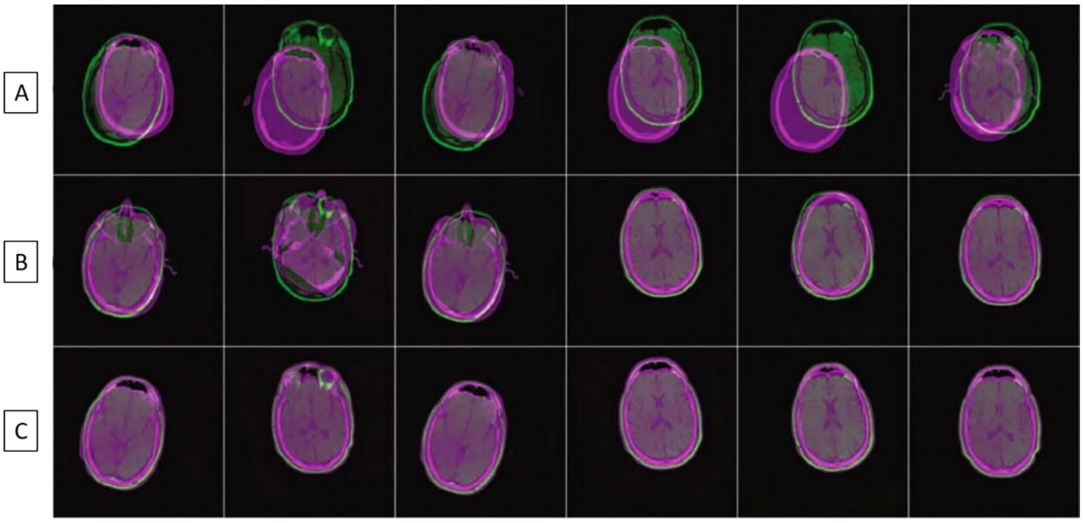

Figure 38: Initial fixed CT (purple) and moving MRI (green) images (A). Registration output without and with the symmetry alignment (B and C). Adapted from “A deep learning based framework for the registration of three dimensional multi-modal medical images of the head”, Islam KT, Wijewickrema S, O’Leary S, 2021, Sci Rep, 11(1). CC BY 4.0 (2021).

2.4.3. Segmentation

Segmentation is also a task where DL has been introduced (271), leading to fast, reproducible and robust delineations. Metastases lesions were segmented by Xue et al. (272) with a 3D fully convolutional network. 1652 brain metastases patients, collected from institution 1 (1201 patients), institution 2 (231 patients) and institution 3 (220 patients), were included. The acquired MRI were 3D T1. For each institution, cross validation was performed splitting the dataset into 4 folds. Three were used for training, wherein random 10% were assigned to validation. The remaining fold was dedicated to test. VDSC performances were equal to 0.85 +/- 0.08, 0.84 +/- 0.07 and 0.83 +/- 0.06 for institutions 1, 2 and 3 respectively. Figure 39 presents qualitative segmentation qualities, for 4 cases.

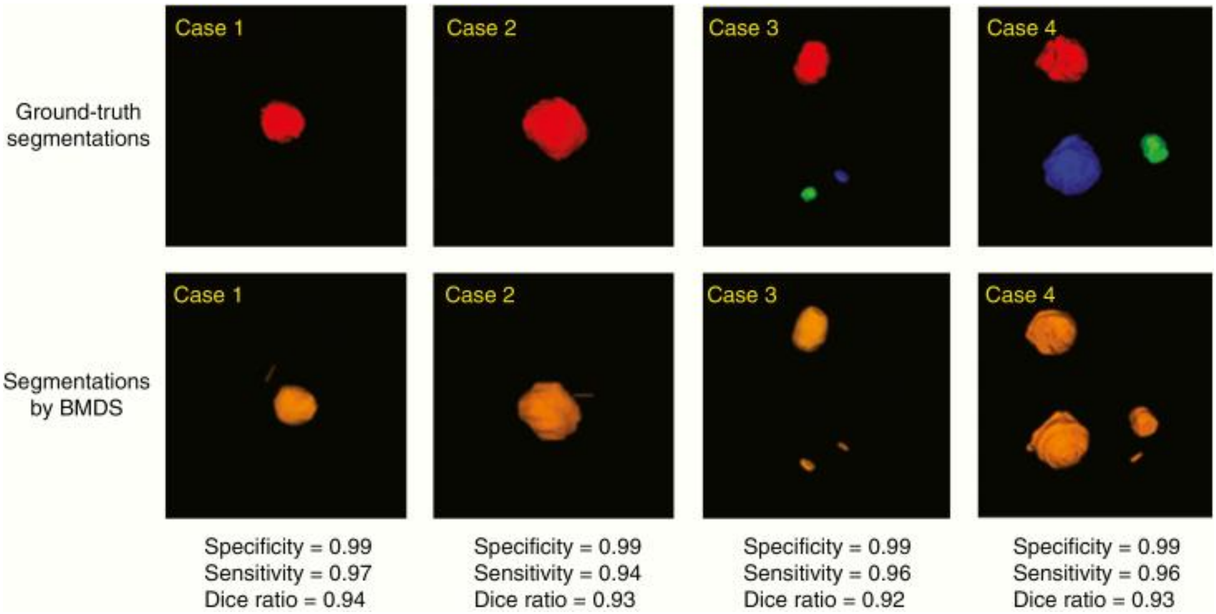

Figure 39: Qualitative performances of Brain Metastases Detection and Segmentation network (BMDS) for four given cases. First and second rows respectively present ground truth and BMDS predicted segmentations. Reprinted from “Deep learning-based detection and segmentation-assisted management of brain metastases”, 22(4), Xue J, Wang B, Ming Y, Liu X, Jiang Z, Wang C, et al., 505-514, Copyright 2020, with permission from Oxford University Press.

Regarding OAR, Mlynarski et al. (273) proposed a modified version of the 2D UNet to segment the brain, hippocampus, brainstem, pituitary gland, eyes, lenses, optic nerves and chiasm from T1 MRI. VDSC comprised between 0.58 and 0.97 were respectively obtained for the pituitary gland and the brain. Similarly based on MRI, Dolz et al. (274) achieved better performance, i.e. VDSC respectively equal to 0.83 +/- 0.06, 0.76 +/- 0.06, 0.77 +/- 0.08, 0.78 +/- 0.05, 0.80 +/- 0.06 for chiasm, pituitary gland and stalk, left and right optic nerves.

Note that hybrid methods, combining different tasks, have also been investigated. Estienne et al. (275) proposed to associate an encoder and two decoders to generate both the reference and moving 3D MRI segmentation masks and the corresponding displacement fields.

Regarding AI-based segmentation commercially available, various software including for instance ART-Plan (TheraPanacea, Paris, France), RayStation (RaySearch Laboratories, Stockholm, Sweden), MVision AI (MVision AI Oy, Helsinki, Finland), are already available.

2.4.4. Dosimetry

DL has been applied to dosimetry enabling a fast and reproducible process. An original two-step method was proposed by Chen et al. in 2019 (276) for head and neck patients treated with step and shoot. First, a residual network predicted a coarse dose map from CT and associated segmentations. Then, a convolution was applied to generate the final fine dose map. Lowest 3%/3mm global gamma index was obtained for the larynx, and was equal to 75.3% +/- 11.9%, highlighting the presented approach potential. The dose map computation time was lower than 10 seconds. More recently, DeepDose was proposed by Kontaxis et al. (277) which is a 3D UNet, to predict dose maps for IMRT-treated prostate patients. 5 input maps were required, namely a ray tracing-based segment mask representing different field sizes effects on scattering in case of non-square beams, distance from source, central beam distance, radiological depth and volume densities. Mean 3%/3mm gamma index with 10% dose threshold of 99.9% +/- 0.3% was achieved. Testing a patient was a fast process and lasted approximately one minute. Low computation times are highly important in oncology, especially for online adaptive radiotherapy where the patient is lying on the table, and a plan recalculation is performed based on the anatomy of the day. Figure 40 illustrates the ground truth and predicted dose maps comparison for a transversal slice.

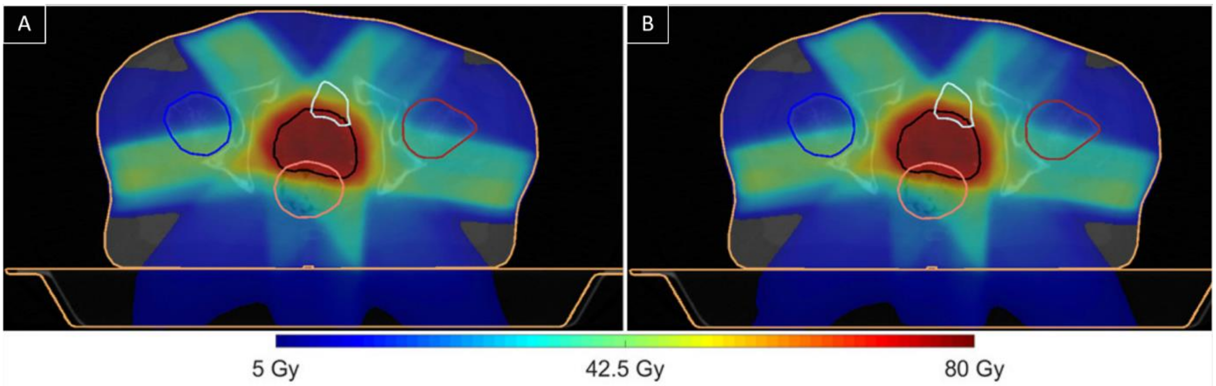

Figure 40: Axial slices of the reference (A) and predicted (B) dose distributions. Femoral heads (blue and red lines), PTV (black line), rectum (orange line) and bladder (white line) segmentations are also presented. From "DeepDose: Towards a fast dose calculation engine for radiation therapy using deep learning", 65(7), Kontaxis C, Bol GH, Lagendijk JJW, Raaymakers BW, Phys Med Biol., 2020. DOI: 10.1088/1361-6560/ab7630. © Institute of Physics and Engineering in Medicine. Adapted by permission of IOP Publishing. All rights reserved.

Concerning clinics, RayStation (RaySearch Laboratories, Stockholm, Sweden) was the first to propose in 2019 a ML-based treatment plan option, which was tested on 100 prostate cancer patients, and recently reported (278). The first 50 patients were all treated with human-derived plans. They were retrospectively compared to ML-derived plans. Concerning the remaining 50 patients, both plans were generated for each patient, the best being blindly assessed by a physician. 72% of the ML-derived plans were finally selected over human ones, suggesting their high accuracy.

2.4.5. QA

QA integration into radiotherapy has been strongly advised by IAEA (279). QA is a task which can be automated via AI. A pioneer research study presented by Tomori et al. (280) consisted in predicting gamma indexes, for prostate patients QA time saving. 2D convolutional neural network predictions were obtained from verification plans on a QA phantom and were compared to measurements deriving from GAFCHROMIC EBT3 films. Global 2%/2mm predicted and measured pass rates with 30% dose threshold were respectively equal to 86.78% +/- 2.55% and 88.03% +/- 2.00%. The associated Spearman coefficient was 0.62 ($p < 0.01$). More generally, all the prediction-measurement comparisons led to a Spearman coefficient superior or equal to 0.51 (p=0.02) except for the 3%/3mm criterion (Spearman coefficient = 0.32, p = 0.16), suggesting the correlation between the two values.

2.4.6. Patient set up

Lastly, patient set up task has been impacted by AI. Studies investigated approaches to improve its imaging quality via DL. In 2018, Kida et al. (281) proposed a modified 2D UNet to learn a correspondence between prostate CBCT and original CT, and generate CBCT_1. They compared it to a CBCT_2 enhancement approach based on a scatter a priori correction. The spatial non-uniformity root mean square differences between CT and original CBCT were equal to 109HU and 57HU in fat and muscle respectively. Concerning CT and CBCT_2 same tissues, differences of 14HU and 7HU were achieved. Differences between CT and proposed improved CBCT_1 were 13HU and 11HU for the same tissue categories, highlighting a spatial uniformity increase, at least in fat. Figure 41 presents qualitative comparison of the three image types.

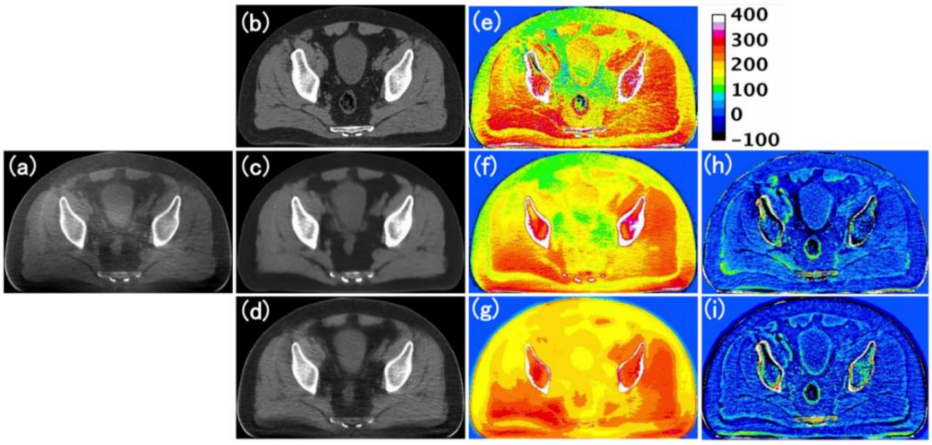

Figure 41: Original CBCT (a), planning CT (b), CBCT_1 (c), CBCT_2 (d), planning CT – original CBCT (e), CBCT_1 – original CBCT (f), CBCT_2 – original CBCT (g), planning CT – CBCT_1 (h), planning CT – CBCT_2 (i). Adapted from "Cone Beam Computed Tomography Image Quality Improvement Using a Deep Convolutional Neural Network", Kida S, Nakamoto T, Nakano M, Nawa K, Haga A, Kotoku J, et al., 2018, Cureus, 10(4). CC (2018).

More recently, Zhang et al. (282) developed a 2.5D GAN with feature mapping which was trained and validated on 150 pairs of pelvic CT-CBCT couples. The derived model was tested on 15 patients, leading to MAE of 43.8HU +/- 6.9HU and 23.6HU +/- 4.5HU for CT-original CBCT and CT- DL CBCT comparisons respectively. Similarly, a test on 10 head and neck patients, was conducted and led to 32.3HU +/- 5.7HU and 24.1HU +/- 3.8HU MAE, suggesting the considerable CBCT quality improvement.

3. Pseudo CT

3.1. Introduction

Generating pCT from MRI has been widely studied these past few years, implying pCT are placed in the MRI spatial frame. No CT/MRI registration is required, theoretically leading to a decrease in the global treatment error. Additional advantages are costs reduction and an improved patient comfort. However, clinical reality, including MRI artefacts, MRI contraindications (e.g. pacemakers wearing patients), claustrophobia, results in a non-straightforward hospitals implementation.

Numerous pCT applications exist. As previously mentioned, online ART, i.e. a re-planning when the patient is on the couch, requires the transformation of the online MRI into a CT to enable a dose calculation and a patient treatment (283). A second application refers to combined Positron Emission Tomography/MRI devices, which are nuclear medicine imaging technologies, and in which the attenuation map computation is needed for uptake quantification (284). Lastly, a recent study published by Zijlstra et al. (285) presented pCT generated in the context of orthopaedics.

In section 3.2., commonly reported metrics for pCT quality evaluation are presented. Section 3.3. focuses on three main pCT generation techniques as well as their associated performances.

3.2. Evaluation tools as summarized by Vandewinckele et al. (1)

3.2.1. Intensity-based metrics

3.2.1.1. Mean Error (ME) and MAE

According to Largent al. (286), ME is calculated as:

$$ME = \frac{1}{N}\sum_{i=1}^{N} pCT_i - CT_i \qquad \text{Equation 61}$$

MAE is defined as:

$$MAE = \frac{1}{N}\sum_{i=1}^{N} |pCT_i - CT_i| \qquad \text{Equation 62}$$

With, for both formulas, N, CT_i, pCT_i the total number of voxels of the considered ROI, CT intensity of voxel i, pCT intensity of voxel i.

For brain pCT evaluation, MAE is commonly computed within the head contour (232,260,261), which is also the case for ME (261,287–289). Additionally, several studies reported these metrics in bone (261,288,289), air and water tissues (261), defined based on a thresholding approach. ME and MAE have the advantage to be easily computed. However, since they are voxel-wise metrics, they are impacted by the registration quality. Indeed, satisfying fusions imply a good CT/pCT correspondence leading to lower ME/MAE than poor quality registrations, where different tissues are superimposed.

Most of the studies report the MAE based on the cohort, reflecting a group performance. Thus, no information regarding mis-reconstructed patients is provided, preventing to elaborate MRI-only radiotherapy workflow exclusion criteria. A possible way to overcome this issue consists in reporting mean and standard MAE, as well as minimum and maximum.

3.2.1.2. Peak Signal to Noise Ratio (PSNR)

PSNR is a similarity metric mostly applied to quantify images reconstruction performance, in case of loss-associated image compression for instance (290). It is calculated as follows:

$$PSNR = 10 \log_{10} \frac{L^2}{MSE} \qquad \text{Equation 63}$$

With L the maximum gray value intensity in the CT (291).

High PSNR reflects low MSE, and thus high correspondence between CT and pCT. On the contrary, low PSNR implies low CT/pCT similarity. One major PSNR limitation is its low correlation with human image visualization and representation (292).

3.2.1.3. Structural SImilarity Metric (SSIM)

To overcome the issue raised by PSNR, SSIM has been introduced. Malpica et al. presented a detailed approach of such metric calculation (293). Let CT_f and pCT_f represent the two full images to be compared. The calculation is performed in a sliding-window fashion, usually of size 11x11. Let CT and pCT represent the CT_f and pCT_f extracted windows respectively. First, luminance similarity l is defined as:

$$l(CT, pCT) = \frac{2\mu_{CT_f}(CT, pCT)\mu_{pCT_f}(CT, pCT) + C_1}{\mu_{CT_f}(CT, pCT)^2 + \mu_{pCT_f}(ct, pCT)^2 + C_1} \qquad \text{Equation 64}$$

Where C_1 refers to a constant. μ_I associated to image I is computed as:

$$\mu_I(CT, pCT) = \sum_{p=-P}^{P} \sum_{q=-Q}^{Q} w(p, q) I(CT + p, pCT + q)$$

Where w is a weighting function such that: $\sum_{p=-P}^{P}\sum_{q=-Q}^{Q} w(p,q) = 1$.

Second function is the contrast similarity c, and is computed as:

$$c(CT,pCT) = \frac{2\sigma_{CT_f}(CT,pCT)\sigma_{pCT_f}(CT,pCT) + C_2}{\sigma_{CT_f}(CT,pCT)^2 + \sigma_{pCT_f}(CT,pCT)^2 + C_2} \qquad \text{Equation 65}$$

With C_2 is a constant. σ_I is the contrast of image I obtained calculating the standard deviation, as follows:

$$\sigma_I^2(CT,pCT) = \sum_{p=-P}^{P}\sum_{q=-Q}^{Q} w(p,q)[I(CT+p,pCT+q) - \mu_I(CT,pCT)]^2$$

Lastly, the structural similarity function s is defined:

$$s(CT,pCT) = \frac{2\sigma_{CT_f\,pCT_f}(CT,pCT) + C_3}{\sigma_{CT_f}(CT,pCT) + \sigma_{pCT_f}(CT,pCT) + C_3} \qquad \text{Equation 66}$$

With C_3 a constant and $\sigma_{CT_f\,pCT_f}$ is calculated as:

$$\sigma_{CT_f\,pCT_f}(CT,pCT) = \sum_{p=-P}^{P}\sum_{q=-Q}^{Q} w(p,q)\left[CT_f(CT+p,pCT+q) - \mu_{CT_f}(CT,pCT)\right][pCT_f(CT+p,pCT + q) - \mu_{pCT_f}(CT,pCT)]$$

Structural comparison relies on the ZMUV approach, obtained subscripting the mean μ to the window, and then dividing it by the standard deviation σ.

The final SSIM is finally obtained as follows:

$$SSIM(CT,pCT) = l(CT,pCT)\;c(CT,pCT)\;s(CT,pCT) \qquad \text{Equation 67}$$

SSIM is then iteratively calculated for other windows. The final value ranges from -1 to 1, respectively referring to low and high degrees of similarity between CT and pCT. This metric has been reported to face several limitations when used in the medical imaging field, such as undervaluing distortions next to certain edges for instance (294).

3.2.1.4. VDSC

Last intensity-based metric is VDSC and has been previously defined. It is commonly assessed in bone regions. The latter is commonly obtained thresholding the image, with values of 200HU (232,288,295), 250HU (261,289), 300HU (296) and 500HU (297).

3.2.2. Dosimetry-based metrics

Most of the dosimetry-based studies (232,255,261,287,289,295,298) transferred the treatment plan from the initial brain CT onto the pCT for evaluation, enabling a direct estimation of the proposed model performance. Tang et al. (299) performed both the plan transfer and the optimization on the pCT, investigating the dosimetry differences between the two approaches. Lastly, Prabhakar et al. (300) and Weber et al. (301) generated treatment plans optimized on pCT. The most common evaluation metrics used to assess the resulting dosimetry performance are presented in the following sections.

3.2.2.1. Gamma index

Gamma index summarizes global dosimetry performances. More precisely, it enables an evaluation of the dosimetry agreement between two dose maps, in the present case a reference CT and a testing pCT. A 1D illustration is presented in Figure 42, and was provided by Hussein et al. (302).

First, two acceptability criteria δr and δD are defined, respectively associated to the distance (in mm) and the dose (in Gy). They are reported on the graph, respectively on the r and D axes. Let the reference CT point be represented as a blue cross point at the centre of the figure. The tested pCT-based dose curve corresponds to the blue curved line. To simplify the problem, only three pCT tested points are defined: $D_E(r_R)$, $D_E(r_E)$ and $D_E(r_{E1})$. The gamma index algorithm iteratively evaluates the distances between the reference point and the different tested points. A reference point passes the test if it exists a combination (reference point, tested point) which satisfies the two criteria previously defined, i.e. a spatial distance lower than δr and a dose difference lower than δD. More concretely, it implies that only a tested point located within the ellipse (blue dotted line), representing the combination of the Euclidian distance and dose difference criteria, passes the test. In Figure 42, ($D_R(r_R)$, $D_E(r_E)$) passes while ($D_R(r_R)$, $D_E(r_R)$) and ($D_R(r_R)$, $D_E(r_{E1})$) fail.

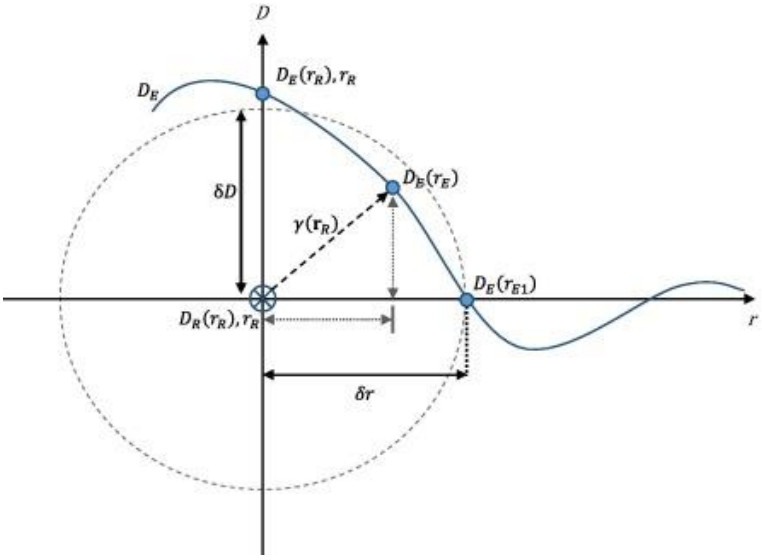

Figure 42: 1D gamma index calculation principle. The cross point, blue line, dotted line represent the reference point, the tested curve and the acceptance ellipse respectively. Reprinted from “Challenges in calculation of the gamma index in radiotherapy – Towards good practice”, 36, Hussein M, Clark CH, Nisbet A, Phys Med., 1-11, Copyright 2017, with permission from Elsevier.

More formally, the gamma value γ for a CT reference point R_{CT} is calculated as follow (302):

$$\gamma(R_{CT}) = \min\left(\sqrt{\frac{\Delta \mathrm{D}^2(R_{CT}, R_{pCT})}{\delta \mathrm{D}^2} + \frac{\Delta \mathrm{r}^2(R_{CT}, R_{pCT})}{\delta \mathrm{r}^2}};\ \forall\ R_{pCT}\right) \qquad \text{Equation 68}$$

With R_{pCT} the evaluated pCT dose map point, ΔD the dose difference between the reference and tested points doses, Δr the distance between the reference point and the tested point, δD the dose acceptance criterion, δr the distance acceptance criterion.

Results can either be presented with gamma values maps. Additionally, a gamma index can be reported, corresponding to the percentage of points passing the test, i.e. points with gamma values γ lower than 1. Thus, the objective is to obtain a gamma index close to 100%.

Typical δr values are equal to 1mm, 2mm, 3mm. Similarly, widely used δD values are 1%, 2% and 3% of the dose. In a global gamma index approach, the dose percentage is applied to the reference dose matrix maximum. For a local gamma index, the dose percentage of the reference point dose is considered. As a result, low doses located in OAR easily pass the gamma index criterion for a global approach, leading to a less restrictive test than local. Applying a dose threshold on the CT dose map is commonly performed to disregard all these low dose regions and focus on target volumes. If the latter is small, it is advantaged with low thresholds, since many points are considered. A dose error in the target volume would be compensated by the surrounding healthy tissues gamma indices. On the contrary, large PTV benefit from high dose thresholds, since a target volume dose error is compensated by the remaining PTV points.

3.2.2.2. DVH differences

Relative DVH-based differences ΔD are commonly performed between CT and pCT with:

$$\Delta D = 100 \times \frac{|D_{CT} - D_{pCT}|}{D_{CT}} \quad \text{Equation 69}$$

Where D_{CT} and D_{pCT} the DVH-derived metrics associated to CT and pCT respectively.

PTV differences are usually based on D95% and D98% metrics, referring to minimum doses received by 95% and 98% of the PTV respectively. Concerning optic nerves, chiasma, brainstem and lenses, the metric is Dmax, representing the maximum dose received by the structure. Lastly, cochlea DVH differences are assessed between Dmean, i.e. the cochlea structures mean dose.

DVH differences are accurate for a specific structure dose performance investigation. However, contrary to gamma values maps, no spatial information is provided.

3.2.2.3. IGRT accuracy

IGRT accuracy has been increasingly computed, since set up performances assessment is vital for an end-to-end MRI-only radiotherapy workflow. Two main approaches have been described so far in the literature. The first consists in rigidly registering CBCT to CT and pCT (289,298). Second approach is 2D, and relies on the rigid registration of 2D kV images to DRR extracted from CT and pCT (298). In both approaches, registrations are compared to assess discrepancies in the six degrees of freedom, namely translations along x, y, z and pitch, roll and yaw rotations.

3.3. Literature review

3.3.1. Bulk Density Assignment (BDA)

Bulk Density Assignment (BDA) first requires segmenting the MRI into categories of tissues. A pre-defined ED is then assigned to each delineated area.

Several studies included one segmented category, i.e. water, such as Prabhakar et al. in 2007 (303). A total of 25 brain tumor patients were enrolled in the study. Three different plans were compared, namely CT without heterogeneity correction (method 1), CT combined with heterogeneity correction (method 2) and MRI-only (method 3). In method 3, a water-equivalent pCT was generated assigning the value of 0HU to the external contour. Plans were calculated with pencil beam. Mean doses of 98.85% +/- 0.76%, 98.97% +/- 0.70% and 98.89% +/- 0.71% were obtained for methods 1, 2, 3 respectively leading to a p-value of 0.492 and a non-significance of the observed differences.

Similarly, in 2008, Weber et al. (301) collected data from ten glioblastoma patients to compare different approaches, namely a dosimetry based on CT without heterogeneity correction (method 1), a dosimetry based on CT with heterogeneity correction (method 2) and a MR-based dosimetry (method 3). The latter approach consisted in assigning a water density to the MRI. Maximum percentage differences between doses per fraction for methods 1-2, 1-3, 2-3 were equal to 0.4% +/- 0.1%, 0.3% +/- 0.9%, -0.01% +/- 0.4% proving the feasibility of the MR-only radiotherapy workflow.

Yet, assigning ED to only one tissue category is not optimal, as proved by Kristensen et al. (304). They investigated three different dosimetry methods, namely CT-based (method 1), MR-based with unit water density assigned (method 2) and MR-based with bone and water assigned densities (method 3). Bone assigned value was set to 1.61g/cm^3. Plans were optimized on CT combined with a heterogeneity correction. The workflow was tested for eleven patients with brain tumors. The isocentre dose differences between methods 1 and 2, 1 and 3, led to percentages of 1.4% +/- 0.4% and 0.4% +/- 0.3% respectively, suggesting the dual ED assignment superiority.

Stanescu et al. (305) also applied a two categories-based pCT generation to compare the standard MRI and CT-based workflow to an MR-only process. Brain, scalp and bone of four patients suffering from glioblastoma multiforme were semi-automatically contoured. Brain and scalp were considered as water equivalent. Bone was assigned the density of 1.47g/cm^3. Treatment plans were optimized using CT and MRI data, before being transferred to the generated pCT. DVH differences were computed for the OAR and did not exceed 1% potentially suggesting the equivalence of the two compared approaches. In 2008, they performed a similar study (306) and enrolled the same types of patients. Comparison was made between the standard workflow based on both CT and MRI plan uncorrected from distortion and the MRI-only distortion-corrected approach. The latter was generated applying a distortion correction on the MRI and assigning to scalp, brain and bone regions the ED corresponding to 0HU, 0HU and 1000HU. As in their previous study, treatment plans were not optimized on pCT. Differences in Dmax between both methods were below 1% for all the four patients, proving the MRI-based workflow feasibility.

The bulk density approach presents several disadvantages. First, it is based on the contouring of various tissues categories, raising the issue of quality, reliability and reproducibility, even though an increasing number of software currently offer the ability to automatically delineate. Moreover, segmenting the bone from conventional MR sequences, is not straightforward. Indeed, bone T1 and T2 range from 0.001 to 1ms (307), i.e. are low. Thus, standard MRI sequences cannot capture contrasted bony structures. One possibility to overcome this issue may consist in acquiring exotic sequences such as Utrashort Echo Time (UTE), presenting TE ranging from 0.008ms to 0.50ms (308). Recently, Zero Echo Time MR sequences have been introduced, offering a TE of 8µs and a better SNR than UTE (309). In this case, the bulk density approach can be time-consuming since it can require extra MRI, not routinely acquired.

In 2016, Philips presented its first commercially available Magnetic Resonance for Calculating ATtenuation (MRCAT) software for prostate tumors location. The MRI input images are T1-weighted mDixon and T2-weighted Turbo Spin Echo. Dixon approach enables four contrasts MRI, namely in phase, i.e. water + fat, opposed phase, i.e. water – fat, water-only and fat-only. The mDixon approach proposed by Philips is a modified version of Dixon, relying on a single image reconstruction to speed up the acquisition process. The bulk density assignment relies on the segmentation of five categories, namely air, compact bone, spongy bone, fat and water. Bratova et al. (310) investigated the quality of the computed image, compared with a water equivalent pCT involving 10 prostate tumors patients. The 2D gamma indexes were larger than 79.4% and 97.5% with a water equivalent approach and MRCAT respectively, for 7 patients out of 10. It proved the dosimetry accuracy of such approach. In 2019, Philips launched MRCAT Pelvis (Philips Healthcare, Best, The Netherlands), suitable for both male and female patients. An evolution of the generation method has been recently proposed to enhance pCT quality and consists in continuous bulk density assignments (311). Indeed, water and fat Dixon MRI are used to define water and fat clusters and associated centres. The continuous function passing through the two clusters centres is assessed, enabling the ED assignment of intermediate intensity voxels located outside of water and fat clusters. Regarding bone, the final assigned ED value depends on the distance between the intensity point, which does not belong to the water-rich and fat-rich tissues clusters, and the mapping function previously determined. The whole cohort was composed of 77 prostate, 43 rectum and 27 gynaecological patients. A dosimetry analysis involving IMRT and VMAT plans was performed for further evaluation. The reported MAE on 12 patients was equal to 38.8HU +/- 4.0HU for the body contour. PTV and OAR averages DVH differences were all below 1%, highlighting the accuracy of the proposed method.

Regarding Elekta (Stockholm, Sweden), as mentioned in Section 1.4.2., pCT are required in case of online ART. In the *Adapt to shape* approach, bulk ED are assigned to each contour. They correspond to the ED mean wrapped by the deformed contours, potentially adapted to the online MRI-derived patient anatomy, on the planning CT (191). Recently, Snyder et al. (312) reported the commissioning of the Elekta Unity (Elekta, Stockholm, Sweden), including a dosimetry test of the *Adapt to shape* method. It was compared to another technique, namely *Adapt to position*, which does not require a pCT generation. The latter only accounts for translations errors and relies on the rigid registration of the planning CT with the online MRI. The isocentre position is then updated, before re-optimizing or re-launching the dose calculation. This approach does not necessitate new contours, and thus does not consider the patient anatomy “of the day”. For a head and neck irradiation delivered with IMRT, global 3%/2mm gamma indices with 10% dose threshold were equal to 99.6% and 99.7% respectively with *Adapt to position* and *Adapt to shape* approaches. Corresponding dose re-calculations durations were respectively 171s and 460s.

Siemens has developed their syngo.via RT Image Suite which presents the ability to generate pCT for brain and pelvis. Regarding brain tumors, four MR sequences are required, namely T1 Vibe Dixon, T2 SPACE, PETRA and FLASH Gradient Echo. They are then classified into water, fat, white matter, gray matter, bone. Concerning pelvis, a T1 Vibe Dixon is acquired before segmenting water, fat, bone and air. Both types of pCT generation are reported to last approximately one minute (313).

3.3.2. Atlas and multi-atlas

In the pCT generation context, an atlas represents a combination of a CT and a MRI for a given patient. Figure 43 presents the global pCT generation workflow (314). First, the CT atlas(es) are rigidly registered to the MRI atlas(es) (step 1). Second step consists in applying a deformable transformation to register the MRI atlas(es) onto the patient MRI (step 2). The output deformation fields are then applied to the CT atlases to create multiple pCT (step 3). Finally, these various pCT are combined to generate a single pCT (step 4). Various combinations approaches have been proposed in the literature, such as the median (315), mean (316) or a weighted average based on similarity (317).

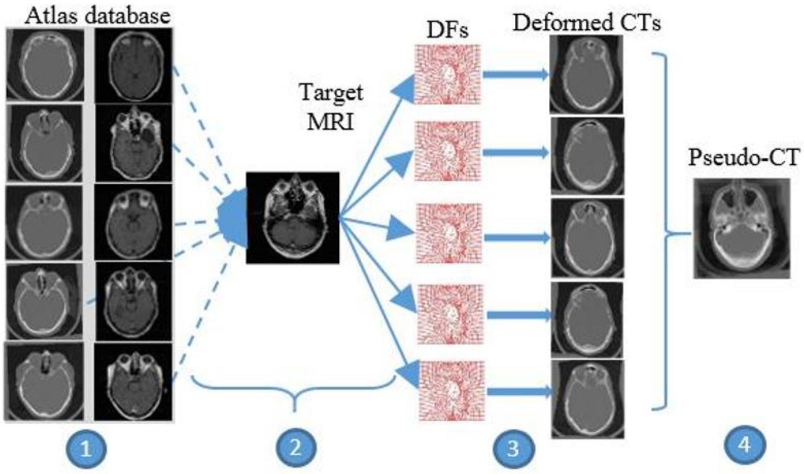

Figure 43: Scheme of the multi-atlas workflow, including the registration of the CT atlases on the MRI atlases (step 1), deformable registration of the MRI atlases onto the target MRI (step 2), applying the output deformation fields onto the CT atlases (step 3) and combining the pCT to obtain the final pCT (step 4). Reprinted from “Pseudo-CT Generated Through Multi-Metric Image Registration and Atlas Fusion (Application to T1-Weighted Brain MRI)”, Boukellouz W, Moussaoui A, Taleb-Ahmed A, 2017. Copyright © 2017, IEEE.

The simplest approach consists in constituting atlases cohorts composed of a single CT and MRI couple. It is referred as single atlas method, and has been investigated in 2016 by Demol et al. (318). They enrolled twenty-two brain tumors patients, and selected the atlas patient based on densities, artefacts criteria. The different pCT generations were applied. First, a 5x5x5 patch was centred on each MRI voxel of interest. The associated atlas CT wrapped voxels were used to generate the pCT values in a weighted average fashion based on the inverse distance. Second, a 9x9x9 box centred on each patient MRI voxel was applied, combined with a 10% threshold to select the voxels under consideration. These voxels were then weighted averaged (as previously performed) from the warped atlas CT to obtain the final pCT voxel value. Concerning dosimetry, no optimization on the pCT was performed. Soft tissues presented a ME inferior to 50HU. The D98 PTV differences were respectively equal to -0.5% +/- 0.38% and -0.1% +/- 0.35% for methods based on 5x5x5 and 9x9x9 patches highlighting the superiority of the latter.

To enhance image qualities, a noise removal filter was applied by Demol et al. (318) on patients CT and pCT and on atlas images by Chegeni et al. (315). They selected 10 glioblastoma and 10 other brain tumors patients. In each category, one patient was kept apart, and the nine remaining patients were assigned to the atlases cohort. A 2D deformable registration inspired from Demons was used to deform the atlas MRI onto the incoming MRI. The combination of all warped CT into a final pCT was performed via the use of the median per pixel. Concerning the glioblastoma group, T1 MRI and CT-based brain windowing combination led to a MAE of 80HU +/- 10.8HU and bone VDSC of 0.83 +/- 0.05. Concerning the other tumor group, 80HU +/- 9.7HU and 0.85 +/- 0.04 were respectively achieved for the MAE and bone VDSC.

The superiority of multi-atlas over single atlas was proved by Uh et al. (316), in 2014. Twelve and 14 brain tumor patients were respectively included in the multi-atlas and testing cohorts. They adopted five different approaches to generate the pCT. First, single atlas and 6-atlas strategies were adopted, combined with an arithmetic mean final voxel combination for the latter. Second, a method reported as “Pattern Recognition with Gaussian Process” by the authors was used. For a given voxel under consideration, all the neighbouring voxels encompassed by a square matrix were considered for the weighted average to compute the final pCT value. The weights were derived from comparisons between patches extracted from atlases and patient MRI. This approach was either applied with 6 or 12 atlases patients. Lastly, a water equivalent pCT was obtained assigning a 0HU to the body outline. Dosimetry was performed transferring the CT dosimetry plan to pCT. Reported root mean square differences were respectively equal to 391HU +/- 30HU, 346HU +/- 81HU, 224HU +/- 36HU, 219HU +/- 35HU and 207HU +/- 33HU for water, single atlas, 6-atlase approach, pattern recognition with gaussian process based on 6 or 12 atlases patients. D95 and Dmean PTV DVH differences were respectively inferior to 1.8% and 1.3% for all the methods based on the arithmetic mean and pattern recognition with gaussian process.

The strategy of pCT combination is key in multi-atlas strategies, and has been explored by Sjolund et al. (319). To this aim, ten patients treated for a brain tumor with stereotactic radiosurgery were retrieved. An important preprocessing step consisted in masking the frame, which was highly visible on CT, to avoid affecting final pCT quality. Five different pCT generation strategies have been proposed, including the standard mean per voxel combination (method 1) and median per voxel combination (method 2). The authors additionally proposed a new combination approach, consisting in iteratively calculating the various warped CT means per voxel, registrating the warped CT to this mean image, re-estimating a new mean, etc (method 3). A similar approach based on the median was investigated (method 4). Lastly, authors compared pCT quality to the lowest MAE single deformed CT (method 5). Reported foreground MAE were respectively equal to 123.2HU +/- 18.6HU, 113.4HU +/- 17.6HU, 117.9HU +/- 20.3HU, 114.5HU +/- 20.5HU and 125.6HU +/- 13.4HU for methods 1, 2, 3, 4 and 5 respectively. Method 2 was reported to significantly outperform method 3 (p=0.0098) and method 5 (p=0.0029).

Boukellouz et al. (320) also compared mean and median fusion approaches based on eleven patients with brain CT/T2 images with previous contrast agent injection. A leave one out cross validation was used to assess all uncorrected pCT. Then, a HU-correction method was used consisting in applying weights based on MRI intensities, tissues classes and information retrieved spatially to uncorrected pCT voxels. Normalized cross correlations between CT and pCT averaged for all patients were equal to 0.92 and 0.91 for the mean and median approaches respectively. In addition, reported bone VDSC were 0.68 and 0.72 for the two strategies, which seemed to be in agreement with Sjolund et al. (319).

More generally, atlas and multi-atlas methods are limited, since the total computation time is directly proportional to the number of patients constituting the atlases. Thus, pCT quality is a balance between clinical acceptable computing durations and atlases cohort sizes. In addition, the atlas patients choice should be performed with caution, as it should be heterogeneous and representative of the global patient morphologies diversity. To our knowledge, no CE-marked software based on such approach is available.

3.3.3. DL

In 2017, a study performed by Han (260) was one of the first to explore brain pCT generation with DL. Eighteen patients who underwent stereotactic radiosurgery were selected providing CT/T1 couples, acquired with the same devices for the whole cohort, which were rigidly registered. The network consisted in a modified version of the 2D UNet architecture, initially designed for classification tasks. The optimization strategy was on Adam enabling to backpropagate the error derived from the MAE loss function. A 6-fold cross validation was performed, assigning 5 and 1 folds to the training and testing cohorts respectively. Achieved MAE and MSE were respectively equal to 84.8HU +/- 17.3HU and 188.6HU^2 +/- 33.7HU^2.

The 2D UNet architecture has indeed been proved to be highly efficient, since it is equivalent or even moderately superior to 3D. Neppl et al. (287) retrieved 89 patients treated for brain lesions with photons. Investigated networks were the 2D and 3D UNets. A dosimetry analysis was performed transferring treatment plans to the pCT. 57, 28 and 4 patients were respectively assigned to the training, validation and testing sets. Reported head MAE ranged from 82HU to 135HU and from 82HU to 147HU for the 2D and 3D UNets respectively. In parallel, a dosimetry analysis was performed applying VMAT treatment delivery. Concerning the two architectures, 3D global 1%/1mm and 2%/2mm gamma indexes with a 20% dose threshold were respectively superior to 95% and 98%.

However, the lowest MAE reported in the literature were achieved by Kazemifar et al. (232) with a 2D cycle GAN. Seventy-seven brain tumor patients with sizes comprised between 1.1cm^3 and 42.4cm^3 and treated with radiotherapy were enrolled. 82% of the data were assigned to training and validation, and used to perform a 5-fold cross validation. The testing cohort was composed of the remaining data. The GAN generator and discriminator were respectively the UNet and the combination of six convolutional and 5 dense layers. The corresponding loss functions were the mutual information and the binary cross entropy. Concerning dosimetry, VMAT CT original plans were transferred to pCT and re-calculated. MAE computed within head contour was equal to 47.2HU +/- 11.0HU. Air and bone VDSC were respectively equal to 70% +/- 7% and 80% +/- 6%. 99.2% +/- 0.8% and 94.6% +/- 2.9% were respectively achieved for 2%/2mm and 1%/1mm 3D gamma indexes. No information regarding the global/local approach, as well as dose threshold was provided. Li et al. (321) adopted the same type of architecture, namely a 2D cycle GAN, but retrieved data from two hospitals, and three different MRI protocols, leading to a highly heterogeneous cohort. Hospital 1 provided couples of CT and T2 images of 28 patients, acquired with the same devices. Concerning hospital 2, 19 and 24 couples of CT/T1 FLAIR and CT/T1-Gd were collected, with varying MRI acquisition devices and parameters. Four different scenarios were studied to train and test the network. Scenario 1 consisted in training and testing the network with hospitals 1 and 2 data respectively. The testing cohort was either composed of 5 T1 FLAIR or 6 T1-Gd patients. Scenario 2 was the reverse of scenario 1, i.e. training and testing on images from hospitals 2 and 1. The training cohort enrolled 14 T1 FLAIR or 18 T1-Gd. Scenario 3 was an hybrid approach, for which the training was based on the 28 patients from hospital 1 and either 14 T1 FLAIR or 18 T1-Gd MRI from hospital 2. Corresponding testing sets was composed of 5 T1 FLAIR or 6 T1-Gd respectively. Lastly, Scenario 4 was based on transfer learning. A first training with the whole hospital 1 data led to a model which was then re-trained with 14 T1 FLAIR or 18 T1-Gd. The associated testing cohorts were constituted of 5 T1 FLAIR and 6 T1-Gd patients respectively. MAE of 126.81HU +/- 14.97HU, 95.13HU +/- 8.70HU, 94.17HU +/- 8.07HU and 74.56HU +/- 8.61HU were respectively achieved for scenarios 1, 2, 3 and 4 when the sub-dataset of T1-

FLAIR was considered. Regarding T1-Gd data from hospital 2, MAE equal to 109.94HU +/- 5.67HU, 105.15HU +/- 14.01HU, 88.50HU +/- 24.93HU and 74.89HU +/- 15.64HU, suggesting the superiority of the transfer learning-based approach.

The previously mentioned study built a DL model, including T2, T1-FLAIR and T1-Gd, raising the issue of the best suited MRI sequence for brain pCT generation. Massa et al. (297) recently investigated the topic, collecting 92 patients with a CT , T1, T1-Gd, T2 FLAIR and T2 where fat signal has been removed (T2 FatSat). The selected network architecture was a modified version of the UNet. The network optimization relied on Adam, combined with a learning rate of 10^{-4}. Regarding the training and testing sets, they were composed of 81 and 11 patients respectively. An 8-fold cross validation was further performed among training cohort patients. Head SSIM were respectively equal to 0.6107 +/- 0.0395, 0.6453 +/- 0.0466, 0.635 +/- 0.033, 0.6291 +/- 0.0291 for T2 FLAIR, T1, T1-Gd, T2 FatSat. In the same order, reported rates for PSNR in air were 43.813 +/- 1.57, 42.98 +/- 1.668, 43.241 +/- 1.364 and 43.077 +/- 1.29, proving the equivalence of the MRI sequences.

Instead of providing single MRI sequences to the network, Koike et al. (255) decided to combine T1, T2 and T2 FLAIR images from 15 patients into a 3-channel input. However, to enable performances comparisons, single input composed of T1 MRI were also used to feed the network. The latter was a 2D conditional GAN, composed of a UNet and a PatchGAN-based architectures as generator and discriminator respectively. To overcome the low number of patients, data were synthetically augmented via flips, translations, zooms and rotations. 3DCRT and VMAT plans, both combined with dose boosts, were transferred to pCT re-calculated with the TPS Eclipse (Varian Medical System, Palo Alto, USA). MAE for single and 3 channel-based input were respectively equal to 120.1HU +/- 20.4HU and 108.1HU +/- 24.0HU in the head, 399.4HU +/- 51.8HU and 366.2HU +/- 62.0HU. 2D 1%/1mm gamma indices were 94.2% +/- 4.9% and 95.3% +/- 4.7% for single and multi-channels approaches. Relative differences between CT and pCT DVH of PTV and OAR were all inferior to 1%. It highlighted a slight increase in the overall pCT quality when combining multiple MRI as input.

Maspero et al. (295) outperformed Koike et al. (255), since three 2D conditional GAN were trained in parallel, one for each view, instead of one. A modified 2D UNet and PatchGAN represented the generator and discriminator respectively. Predicted pCT were finally combined calculating the median for each voxel. 60 paediatrics were retrieved, leading to 15452, 11584 and 17456 slices in the axial, sagittal and coronal planes respectively. CT parameters, including tube kilovoltage, current, exposure time varied for each patient, ranging from 90kVp to 120KVp, 94mA to 324mA and 923ms to 1712ms respectively. Concerning MRI, field strengths were 1.5T and 3T for 36 and 24 patients respectively. Contrast agent was injected for only 22 patients. In addition, VMAT plans, composed of 1 or 2 arcs, were transferred and re-calculated on pCT using a MC dose algorithm. Coronal, sagittal, axial and multi-views respectively resulted in head MAE of 69HU +/- 15HU, 70HU +/- 15HU, 73HU +/- 17HU and 61HU +/- 14HU. The multi-views additionally led to 3D global 2%/2mm gamma index of 99.5% +/- 0.8% when applying a 10% dose threshold, proving the efficiency of the approach.

Dinkla et al. (261) also adopted a 2.5D approach to build a DL model from a 10-layer convolutional neural network. Kernel dilations were used for layers 2 to 7, with a factor varying from 2 to 32. The optimization process was based on Adam. Regarding data, 52 patients treated with head radiotherapy were enrolled, providing couples of CT/T1 acquired on the same devices. 26 and 26 patients were assigned to the training and testing sets. A two fold-cross evaluation was performed, leading to predictions of 52 pCT in total. A dosimetry analysis was performed transferring and re-calculating CT plans to pCT with MC dose algorithm. Treatment deliveries were VMAT and IMRT for 43 and 9 patients respectively. Head, bone, water and air ME were respectively equal to 13HU +/- 9HU, 75HU +/- 41HU, -2HU +/- 3HU, -72HU +/- 27HU. Head MAE was reported to be 67HU +/- 11HU. VDSC of 0.85 +/- 0.04 and 0.71 +/- 0.07 were achieved for bone and air. 2%/2mm gamma indices were 99.1% +/- 0.8% and 99.8% +/- 0.7% when applying dose thresholds of 10% and 50% respectively.

Similar performances in terms of MAE were achieved by Tang et al. (299), but with significantly less brain tumors patients in the cohort, i.e. 37. The network consisted in a conditional GAN, composed of a modified UNet and a series of 5 convolutional layers for the generator and discriminator respectively. Adam optimizer combined with a learning rate of 0.0002 were applied. Dataset was split into training (27 patients) and testing sets (10 patients). Five fold-cross validation was performed among training set patients, resulting in 4-fold training set and 1-fold validation set. Regarding dosimetry analysis, VMAT plans, as well as segmentations, were transferred to pCT. Obtained head MAE and PSNR were respectively equal to 60.52HU +/- 13.32HU and 49.23dB +/- 1.92dB over all 5 folds. Relative mean DVH differences for PTV and OAR were all comprised between -0.77% and 1.33%. Mean gamma indices of 99.76% and 97.25% were respectively achieved for 3%/3mm and 2%/2mm criteria. In parallel, pCT-based treatment plans optimization was performed, leading to means equal to 99.96% and 97.99% for 3%/3mm and 2%/2mm criteria respectively. It might suggest the potential equivalence of the two optimization approaches.

As one can notice, pCT evaluations studies commonly adopt a hybrid approach, i.e. report both intensity-based and dosimetry metrics. To our knowledge, Liu et al. (298) were the first to propose in 2021 a study reporting only dosimetry performances. 12 brain tumor patients were enrolled, and more precisely, 6 and 6 patients treated with conventional and stereotactic radiotherapy respectively. CT and T1-Gd were acquired on the same devices. A GAN composed a 9 residual block-generator and a 5 convolution layer-discriminator was used to generate pCT. Concerning dosimetry, 3 patients benefited from a boost, resulting in a total of 15 plans. Treatment deliveries were VMAT and DYNARC for 13 and 2 plans respectively. No optimization on pCT was performed. Reported conformity indices for CT and pCT plans were respectively equal to 1.14 +/- 0.40 and 1.12 +/- 0.40 ($p > 0.05$). 2D 2%/2mm and 1%/1mm gamma indices were respectively 99.9% +/- 0.2% and 99.0% +/- 1.5% when applying a 10% dose threshold. PTV D95% and D99% mean errors were respectively equal to 0.10Gy +/- 0.04Gy and 0.05Gy +/- 0.04Gy. Concerning OAR, D0.035cc were all inferior to 0.13Gy +/- 0.04Gy, proving the high similarity between CT and pCT plans.

DL approaches have been proved to be highly efficient for the pCT generation task. They are currently implemented into two main software. First, MRI Planner (Spectronic Medical AB, Helsingborg, Sweden) is already commercially available for head, head and neck and pelvis. It first relies on the estimation of a transfer function parameters via DL. The spatially resampled patient MRI is used as input, while a 3D matrix containing information of both tissues labels categories and the transfer function parameters constitute the output. According to Cronholm et al. (322), the 3D networks benefit from residual connections and large receptive fields, ensuring high pCT quality. The derived transfer function is finally applied onto the MRI, to predict the final pCT. Recently, Lerner et al. (289) evaluated MRI Planner v2.2 on 10 glioma patients, and 10 additional patients with other brain tumors. CT and Dixon MR images were acquired on the same devices. VMAT plans containing 1 to 4 arcs were used for planning, without optimizing on pCT. Head MAE and ME were respectively equal to 62.2HU +/- 4.1HU and -5.6HU +/- 4.6HU. 3D global 3%/3mm, 2%/2mm and 1%/1mm gamma indices were 100.0% +/- 0.0%, 99.8% +/- 0.2% and 99.1% +/- 0.6% with a 15% dose threshold.

In July 2021, GE Healthcare officially announced its collaboration with Spectronic Medical AB, for a fully dedicated DL radiotherapy framework, involving AIR Recon DL and MRI Planner technologies to respectively enhance MRI quality and generate pCT.

Second, MRCAT Brain (Philips Healthcare, Best, The Netherlands) is currently under development and aims at generating pCT from 3D T1 mDixon MRI. It has been reported that mean dose differences in the PTV for CT and pCT were lower than 1% for 95% of the evaluation cohort (323).

Finally, Table 7 presents a non-exhaustive DL-based studies summary.

Table 7: DL studies to compute pCT from brain MRI. Several network architectures are reported such as GAN or convolutional neural network. The main adopted MR sequences are T1 and T1-Gd. Finally, the achieved results are presented in terms of MAE, gamma indices and DVH computed on OAR and PTV.

Authors	**Network architecture**	**Dataset size**	**MR sequence**	**MRI bias field correction**	**MRI standardization**	**Image comparison / Dosimetry results**
Han et al. (2017) (260)	2D UNet	Training: 15 Testing: 3	3D T1	N3 filter (324)	Dynamic histogram warping (325)	Head MAE=85HU+/-17HU
Wolterink et al. (2017) (326)	2D cycle GAN	Training: 18 Testing: 6	3D T1	-	-	Head MAE=74HU+/-2HU
Nie et al. (2017) (251)	3D GAN	16	-	-	ZMUV	MAE=93HU+/-14HU
Kläser et al. (2018) (327)	HighResNet	20	T1 and T2	-	-	Head MAE=69HU+/-15HU

Dinkla et al. (2018) (261)	2.5D Dilated CNN	Training: 26 Testing: 26	3D T1	-	Linear scaling	Head MAE=67HU+/-11HU 1%/1mm Gamma index=91.1%+/-3.0% 2%/2mm Gamma index=95.8%+/-2.1% 3%/3mm Gamma index=99.3%+/-0.4%
Emami et al. (2018) (328)	2D GAN	Training: 12 Testing: 3	3D T1-Gd	-	-	Head MAE=89HU+/-10HU
Kläser et al. (2019) (329)	HighResNet	20	T1 and T2	-	-	Head MAE=70HU+/-32HU
Liu et al.	2D	Training:	3D T1-	-	-	Head MAE

(2019) (296)	convolutional autoencoder	40 Testing: 10	Gd			=75HU+/-23HU Difference PTV D_{max}=1.39%+/-1.31% 3%/3mm Gamma index=99.2%
Kazemifar et al. (2019) (232)	2D GAN	Training: 54 Validation: 9 Testing: 14	2D T1-Gd	-	-	Head MAE=47HU+/-11HU DVHs (mean, min, max, 95%, 5% doses) for the PTV and OAR < 1% 3D 2%/2mm gamma index=99.2%+/-0.8% 3D 1%/1mm gamma index=94.6%+/-

						2.9%
Lei et al. (2019) (330)	3D cycle GAN	Training: 23 Testing: 1	T1	-	-	Head MAE=56HU+/-9HU

4. Dosimetry-driven quality measure of brain pseudo Computed Tomography generated from DL for MRI-only radiotherapy treatment planning – Study 1

4.1. Context

A growing number of DL-based pCT generation studies have been performed these past few years. High variability in terms of included patients in the cohorts was reported, with rates ranging from 12 (298) to 92 (297). Same conclusion was obtained when analysing the MRI sequences used as inputs of the networks, which were either T1 (260) or T1-Gd (328) or T2 FLAIR (321) or T2 (327) or T2 FatSat (297). In addition, the use of MRI as network input implies the need to standardize their intensities to a same reference scale. Few studies reported the use of a standardization method (260,261). Thus, it remains unclear if this step significantly improves the final pCT quality. Similarly, bias field filters are optionally applied to remove intensities inhomogeneities. Lastly, no consensus was adopted regarding the best suited network architectures, varying between UNet (260), HighResNet (327), GAN (251), cycle GAN (326).

As a result, it appeared of most importance to investigate these parameters, namely the size of the training set, MRI sequence, MRI standardization, bias field filter application and network architecture. To do so, a DL-based pCT generation pipeline previously developed by TheraPanacea (Paris, France) was re-used, and this study benefitted from the algorithm author expertise in DL. Varying training set sizes, equal to 242, 121, 60, 30 and 15 patients, with validation and testing cohorts remaining unchanged and composed of 81 and 79 patients respectively, were retrieved. The default network architecture was the modified 3D HighResNet. Regarding MRI sequences, two networks were trained in parallel, considering either T1 or T1-Gd sequences. Concerning standardization, three different strategies were investigated. Their performances were compared against the ones achieved when using the original MRI. Concerning the network architecture impact, the default HighResNet was compared with a 3D UNet based on the same patients split as previously described.

pCT quality was evaluated with MAE, and dosimetry-derived analysis relying on plans transfer and dose re-calculation on pCT. PTV-based DVH differences and global 1%/1mm, 2%/2mm and 3%/3mm gamma indices without dose threshold were computed. Finally, optimized parameters combination was assessed, paving the way for highly efficient pCT generation.

Note that the study presented below has been adapted from the original published version.

4.2. Abstract

Purpose: This study aims at evaluating the impact of key parameters on the pCT quality generated from MRI with a 3D convolutional neural network.

Methods: 402 brain tumor cases were retrieved yielding to associations of 182 CT/T1, 180 CT/T1-Gd and 40 CT/T1/T1-Gd. A 3D convolutional neural network was used to map T1 or T1-Gd into CT and evaluate the importance of different components. First, the training set size influence on the testing set accuracy was assessed. Moreover, we evaluated the MR sequence impact, using T1 only and T1-Gd only cohorts. Then, we investigated four MRI standardization approaches, namely HB, ZMUV, WS and no standardization (NS) based on training, validation and testing cohorts composed of 242, 81 and 79 patients cases respectively, as well as a bias field correction influence. Finally, two networks, namely modified HighResNet and 3D UNet, were compared to evaluate the architecture impact on the pCT quality. The MAE, gamma indices and dose volume histograms were used as evaluation metrics.

Results: Generating models using all the available cases for training led to higher pCT quality. The T1 and T1-Gd models indicated maximum differences in gamma indices means of 0.07 percent point. The MAE obtained with WS was 78HU +/- 22HU, which slightly outperformed HB, ZMUV and NS ($p<0.0001$). Regarding the network architectures, 3%/3mm global gamma indices without dose threshold equal to 99.83%+/-0.19% and 99.74%+/-0.24% were obtained for HighResNet and 3D UNet respectively.

Conclusion: Our best pCT were generated using more than 200 samples in the training dataset, while training with T1 only and T1-Gd only did not significantly affect the performance. Regardless of the preprocessing applied, the dosimetry quality remained equivalent and relevant for a potential use in clinical practice.

4.3. Introduction

MRI has become prevalent in radiotherapy planning due to its excellent soft tissue contrast compared to CT. During a brain tumor radiotherapy process, MRI and CT are complementary and play a key role in segmentation and segmentation/dosimetry respectively. Yet, dealing with multiple imaging modalities requires to co-register them, leading to errors up to 2mm (4), and target volumes margins increase.

To address this limitation, numerous approaches have been developed to generate a pCT from MRI (331,332). First, the bulk density approach (333,334) assigns specific ED to pre-segmented MRI relying however on the labelling quality. Second, the multi-atlas method constitutes a multiple “atlases” database representing rigidly co-registered pairs of CT and MRI acquired from different patients. The incoming MRI is first aligned to the atlases MRI through a deformable registration. The resulting deformation fields are then applied to the atlases CT which are finally combined to generate the pCT (318,319). Due to the computational complexity of deformable registrations, the multi-atlas approach is time-consuming. To mitigate these limitations, DL methods have been recently introduced, reporting promising results (232,261). Compared to the other approaches, DL-based methods efficiently exploit large databases to learn a direct mapping from MRI to CT. A deep convolutional neural network consists in a composition of convolutional filters and simple non-linear functions organized in layers. The parameters of the convolutional neural network are learned using pairs of MRI/CT training data via empirical risk minimization and stochastic gradient descent. DL-based methods benefit from highly efficient Graphical Processing Unit (GPU) implementations which reduce the inference time of the pCT of several orders of magnitude compared to atlas-based methods. Based on a NVIDIA Titan X GPU, Han et al. (260) reported durations of 9 seconds and 10 minutes for the DL and atlas-based approaches respectively.

However, there is still no consensus regarding: 1) the optimal training set size, 2) the best-suited MR sequence, 3) the optimal MR standardization preprocessing, 4) the use of an inhomogeneity correction and 5) the best suited network architecture (Table 7). Additionally, there is no discussion about the approach to evaluate the generated pCT.

Indeed, training datasets sizes ranging from 12 (298) to 92 patients (297) have been reported, raising the issue of the minimal number of training patients required to ensure a satisfying generalization to unseen examples. Moreover, most of the studies used either T1 or contrast T1-Gd. However, the benefit of using a contrast agent in terms of pCT quality is still unclear.

Additionally, only few studies applied MRI intensities standardization as preprocessing. Yet, it can improve the pCT quality (81). A similar question concerns the bias field correction, as only Han (260) applied it. Finally, several convolutional neural network architectures have been used in the literature, such as HighResNet (327,329) and UNet (260) for instance, without systematically comparing them.

An additional aspect which it is not explicitly discussed in these works is the influence of these parameters on a dosimetry-based pCT evaluation. Numerous studies report their performances using peak signal-to-noise ratio or MAE metrics (251,260,326) which can possibly be irrelevant to the real clinical scenario.

This study aims at evaluating the impact of significant parameters, namely the training dataset size, the input MR sequence, the standardization strategy, the application of an inhomogeneity correction and the network architecture, on the computed pCT accuracy and the associated clinical dosimetry. The pCT evaluation is based on both the MAE and clinical criteria, namely the global 1%/1mm, 2%/2mm and 3%/3mm gamma indices with no dose threshold and differences in DVH of the PTV.

4.4. Materials and Methods

4.4.1. Images acquisition and preprocessing

402 institutional patients treated between 2006 and 2017 for brain tumors were included in this retrospective study. For all of the patients, the delay between the planning CT and T1 or T1-Gd MR acquisitions did not exceed eight days. The dataset was composed of 182 CT/T1, 180 CT/T1-Gd and 40 CT/T1/T1-Gd paired images. It did not contain severe MRI artefacts. However, patients presenting discontinuous skulls owing to surgery were not excluded from the cohort.

All the CT were acquired with a Sensation Open scanner (Siemens Healthineers, Erlangen, Germany) using a 120kVp tube voltage. The slice thickness was equal to 1mm, 2mm, 3mm and 5mm for 3, 45, 353 and 1 patients cases respectively. The native X and Y voxel sizes were included in [0.50mm; 0.70mm], [0.70mm; 0.90mm] and [0.90mm; 1.10mm] for 208, 76 and 118 patients respectively.

The MRI were all acquired with GE Healthcare devices (GE Healthcare, Milwaukee, Wisconsin, USA). Two patients cases' MR sequences were from external institutes and were acquired on two different 1.5T devices: Optima MR360 and Discovery MR450. The remaining

MRI were institutional images, acquired on a 3T Discovery MR750w (224 patient cases), a 1.5T Optima MR450w (9 patient cases) and a 1.5T Signa Excite (167 patient cases). Only 3D axial T1-weighted images with or without a gadolinium injection were used. Initial slice thicknesses were included in [1mm; 1.2mm], [1.4mm; 2mm], [3mm; 3.2mm] and equal to 5mm for 234, 10, 157, 1 patients respectively. Regarding the native X and Y voxel sizes, they were included in [0.44mm; 0.50mm], [0.50mm; 0.58mm] and equal to 0.94mm for 325, 73 and 4 patients respectively.

For each patient, the CT was first rigidly registered to the T1 or T1-Gd images using the Drop library (https://github.com/biomedia-mira/drop2). Then, the images were linearly resampled to a 1mm×1mm×1mm voxel size, before harmonizing the volumes to 300x300x242 voxels. Volumes were placed in a reference frame, re-arranging the matrices axes. Both the MRI intensities and the CT HU were clipped, to 0.1 and 99.9 percentiles and [-1000HU, 1800HU] respectively. The maximum HU was empirically determined based on CT intensity histograms. HU were finally rescaled between [-1, 1].

Lastly, 60%, 20% and 20% of the patients were randomly parsed into training, validation and testing sets, provided that the T1 and T1-Gd were equal in proportion. Patients with all CT, T1 and T1-Gd images were automatically assigned to the testing set, to be used for the dosimetry-based evaluation.

4.4.2. Standardization strategies

Three different approaches were adopted to standardize the MRI. The first approach was HB with a percentiles list of [10, 20, 30, 40, 50, 60, 80, 90]. The second approach was ZMUV, applied in every patient head region. The last method, namely WS, required brain masks generation, which were first extracted with the HD-BET tool (335). The MR images were then normalized with the intensity-normalization package (81).

4.4.3. Network architectures

Following popular choices of network architectures in the literature, we decided to use the modified version of the 3D HighResNet presented by Li et al. (237) (section 2.3.5.2.) and the 3D UNet (336), respectively composed of 0.8 million parameters and 15 million parameters.

To optimize the network parameters, we used the MAE loss function. Due to memory constraints, patches of size 96x96x96 voxels and 136x136x136voxels were used as input of the HighResNet and 3D UNet respectively. During inference, the 3D MRI were divided into patches to reconstruct the whole pCT. A patch margin of length 5 and 1 voxels for the HighResNet and 3D UNet respectively, was applied leading to predictions inside sub-patches of size 86 x 86 x 86 and 134 x 134 x 134. The motivation of the margins is to guarantee a smooth transition between patches prediction. Note that patches overlapped, contrary to sub-patches. The overlap process is illustrated in Figure 44.

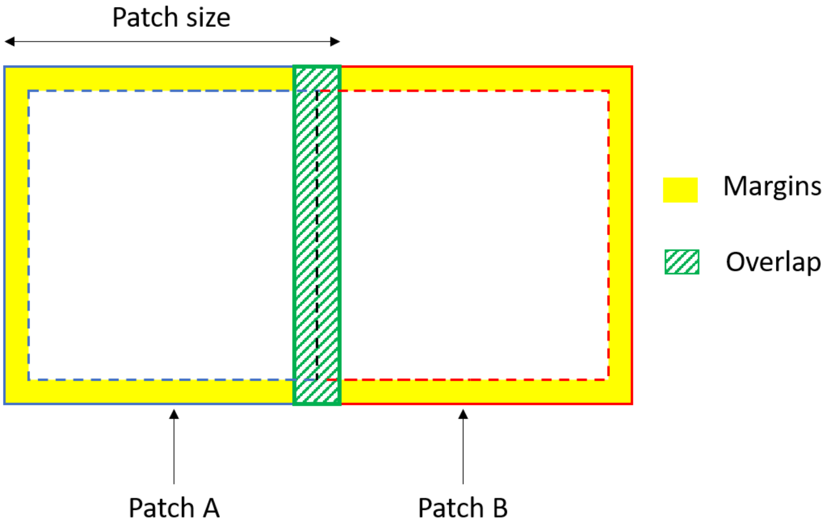

Figure 44: Adjacent patches overlap. The yellow zone corresponds to the margins. White sub-patches included in the initial patches represent the prediction region. Patch overlap, corresponding to the green area, occurs for the patches, but not for the sub-patches.

For both networks, the learning rate was set to 0.001. Early stopping on the validation set was used as stopping criterion to assess the convergence of the convolutional neural network. Dropout was used after the penultimate layer during training with a probability of 0.5.

Note that no data augmentation was used in this study.

4.4.4. Impact of key parameters

The first experiment consisted in quantifying the impact of the training set size. Five different HighResNet networks were trained using 242 (121 T1-121 T1-Gd), 121 (61 T1–60 T1-Gd), 60 (30 T1–30 T1-Gd, 30 (15 T1–15 T1-Gd) and 15 (8T1–7T1-Gd) patients respectively in the training set. The validation and testing cohorts were the same for all the training set sizes and included 81 (41 T1–40 T1-Gd) and 79 (39 T1–40 T1-Gd) cases respectively. All the MR images were standardized using the HB method.

A second experiment was conducted to determine the best suited T1 input sequence to generate pCT. We constituted two HB-standardized cohorts: 1) a T1-only cohort with 134, 44 and 40 T1 MRI cases for the training, validation and testing sets respectively, 2) a T1-Gd-only cohort with 133, 44 and 40 patients cases respectively. The cases included in the two testing cohorts were the same, for a fair comparison. For this experiment, different T1 and T1-Gd histograms templates were computed for the HB standardization, based on the 134 and 133 patients included in the training cohorts. This experiment was based on the HighResNet.

The third experiment assessed the role of the MRI standardization using 242 (121 T1–121 T1-Gd), 81 (41 T1–40 T1-Gd) and 79 (39T1 –40 T1-Gd) cases in the training, validation and testing sets respectively. The HighResNet architecture was used in this experiment. Four different approaches were investigated: HB, ZMUV, WS and no standardization (NS).

The fourth experiment was performed to evaluate the role of the bias field correction, using HighResNet. Bias field is a low frequency contamination field resulting in intensities inhomogeneity. To correct from this signal, the N4 filter (337) was optionally applied on MR images. N4 is an improved version of the original nonparametric nonuniform intensity normalization N3 (324), expressing the bias field via a multiplicative approach as follows:

$$v(x) = u(x)f(x) + n(x) \qquad \text{Equation 70}$$

With v, u, f, n, x are the biased image, the originally pure image, the bias field, noise and a considered point.

After applying the log function and considering probability densities functions, Equation 70 is further expressed as:

$$V(\hat{v}) = U(\hat{v}) * F(\hat{v}) \qquad \text{Equation 71}$$

With $\hat{v}$, V, U, F the log function applied to v, the probability density functions of v, u, f.

N3 iteratively deconvolves a sub-Gaussian from V and estimates the bias field, until convergence (338). N4, the improved N3 version, relies on a new B-spline based approximator enabling smaller distance between control points, and a new optimization approach, based on the assessment of a residual bias field at each iteration (337).

The best standardization technique defined by experiment 3 was used here. The training, validation and testing sets were those used in experiment 3.

The last experiment was conducted to analyse the influence of the network architecture on the quality of the generated pCT. To this aim, the HighResNet used in the previous experiments and the 3D UNet, were trained, validated and tested. Best preprocessing strategies highlighted by the third and fourth experiments were applied. The split of the dataset was the same as experiment 3.

A summary of the experiments is presented in Figure 45.

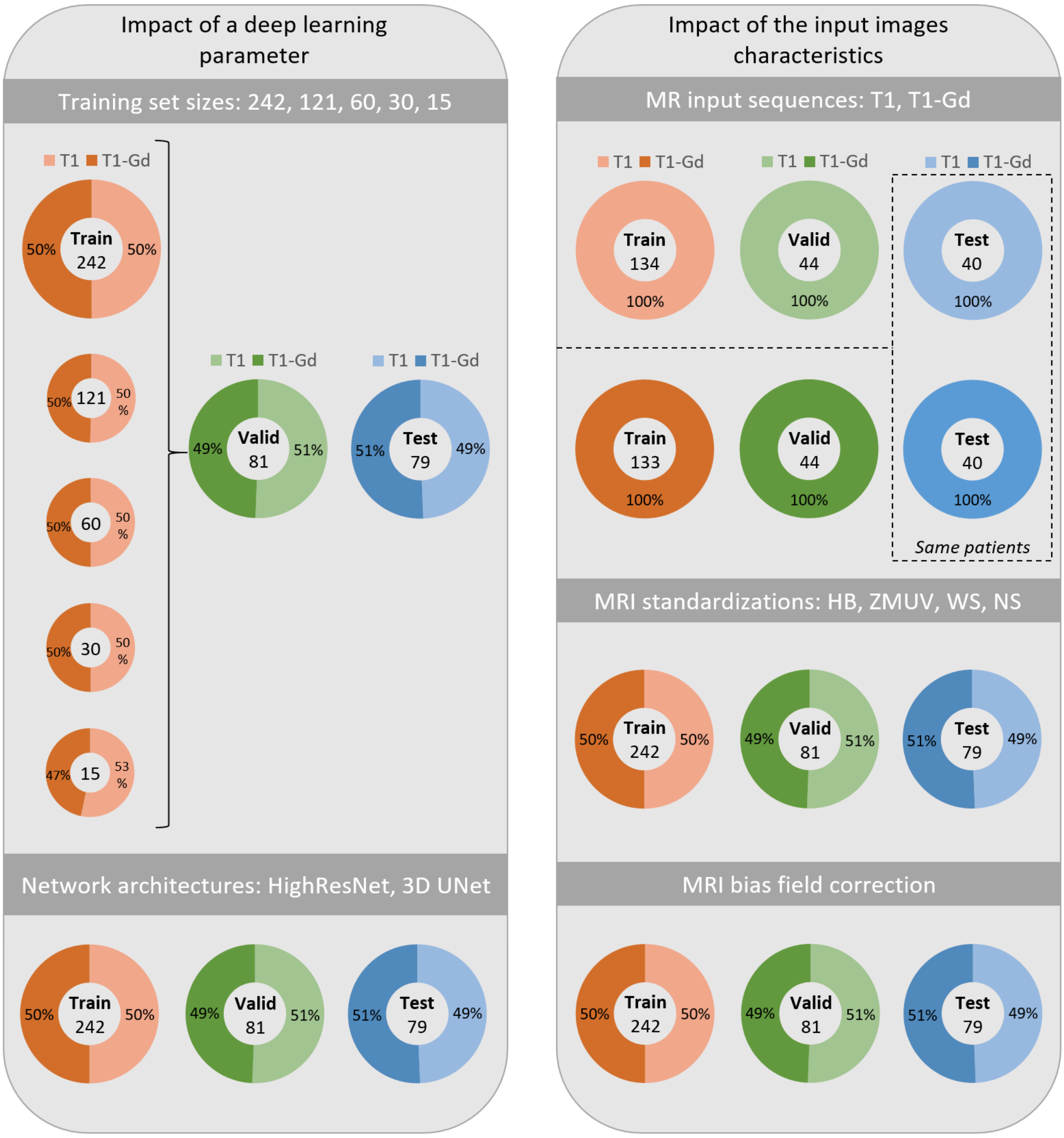

Figure 45: Data repartition into training, validation and testing sets for the experiments regarding the training set sizes, the MR sequences, namely T1 and contrast T1-Gd, the MRI standardizations, namely HB, ZMUV, WS and NS, the bias field correction and the network architectures, namely the HighResNet and 3D UNet.

4.4.5. Evaluation criteria

First, the initial CT and the pCT were compared using the MAE. It was computed in four different areas: whole head, air, bone and water. The head was segmented using the Otsu approach, described in (339). The other regions were obtained thresholding the CT: $x \leq -200HU$, $-200HU < x < 250HU$ and $250HU \leq x$ for the air, water and bone regions respectively. The MAE was calculated from the 3D intensities volumes or the 3D ED volumes obtained applying the HU-ED calibration curve.

Furthermore, we evaluated the pCT quality in terms of dose prediction for all the experiments, except the first one, by computing metrics used in clinics. 3D global 1%/1mm, 2%/2mm and 3%/3mm gamma indices were considered, and no dose threshold was applied. In addition, relative differences between CT and pCT DVH ($D_{02\%}$, $D_{50\%}$, $D_{95\%}$ and $D_{98\%}$) of the PTV were calculated.

The dosimetry plans from the original CT were transferred and recalculated on the pCT, with the pencil beam dose calculation algorithm implemented in iPlan RT 4.5 Dose (Brainlab, Munich, Germany) (340). The default grid size was set to 2mm. It is worth noting the grid was adaptive, meaning that it became finer for small object. This approach was combined with a ray-tracing technique which was applied during the radiological path length calculation. These two approaches resulted in a speed up of the dose calculation. For this dosimetry analysis, a subset cohort of the testing set, corresponding to cases whose dosimetry had been realized with iPlan, was used. It was composed of 39 grades III and IV glioma patients cases (19 T1 - 20 T1-Gd) treated with a sliding window IMRT approach, delivered with a 6 MV beam. 18, 11, 7, 2 and 1 patients cases were treated with 5, 6, 7, 8 and 10 beams respectively. An illustration of the overall workflow is presented in Figure 46.

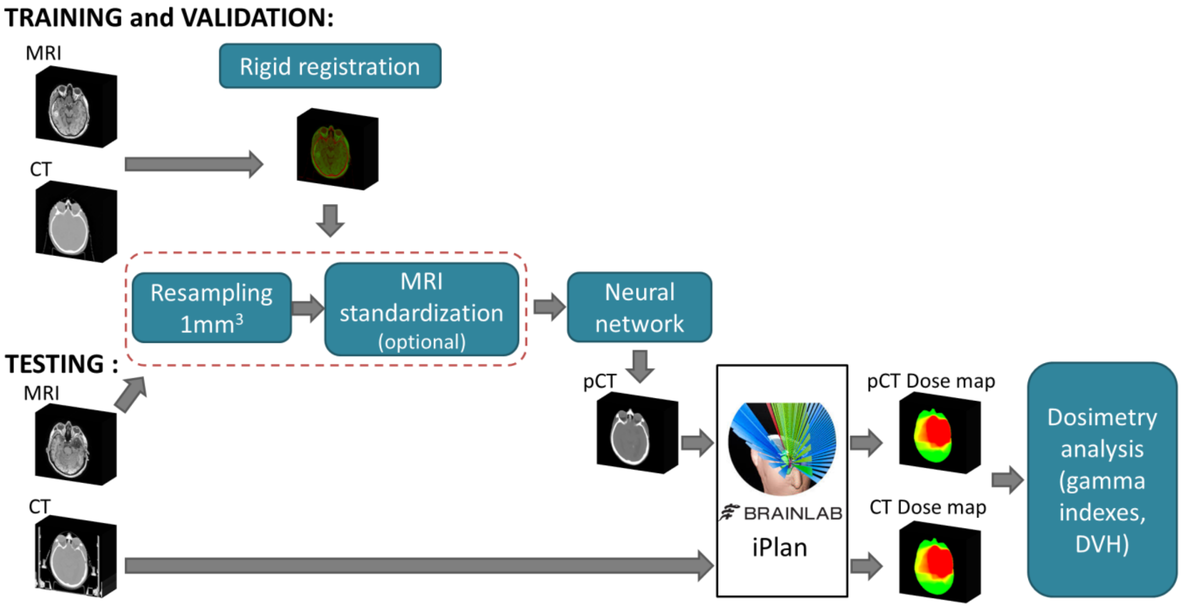

Figure 46: General workflow of the study. In a first step (training and validation) the parameters of the network and other hyperparameters are optimized using preprocessed paired CT and MRI. The second step is the test: an MRI is preprocessed before passing through the trained network and generating the pCT. For the evaluation, the pCT and the original CT are compared based on the MAE metric and a dosimetry analysis based on DVH differences and gamma indexes.

Two-sided paired Wilcoxon tests, with a significance level set to 0.05, were performed as statistical analysis.

Only results computed on the testing set are reported.

4.5. Results

Figures 47.A and 47.B present examples of MRI, CT and pCT with soft tissues and bone windows and levels respectively. They were extracted from the third experiment, using the HighResNet and the HB intensities standardization. The first line corresponds to a low MAE case (head MAE=64HU) and the second line to a high MAE case (head MAE=110HU). Some air and bone areas appear to be less accurately reconstructed, as highlighted by the squares.

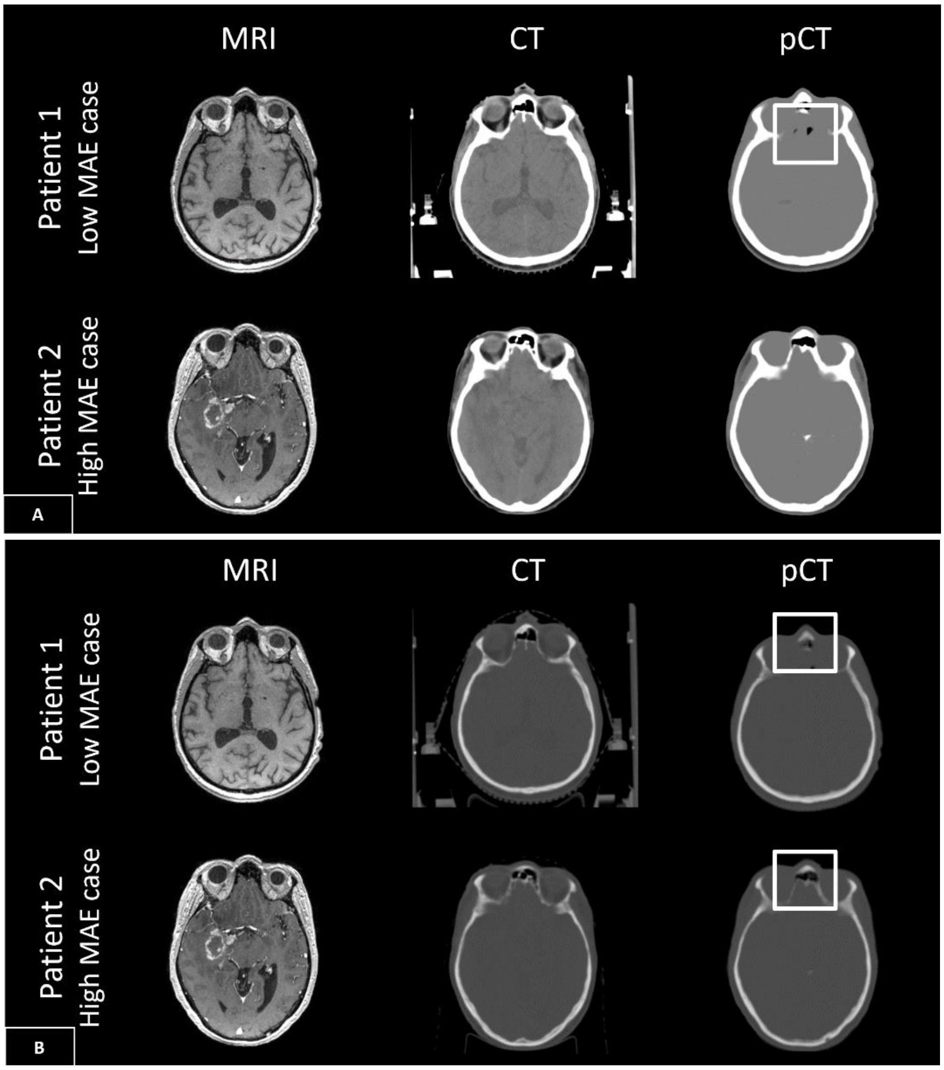

Figure 47: (From left to right) MRI, original CT, and pCT with soft tissue (A) and bone (B) windows widths and levels, respectively, for 2 patients. The squares highlight some of the incorrect reconstructed areas.

The intensity-based MAE obtained from different training set sizes, is displayed in Figure 48. For the head area, increasing the training dataset resulted in a decrease of the MAE mean+/-standard deviation from 189HU+/-28HU for the 15 patients-training set model to 92HU+/-23HU corresponding to the 242 patients-training set model. Bone and air regions reported the highest MAE.

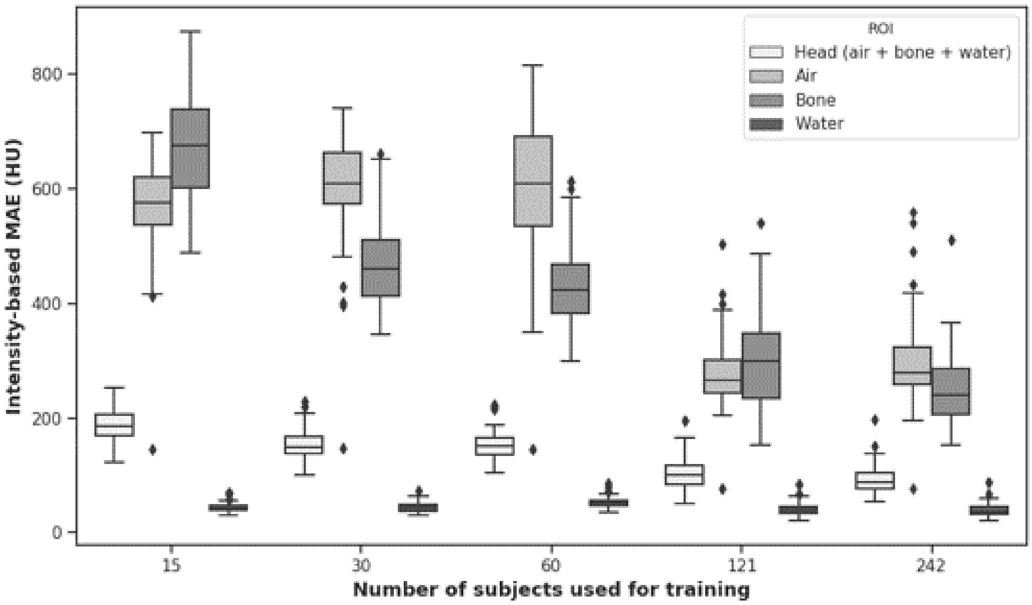

Figure 48: Evolution of the MAE in HU when modifying the number of training subjects. The boxplot corresponds to the first and third MAE quartiles with the MAE median in the middle; the whiskers correspond to the range of the MAE distribution after excluding the outliers.

Differences between all the training size models were significant for the head region (p<0.0001) except between 30 and 60 patients (Table 8).

Table 8: Statistical analysis based on the MAE and concerning the training set size experiment.

	MAE head
p-value 242 patients model – 121 patients model	**<0.0001**
95% Confidence interval 242 patients model – 121 patients model	[-10.31, -6.64]
p-value 242 patients model – 60 patients model	**<0.0001**
95% Confidence interval 242 patients model – 60 patients model	[-61.64, -56.16]
p-value 242 patients model – 30 patients model	**<0.0001**
95% Confidence interval 242 patients model – 30 patients model	[-62.10, -55.82]
p-value 242 patients model – 15 patients model	**<0.0001**
95% Confidence interval 242 patients model – 15 patients model	[-101.27, -90.68]
p-value 121 patients model – 60 patients model	**<0.0001**
95% Confidence interval 121 patients model – 60 patients model	[-52.53, -47.35]
p-value 121 patients model – 30 patients model	**<0.0001**

95% Confidence interval 121 patients model – 30 patients model	[-52.85, -47.85]
p-value 121 patients model – 15 patients model	**<0.0001**
95% Confidence interval 121 patients model – 15 patients model	[-91.82, -82.51]
p-value 60 patients model – 30 patients model	0.84
95% Confidence interval 60 patients model – 30 patients model	[-2.05, 1.52]
p-value 60 patients model – 15 patients model	**<0.0001**
95% Confidence interval 60 patients model – 15 patients model	[-41.08, -32.17]
p-value 30 patients model – 15 patients model	**<0.0001**
95% Confidence interval 30 patients model – 15 patients model	[-39.90, -33.08]

The ED-based MAE is presented in Figure 49, to more accurately assess the pCT quality with respect to its clinical use. A similar behaviour is observed, with a head MAE decrease from 0.10+/-0.01 to 0.05+/-0.01 when increasing the training set size from 15 to 242.

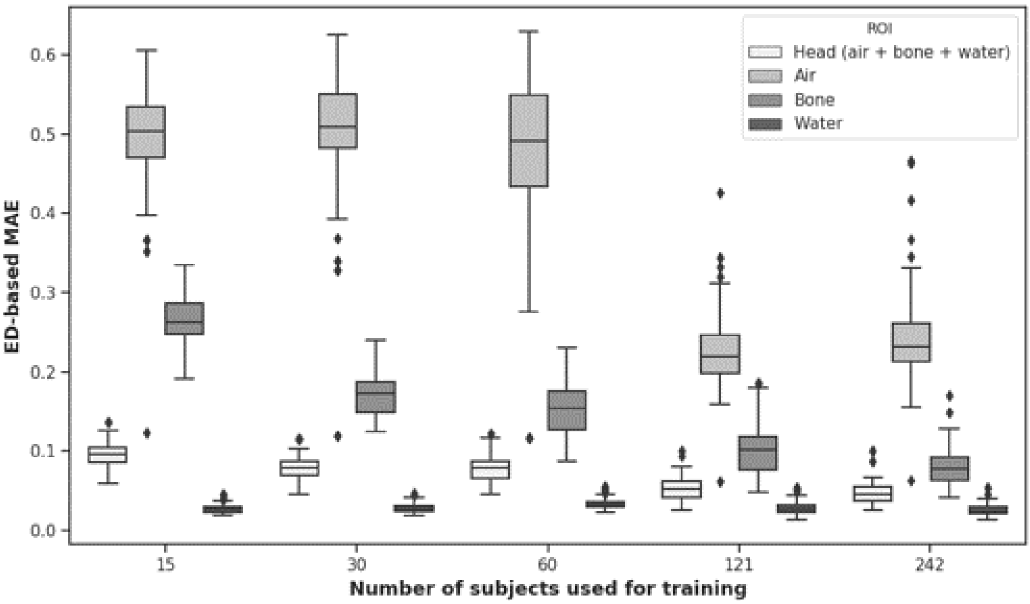

Figure 49: Evolution of the MAE in ED when modifying the number of training subjects. Same boxplot representation as in Figure 48 was used.

Table 9 presents the MAE, gamma indices, DVHs differences and Wilcoxon tests values derived from the T1-only and T1-Gd only models. The maximum differences between the T1 and T1-Gd models obtained for the head MAE means and gamma indices means were equal to 3HU and 0.07 percent point (pp) respectively.

Table 9: Means+/-standard deviations of the MAE, gamma indices, DVH differences computed for the PTV and statistical analysis derived from the T1 and T1-Gd cohorts comparison.

	T1 only	**T1-Gd only**	**p-value**	**95% Confidence interval**
MAE head	84HU+/-25HU	87HU+/-28HU	**0.0047**	[-3.93, -0.76]
MAE air	274HU+/-63HU	306HU+/-74HU	**<0.0001**	[-36.51, -22.37]
MAE bone	228HU+/-63HU	236HU+/-71HU	0.066	[-11.38, 0.48]
MAE water	38HU+/-11HU	38HU+/-12HU	0.82	[-0.83, 0.73]
1%/1mm gamma index	97.87%+/-1.16%	97.94%+/-1.07%	0.59	[-0.12, 0.05]
2%/2mm gamma index	99.60%+/-0.33%	99.63%+/-0.30%	0.50	[-0.05, 0.02]
3%/3mm gamma index	99.84%+/-0.18%	99.85%+/-0.18%	0.44	[-0.03, 0.01]
Difference PTV $D_{02\%}$	0.20%+/-0.15%	0.15%+/-0.09%	**0.0041**	[0.02, 0.08]
Difference PTV $D_{50\%}$	0.20%+/-0.15%	0.13%+/-0.08%	**0.015**	[0.02, 0.12]
Difference PTV $D_{95\%}$	0.20%+/-0.17%	0.14%+/-0.10	**0.012**	[0.02, 0.12]
Difference PTV $D_{98\%}$	0.27%+/-0.37%	0.22%+/-0.41%	**0.026**	[0.01, 0.12]

Means+/-standard deviations of the MAE, gamma indices and DVH differences obtained for the standardization experiment are provided in Table 10. The statistical analysis is presented in Table A1 in Supplementary Materials. WS led to a head MAE of 78HU+/-22HU, which was significantly lower than the three other methods (p-values<0.0001). Regarding the dosimetry, global 3%/3mm gamma indices of 99.86%+/-0.16%, 99.83%+/-0.19%, 99.85%+/-0.17%, 99.86%+/-0.18% were respectively achieved for the HB, ZMUV, WS and NS approaches.

Table 10: Means+/-standard deviations of the MAE, gamma indices and DVH differences computed for the PTV derived from the HB, ZMUV, WS and NS cohorts.

	HB	**ZMUV**	**WS**	**NS**
MAE head	92HU+/-23HU	83HU+/-22HU	78HU +/- 22HU	96HU+/-23HU
MAE air	297HU+/-73HU	284HU+/-62HU	253HU+/-65HU	313HU+/-68HU
MAE bone	251HU+/-61HU	214HU+/-55HU	199HU+/-54HU	252HU+/-60HU
MAE water	39HU+/-11HU	38HU+/-12HU	36HU+/-11HU	43HU+/-11HU
1%/1mm gamma index	97.94%+/- 1.06%	97.90%+/- 1.10%	98.08%+/- 1.01%	97.80%+/- 1.17%
2%/2mm gamma index	99.63%+/- 0.28%	99.61%+/- 0.30%	99.64%+/- 0.29%	99.61%+/- 0.31%
3%/3mm gamma index	99.86%+/- 0.16%	99.83%+/- 0.19%	99.85%+/- 0.17%	99.86%+/- 0.18%
Difference PTV $D_{02\%}$	0.22%+/-0.17%	0.22%+/-0.16%	0.20%+/-0.13%	0.24%+/-0.20%
Difference PTV $D_{50\%}$	0.24%+/-0.16%	0.23%+/-0.16%	0.21%+/-0.13%	0.27%+/-0.17%

Difference PTV $D_{95\%}$	0.27%+/-0.31%	0.21%+/-0.17%	0.19%+/-0.15%	0.32%+/-0.32%
Difference PTV $D_{98\%}$	0.38%+/-0.58%	0.27%+/-0.35%	0.20%+/-0.17%	0.38%+/-0.46%

Regarding the fourth experiment based on the combination of the HighResNet with the WS standardization, means+/-standard deviations of the MAE and dosimetry metrics are presented in Table 11. Applying the bias field correction led to a head MAE of 81HU+/-22HU. Concerning the DVH D02%, differences equal to 0.15%+/-0.12% and 0.20%+/-0.13% were achieved with and without the application of the N4 filter respectively (p-value=0.026).

Table 11: Means +/- standard deviations of the MAE, gamma indices, DVH differences of the PTV and statistical analysis derived from the WS and WS combined with a N4 bias field correction cohorts comparison.

	WS	**WS & N4**	**p-value**	**95% Confidence interval**
MAE head	78HU+/-22HU	81HU+/-22HU	**<0.0001**	[-4.79, -2.57]
MAE air	253HU+/-65HU	244HU+/-62HU	**<0.0001**	[5.23, 11.84]
MAE bone	199HU+/-54HU	230HU+/-56HU	**<0.0001**	[-35.81, -27.07]
MAE water	36HU+/-11HU	34HU+/-10HU	**<0.0001**	[2.02, 2.91]
1%/1mm gamma index	98.08%+/-1.01%	97.92%+/-1.06%	**0.0035**	[0.04, 0.19]
2%/2mm gamma index	99.64%+/-0.29%	99.60%+/-0.32%	**0.0026**	[0.01, 0.06]
3%/3mm gamma index	99.85%+/-0.17%	99.83%+/-0.19%	**0.012**	[0.00, 0.03]
Difference PTV $D_{02\%}$	0.20%+/-0.13%	0.15%+/-0.12%	**0.026**	[0.00, 0.13]
Difference PTV $D_{50\%}$	0.21%+/-0.13%	0.13%+/-0.10%	**0.0017**	[0.03, 0.15]

Difference PTV $D_{95\%}$	0.19%+/-0.15%	0.11%+/-0.12%	**0.0034**	[0.03, 0.14]
Difference PTV $D_{98\%}$	0.20%+/-0.17%	0.13%+/-0.13%	**0.0088**	[0.02, 0.14]

Table 12 provides the MAE and dosimetry values for the last experiment, which was conducted to compare the HighResNet with the 3D UNet. For both networks, the WS MRI standardization and the N4 filter were applied. Means+/-standard deviations obtained for the head MAE were equal to 81HU+/-22HU and 90HU+/-21HU for the HighResNet and 3D UNet respectively (p-value<0.0001). Significantly higher gamma indices were obtained with the HighResNet (p-value<0.0001), with a pass rate of 97.92%+/-1.06% for the most restrictive global 1%/1mm criterion.

Table 12: Means+/-standard deviations of the MAE, gamma indices, DVH differences computed for the PTV and statistical analysis derived from the WS combined with a N4 bias field correction and the initial HighResNet against WS associated with N4 and the 3D UNet cohorts comparison.

	WS & N4 & HighResNet	**WS & N4 & 3D UNet**	**p-value**	**95% Confidence interval**
MAE head	81HU+/-22HU	90HU+/-21HU	**<0.0001**	[-9.39, -6.99]
MAE air	244HU+/-62HU	266HU+/-66HU	**<0.0001**	[-27.18, -15.56]
MAE bone	230HU+/-56HU	209HU+/-54HU	**<0.0001**	[16.91, 25.79]
MAE water	34HU+/-10HU	49HU+/-11HU	**<0.0001**	[-15.81, -14.09]
1%/1mm gamma index	97.92%+/-1.06%	97.28%+/-1.46%	**<0.0001**	[0.42, 0.79]
2%/2mm gamma index	99.60%+/-0.32%	99.39%+/-0.47%	**<0.0001**	[0.10, 0.24]

3%/3mm gamma index	99.83%+/0.19%	99.74%+/0.24%	**<0.0001**	[0.03, 0.11]
Difference PTV $D_{02\%}$	0.15%+/-0.12%	0.33%+/-0.21%	**<0.0001**	[-0.28, -0.11]
Difference PTV $D_{50\%}$	0.13%+/-0.10%	0.29%+/-0.19%	**<0.0001**	[-0.22, -0.10]
Difference PTV $D_{95\%}$	0.11%+/-0.12%	0.28%+/-0.18%	**<0.0001**	[-0.24, -0.13]
Difference PTV $D_{98\%}$	0.13%+/-0.13%	0.31%+/-0.18%	**<0.0001**	[-0.26, -0.15]

4.6. Discussion

This study aimed at evaluating the impact of key parameters of brain pCT generation from T1 or T1-Gd images, namely the training set size, the MR input sequence, the standardization strategy, the application of a bias field correction and the network architecture. Best results were achieved when combining the WS MRI standardization with an inhomogeneity correction, the HighResNet, and all our 242 training patients cases. This suggests that more training cases could lead to further improvements. In addition, computing the gamma indices and DVH performances associated with experiment 1 would have been of great interest to assess the impact of varying training set size on dosimetry.

Regarding the MR sequences experiment, a difference of 3HU was observed between the head MAE means of the T1 only and T1-Gd only models, suggesting that the contrast agent resulted in a negligible pCT improvement, although a statistically significant difference was proved (p=0.0047). The DVH differences led to a similar conclusion, as only 0.07pp maximum difference between the two models means was obtained. We conducted an extra experiment to evaluate the potential benefit of the T2 FLAIR MR sequence. 134, 44 and 40 patients were included in the training, validation and testing sets respectively. The preprocessing described for the T1 only and T1-Gd only cohorts was similarly applied. A mean MAE+/-standard deviation of 115HU+/-22HU was obtained within the head area. Differences with the T1 only and T1-Gd only cohorts were found significant ($p<0.001$). Thus, T2 FLAIR appeared to generate largest pCT intensities-linked errors. It could be attributed to the lower contrast contained in T2 FLAIR images compared to T1/T1-Gd images. A second interpretation could be the slice thickness which was larger for most of the T2 FLAIR images compared to T1/T1-Gd images, resulting in a less informative spatial sampling. Future work includes the comparison of T1 and unusual sequences, such as zero echo time in which bone areas are more visible, to assess which combination of MRI sequences is optimal for an accurate pCT reconstruction in radiotherapy.

The third experiment concerned the MRI standardization and used the HighResNet as network architecture. A mean+/-standard deviation of 78HU+/-22HU was obtained for the head MAE when applying the WS standardization, which slightly outperformed HB, ZMUV and NS ($p<0.0001$). Largest errors were located in the air and bone areas, with MAE of 253HU+/-65HU and 199HU+/-54HU respectively and seemed to correspond to misaligned regions or areas with high dose gradients.

Dinkla et al. (261) reported competitive head MAE of 67HU+/-11HU. All the CT and MR images used in their study were acquired on the same device. In this work, MR images were acquired from five different devices. Table A2 in Supplementary Materials presents the composition of the training, validation and test sets in terms of MR devices. As one can notice, most of the MRI of the training set, namely 133, were acquired with the DISCOVERY MR750w - 3Teslas (T) device. To analyse the impact of this unbalance, the test set was split into two subsets: MRI from the DISCOVERY MR750w - 3T (57 patients) and MRI from the SIGNA EXCITE – 1.5T (21 patients). The default HB standardization and HighResNet were used for this experiment. Mean head MAE+/-standard deviation led to 86HU+/-22HU for the DISCOVERY MR750w - 3T and 106HU+/-16HU for the SIGNA EXCITE – 1.5T ($p<0.0001$). It showed pCT computed from the DISCOVERY MR750w - 3T device were of higher quality since more MRI acquired with such device composed the training set and since 3T devices offer a better images resolution. Thus, we think that the composition of the training set had a non-negligible impact on the generated pCT. Comparing the literature MAE is however not trivial due to the use of heterogeneous datasets, suggesting the need for publicly available datasets.

Concerning the dosimetry analysis, negligible differences were observed between the different standardization approaches. Regarding WS, a mean+/-standard deviation of 99.85%+/-0.17% was obtained for the 3%/3mm gamma index, which was not significantly different from the ZMUV, HB and NS gamma indices (p-values$\geq$0.14). These non-significant dosimetry results can be attributed to the non-linearity of both the HU-ED curve and the radiation matter interactions effects. Several studies reported dosimetry evaluations for brain pCT generated with a DL-based approach. Dinkla et al. (261) achieved 91.1%+/-3.0%, 95.8%+/-2.1% and 99.3%+/-0.4% for 1%/1mm, 2%/2mm and 3%/3mm head gamma indices with no threshold. A similar performance was obtained by Liu et al. (296) who reported 99.2% for the 3%/3mm gamma index. Recently, Kazemifar et al. (232) achieved state of-the-art 1%/1mm and 2%/2mm gamma indices of 94.6%+/-2.9% and 99.2%+/-0.8%. Eventually, dosimetry analyses are crucial as they are the only relevant metric for a use in clinics.

Fourth experiment evaluated the role of an inhomogeneity correction combined with the HighResNet and the WS standardization. Although a slight increase of 3HU of the mean head MAE was obtained when applying the N4 filter, the DVH metrics analysis showed a negligible decrease in the means up to 0.08pp, suggesting the potential use into clinics of both approaches even if significant p-value inferior or equal to 0.026 were achieved. It could be justified by an acceptable MRI quality or the network ability to handle this issue.

The last experiment was the evaluation of two different network architectures, namely the HighResNet and the 3D UNet. For each model, the WS standardization and the N4 filter were applied. Mean head MAE+/-standard deviation were equal to 81HU+/-22HU and 90HU+/-21HU for the HighResNet and 3D UNet respectively. The lower HighResNet MAE may be attributed to two major advantages: the dilated convolution filters which enable a large spatial context while retaining the full image resolution and the residual connections which regularize the optimization of the model. Regarding the dosimetry, 3%/3mm gamma indices equal to 99.83%+/-0.19% and 99.74%+/-0.24% were obtained for the HighResNet and the 3D UNet respectively. As a result, no significant clinical impact was observed between the two architectures, even if significant differences were suggested by the Wilcoxon tests ($p<0.0001$). In the literature, a lower MAE of 47HU+/-11HU was reported by Kazemifar et al. (232) using a 2D GAN. In the context of pCT generation, a GAN corresponds to the training of a second auxiliary neural network which learns a loss function to estimate the distance between a pCT and the distribution of all the true possible CT. This data-driven loss function is used to train the main neural network that learns the mapping from MRI to pCT. Therefore, pCT produced by a GAN are not guaranteed to respect the anatomy of the patient. To mitigate this issue, CycleGAN using an additional cycle-consistency penalization have been proposed (326,330). However, the cycle-consistency implies a one-to-one mapping between the MRI and CT, which is not realistic and can lead to artefacts in the pCT (341). As a result, further investigation of the errors specific to GAN and CycleGAN is needed for their clinical use in radiotherapy and is beyond the scope of this paper. The loss function used to train the network has a knock-on effect on the pCT quality. Here, the MAE was chosen since it was found to generate less blurry images than the MSE during preliminary experiments. Kazemifar et al. (232) trained two 2D GAN based on the MAE and the mutual information loss functions and obtained head MAE means+/-standard deviation of 60HU+/-22HU and 47HU+/-11HU respectively. Therefore, exploring different loss functions is of interest as it can heavily impact intensities-linked errors. Based on all the dosimetry results, very small discrepancies were obtained between all the preprocessing applied. For instance, 3%/3mm gamma indices equal to 99.83%+/-0.19% and 99.85%+/-0.17% were achieved for the experiments based on the combination of the HighResNet with the WS standardization and optionally applying the N4 filter (Table 11). Although a significant p-value of 0.012 was obtained, no major clinical impact is expected. As a result, it suggests that the proposed pCT generation method may be suitable for an introduction into clinics, regardless of the preprocessing applied.

The dose calculation algorithm used in this study was in pencil beam. An extra experiment was conducted to evaluate its relevance against MC, considered as more accurate in taking into account heterogeneities (342,343). Since the latter is not commissioned in our institution for IMRT plans, we constituted an additional cohort of 8 brain tumor patients treated with arctherapy. 4 out 8 patients had a CT and a T1 MRI, the rest had a CT and a T1-Gd MRI. The preprocessing previously described in section 4.4. was similarly applied and pCT were generated. A dosimetry was performed on the pCT with the two different dose algorithms. No significant differences were observed for the DVH differences analysis ($p \geq 0.27$). A similar conclusion was obtained for the 3%/3mm and 2%/2mm gamma indices ($p \geq 0.40$). Concerning the 1%/1mm criterion, 98.94%+/-0.68% and 98.40%+/-0.84% gamma pass rates were achieved for the pencil beam and MC algorithms respectively ($p=0.0078$). As a result, pencil beam approach is a reliable technique for the head localization, due to the absence of large inhomogeneities.

Regarding the dataset, it was composed of 402 cases. To our knowledge, it is the largest cohort ever used in the head pCT generation field. Previous studies involved up to 92 patients (297).

Our data were split into independent sets, namely training, validation and testing. Note that most of the published studies lack a validation set (260,261,296,326,328,330), potentially leading to biased results.

To better exploit this large dataset, data augmentation could have been investigated. It aims at synthetizing artificial new images from existing ones to increase a dataset size. Regarding the most standard operations, Han (260) proposed random 3D translations of maximum 20 pixels and flips. The latter were similarly used in the left/right direction by Maspero et al. (295). More interestingly, data augmentation specific to the studied domain can be performed, to simulate for instance different noise magnitudes derived from a scanner. In 2019, Lei et al. (330) proposed to apply elastic deformations, which may represent a variation in patients anatomies. MRI-only radiotherapy can remove isotropic 2mm of errors due to registration errors (4), potentially resulting in GTV to PTV margins reduction. Head tumors located at the base of the skull and treated with 3DCRT or IMRT may benefit from this observation, since current ESTRO ACROP guidelines (344) recommend the application of large PTV margins ranging from 3mm to 5mm. However, MRI-only radiotherapy workflows heavily rely on the MRI quality, potentially suffering from distortions. The latter are either classified as geometric, i.e. impacting the voxel location, or intensity-based. In 2016, Weygand et al. (345) investigated geometric distortions quantification, and concluded that accuracy up to 2mm was achievable after applying a correction algorithm. Therefore, establishing a reliable quality assurance (346,347) is the key to unlock the full potential of radiotherapy.

Several limitations are present in this study. First, our DL pipeline necessitated paired images, and thus an intermodality registration which can introduce errors in the training set. To evaluate this error, an experienced radiologist placed three landmarks both on the CT and the MRI of ten patients. Registrating the CT onto the MR led to a mean distance error+/-standard deviation of 3.0mm+/-1.1mm. Further investigation may focus on rigid registration errors and evaluate different algorithms, such as the FLIRT (348,349) tool for comparison. Second, no analysis of the interplay effect of preprocessing steps and networks architecture was performed. Indeed, the use of a bias field correction and the selection of WS as the best standardization was based on experiments performed using HighResNet. This may have introduced a bias in the comparison between HighResNet and 3D Unet.

4.7. Conclusion

In this study, we aimed at optimizing relevant parameters to achieve high quality pCT for MR-only radiotherapy. The large variety of imaging devices and the considerable patients number constituting the training set appear to have a great impact on the pCT quality. All the remaining parameters, such as the MR sequence, intensities standardization, bias field correction, network architecture, have minor dosimetry influence as the gamma indices and DVH differences remained clinically convincing for every technique in our cohort. It suggests the efficiency of the model and its possible introduction into clinics. Future work includes the previously developed adults-based model testing on unseen paediatric patients treated with protontherapy, potentially representing a dataset more sensitive to HU.

5. Assessment of the best dosimetry metrics to characterize quality of pCT generated from MRI for brain radiotherapy – Study 2

5.1. Context

Dosimetry analyses are widely used in the literature to quantify pCT qualities. However, no clear guidelines exist regarding the optimal metrics to select. Thus, for gamma index for instance, criteria corresponding to 1%/1mm (232,287), 2%/2mm (255,299) and 3%/3mm (289,296) are commonly reported, regardless of the treatment technique. A similar heterogeneity is observed for applied dose thresholds, ranging from 0% to 90% (261,295). Lastly, the approach is either global (289,295) or unspecified (261,298). To our knowledge, no study has performed local gamma indices. All this variability deriving from the absence of recommendations results in difficulties when comparing inter-studies performances.

In addition, note that precisely detailed cohort description, including the tumor volumes, the tumor locations, etc, are rarely reported. However, some gamma indices criteria were proved to be advantaged in some situations, such as a global gamma index without dose threshold calculated in case of small target volumes. Indeed, target dose errors represented a small number of voxels, easily compensated with low doses surrounding healthy tissues.

Thus, with the collaboration of two medical physicists, the defined goals were to assess the optimal dosimetry criteria, i.e. presenting no correlation with PTV volumes, and to raise awareness about each metric limitation. To do so, training and validation sets were re-used from Study 1, and a new testing cohort was collected to exhaustively reproduce clinical situations in terms of tumor locations and treatment delivery. The new testing cohort was composed of 71 brain radiotherapy patients treated with either DYNARC (17 patients), or 3DCRT (27 patients) or VMAT (27 patients). The best preprocessing derived from Study 1, namely WS standardization combined with a bias field correction, was applied to input MRI. The previously presented modified HighResNet was used to generate pCT. Evaluation relied on MAE, VDSC calculation as well as dosimetry. With the help of a research engineer, plans were re-calculated either with pencil beam (17 DYNARC patients) or collapsed cone (27 3DCRT and 27 VMAT patients). First, DVH differences for PTV and OAR were calculated. Second, global 1%/1mm with 0%, 10%, 20%, 50% and 80% dose thresholds were assessed as well as local 1%/1mm with 0% and 10% dose thresholds. To estimate the PTV-unbiased metrics, Spearman correlations between the previous gamma indices and PTV volumes were performed.

The PTV volume-uncorrelated gamma indices were further analysed computing their Spearman correlations with intensity-based metrics, namely head/air/bone MAE and air/bone VDSC.

Lastly, the network training was stopped twice before reaching the best model, to simulate extreme errors scenarios. The previously selected gamma indices, as well as DVH differences for PTV and OAR, were evaluated to assess clinical impacts of these generation errors.

5.2. Abstract

Purpose: Generating pCT from MRI has been increasingly investigated these past few years, commonly reporting the MAE metric and gamma indices for evaluation. This study aimed at defining the dosimetry metrics the less sensitive to PTV volume, analysing whether such criteria correlate with intensity-based metrics and assessing the clinical impact of low, intermediate and high pCT qualities.

Methods: 242 and 81 brain radiotherapy patients were used to train and validate a modified HighResNet network to generate pCT. The testing cohort contained 17, 27 and 27 patients respectively treated by DYNARC, 3DCRT and VMAT. Spearman correlations between the PTV volumes and 1%/1mm global gamma indices with 0%, 10%, 20%, 50%, 80% dose thresholds, local gamma indices with 0%, 10% dose thresholds were investigated to define PTV unbiased dosimetry criteria. The correlations of these criteria with MAE and VDSC were calculated. Additional pCT were computed stopping the network training earlier to generate low and intermediate qualities and evaluate their impact on dosimetry metrics.

Results: Based on their low treatment-wise correlations with PTV volumes ($|\rho| \leq 0.53$), global gamma index with 50% dose threshold and local criterion with 10% dose threshold were selected as optimal criteria. Their absolute correlations with head MAE were intermediate to strong ($|\rho| \geq 0.43$, $p \leq 0.026$). For worst and best models, pass rates ranging from 70.00%+/-4.97% to 83.82%+/-5.60% and 63.10%+/-4.90% and 78.20%+/-6.19% were obtained for previously selected global and local criteria respectively.

Conclusion: Global and local 1%/1mm gamma indices should be preferentially computed, with avoidance of 0% and 80% extreme dose thresholds, for gamma indices-based analyses independent of PTV volumes. Extreme pCT errors generation did not result in large dosimetry discrepancies with original CT.

5.3. Introduction

Brain tumor radiotherapy treatments are based on two complementary imaging modalities, namely CT and MRI. The former, whose intensities are the HU, is easily linked to the electron densities via a calibration curve and is used for dose calculation. The latter has the advantage to present an excellent soft tissue contrast, enabling a higher quality of annotation of target volumes when combined with contrast agent. However, Ulin et al. (4) proved that the registration process used to place the different images in the same spatial frame, induces errors up to 3mm currently taken into consideration with increased PTV margins.

Generating pCT from brain MRI appears to be the most reliable and state of the art approach to overcome this issue. Methods based on DL have proved high efficiency in such a task (350). The process consists in an ensemble of layers with parameters to optimize, namely weights and biases. First, the forward cycle relies on passing the input through the network and calculating a loss function to evaluate the distance between the prediction and the ground truth. Commonly reported loss functions are MAE (255,260,299), MSE (296,297), L1-norm (261,295), and mutual information (232). The error is then backpropagated leading to parameters adjustments. Mostly reported networks architectures are UNet (260,287), GAN (232,251,255,295,298,299,328) and cycle GAN (321,326,351). DL requires large databases to ensure accuracy and robustness (232,261) and offers an inference computation time independent of the cohort size, contrary to other techniques, such as multi-atlas.

Recently, Vandewinckele et al. (1) investigated the clinical implementation of AI-based tools, such as the segmentation, treatment planning and pCT generation. More precisely, pros and cons of commonly used evaluation metrics for synthetic CT were summarized. They are briefly reminded here. First, MAE and ME representing a voxel-wise distance have the advantage to be easily implemented, but result in difficult inter-studies comparisons, and often hide spread in HU errors at a patient scale. Peak signal to noise ratio is calculated dividing the maximum of a signal by the power of the corrupting noise that affects its quality, leading to a loss of spatial information. Structural similarity metric measures the similarity between two images and has the advantage to take into account voxels interdependencies. VDSC characterizes the overlapping between two structures and has been widely considered for autocontouring tools evaluation (352–355). VDSC corresponding to bony structures has been highlighted as a relevant metric for dose calculation by the authors but suffers from being dependant on the threshold used for structure delineation and on the structure volume.

Regarding dose metrics, DVH differences have the advantage to be easily accessible. However, interpretation should be performed with caution, since pCT failure is not the only cause leading to DVH differences (356). Concerning gamma indices, difficulties are encountered when comparing pass rates across studies, partly due to the small information provided about gamma indices calculation methodology (whether it is global or local and if a dose threshold is applied, etc). Note that the two main pCT dosimetry approaches, either consisting in transferring the initial plan on the pCT without parameter change or re-optimizing directly on pCT, may not result in the same DVH and gamma indices performances (299). Finally, matching accuracy metrics have been proposed, aiming at assessing the potential pCT use for patient set up. However, residual misregistrations are difficult to analyse as no ground truth exists.

In the literature, most of the pCT quality investigations have been based on the calculation of the MAE (232,251,255,260,261,296,326–330,350), ME (261,287–289) and on a dosimetry analysis (232,255,261,287,289,295,296,298,299). Regarding the latter, most commonly used metrics involved 1%/1mm (232,255,261,287,289,298), 2%/2mm (232,255,261,289,295,298,299) and 3%/3mm (255,261,289,295,296,299) gamma indices criteria with the application of dose thresholds equal to 0% (261,295), 1% (261), 10% (261,295,298), 15% (289), 20% (287), 50% (261), and 90% (261,295). Yet, depending on the approach used to calculate the index, namely global or local, the dose threshold optionally applied, the pass rate meaning drastically changes. This information was provided by few studies (287,289,295,298), all reporting a global gamma index calculation. Thus, this article throws light on the true dosimetry criteria signification and investigates which gamma metrics are more suited for brain pCT quality assessment for radiotherapy and their link with intensity-based metrics.

This study approach is three-fold: to propose PTV volume-unbiased dosimetry metrics for photon radiotherapy, investigate their correlations with intensity-based metrics and evaluate these metrics on pCT of various qualities.

5.4. Materials and Methods

5.4.1. Data acquisition and split

The training and validation sets were re-used from a previously published study (238) evaluating the impact of training dataset size, image preprocessing and network parameters on the pCT quality for brain tumor patients.

In brief, 242 patients, composed of 121 T1 and 121 T1 Gd, were used as training set of the network. The validation set contained 81 patients (41 T1 and 40 T1Gd). All CT were acquired with a Sensation Open (Siemens Healthineers, Erlangen, Germany), associated with a 120kVp tube kilovoltage. Slice thicknesses were equal to 1mm, 2mm, 3mm and 5mm for 2, 4, 316 and 1 patient respectively. MRI were exclusively GE Healthcare devices (Chicago, United States) and were composed of Discovery MR750w (167 patients), Optima MR450w (8 patients), Signa Excite (146 patients), Optima MR360 (1 patient) and Discovery MR450 (1 patient). Slice thicknesses were comprised between (1mm; 1.2mm), (1.4mm, 2mm), (3mm, 3.2mm) and equal to 5mm for 175, 10, 137 and 1 patient.

Seventy-one patients were retrospectively included in the testing dataset. Six different radiation locations composed the cohort, namely cerebellum (4 patients), Whole Brain (WB, 26 patients), frontal (14 patients), occipital (3 patients), parietal (12 patients), and temporal (12 patients). DYNARC, 3DCRT, and VMAT techniques were used to deliver the treatment. Table 13 presents the distribution of patients, both in terms of radiation location and treatment technique, and the associated prescribed doses and PTV volumes.

Table 13: Number of patients per radiation location, i.e. cerebellum, WB, frontal, occipital, parietal, and temporal and treatment delivery technique namely DYNARC, 3DCRT, and VMAT. Prescribed doses and PTV volumes are also reported.

	Cerebellum	**WB**	**Frontal**	**Occipital**	**Parietal**	**Temporal**	**Prescribed dose (Gy)**	**PTV Volume (cc)**
Number of DYNARC patients	2	0	5	3	2	5	32.0 +/- 6.9	25 +/- 41
Number of 3DCRT patients	0	26	0	0	1	0	29.3 +/- 2.7	1663 +/- 335
Number of VMAT patients	2	0	9	0	9	7	51.5 +/- 9.4	236 +/- 135
Prescribed dose (Gy)	34.63 +/- 4.42	29.23 +/- 2.72	46.07 +/- 13.28	25.59 +/- 6.69	46.17 +/- 11.74	45.87 +/- 12.47	-	-
PTV volume (cc)	74 +/- 109	1718 +/- 174	187 +/- 188	6 +/- 6	167 +/- 110	182 +/- 142	-	-

Sixty-nine patients had one PTV, the remaining patients had two PTV. PTV volumes were respectively equal to 25cm^3 +/- 41cm^3, 1663cm^3 +/- 335cm^3 and 236cm^3 +/- 135cm^3 for DYNARC, 3DCRT and VMAT reflecting local practices. Volumes were comprised between (1cm^3; 100cm^3), (100cm^3; 250cm^3), (250cm^3; 500cm^3), (500cm^3; 1000cm^3), (1000cm^3; 1800cm^3) and (1800 cm^3; 2500 cm^3) for 22, 10, 13, 2, 17 and 9 PTV respectively. Mean +/- standard deviation prescribed doses were equal to 32.0Gy +/- 6.9Gy, 29.3Gy +/- 2.7Gy and 51.5Gy +/- 9.4Gy for the DYNARC, 3DCRT and VMAT treatments respectively. Additional information concerning imaging CT and MRI parameters are provided in Table 14.

Table 14: CT and MRI acquisition information concerning devices, acquired images types and voxels sizes.

	CT		**MRI**	
Imaging device	**Siemens Healthineers, Erlangen, Germany**	Sensation Open (71 patients)	**GE Healthcare (Chicago, United States)**	Discovery MR 750w (21 patients) Optima MR 450w (21 patients) SIGNA Architect (3 patients) SIGNA Artist (1 patient) Signa EXCITE (6 patients) SIGNA Premier (1 patient) Signa HDxt (3 patients) SIGNA Explorer (3 patients) Discovery MR 750 (2 patients)
			Siemens Healthineers (Erlangen, Germany)	MAGNETOM Aera (4 patients) MAGNETOM Avantofit (2 patients) MAGNETOM Avanto (2 patients) MAGNETOM Lumina (1 patient)
			Canon Medical Systems (Ōtawara, Japan)	Vantage Elan device (1 patient)
	Tube kilovoltage	120kVp	**Echo times**	(2.0ms; 6.0ms) (69 patients) (10.0ms; 17.0ms) (2 patients)
			Repetition times	(6.0ms; 800ms) (62 patients) (800ms; 1600ms) (4 patients) (1600ms; 2300ms) (5 patients)
			Magnetic field strength	1.5T (43) 3T (28)
Image type	Unenhanced contrast CT (71 patients)		T1 (38 patients) T1-Gd (33 patients)	
Slice thicknesses	2mm		(0.50mm; 1.90mm) (63 patients) (1.90mm; 3.40mm) (6 patients) (3.40mm; 5.00mm) (2 patients)	
In-plane voxels sizes	(0.68mm; 0.78mm) (17 patients) (0.78mm; 0.88m) (26 patients)		(0.43mm; 0.60mm) (62 patients) (0.60; 0.80) (1 patient) (0.80; 1.06mm) (8 patients)	

	(0.88mm; 0.98mm) (28 patients)	

5.4.2. Images preprocessing

CT were first rigidly registered to the MR using the drop2 library (https://github.com/biomedia-mira/drop2). All the images were then spatially resampled to a 1mm x 1mm x 1mm voxel size, before harmonizing them to a size of 300x300x242 voxels. Axes were re-arranged, to place the images in the same standard frame. CT and MRI intensities were respectively clipped between [-1000HU; 1800HU] and [0.1 percentile; 99.9 percentile]. HU were re-scaled between -1 and 1 for network stability. A N4-based filter (357) obtained from SimpleITK library was applied on the MRI to correct from bias field. Lastly, a brain extraction achieved with HD-BET tool (358) was performed on the MRI before applying a WS-based standardization (359) via the intensity-normalization package (81).

5.4.3. pCT generation

The architecture was detailed in a previous study (238) and was a modified 3D HighResNet version adapted from Li et al. (237). A constant learning rate of 0.001 was used. The selected loss function was the MAE. Early stopping was applied to identify convergence, meaning if no improvement in the validation set MAE computation was obtained during 8 consecutive epochs, the training was automatically stopped. To simulate three distinct pCT qualities, the training was stopped at epochs 3, 14 and 48 corresponding to validation set head MAE equal to 186.81HU +/- 28.12HU, 149.09 +/- 24.09HU, 100.05HU +/- 29.22HU respectively. Epoch 48 represented the best achieved image quality, as obtained using early stopping.

5.4.4. pCT evaluation

pCT were first evaluated with intensity-based metrics. The MAE was computed in four areas:

- the whole head region was computed via the Otsu approach (360) applied on 38 T1 and 33 T1-Gd MRI sequences.
- sub-volumes corresponding to the air, water and bone regions were respectively

obtained applying the same HU thresholds as in Study 1.

- VDSC within the three previous areas were also computed, between CT and pCT contours and relative confusion matrices were obtained, as proposed in Massa et al. (297), for each patient to accurately quantify the misclassifications related to the air, bone and water tissue categories.

Second, a dosimetry analysis was performed to further assess the pCT quality. The workflow consisted in rigidly registering the pCT to the CT used for treatment planning before transferring the treatment plan and re-calculating the dose. The 27 3DCRT and 27 VMAT plans were calculated with RayStation Research 8.B. (RaySearch Laboratories, Stockholm, Sweden) with a collapsed cone approach. The 17 DYNARC dosimetries were computed using the pencil beam algorithm from iPlan RT 4.5 Dose (Brainlab, Munich, Germany). From cumulative DVH, D95% and D98% were extracted for the PTV, respectively representing the dose received by 95% and 98% of the volume. Maximum doses (Dmax) were calculated for lenses, optic nerves, chiasm, brainstem. Mean doses (Dmean) were calculated for cochleae (361). Note that the OAR DVH evaluation was performed on a sub-cohort of the testing set enrolling all the 45 patients with complete segmentation, and more precisely 15 ARCDYN, 4 3DCRT and 26 VMAT cases. Relative and absolute DVH differences were respectively computed for PTV and OAR. 3D global 1%/1mm gamma indices with dose thresholds equal to 0%, 10%, 20%, 50% and 80% were computed. Additional 3D local gamma indices were obtained applying 0% and 10% dose thresholds.

Several statistical analyses were performed. First, Spearman’s tests were done to assess correlations between global 1%/1mm gamma indices with 0%, 10%, 20%, 50%, 80% thresholds, local 1%/1mm gamma indices with 0%, 10% thresholds and PTV volumes treatment-technique wise and cohort-wise. PTV-unbiased criteria, i.e. presenting small correlations with PTV volume, MAE and VDSC correlations were calculated to establish a relationship between intensity-based and dosimetry metrics. Intermediate, strong and very strong correlations respectively corresponded to significant coefficients comprised between [0.4; 0.70[, [0.70; 0.90[and [0.90; 1.00[, inspired by (362). Lastly, statistical paired Wilcoxon tests were performed between treatment-wise head MAE performances to evaluate the observed HU differences significance. All significance thresholds were set to 5%.

Note that all the reported results were derived from the testing set.

5.5. Results

5.5.1. PTV-unbiased dosimetry metrics determination

Table 15 provides the Spearman's correlation coefficients between dosimetry criteria and PTV volumes for the best model, i.e. model 48. Global 1%/1mm gamma index with 0% threshold presented intermediate to strong correlations with PTV volumes, respectively equal to -0.71 ($p=0.0015$), and -0.65 ($p=0.00012$) for the DYNARC and VMAT techniques. Correlations between the two variables remained significant for the 3DCRT technique when an 80% threshold was applied ($\rho=-0.44$, $p=0.021$). As a result, these gamma indices metrics were not selected for the rest of the study.

In addition, not applying a dose threshold implies to consider all the points within the head, and thus give weight to non-clinically relevant areas. Thus, the local 1%/1mm gamma index without dose threshold was disregarded.

Figure 50 shows the remaining PTV-unbiased gamma indices versus PTV volume. A dependency between PTV volumes and global 1%/1mm gamma index with 10% and 20% dose thresholds variables was visible (Figures 50.A and 50.B). On the contrary, thresholding with 50% appeared to reduce this effect (Figure 50.C). This observation was confirmed by Spearman's tests resulting in values equal to -0.83, -0.81 and -0.77 ($p < 0.0001$) for correlations between PTV volumes and global 1%/1mm gamma index with 10%, 20% and 50% dose threshold respectively, regardless of the technique. Thus, among the considered global criteria, the 1%/1mm gamma index with 50% dose threshold was arbitrary selected for the next study steps. It was combined with the local 1%/1mm gamma index with 10% dose threshold, reporting no correlation with DYNARC, 3DCRT and VMAT ($|\rho| \leq 0.35$, $p \geq 0.075$).

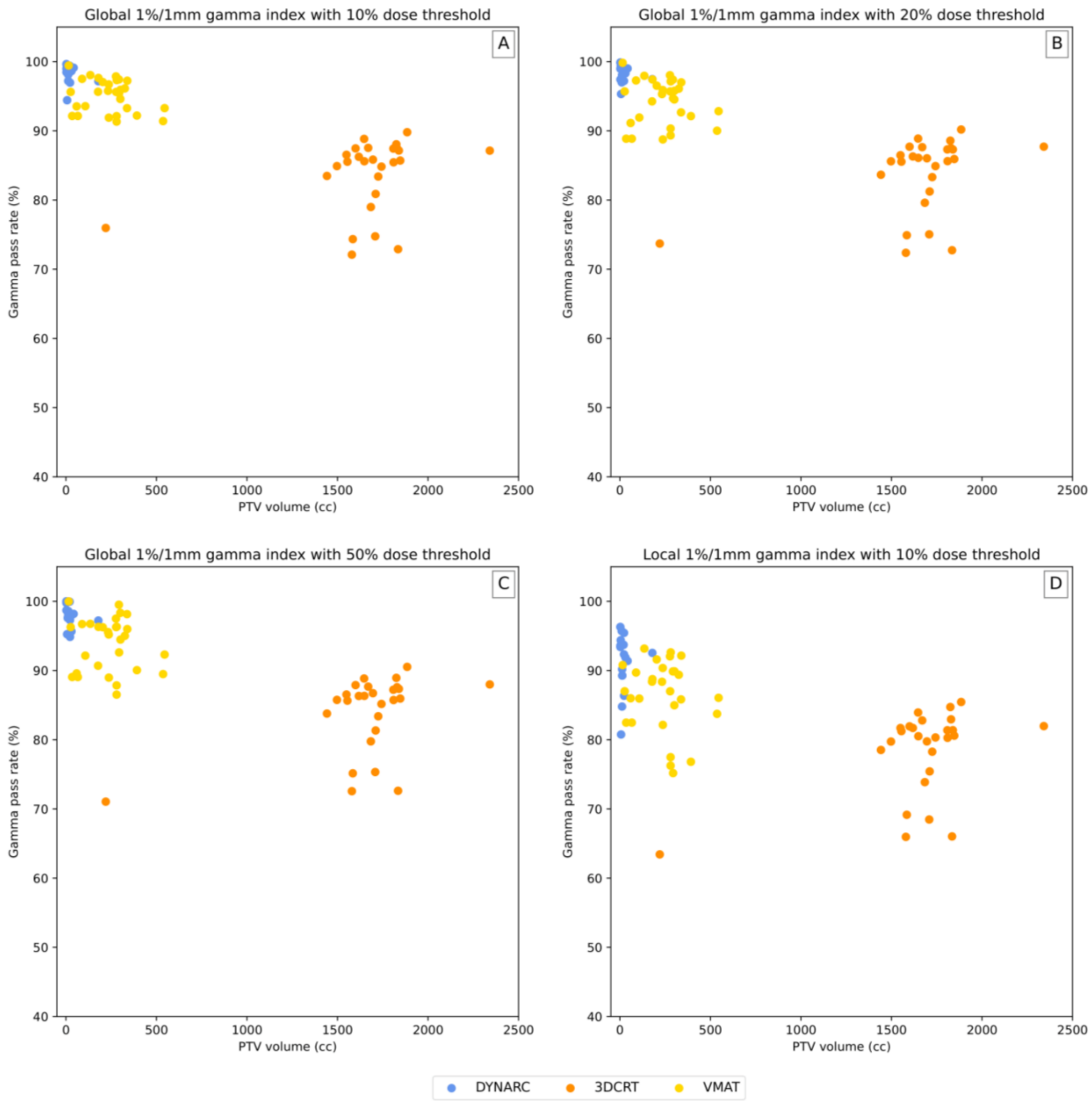

Figure 50: Pass rates against PTV volumes for four dosimetry criteria, namely global 1%/1mm 10% (A), 20% (B) and 50% (C) dose thresholds and local 1%/1mm 10% dose threshold (D). Treatment techniques are also provided, i.e. DYNARC (blue), 3DCRT (orange) and VMAT (yellow).

5.5.2. Correlation with intensity-based metrics

Correlations between global 1%/1mm gamma index with 50% dose threshold, and head and bone MAE respectively resulted in rates equal to -0.43 and -0.40 for VMAT technique ($p \leq 0.038$, Table 15). On the contrary, DYNARC presented insignificant correlations between the same criteria ($p \geq 0.39$). Regarding the correlation between the global 1%/1mm gamma index with 50% dose threshold and air/bone VDSC, only 3DCRT bone VDSC showed a significant correlation of 0.52 (p=0.0054). Intermediate and strong correlations equal to -0.52 (p=0.034), -0.60 (p=0.00084) and -0.72 (p<0.0001) were respectively achieved for DYNARC, 3DCRT and VMAT when analysing the local 1%/1mm gamma with 10% dose threshold with head MAE. Correlations between local 1%/1mm gamma index with 10% dose threshold and VDSC bone led to 0.54 (p=0.025), 0.55 (p=0.0032) and 0.74 (p<0.0001) values for DYNARC, 3DCRT and VMAT respectively.

Table 15: Spearman’s coefficients and corresponding p-values for 1%/1 mm gamma indices, PTV volume, MAE, and VDSC correlations based on model 48. Bold rates correspond to statistically significant correlations (p<0.05).

	Delivery technique	**Correlation coefficient (p-value)**
1%/1mm global no threshold & PTV	**All**	**-0.91 (< 0.0001)**
	DYNARC	**-0.71 (0.0015)**
	3DCRT	0.093 (0.65)
	VMAT	**-0.65 (0.00012)**
1%/1mm global threshold 10% & PTV	**All**	**-0.83 (< 0.0001)**
	DYNARC	-0.21 (0.42)
	3DCRT	0.32 (0.11)
	VMAT	-0.23 (0.23)
1%/1mm global threshold 20% & PTV	**All**	**-0.81 (<0.0001)**
	DYNARC	-0.13 (0.63)
	3DCRT	0.36 (0.068)
	VMAT	-0.081 (0.67)
1%/1mm global threshold 50% & PTV	**All**	**-0.77 (<0.0001)**
	DYNARC	**-0.53 (0.028)**
	3DCRT	0.38 (0.048)
	VMAT	-0.040 (0.84)
1%/1mm global threshold 80% & PTV	**All**	**-0.69 (<0.0001)**
	DYNARC	-0.48 (0.052)
	3DCRT	**0.44 (0.021)**
	VMAT	-0.068 (0.72)
1%/1mm local no threshold & PTV	**All**	**-0.33 (0.0049)**

	DYNARC	0.18 (0.49)
	3DCRT	0.29 (0.14)
	VMAT	0.26 (0.18)
1%/1mm local threshold 10% & PTV	**All**	**-0.67 (<0.0001)**
	DYNARC	-0.26 (0.30)
	3DCRT	0.35 (0.075)
	VMAT	-0.14 (0.48)
1%/1mm global threshold 50% & Head MAE	**All**	**-0.27 (0.025)**
	DYNARC	-0.22 (0.39)
	3DCRT	**-0.58 (0.0014)**
	VMAT	**-0.43 (0.026)**
1%/1mm global threshold 50% & Air MAE	**All**	**-0.37 (0.0016)**
	DYNARC	-0.34 (0.19)
	3DCRT	**-0.51 (0.0060)**
	VMAT	-0.18 (0.37)
1%/1mm global threshold 50% & Bone MAE	**All**	-0.20 (0.099)
	DYNARC	-0.16 (0.53)
	3DCRT	-0.27 (0.18)
	VMAT	**-0.40 (0.038)**
1%/1mm global threshold 50% & VDSC air	**All**	**0.24 (0.043)**
	DYNARC	0.42 (0.097)
	3DCRT	0.30 (0.13)
	VMAT	0.16 (0.41)
1%/1mm global threshold 50% & VDSC bone	**All**	0.038 (0.75)
	DYNARC	0.15 (0.55)
	3DCRT	**0.52 (0.0054)**
	VMAT	0.34 (0.087)
1%/1mm local threshold 10% & Head MAE	**All**	**-0.47 (<0.0001)**
	DYNARC	**-0.52 (0.034)**
	3DCRT	**-0.60 (0.00084)**
	VMAT	**-0.72 (< 0.0001)**
1%/1mm local threshold 10% & Air MAE	**All**	**-0.55 (<0.0001)**
	DYNARC	-0.38 (0.13)

	3DCRT	**-0.55 (0.0032)**
	VMAT	**-0.54 (0.0037)**
1%/1mm local threshold 10% & Bone MAE	**All**	**-0.39 (0.00081)**
	DYNARC	**-0.50 (0.042)**
	3DCRT	-0.31 (0.11)
	VMAT	**-0.66 (0.00017)**
1%/1mm local threshold 10% & VDSC air	**All**	**0.42 (0.00023)**
	DYNARC	0.22 (0.40)
	3DCRT	0.34 (0.082)
	VMAT	**0.56 (0.0021)**
1%/1mm local threshold 10% & VDSC bone	**All**	**0.27 (0.022)**
	DYNARC	**0.54 (0.025)**
	3DCRT	**0.55 (0.0032)**
	VMAT	**0.74 (< 0.0001)**

5.5.3. Extreme errors scenarios impact on intensity-based metrics

As presented in Table 16, head MAE equal to 172.1HU +/- 19.2HU, 185.0HU +/- 20.8HU and 195.0HU +/- 28.9HU were obtained for DYNARC, 3DCRT and VMAT patients based on model 3 (worst quality). Wilcoxon statistics resulted in p-values respectively equal to 0.073 and 0.0083 when comparing DYNARC/3DCRT and DYNARC/VMAT. Regarding model 14, observed differences between the three delivery techniques were all found significant ($p \leq 0.039$). Model 48 led to 104.8HU +/- 30.7HU, 109.2HU +/- 22.8HU and 116.8HU +/- 24.9HU for the same treatment techniques ($p \geq 0.11$).

VDSC for the air region were equal to 0.41 +/- 0.09, 0.48 +/- 0.11 and 0.68 +/- 0.08 for models 3, 14 and 48 respectively, considering all the 71 patients. Relatively small improvement through epochs was observed for VDSC in water, with rates equal to 0.89 +/- 0.02, 0.91 +/- 0.02 and 0.93 +/- 0.02 for models 3, 14 and 48 respectively.

Table 16: Technique-wise MAE and VDSC corresponding to models 3, 14 et 48.

	Delivery technique	**Model 3**	**Model 14**	**Model 48**
Head MAE (HU)	**All**	185.7 +/- 25.2	153.5 +/- 23.5	111.0 +/- 25.8
	DYNARC	172.1 +/- 19.2	140.1 +/- 21.7	104.8 +/- 30.7
	3DCRT	185.0 +/- 20.8	152.7 +/- 19.7	109.2 +/- 22.8
	VMAT	195.0 +/- 28.9	162.7 +/- 24.8	116.8 +/- 24.9
Air MAE (HU)	**All**	569.3 +/- 69.7	519.3 +/- 66.1	340.0 +/- 73.7
	DYNARC	519.5 +/- 51.4	461.1 +/- 41.0	305.0 +/- 57.4
	3DCRT	581.3 +/- 64.9	531.0 +/- 55.0	351.6 +/- 61.8
	VMAT	588.8 +/- 71.3	544.2 +/- 68.7	350.4 +/- 87.9
Bone MAE (HU)	**All**	540.3 +/- 79.4	407.3 +/- 66.0	245.9 +/- 62.8
	DYNARC	499.4 +/- 70.5	370.3 +/- 62.3	231.6 +/- 83.0
	3DCRT	539.9 +/- 58.0	406.5 +/- 57.3	242.0 +/- 54.9
	VMAT	566.5 +/- 93.5	431.6 +/- 67.5	258.9 +/- 55.3
Water MAE (HU)	**All**	50.9 +/- 8.2	47.5 +/- 8.0	51.1 +/- 13.0
	DYNARC	51.4 +/- 5.9	47.6 +/- 7.0	51.3 +/- 12.7
	3DCRT	51.0 +/- 9.6	47.1 +/- 9.8	49.4 +/- 14.4
	VMAT	50.4 +/- 8.2	47.7 +/- 6.7	52.6 +/- 12.1
VDSC air	**All**	0.41 +/- 0.09	0.48 +/- 0.11	0.68 +/- 0.08
	DYNARC	0.45 +/- 0.08	0.54 +/- 0.09	0.70 +/- 0.07
	3DCRT	0.40 +/- 0.08	0.46 +/- 0.11	0.67 +/- 0.08
	VMAT	0.40 +/- 0.09	0.47 +/- 0.11	0.69 +/- 0.09
VDSC bone	**All**	0.56 +/- 0.10	0.71 +/- 0.07	0.83 +/- 0.06
	DYNARC	0.53 +/- 0.13	0.70 +/- 0.08	0.82 +/- 0.08
	3DCRT	0.58 +/- 0.08	0.72 +/- 0.07	0.84 +/- 0.06
	VMAT	0.56 +/- 0.10	0.69 +/- 0.07	0.83 +/- 0.05
VDSC water	**All**	0.89 +/- 0.02	0.91 +/- 0.02	0.93 +/- 0.02
	DYNARC	0.89 +/- 0.02	0.91 +/- 0.02	0.93 +/- 0.02
	3DCRT	0.89 +/- 0.01	0.91 +/- 0.02	0.94 +/- 0.02
	VMAT	0.89 +/- 0.02	0.90 +/- 0.02	0.93 +/- 0.02

The mean confusion matrix obtained for all the testing set patients is presented in Table A3.

5.5.4. Extreme errors scenarios impact on dosimetry-based metrics

Dosimetry analyses are provided in Table 17. For comparison purpose, global 1%/1mm gamma index with 10% dose threshold is additionally presented to the selected gamma indices. The best improvement through epochs regarding the global 1%/1mm gamma index criterion with 50% dose threshold corresponded to 3DCRT and led to pass rates equal to 70.00% +/- 4.97%, 75.79% +/- 4.77% and 83.82% +/- 5.60% for models 3, 14 and 48 respectively. However, in most of the cases, the highest pass rates were obtained for the global 1%/1mm gamma index with 10% dose threshold, and corresponded to 94.01% +/- 9.53%, 97.17% +/- 2.22% and 98.35% +/- 1.30% for models 3, 14 and 48 respectively, combined with a DYNARC treatment technique. Pass rates of 78.93% +/- 12.35%, 84.20% +/- 12.53% and 91.26% +/- 4.16% were respectively obtained for models 3, 14 and 48, for the local 1%/1mm 10% dose threshold gamma index for the DYNARC patients. Regarding VMAT, the same criterion led to indices of 76.63% +/- 6.74%, 80.79% +/- 6.56% and 86.56% +/- 5.24% for models 3, 14 and 48 respectively. Mean PTV D98 differences ranged from 0.33% to 0.66% for the three models and techniques. Regarding OAR, maximum mean deviation was equal to 0.31Gy for chiasma, lenses and optic nerves. For brainstem and cochleas, maximum mean deviations were respectively equal to 1.61Gy and 1.30Gy (Model 3).

Table 17: Gamma indices (1%/1mm local 10% threshold) and DVH differences for models 3, 14 et 48. Results are technique-wise presented. Mean DVH difference inferior to 1% for the PTV or 0.30Gy for the OAR is represented as green. Orange values correspond to mean DVH differences comprised between 1% and 2% for the PTV and between 0.30Gy and 0.60Gy for the OARs. Values superior to 0.60Gy for the OAR are presented in red.

	Delivery technique	**Model 3**	**Model 14**	**Model 48**
1%/1mm global 10% threshold (%)	**DYNARC**	94.01 +/- 9.53	97.17 +/- 2.22	98.35 +/- 1.30
	3DCRT	69.30 +/- 4.32	75.16 +/- 4.45	83.64 +/- 5.22
	VMAT	92.34 +/- 3.65	93.73 +/- 3.17	95.19 +/- 2.38
1%/1mm global 50% threshold (%)	**DYNARC**	82.92 +/- 23.06	92.13 +/- 13.78	98.04 +/- 1.66
	3DCRT	70.00 +/- 4.97	75.79 +/- 4.77	83.82 +/- 5.60
	VMAT	89.67 +/- 5.51	91.84 +/- 5.27	93.99 +/- 3.81
1%/1mm local 10% threshold (%)	**DYNARC**	78.93 +/- 12.35	84.20 +/- 12.53	91.26 +/- 4.16
	3DCRT	63.10 +/- 4.90	68.90 +/- 4.91	78.20 +/- 6.19
	VMAT	76.63 +/- 6.74	80.79 +/- 6.56	86.56 +/- 5.24

PTV D95 relative differences (%)	**DYNARC**	0.68 +/- 0.50	0.50 +/- 0.45	0.29 +/- 0.22
	3DCRT	0.22 +/- 0.20	0.38 +/- 0.52	0.29 +/- 0.32
	VMAT	0.28 +/- 0.21	0.23 +/- 0.17	0.25 +/- 0.24
PTV D98 relative differences (%)	**DYNARC**	0.64 +/- 0.46	0.49 +/- 0.45	0.64 +/- 1.02
	3DCRT	0.51 +/- 0.79	0.66 +/- 0.72	0.54 +/- 0.61
	VMAT	0.42 +/- 0.50	0.36 +/- 0.52	0.33 +/- 0.47
***Brainstem Dmax absolute differences (Gy)**	**DYNARC**	0.02 +/- 0.03	0.03 +/- 0.08	0.01 +/- 0.02
	3DCRT	1.61 +/- 0.32	0.18 +/- 0.13	0.31 +/- 0.20
	VMAT	0.22 +/- 0.19	0.17 +/- 0.16	0.09 +/- 0.08
***Chiasm Dmax absolute differences (Gy)**	**DYNARC**	0.00 +/- 0.01	0.01 +/- 0.05	0.01 +/- 0.01
	3DCRT	0.05 +/- 0.01	0.09 +/- 0.03	0.07 +/- 0.04
	VMAT	0.20 +/- 0.29	0.14 +/- 0.22	0.09 +/- 0.12
***Lens Left Dmax absolute differences (Gy)**	**DYNARC**	0.01 +/- 0.03	0.02 +/- 0.05	0.00 +/- 0.01
	3DCRT	0.22 +/- 0.32	0.09 +/- 0.11	0.22 +/- 0.25
	VMAT	0.05 +/- 0.08	0.06 +/- 0.08	0.06 +/- 0.10
***Lens Right Dmax absolute differences (Gy)**	**DYNARC**	0.01 +/- 0.01	0.00 +/- 0.01	0.00 +/- 0.00
	3DCRT	0.31 +/- 0.50	0.23 +/- 0.45	0.22 +/- 0.31
	VMAT	0.09 +/- 0.14	0.07 +/- 0.10	0.08 +/- 0.16
***Optic Nerve Left Dmax absolute differences (Gy)**	**DYNARC**	0.01 +/- 0.03	0.03 +/- 0.09	0.00 +/- 0.01
	3DCRT	0.15 +/- 0.15	0.15 +/- 0.11	0.12 +/- 0.07
	VMAT	0.25 +/- 0.38	0.19 +/- 0.30	0.12 +/- 0.21
***Optic Nerve Right Dmax absolute differences (Gy)**	**DYNARC**	0.00 +/- 0.01	0.01 +/- 0.02	0.00 +/- 0.01
	3DCRT	0.30 +/- 0.17	0.28 +/- 0.13	0.17 +/- 0.04
	VMAT	0.22 +/- 0.34	0.17 +/- 0.31	0.11 +/- 0.19
***Cochlea Left Dmean absolute differences (Gy)**	**DYNARC**	0.04 +/- 0.11	0.02 +/- 0.04	0.01 +/- 0.01
	3DCRT	1.08 +/- 0.24	0.22 +/- 0.22	0.11 +/- 0.01
	VMAT	0.19 +/- 0.30	0.05 +/- 0.09	0.06 +/- 0.10
***Cochlea Right Dmean absolute differences (Gy)**	**DYNARC**	0.02 +/- 0.03	0.02 +/- 0.03	0.01 +/- 0.02
	3DCRT	1.30 +/- 0.24	0.34 +/- 0.45	0.21 +/- 0.29
	VMAT	0.11 +/- 0.18	0.12 +/- 0.22	0.09 +/- 0.17

*: Metrics computed on 45 patients only, who had complete OAR segmentations.

5.6. Discussion

This study aimed at analysing correlations between dose metrics and volumes of PTV technique-wise, establishing correlations with intensity-based metrics and evaluating the clinical impact of extreme errors scenario.

Two evaluation metrics were found to be complementary for an accurate characterization of the pCT quality namely global 1%/1mm gamma index with 50% dose threshold and local 1%/1mm with 10% dose threshold. The latter is a highly stringent criterion, combined with the former being less restrictive and more clinical-related. The intermediate 50% threshold was arbitrary selected since it was not excessively high, to avoid advantaging large target volumes and not extremely low, preventing small target lesions to benefit from it.

They led to absolute correlations with PTV volumes inferior to 0.53 (Table 15). For at least one delivery technique, an intermediate correlation superior or equal to 0.40 was obtained between global 1%/1mm gamma index with 50% dose threshold and head/air/bone MAE (Table 15). Absolute coefficients higher than 0.52 were achieved between 1%/1mm local with 10% dose threshold and head MAE for DYNARC, 3DCRT and VMAT techniques, suggesting moderate to strong correlations. Regarding correlations with VDSC bone, absolute rates superior or equal to 0.54 were obtained for DYNARC, 3DCRT and VMAT techniques. Thus, the previously proposed gamma indices criteria were proved to be globally linked to MAE and VDSC performances, enhancing the need for future studies to compute and report it.

Finally, three different pCT qualities were simulated stopping the network training at three different epochs, namely 3, 14 and 48, the latter representing the best achieved model. The head MAE for the entire testing cohort were equal to 185.7HU +/- 25.2HU, 153.5HU +/- 23.5HU and 111.0HU +/- 25.8HU for models 3, 14 and 48 respectively. The lowest obtained MAE was higher than the one obtained in our previous study (238), which was 81HU +/- 22HU. This discrepancy may derive from pCT evaluation performed on T1 MRI masks since the new testing cohort did not contain T2 weighted MRI sequences (T2). T1 examinations are usually larger than T2, leading to evaluations including critical areas such as teeth or neck, and thus decreasing the MAE. In addition, these discrepancies can also be associated to the heterogeneous MRI derived from various devices.

The extreme errors pCT dosimetry investigation was evaluated with the two previously defined optimal metrics. Global 1%/1mm gamma indices with 50% dose threshold led to pass rates for models 3 and 48 comprised between 82.92% +/- 23.06% and 98.04% +/- 1.66%, 70.00% +/- 4.97% and 83.82% +/- 5.60%, 89.67% +/- 5.51% and 93.99% +/- 3.81% for DYNARC, 3DCRT, VMAT respectively. Concerning the local 1%/1mm gamma index with 10% dose threshold, the lowest and highest pass rates were respectively equal to 63.10% +/- 4.90% (model 3, 3DCRT) and 91.26% +/- 4.16% (model 48, DYNARC). Globally, mean percentage points gains when comparing models 3 and 48 were respectively equal to 12%, 15% and 10% for DYNARC, 3DCRT and VMAT when considering the previous gamma criterion. As a result, no large errors were obtained for all the three models being of promising prognostic for patients. However, caution is needed for double irradiation patient cases, since it necessitates to re-use a proper previous dose map.

PTV-based D95 were found to be globally equivalent for all models, ranging from 0.28% +/- 0.21% to 0.25% +/- 0.24% for models 3 and 48 respectively, VMAT technique. Thus, as gamma indices analysis, no large error with original CT was highlighted. This observation was consolidated with the OAR DVH analysis. Indeed, all mean deviations were found to be below or equal to 0.31Gy except for brainstem and cochleas, indicating minor deviations with the real CT. 3DCRT patients were treated with two lateral beams. Thus, structures located in the beam paths, namely cochleas and brainstem, received a higher amount of dose. A pCT error generation had more impact in these areas, leading to superior DVH differences for cochleas and brainstem for model 3.

Collapsed cone dose algorithm was used for all techniques, except for DYNARC. To investigate the impact of dose algorithm on dosimetry metrics, DYNARC patients were re-calculated with Monte Carlo via the same iPlan RT 4.5 Dose (Brainlab, Munich, Germany) treatment planning software, also commissioned in our radiotherapy department. Global 1%/1mm gamma index with 50% dose threshold led to pass rates equal to 71.93% +/- 6.67%, 77.06% +/- 5.28% and 79.76% +/- 3.85% for models 3, 14 and 48 respectively. For these same models, local 1%/1mm gamma indices with 10% dose threshold were equal to 52.17% +/- 6.01%, 54.32% +/- 5.44% and 55.60% +/- 5.54%, highlighting a saturation. It may be attributed to the high accuracy of Monte Carlo approach for dose calculation since it considers heterogeneities. Thus, applying such dose algorithm highly penalizes pCT mis-generated regions, leading to lower pass rates.

Local 1%/1mm gamma indices with 10% dose threshold equal to 63.10% +/- 4.90% and 76.63% +/- 6.74% were respectively achieved for 3DCRT and VMAT techniques, concerning model 3, using the same collapsed cone dose engine. Regarding global 1%/1mm gamma index with 50% dose threshold, model 14 resulted in pass rates of 75.79% +/- 4.77% and 91.84% +/- 5.27% for 3DCRT and VMAT respectively. For the same techniques, model 48 led to values of 83.82% +/- 5.60% and 93.99 %/- 3.81% respectively. As a result, VMAT might be less sensitive to generation errors than 3DCRT, which may be due to the continuous treatment delivery and the multiple entry points, compared to the 3DCRT two lateral beams used for WB treatments in the present study. To investigate this point, five patients initially treated with 3DCRT were re-planned with VMAT. The PTV volumes were comprised between 1442cc and 1600cc. Achieved global VMAT 1%1mm gamma indices with 50% dose threshold were equal to 79.99% +/- 5.38% and 87.22% +/- 6.04%, for models 3 and 48 respectively. These rates were equal to 69.67% +/- 6.28% and 81.01% +/- 6.75% when using 3DCRT. Regarding local 1%/1mm gamma index with 10% dose threshold, models 3 and 48 respectively resulted in pass rates of 70.34% +/- 6.82% and 80.11% +/- 7.56% when delivering the dose with VMAT. Based on 3DCRT, these pass rates were equal to 63.23% +/- 6.02% and 75.36% +/- 7.33%, validating the initial hypothesis.

In the literature, Dinkla et al. (261) reported pass rates of 97.0% +/- 2.2% and 98.8% +/- 2.2% when considering 1%/1mm gamma index with 20% and 50% dose thresholds, based on 43 VMAT and 9 IMRT plans calculated with Monte Carlo. The global or local gamma criterion was unspecified. Regarding Lerner et al. (363), they used VMAT technique (1 to 4 arcs) to re-calculate the dose of the 20 patients whose PTV sizes ranged from 2cm^3 to 448cm^3. They achieved a rate of 99.1% +/- 0.6% for the 3D global 1%/1mm gamma index with a 15% dose threshold. Note that for an accurate pCT evaluation, target volumes should be reported, to avoid biases, as well as precise details about gamma index type.

One strength of the present study was the cohort diversity used for testing. First, three delivery techniques, namely DYNARC, 3DCRT and VMAT, and two dose algorithms, i.e. collapsed cone and pencil beam, were represented enabling a representation of intra-institutions clinical practices variability. In the literature, most studies focused on single MRI device (261,287,289,298), treatment delivery (232,261,289,296) and dose algorithm (261,289,296,298,299). However, note that this strength could also be viewed as a limitation, since the heterogeneity increased the varying parameters number resulting in non-straightforward performance comparisons. In addition, three different pCT categories were generated corresponding to a training model stopped at epochs 3, 14 and 48. This approach was investigated and optimized to generate realistic DL-derived pCT errors. To our knowledge, it is the first study to generate a large variety of pCT qualities based on such approach, highlighting the novelty approach.

This study suffers from certain limitations. First, some parameters, such as treatment technique and PTV volume, are highly correlated, owing to institution practises. Indeed, small volumes were treated with DYNARC and VMAT, while large volumes corresponded to 3DCRT. Thus, dissociating the two parameters when analysing differences between models is not trivial, and should be performed with caution. Similarly, the Monte Carlo algorithm was not commissioned for the 3DCRT and VMAT treatments, making some conclusions difficult to draw. Second, the previous presented errors related to CT/pCT deviations are not the radiotherapy workflow final error. They need to be combined with delivery errors, explicitly described in the 31st International Atomic Energy Agency report (178). Multileaf collimator stability and beam monitor stability uncertainties were for instance equal to several percents and 2% respectively. According to the American Association of Physicists in Medicine Report No. 85 (364), a final dose error exceeding 5% leads to changes in the tumor control probability comprised between 10% and 20%, and impacts of 20% to 30% on complication rates in normal tissue.

5.7. Conclusion

Pseudo CT were generated with DL from MRI, for brain radiotherapy patients. It was proved that DVH differences, global and local gamma indices without extreme dose thresholds, such as 0% or 80%, should be reported for an unbiased study. The complementary role of global 1%/1mm gamma index with 50% dose threshold and local 1%/1mm with 10% dose threshold to quantify the pCT quality independently of PTV volume was highlighted, as well as significant correlation with intensity-based metrics. Three different image qualities were obtained with various model training durations to simulate different extreme error scenarios, resulting in non-extreme dosimetry errors.

These recommendations should contribute to standardise the pCT assessment in literature, hence facilitating the pCT integration into radiotherapy routine.

6. Assessment of the generalizability to pediatric protontherapy of Study 1 pCT model – Study 3

6.1. Introduction

Craniopharyngioma were explored in this third study, performed during the internship of a third-year bachelor student, under my supervision. These tumors are Grade I benign neoplasms located in the sella and suprasellar regions, i.e. around the pituitary gland in the centre of the base of skull. They are close to optic nerves, chiasma, cavernous sinuses and brainstem. The associated global occurrence ranges from 2% to 5% (365), and more specifically from 6% to 13% for paediatrics (366). Regarding the latter case, highest incident rate is comprised between 8 and 10 years (367). These neoplasms present two different forms, namely adamantinous and squamous-papillary, both differentiable on the MRI (368). The latter and former forms are respectively more common in paediatrics and adults. Regarding paediatrics, tumors appear to derive from epithelial cells. Current treatments rely on a global or partial surgery, representing a challenge owing to the direct proximity of pituitary gland and hypothalamus. Surgery is then followed by radiotherapy. Current delivery techniques, as reported by Kortmann et al. (369), are either 3D stereotactic radiotherapy (370,371) or IMRT (372), with doses ranging from 30Gy to 56Gy.

However, photon-based radiotherapy is well-known to cause damages to healthy tissues surrounding target volumes, due to the deep depth after the maximum dose deposition peak required to attenuate commonly used clinical energies beams. For paediatrics, patient safety and ALARA principle are even more crucial, owing to the numerous reported side effects, such as hypothalamus-based dysfunction, cognitive dysfunction, metastases, etc (373).

As a result, it appears of great interest to use protontherapy in paediatrics with craniopharyngioma, to ensure the most local dose deposition. Indeed, body D50% respectively equal to 6.32Gy +/- 1.65Gy and 3.61Gy +/- 1.07Gy were reported for IMRT and intensity modulated protontherapy by Beltran et al. (374), proving a dose reduction with protontherapy compared to photon-based radiotherapy.

This study aimed at evaluating the transferability of a pCT model previously generated in adults (Study 1) to children, challenging task owing to the brain anatomy differences between the two groups. Study 1 proved high pCT generation performance was achieved regardless of MRI standardization and bias field filter applied. Thus, the ZMUV-based model developed and optimized on adults was selected here. The testing cohort consisted in unseen paediatric

craniopharyngioma patients treated with protontherapy. Following Study 2 guidelines, pCT quality analysis relied on global and local 1%/1mm gamma indices criteria with thresholds equal to 50% and 10% respectively, as well as DVH differences for PTV and OAR.
In this study, algorithms were re-used from Studies 1 and 2. Data collection, preprocessing and pCT evaluation were the tasks performed by the intern student.

6.2. Materials and Methods

6.2.1. Images acquisition

The training and validation sets were the same as Study 2 cohorts and were respectively composed of 323 adult patients treated with photons, resulting in the collection of 162 CT/T1 and 161 CT/T1-Gd couples.

The testing cohort was composed of eighteen paediatric patients, treated for craniopharyngioma with protontherapy between 2013 and 2018. The cohort included 9 males and 9 females. Mean age was 12.5 +/- 3.8 years. PTV volumes were comprised between [3cm^3; 27cm^3], [27cm^3; 50cm^3] and [50cm^3;78cm^3] for 7, 5 and 4 patients respectively. For the two remaining patients, namely Patients 11 and 12, no information concerning segmented volumes was available.

Unenhanced CT were all acquired on the same dedicated device, namely a Sensation Open (Siemens Healthineers, Erlangen, Germany), associated with a 120kVp and a H20s reconstruction filter. Slice thicknesses were equal to 1mm and 1.5mm for 16 and 2 patients respectively. Transverse voxel sizes were equal to 0.68mm, 0.71mm, 0.77mm, 0.85mm, 0.87mm, 0.91mm and 0.97mm for 9, 2, 1, 1, 2, 2 and 1 patients respectively.

Regarding the acquisition of the T1-Gd sequence, three different devices were included, i.e. Discovery MR 750w (GE Healthcare, Chicago, United States), Optima MR 450w (GE Healthcare, Chicago, United States) and Achieva (Philips Healthcare, Best, The Netherlands) used for 14, 3 and 1 patients respectively. All slice thicknesses were equal to 1mm. In-plane pixel spacings were equal to 0.47mm and 0.86mm for 17 and 1 patients respectively. TR were comprised between [7.21ms; 7.70ms], [7.70ms; 8.18ms] and 30ms for 12, 5 and 1 patients respectively. Lastly, reported TE ranged from 2.88ms to 2.98ms and from 2.98ms to 3.08ms for 12 and 6 patients respectively.

Mean delay between the two acquisitions was 2.0 +/- 3.0 days, obtained including every patient except 11 and 12 whose information were not available.

6.2.2. Images preprocessing

A similar preprocessing as presented in Study 1 was performed. First, CT were rigidly registered onto the MRI with Drop (https://github.com/biomedia-mira/drop2). A 1mm x 1mm x 1mm spatial trilinear resampling was then performed. After applying a padding-based size harmonization to 300x300x242 voxels, images axes re-arrangement occurred, implying interpolation. A clip between [-1000HU; 1800HU] and [0.1th centile; 99.9th centile] was performed for CT and MRI respectively. A HU rescaling between -1 and 1 enabled a higher network stability. Finally, ZMUV MRI standardization was applied.

6.2.3. Generation and evaluation of pCT

The selected network was the architecture described in Studies 1 and 2, i.e. a modified 3D HighResNet (237). The same parameters as presented in Study 1 were adopted. Indeed, the chosen loss function was the MAE. Adam optimizer was used, as well as a learning rate of 0.001. Early stopping criterion was set to 8 epochs. Training, validating and testing were performed on a single graphic card GeForce GTX 1080 Ti.

Regarding pCT evaluation, MAE within the head was computed based on T1-Gd Otsu masks. In addition, MAE within air, bone and water regions were derived with previously presented thresholds, namely corresponding to intensities below -200U, comprised between [-200HU; 250HU] and superior to 250HU respectively.

Second, a dosimetry analysis was performed to further assess pCT quality. To do so, initial CT plans were imported to Isogray 4.2.1 (DOSIsoft) to be re-calculated with pencil beam on CT and pCT. Due to technical constrains, the dosimetry cohort was a sub-cohort of the whole cohort, composed of 16 patients. Prescribed doses were equal to 52.2Gy, 52.7Gy and 54Gy for 8, 1 and 7 patients respectively. All patients had 1 PTV, except Patient 3 who had 2. 3, 4, 5 and 6 beams were used to deliver the double scattering-based treatment for 2, 3, 7 and 4 patients respectively. Based on Study 2 recommendations, global 3D 1%/1mm gamma indices were derived from 3D Slicer 4.10.2 with 50% dose threshold. Local 1%/1mm gamma index combined with a 10% dose threshold was additionally assessed. In addition, PTV-based DVH relative differences were computed for D95% and D98%. Lenses, optic nerves, chiasma and brainstem absolute Dmax differences were assessed for all patients, except for Patient 3 who did not present a brainstem contour. Lastly, cochlea Dmean absolute differences were derived for a complete OAR DVH analysis.

Spearman's statistical tests were performed between gamma indices and PTV volumes to assess their correlation. Same test parameters as in Study 2 were selected concerning the significance threshold, and the different correlation ranges definition.

6.3. Results

Eighty-three seconds were required to generate the whole 3D pCT. Figure 51 presents qualitative results, i.e. MRI, CT, pCT and gamma maps of global 1%/1mm gamma index with 50% dose threshold (Gamma map 1) and local gamma index with a threshold of 10% (Gamma map 2).

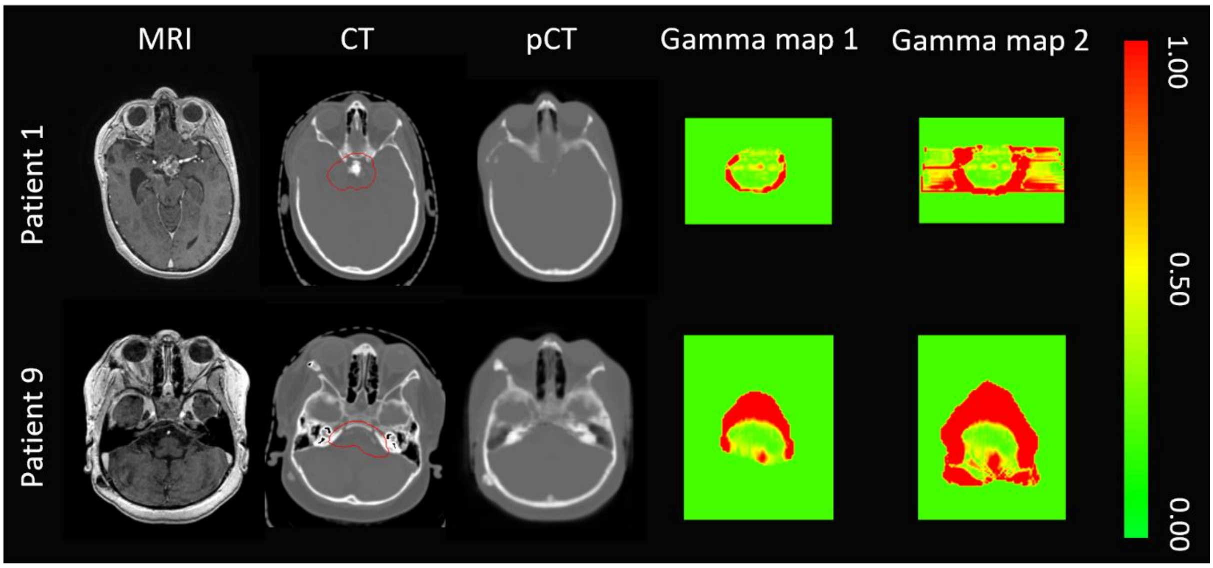

Figure 51: T1-Gd MRI, CT, pCT and corresponding gamma maps of global 1%/1mm gamma index with 50% dose threshold (Gamma map 1) and local gamma indexes with 10% dose threshold (Gamma map 2) for patients 1 and 9. Patients 1 and 9 were respectively 17 and 6 years old. Both patients received a 6-beam based treatment. Red line on the original CT represents the PTV.

Table 18 presents MAE obtained in the head, air, bone and water regions. As one can notice, mean head MAE of 110.6HU +/- 12.1HU was obtained. Patients 18 and 12 respectively presented best and worst head MAE, with values equal to 97.6HU and 138.3HU. Highest region-based discrepancies were observed for the air, exhibiting a MAE of 365.5HU +/- 57.5HU.

Table 18: Head, air, bone and water MAE in HU.

Patient, Age (years)	Head MAE	Air MAE	Bone MAE	Water MAE
Patient 1, 17.2	101.6	312.3	266.4	58.1
Patient 2, 11.2	100.7	484.5	276.0	46.0
Patient 3, 13.2	103.5	470.2	244.5	56.8
Patient 4, 15.8	126.4	360.2	334.1	63.9
Patient 5, 17.4	112.9	303.4	274.8	58.3
Patient 6, 15.0	115.9	434.0	257.0	67.0
Patient 7, 13.1	109.4	344.2	294.7	59.5
Patient 8, 8.5	134.2	383.8	282.7	91.6
Patient 9, 6.0	101.0	344.1	265.8	64.1
Patient 10, 17.3	104.4	321.7	273.2	56.1
Patient 11, 9.6	102.3	395.8	239.1	49.7
Patient 12, 17.3	138.3	413.7	348.7	61.6
Patient 13, 9.8	99.6	286.7	257.7	49.1
Patient 14, 12.0	118.7	346.4	294.5	59.9
Patient 15, 14.8	102.5	304.4	263.1	55.0
Patient 16, 11.1	113.8	331.3	273.3	65.6
Patient 17, 8.3	107.6	380.3	275.0	59.5
Patient 18, 7.0	97.6	326.1	292.0	51.0
Mean	**110.6**	**363.5**	**278.5**	**59.6**
Standard deviation	**12.1**	**57.5**	**27.6**	**9.9**

Table 19 presents gamma indices. Global 1%/1mm criterion combined with a 50% dose threshold led to pass rates of 65.10% +/- 7.50%. Patients 13 and 8 showed the highest and poorest local gamma index performances, with pass rates of 90.46% and 54.29% respectively.

Table 19: Global 1%/1mm gamma indices with 50% dose thresholds and local 1%/1mm gamma indices with 10% dose threshold. Presented results are in percentages.

Patient, Age (years)	**Global 1%/1mm gamma index - threshold 50%**	**Local 1%/1mm gamma index - threshold 10%**
Patient 1, 17.2	68.04	60.33
Patient 2, 11.2	59.55	56.59
Patient 3, 13.2	55.01	55.85
Patient 4, 15.8	69.98	67.35
Patient 5, 17.4	72.57	71.60
Patient 6, 15.0	58.12	71.80
Patient 7, 13.1	61.30	56.71
Patient 8, 8.5	52.27	54.29
Patient 9, 6.0	70.11	59.40
Patient 10, 17.3	63.46	62.68
Patient 13, 9.8	71.64	74.91
Patient 14, 12.0	65.91	71.84
Patient 15, 14.8	57.18	62.18
Patient 16, 11.1	75.00	69.09
Patient 17, 8.3	77.82	65.11
Patient 18, 7.0	63.68	56.97
Mean	**65.10**	**63.54**
Standard deviation	**7.50**	**6.79**

Spearman’s tests between PTV volumes and global 1%/1mm gamma indices with 50% dose threshold resulted in correlation of 0.16 (p=0.53). Correlation between PTV volumes and local 1%/1mm gamma indices with 10% dose threshold were 0.59 (p=0.012).

Table 20 illustrates relative D95% and D98% differences for the PTV. Mean errors of 0.16% +/- 0.17% and 0.21% +/- 0.27% were respectively obtained for D95 and D98 criteria.

Table 20: PTV-based D95% and D98% relative differences, in percentage. Differences below or equal to 1% are depicted in green.

Patient, Age (years)	**Relative differences in PTV D95%**	**Relative differences in PTV D98%**
Patient 1, 17.2	0,02	0,08
Patient 2, 11.2	0,14	0,28
Patient 3, 13.2 (PTV 1)	0,16	0,00
Patient 3, 13.2 (PTV 2)	0,08	0,17
Patient 4, 15.8	0,10	0,06
Patient 5, 17.4	0,12	0,08
Patient 6, 15.0	0,21	0,29
Patient 7, 13.1	0,16	0,00
Patient 8, 8.5	0,73	1,00
Patient 9, 6.0	0,18	0,22
Patient 10, 17.3	0,06	0,02
Patient 13, 9.8	0,06	0,08
Patient 14, 12.0	0,37	0,76
Patient 15, 14.8	0,04	0,06
Patient 16, 11.1	0,13	0,15
Patient 17, 8.3	0,15	0,12
Patient 18, 7.0	0,10	0,18
Mean	**0,16**	**0,21**
Standard deviation	**0,17**	**0,27**

Table 21 provides absolute differences for OAR. Dmax differences of 0.00Gy +/- 0.01Gy and 0.00Gy +/- 0.02Gy were respectively achieved for right and left lenses, representing the best performances. The highest deviations were observed for Dmean right and left cochlea, with mean rates equal to 0.32Gy +/- 0.60Gy and 0.46Gy and 0.68Gy respectively.

Table 21: OAR absolute DVH differences in Gy. Differences below or equal to 0.30Gy are in green. Red values represent differences larger than 0.60Gy.

Patient, Age (years)	**Right lens Dmax**	**Left lens Dmax**	**Right optic nerve Dmax**	**Left optic nerve Dmax**	**Chiasma Dmax**	**Brainstem Dmax**	**Right cochlea Dmean**	**Left cochlea Dmean**
Patient 1, 17.2	0	0	0	0.04	0.04	0.01	0.10	0.01
Patient 2, 11.2	0	0	0.08	0.06	0.03	0.12	1.48	2.11
Patient 3, 13.2	0	0	0.12	0.18	0.07	-	0.03	0.01
Patient 4, 15.8	0	0	0.14	0.09	0.04	0.11	0.09	1.57
Patient 5, 17.4	0	0	0.08	0.13	0.04	0.07	0.08	1.28
Patient 6, 15.0	0	0	0.01	0.13	0.04	0.01	0.07	0.04
Patient 7, 13.1	0	0	0.18	0.14	0.03	0.03	0.02	0.44
Patient 8, 8.5	0	0	0.05	0	0.15	0.05	0.00	0.01
Patient 9, 6.0	0	0	0.05	0.12	0.03	0.03	1.68	0.78
Patient 10, 17.3	0	0	0.08	0.05	0.06	0.11	0.07	0.05
Patient 13, 9.8	0	0	0.02	0	0.05	0.01	0.00	0.00
Patient 14, 12.0	0	0	0.05	0.12	0.05	0.06	0.05	0.04
Patient 15, 14.8	0	0	0.28	0.22	0.03	0.22	0.00	0.00
Patient 16, 11.1	0	0	0.05	0.05	0.02	0.01	0.01	0.01

Patient 17, 8.3	0.05	0	0.05	0.07	0.05	0.05	1.40	1.04
Patient 18, 7.0	0	0.06	0.04	0.11	0.02	0.04	0.01	0.01
Mean	**0.00**	**0.00**	**0.08**	**0.09**	**0.05**	**0.06**	**0.32**	**0.46**
Standard deviation	**0.01**	**0.02**	**0.07**	**0.06**	**0.03**	**0.06**	**0.60**	**0.68**

6.4. Discussion

This study aimed at assessing the generalizability of a DL model previously derived from adults and tested on paediatric patients treated for craniopharyngioma with protontherapy. Acceptable performances were achieved for most of the patients, suggesting the potential model robustness to unseen anatomies.

Concerning head, air, bone and water MAE values of 110.6HU +/- 12.1HU, 363.5HU +/- 57.5HU, 278.5HU +/- 27.6HU and 59.6HU +/- 9.9HU were respectively achieved in this study. Corresponding rates reported in our previous work (Study 1) were respectively 83.5HU +/- 22.3HU, 284.2HU +/- 62.0HU, 214.2HU +/- 55.1HU and 38.0HU +/- 11.5HU, based on the same network and DL model, but tested on 79 adult patients. In the present study, binary masks were derived from T1 acquisitions while it was T2 or T2 FLAIR in Alvarez Andres et al. (238), resulting in larger evaluation region encompassing critical areas such as teeth or neck. In 2019, Spadea et al. (288) also investigated the pCT generation with DL for protontherapy patients. Fifteen couples of CT/T1 images were collected, with delays comprised between 11 to 20 days. A 2.5D approach was adopted to train and validate 3 2D convolutional neural networks in parallel, one for each view. The final pCT value was obtained with the median of the three corresponding voxel values. A leave-one-out cross validation was performed, assigning 14 and 1 patients to the training/validation and testing sets respectively. Finally, head MAE of 65HU +/- 8HU, 59HU +/- 7HU, 60HU +/- 8HU and 54HU +/- 7HU were respectively achieved for axial, coronal, sagittal and combined approaches, outperforming the rates obtained in the current study.

Regarding dosimetry, criteria recommended from Study 2 were analysed, namely global 1%/1mm gamma indices with 50% dose threshold as well as local 1%/1mm criterion with 10%

dose threshold. Corresponding pass rates were respectively equal to 65.10% +/- 7.50% and 63.54% +/- 6.79%. Based on Study 2 best model, rates ranging from 83.82% +/- 5.60% to 98.04% +/- 1.66%, from 78.20% +/- 6.19% to 91.26% +/- 4.16% were obtained for the global 1%/1mm gamma index with 50% dose threshold and local gamma index with 10% dose threshold respectively, when considering all techniques. Discrepancies may be attributed to multiple causes, including differences in the pCT generation approach, the testing cohort size and composition, the treatment type, etc.

In parallel to the previously mentioned VMAT dosimetry study, Maspero et al. (295) investigated pCT generation for protontherapy based on 20 testing patients extracted from an initial cohort composed of 60 brain tumor paediatrics with a mean age of 10 +/- 5 years old.. Regarding treatment delivery, pencil beam scanning relying on three beams placed at 160°, 180° and 200° was used. 3D global 3%/3mm and 2%/2mm gamma indices with a 10% dose threshold led to pass rates of 99.7% +/- 0.6% and 99.6% +/- 1.1% respectively. Similarly, Neppl et al. (287) performed a hybrid dosimetry evaluation, based on VMAT and pencil beam scanning with gantry placed at 45°, 135°, 225° and 315°. The median treatment dose was 60Gy, which were delivered in 30 fractions. 3D global 2%/2mm gamma indices with a 50% dose threshold resulted in 98% +/- 2% and 97% +/- 3% for the 2D and 3D UNets respectively, proving an equivalence of the two networks in this case.

Concerning PTV-based DVH differences, mean equal to 0.16% +/- 0.17% and 0.21% +/- 0.27% were respectively achieved for D95% and D98% criteria. All means were inferior or equal to 1%. Concerning OAR, small differences inferior or equal to 0.30Gy were achieved, leading to rates of 0.06Gy +/- 0.06Gy, 0.05Gy +/- 0.03Gy, 0.08Gy +/- 0.07Gy and 0.09Gy +/- 0.06Gy for brainstem, chiasm, right and left optic nerves respectively. Exceptions occurred for Patients 2, 4, 5, 9 and 17 who presented at least one cochlea volume with a Dmean superior to 0.60Gy. As already noticed in Study 2, cochlea were located on the beam path to reach the target. Thus, a small generation error resulted in large dose differences. In the literature, Kazemifar et al. (375) enrolled 66 brain tumors patients to train and cross-validate a 2D GAN. The generator architecture was a 2D UNet, and the discriminator was a series of six convolution layers. 11 patients composed the testing set. Dosimetry relied on pencil beam scanning treatment to deliver 60Gy. Dose re-calculations were performed either with pencil beam or MC. Regarding pencil beam, CTV-based D95% absolute differences were equal to 0.4% +/- 0.4%. Dmean, mean differences for brainstem, chiasm and optic nerves were comprised between 0.4% +/- 0.6% and 1.2% +/- 1.9%. With MC, D95% for the CTV were equal to 0.5% +/- 0.4%. Dmean differences for OAR ranged from 0.5% +/- 0.5% to 0.8% +/- 1.1%, proving the dosimetry accuracy of both approaches.

Low performance was globally obtained for Patient 8, who was 8.5 years old. Indeed, head MAE was equal to 134.2HU, global 1%/1mm gamma index with 50% dose threshold was 54.29% and PTV-based DVH differences ranged from 0.73% to 1.00%. It was attributed to the poor input MRI quality, which suffers from movement artefact. On the contrary, one of the highest performances was obtained for Patient 13 with head MAE equal to 99.6HU and a local 1%/1mm gamma index with 10% dose threshold of 74.91%. One could have expected this patient to be one of the oldest patients, to have a thicker skull and to reduce anatomical differences between paediatric and adult heads. Yet, this patient was 9.8 years old. It proved the robustness and high DL model ability to adapt to unknown shapes.

This study presented several advantages. First, an adult-based DL model was used on a testing cohort composed of paediatric patients. It is a novel concept in pCT synthesis, relying on transfer learning, i.e. developing a model for a first task, re-using it on a second task with possible parameters adaption. To our knowledge, Li et al. (321) was the only to study to explore such a concept training on a set from hospital 1, testing on a second set from hospital 2 and vice versa. Indeed, hospitals 1 and 2 were composed of T2 and T1-Gd/T1 FLAIR images. Second, this study assessed the pCT quality for a use in protontherapy. Thus, the dose deposition locally occurred, via a spread-out Bragg peak. As a result, any pCT generation error resulted in shifted peak and discrepancies in dose maps. In the literature, relatively few studies evaluated DL-derived brain pCT qualities with protons (287,288,295,375). One last challenging point and successfully overcome was the cohort main histology, namely craniopharyngioma. As previously mentioned, these are tumors located near pituitary gland, bony cavities and optic nerves. Yet, as presented in Table 18, bone areas were poorly reconstructed, with a MAE equal to 278.5HU +/- 27.6HU. Thus, generating satisfying pCT quality in such conditions was an arduous task.

A major limitation was the metrics used to evaluate the pCT. Indeed, global 1%/1mm gamma index with 50% dose threshold and local gamma index with 10% dose threshold criteria were derived from Study 2 recommendations, performed for photon-based treatments. They might not be adapted in this case. This observation was in agreement with the Spearman's test performed between PTV volumes and local 1%/1mm gamma index with 10% dose threshold which proved an intermediate correlation ($|\rho|=0.59$, $p=0.012$). Second, no metric related to

Bragg peak location was used in this study. Yet, it was reported by Pileggi et al. (376) that common dosimetry metrics such as DVH or gamma indices do not provide information concerning proton range shifts. In their study, the latter were derived from 10 central dose maps slices and for each of the three treatment beams (one vertical and two axial). They were calculated as:

$$Range\ shift = R_{pCT}^{80} - R_{CT}^{80} \qquad \text{Equation 72}$$

With R_I^{80}representing the range at 80% of the maximum dose of image I.

Future investigation includes the constitution of a larger cohort, composed of 198 children representing multiple histologies, namely craniopharyngioma (92 patients) ependymoma (26 patients), rhabdomyosarcoma (42 patients) and medulloblastoma (38 patients). The novelty relies on stopping power maps-based model learning, instead of conventional HU, to free from differences in calibration curves and include larger patients in the cohort. The best strategy to predict paediatrics pCT is currently investigated. Four scenarios for network training and validation have been defined, namely a transfer learning without parameters adaption (method 1), transfer learning with parameters adaption relying on a first training on adults cohort before re-training on paediatrics (method 2), a use of mixt cohorts composed of both adults and patients (method 3), and finally purely paediatric cohorts constitution (method 4). Preliminary head MAE results are equal to 143HU +/- 20HU, 102HU +/- 19HU, 106HU +/- 19HU and 107HU +/- 20HU for methods 1, 2, 3 and 4 respectively. It potentially suggests the equivalence of methods 2, 3, 4 in terms of intensity-based performance and the requirement of a dosimetry analysis for thinner methods classification.

6.5. Conclusion

Adult-based DL model applied on unseen paediatrics cohort led to non-extreme pCT generation errors in terms of intensity-based metric, i.e. MAE. Deeper dosimetry analysis relying on protontherapy treatment delivery was performed, confirming the potential clinical pCT use. It proved the network ability to adapt to new anatomies and the probable robustness of the resulting model. Currently, a tripartite collaboration between Gustave Roussy, Centre de Protonthérapie d’Orsay and TheraPanacea is ongoing to further investigate pCT generation in case of protontherapy.

7. Global discussion

This thesis aimed at extensively characterizing and optimizing pCT quality for more accurate image synthesis. Study 1 proved that more than 200 patients were required for network convergence in terms of MAE. Concerning the MRI sequence used as input of the network, standardization of MRI intensities, bias field application and network architecture, no significant impact was proved on dosimetry, suggesting the network ability to overcome these parameters non-uniformity. With the goal to evaluate MRI-based dosimetry non-inferiority compared to conventional CT planning, Study 2 goal provided guidelines regarding optimal dosimetry metrics, namely global 1%/1mm gamma indices with 50% dose threshold, local 1%/1mm with 10% dose threshold and DVH differences for target and OAR volumes. Lastly, an application of Study 1 pCT model onto paediatrics patients treated with protontherapy was performed to evaluate its robustness. Acceptable generalizability was observed except for a few patients, potentially suggesting the model robustness.

As shown in Figure 10.E, some MRI susceptibility artefacts, such as metal-arising, can result in black areas, a signal loss, and a proper pCT prediction network incapacity. Such MRI are commonly excluded from dataset, to ensure fully informative images collection. However, these ideal cohorts lack clinical realism, owing to the high occurrence of body transformations. For instance, in Europe, between 5.5 and 6 million of dental implants are annually placed (377). A recent study performed by Palmér et al. (378) for head and neck pCT generation with MRI Planner (Spectronic Medical AB, Helsingborg, Sweden) overcomes this aspect. Indeed, the training cohort was composed of 80 couples CT/MRI. Artefact imitations were performed during data augmentation. The testing set, composed of 44 patients with CT and T1 images resulted in MAE equal to 67HU +/- 14HU, 38HU +/- 6HU, 195HU +/- 27HU, 198HU +/- 98HU for body, water, bone and air regions respectively. Transferring the VMAT plan to pCT led to mean PTV and OAR-based DVH differences inferior to 0.12Gy. Overcoming artefacts issue may also be done with a multi-atlas approach (319), which relies on multiple atlases without artefacts constitution, and thus enabling a fully pCT reconstruction from incoming MRI.

External contour plays a key role in dosimetry, since no dose calculation are performed outside for most TPS such as RayStation (RaySearch Laboratories, Stockholm, Sweden). In this thesis, external contours were independently computed for CT and pCT, before re-calculating the plans. However, owing to the use of immobilization devices only during CT examinations, anatomic differences existed, especially concerning ears and nose positions. It resulted in non-negligible discrepancies between CT and pCT external contours and different beam attenuations. Limiting this issue was proposed by Liu et al. (298), who used a GAN to generate brain pCT for 12 patients. External contours derived from pCT were re-used on CT, for both VMAT and DYNARC plans calculation. Yet, this approach did not consider variations of CT and pCT source to skin distances, causing different dose depositions. An improved approach was proposed by Persson et al. (379) for MRI Planner (Spectronic Medical AB, Helsingborg, Sweden) evaluation in prostate oncology. To do so, 170 patients were enrolled, with CT and T2 images. MRI Planner for prostate pCT generation was a previous software version relying on atlas method (380). A sub-cohort composed of 28 patients was used to investigate external contours discrepancies impact. Prescribed dose to the prostate was equal to 78Gy. Initial CT contour was used on the pCT. Air and water regions located respectively inside and outside the contour were replaced with water and air. Reported global 1%/1mm gamma indices with 15% dose threshold were equal to 99.96% +/- 0.14% and 98.43% +/- 0.98% for adjusted and original pCT respectively, suggesting the importance of contours harmonization to avoid penalizing gamma indices metrics.

For every patient of this thesis, head stereotactic frames were systematically placed for CT acquisition, since it was radiotherapy imaging, which was different from diagnostic imaging. However, our institutional practices did not include to wear a mask during MRI. Thus, frames were only visible on CT potentially causing streaking artefacts on the top and bottom axial slices. The training, which was based on corresponding CT/MRI patches, could have been biased since the network might try to reconstruct a frame from an empty MRI background. In the literature, Sjölund et al. (319), who computed brain pCT with a multi atlas approach, proposed to manually remove CT frames of patients composing the atlases. In 2017, Han (260) explored brain pCT generation with a 2D UNet. Computating MRI binary masks with Otsu approach was performed, before applying them on CT and setting the background to -1000HU. However, with the perspective of a future clinics implementation, this approach may not be feasible since frames are required on the pCT for the dose calculation.

One last major challenge of MRI-only radiotherapy workflows (not encountered in the thesis since only brain tumors were investigated), concerns organ motion (381). Indeed, many tumors including lungs, liver or pancreas sites, are highly mobile owing to breathing. Ignoring this motion would fatally result in dose delivery errors, such as under-dosages and over-dosages. For abdominal tumors for instance, a real time 2D cine MRI was reported to offer a speed of 4 frames per second leading to a duration of 45seconds for a single sagittal slice composed of tumor (382). Thus, a whole volume acquisition would represent particularly long scan durations, preventing from real time 3D MRI clinics implementation. Partial Fourier techniques have been proposed to overcome this issue, consisting in avoiding the reconstruction of redundant k-space portions representing up to 38% to 44% of the initial k-space size (383).

Various clinical trials have been developed for MRI-only radiotherapy. MR-PROTECT focused on prostate pCT generation based on MRI Planner version 1.1.2 (Spectronic Medical AB, Helsingborg, Sweden). It enrolled a total of 40 patients and global 3%/2mm gamma indices with 15% dose threshold superior to 98% were reported (384). Second clinical trial concerned rheumatology (385). Sacroiliac joints pCT were generated with BoneMRI Pelvic Region software version 1.1 (MRIguidance, Utrecht, The Netherlands), based on a UNet. A total of 30 patients were enrolled. Erosion detection specificity was equal to 96% and 89% for pCT and T1 usually acquired respectively, suggesting improved diagnostics based on pCT.

Conclusion

Nowadays brain tumors radiotherapy deals with multiple imaging types for an accurate treatment delivery. MRI and CT are routinely dually acquired since they are complementary. Introduced errors derived from images registration are currently encountered resulting in the enlargement of PTV margins, the surrounding healthy tissues irradiation and a patient safety decrease. Generating pCT from MRI with DL has been proved to be a fast and reliable approach to overcome this issue.

Lack of knowledge existed regarding the impact of image and network parameters, including the training set size, MRI input sequence, MRI standardization, bias field correction and network architecture. The training set size was proved to play a key role, since small training cohorts led to less qualitative pCT. All other investigated parameters negligibly affected the final pCT dosimetry quality, suggesting the network ability to deal with heterogeneous images.

An additional literature-derived observation was the high heterogeneity in terms of reported dosimetry metrics, preventing from inter-studies comparisons. As a result, one of the most accurate previously described configurations was selected to investigate PTV-unbiased metrics. Global 1%/1mm gamma index with 50% dose threshold, local 1%/1mm with 10% dose threshold and DVH differences for target and OAR volumes should be reported for a precise and PTV volume-unbiased pCT characterization. Intermediate to strong correlations were obtained between gamma indices metrics and head MAE, for at least 2 techniques. Once these evaluation criteria were defined, large pCT errors scenario were simulated to assess their dose impacts. No major dosimetry errors were obtained.

These dosimetry criteria were re-used on a paediatric cohort, in protontherapy. The goal was to evaluate pCT generated from the model previously trained and validated on adults set. It proved an acceptable generalizability, except for a few patients. Future work includes the constitution of a larger paediatric cohort to investigate the most efficient approach to train and validate a network based on novel stopping power maps.

One of the major challenges of AI-derived pCT integration into hospitals is the scepticism from medical clinicians. Indeed, synthetic images intensity errors would directly affect the segmentation quality and dose planning, resulting in potential under or over dosages of target volumes and OAR. One possible method to overcome this scepticism would rely on a solid quality assurance procedure. As mentioned by Vandewinckele et al. (1), pCT should be visually checked to avoid artefacts. An additional pCT algorithm, such as atlas or multi-atlas, can be applied to potentially improve the initial pCT quality.

Finally, note that this work has played a key role during GliopCT clinical trial set up. The latter aims at assessing the non-inferiority of MRI-only radiotherapy dose planning versus the conventional CT-based radiotherapy. It has also the objective to quantify pCT image quality. To do so, 30 consecutive high-grade gliomas patients will be included. The pCT will be derived from T1 or T1-Gd images with 3D GANs developed by TheraPanacea. Regarding dosimetry, relative differences between CT and pCT concerning $D_{2\%}$, $D_{95\%}$, $D_{98\%}$ and D_{mean}/D_{max} will be respectively assessed for PTV and OAR. Finally, 3D global and local 1%/1mm and 2%/2mm gamma indices will be computed. Global and local criteria will be combined with thresholds equal to 20%, 50% and 0%, 10% respectively.

Bibliography

1. Vandewinckele L, Claessens M, Dinkla A, Brouwer C, Crijns W, Verellen D, et al. Overview of artificial intelligence-based applications in radiotherapy: Recommendations for implementation and quality assurance. Radiotherapy and Oncology. 1 déc 2020;153:55-66.

2. Sung H, Ferlay J, Siegel RL, Laversanne M, Soerjomataram I, Jemal A, et al. Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. CA: A Cancer Journal for Clinicians. 2021;71(3):209-49.

3. Johnson KJ, Cullen J, Barnholtz-Sloan JS, Ostrom QT, Langer CE, Turner MC, et al. Childhood Brain Tumor Epidemiology: A Brain Tumor Epidemiology Consortium Review. Cancer Epidemiol Biomarkers Prev. déc 2014;23(12):2716-36.

4. Ulin K, Urie MM, Cherlow JM. Results of a multi-institutional benchmark test for cranial CT/MR image registration. Int J Radiat Oncol Biol Phys. 1 août 2010;77(5):1584-9.

5. Louis DN, Perry A, Reifenberger G, von Deimling A, Figarella-Branger D, Cavenee WK, et al. The 2016 World Health Organization Classification of Tumors of the Central Nervous System: a summary. Acta Neuropathol. juin 2016;131(6):803-20.

6. Mehta M, Vogelbaum MA, Chang S, Patel N. Neoplasms of the central nervous system. DeVita VT Jr, Lawrence TS, Rosenberg SA: Cancer: Principles and Practice of Oncology 9th edition Lippincott Williams & Wilkins. 2011;1700-49.

7. Fisher JL, Schwartzbaum JA, Wrensch M, Wiemels JL. Epidemiology of Brain Tumors. Neurologic Clinics. 1 nov 2007;25(4):867-90.

8. Perkins A, Liu G. Primary Brain Tumors in Adults: Diagnosis and Treatment. AFP. 1 févr 2016;93(3):211-7.

9. Strong M, Garces J. Brain Tumors: Epidemiology and Current Trends in Treatment. Journal of Brain Tumors & Neurooncology. 1 janv 2016;01.

10. Hamberger MJ. Cortical Mapping. In: Kreutzer JS, DeLuca J, Caplan B, éditeurs. Encyclopedia of Clinical Neuropsychology. New York, NY: Springer; 2011. p. 719-21.

11. Park HR, Kshettry VR, Farrell CJ, Lee JM, Kim YH, Won TB, et al. Clinical Outcome After Extended Endoscopic Endonasal Resection of Craniopharyngiomas: Two-Institution Experience. World Neurosurgery. 1 juill 2017;103:465-74.

12. Moussazadeh N, Prabhu V, Bander ED, Cusic RC, Tsiouris AJ, Anand VK, et al. Endoscopic endonasal versus open transcranial resection of craniopharyngiomas: a case-matched single-institution analysis. Neurosurgical Focus. 1 déc 2016;41(6):E7.

13. Song SW, Kim YH, Kim JW, Park C-K, Kim JE, Kim DG, et al. Outcomes After Transcranial and Endoscopic Endonasal Approach for Tuberculum Meningiomas—A Retrospective Comparison. World Neurosurgery. 1 janv 2018;109:e434-45.

14. Bander ED, Singh H, Ogilvie CB, Cusic RC, Pisapia DJ, Tsiouris AJ, et al. Endoscopic endonasal versus transcranial approach to tuberculum sellae and planum sphenoidale meningiomas in a similar cohort of patients. Journal of Neurosurgery. 27 janv 2017;128(1):40-8.

15. Zoli M, Milanese L, Bonfatti R, Faustini-Fustini M, Marucci G, Tallini G, et al. Clival chordomas: considerations after 16 years of endoscopic endonasal surgery. Journal of Neurosurgery. 14 avr 2017;128(2):329-38.

16. Koutourousiou M, Gardner PA, Tormenti MJ, Henry SL, Stefko ST, Kassam AB, et al. Endoscopic Endonasal Approach for Resection of Cranial Base Chordomas: Outcomes and Learning Curve. Neurosurgery. 1 sept 2012;71(3):614-25.

17. InformedHealth.org [Internet]. Cologne, Germany: Institute for Quality and Efficiency in Health Care (IQWiG) [Internet]. 2006. Disponible sur: https://www.ncbi.nlm.nih.gov/books/NBK279427/

18. Ortholan C, Estivalet S, Barillot I, Costa A, Gérard J. Guide des procédures de radiothérapie externe 2007. Cancer / Radiothérapie. 2007;11(6-7):329-30.

19. Stupp R, Mason WP, van den Bent MJ, Weller M, Fisher B, Taphoorn MJB, et al. Radiotherapy plus Concomitant and Adjuvant Temozolomide for Glioblastoma. New England Journal of Medicine. 10 mars 2005;352(10):987-96.

20. Feuvret L, Antoni D, Biau J, Truc G, Noël G, Mazeron J-J. Référentiels d’irradiation des gliomes. Cancer/Radiothérapie. 1 sept 2016;20:S69-79.

21. Yan H, Parsons DW, Jin G, McLendon R, Rasheed BA, Yuan W, et al. IDH1 and IDH2 mutations in gliomas. N Engl J Med. 19 févr 2009;360(8):765-73.

22. Taillandier L, Bauchet L, Figarella-Branger D, Labrousse F, Mohktari K, Duffau H, et al. REFERENTIEL de l’ANOCEF pour les GLIOMES de L’ADULTE. Association des neuro-oncologues d’expression française (Anocef); 2012.

23. Noel G, Taillandier L. Médulloblastomes de l’adulte. Association des neuro-oncologues d’expression française (Anocef); 2016.

24. Chauffert B, Ahle G, Soubeyran P, Bay JO, Casasnovas O, Lamy T, et al. Lymphomes Primitifs du Système Nerveux Central. Association des neuro-oncologues d’expression française (Anocef); 2014.

25. Survival | Brain and spinal cord tumours | Cancer Research UK [Internet]. Disponible sur: https://www.cancerresearchuk.org/about-cancer/brain-tumours/survival

26. Survival Rates for Selected Adult Brain and Spinal Cord Tumors [Internet]. Disponible sur: https://www.cancer.org/cancer/brain-spinal-cord-tumors-adults/detection-diagnosis-staging/survival-rates.html

27. Buzug TM. Computed Tomography. In: Kramme R, Hoffmann K-P, Pozos RS, éditeurs. Springer Handbook of Medical Technology [Internet]. Berlin, Heidelberg: Springer Berlin Heidelberg; 2011. p. 311-42. Disponible sur: https://doi.org/10.1007/978-3-540-74658-4_16

28. Luke AM, Shetty KP, Satish SV, Kilaru K. Comparison of Spiral Computed Tomography and Cone-Beam Computed Tomography. Journal of Indian Academy of Oral Medicine and Radiology. 7 janv 2013;25(3):173.

29. McCollough CH, Leng S, Yu L, Fletcher JG. Dual- and Multi-Energy CT: Principles, Technical Approaches, and Clinical Applications. Radiology. 1 sept 2015;276(3):637-53.

30. Coursey CA, Nelson RC, Boll DT, Paulson EK, Ho LM, Neville AM, et al. Dual-Energy Multidetector CT: How Does It Work, What Can It Tell Us, and When Can We Use It in Abdominopelvic Imaging? RadioGraphics. 1 juill 2010;30(4):1037-55.

31. Krestel E. Imaging systems for medical diagnostics [Internet]. 1990. Disponible sur: http://inis.iaea.org/Search/search.aspx?orig_q=RN:22037579

32. Michael G. X-ray computed tomography. Phys Educ. 1 nov 2001;36(6):442.

33. Cunningham IA, Judy PF. Computed Tomography. In 2000.

34. Ulzheimer S, Freund J. The Stellar Detector. White Paper - Siemens. 2013;

35. Chetih N, Messali Z. Tomographic image reconstruction using filtered back projection (FBP) and algebraic reconstruction technique (ART). In 2015. p. 1-6.

36. Willemink MJ, Noël PB. The evolution of image reconstruction for CT-from filtered back projection to artificial intelligence. Eur Radiol. mai 2019;29(5):2185-95.

37. Stiller W. Basics of iterative reconstruction methods in computed tomography: A vendor-independent overview. European Journal of Radiology. 1 déc 2018;109:147-54.

38. Xue Z, Antani S, Long LR, Demner-Fushman D, Thoma GR. Window Classification of Brain CT Images in Biomedical Articles. AMIA Annu Symp Proc. 3 nov 2012;2012:1023-9.

39. Ohkubo M, Wada S, Ida S, Kunii M, Kayugawa A, Matsumoto T, et al. Determination of point spread function in computed tomography accompanied with verification. Medical Physics. 2009;36(6Part1):2089-97.

40. Verdun FR, Racine D, Ott JG, Tapiovaara MJ, Toroi P, Bochud FO, et al. Image quality in CT: From physical measurements to model observers. Phys Med. déc 2015;31(8):823-43.

41. Kayugawa A, Ohkubo M, Wada S. Accurate determination of CT point-spread-function with high precision. J Appl Clin Med Phys. 8 juill 2013;14(4):3905.

42. Lin E, Alessio A. What are the basic concepts of temporal, contrast, and spatial resolution in cardiac CT? J Cardiovasc Comput Tomogr. déc 2009;3(6):403-8.

43. Rodríguez-Sánchez Á, Thompson A, Körner L, Brierley N, Leach R. Review of the influence of noise in X-ray computed tomography measurement uncertainty. Precision Engineering. 1 nov 2020;66:382-91.

44. Duan X, Wang J, Leng S, Schmidt B, Allmendinger T, Grant K, et al. Electronic noise in CT detectors: Impact on image noise and artifacts. AJR Am J Roentgenol. oct 2013;201(4):W626-632.

45. Kuo Y, Lin Y-Y, Lee R-C, Lin C-J, Chiou Y-Y, Guo W-Y. Comparison of image quality from filtered back projection, statistical iterative reconstruction, and model-based iterative reconstruction algorithms in abdominal computed tomography. Medicine (Baltimore). août 2016;95(31):e4456.

46. Shuman WP, Green DE, Busey JM, Kolokythas O, Mitsumori LM, Koprowicz KM, et al. Model-Based Iterative Reconstruction Versus Adaptive Statistical Iterative Reconstruction and Filtered Back Projection in Liver 64-MDCT: Focal Lesion Detection, Lesion Conspicuity, and Image Noise. American Journal of Roentgenology. 1 mai 2013;200(5):1071-6.

47. Barrett JF, Keat N. Artifacts in CT: Recognition and Avoidance. RadioGraphics. 1 nov 2004;24(6):1679-91.

48. Boas FE, Fleischmann D. CT artifacts: causes and reduction techniques. Imaging in Medicine. 2 avr 2012;4(2):229-40.

49. Dotson B, Lambert JW, Wang ZJ, Sun Y, Ohliger MA, Winklhofer S, et al. Benefit of iodine density images to reduce out-of-field image artifacts at rapid kVp switching dual-energy CT. Abdom Radiol. 1 mars 2017;42(3):735-41.

50. Davis G, Jain N, Elliott J. A modelling approach to beam hardening correction. Proceedings of SPIE - The International Society for Optical Engineering. 18 sept 2008;7078:70781E.

51. Monnin P, Sfameni N, Gianoli A, Ding S. Optimal slice thickness for object detection with longitudinal partial volume effects in computed tomography. Journal of Applied Clinical Medical Physics. 2017;18(1):251-9.

52. Keall P, Vedam S, George R, Williamson J. Respiratory regularity gated 4D CT acquisition: Concepts and proof of principle. Australasian physical & engineering sciences in medicine / supported by the Australasian College of Physical Scientists in Medicine and the Australasian Association of Physical Sciences in Medicine. 1 oct 2007;30:211-20.

53. Morton N, Sykes J, Barber J, Hofmann C, Keall P, O'Brien R. Reducing 4D CT imaging artifacts at the source: first experimental results from the respiratory adaptive computed tomography (REACT) system. Phys Med Biol. avr 2020;65(7):075012.

54. Subhas N, Primak AN, Obuchowski NA, Gupta A, Polster JM, Krauss A, et al. Iterative metal artifact reduction: Evaluation and optimization of technique. Skeletal Radiol. 1 déc 2014;43(12):1729-35.

55. Zhang Y, Yu H. Convolutional Neural Network Based Metal Artifact Reduction in X-Ray Computed Tomography. IEEE Transactions on Medical Imaging. juin 2018;37(6):1370-81.

56. Dance D, Christofides S, Maidment A, McLean I. Diagnostic Radiology Physics : A Handbook for Teachers and Students - International Atomic Energy Agency. 2014.

57. Bojorquez JZ, Bricq S, Acquitter C, Brunotte F, Walker PM, Lalande A. What are normal relaxation times of tissues at 3 T? Magn Reson Imaging. janv 2017;35:69-80.

58. Hornak J. The basics of MRI. In 2010.

59. Ferreira P, Gatehouse P, Firmin D. Cardiovascular magnetic resonance artefacts. Journal of cardiovascular magnetic resonance : official journal of the Society for Cardiovascular Magnetic Resonance. 22 mai 2013;15:41.

60. Markl M, Leupold J. Gradient echo imaging. Journal of Magnetic Resonance Imaging. 2012;35(6):1274-89.

61. Ernst RR, Anderson WA. Application of Fourier Transform Spectroscopy to Magnetic Resonance. Review of Scientific Instruments. 1 janv 1966;37(1):93-102.

62. Chavhan GB, Babyn PS, Thomas B, Shroff MM, Haacke EM. Principles, Techniques, and Applications of T2*-based MR Imaging and Its Special Applications. Radiographics. sept 2009;29(5):1433-49.

63. Tang MY, Chen TW, Zhang XM, Huang XH. GRE T2∗-Weighted MRI: Principles and Clinical Applications. Biomed Res Int. 2014;2014:312142.

64. Cakirer S, Karaarslan E, Arslan A. Spontaneously T1-hyperintense lesions of the brain on MRI: a pictorial review. Curr Probl Diagn Radiol. oct 2003;32(5):194-217.

65. Xiao Y-D, Paudel R, Liu J, Ma C, Zhang Z-S, Zhou S-K. MRI contrast agents: Classification and application (Review). Int J Mol Med. nov 2016;38(5):1319-26.

66. Wahsner J, Gale EM, Rodríguez-Rodríguez A, Caravan P. Chemistry of MRI Contrast Agents: Current Challenges and New Frontiers. Chem Rev. 23 janv 2019;119(2):957-1057.

67. Rohrer M, Bauer H, Mintorovitch J, Requardt M, Weinmann H-J. Comparison of magnetic properties of MRI contrast media solutions at different magnetic field strengths. Invest Radiol. nov 2005;40(11):715-24.

68. Montagne A, Toga AW, Zlokovic BV. Blood-Brain Barrier Permeability and Gadolinium: Benefits and Potential Pitfalls in Research. JAMA Neurol. janv 2016;73(1):13-4.

69. Krupa K, Bekiesińska-Figatowska M. Artifacts in Magnetic Resonance Imaging. Pol J Radiol. 23 févr 2015;80:93-106.

70. Budrys T, Veikutis V, Lukosevicius S, Gleizniene R, Monastyreckiene E, Kulakiene I. Artifacts in magnetic resonance imaging: how it can really affect diagnostic image quality and confuse clinical diagnosis? Journal of Vibroengineering. 2018;20(2):1202-13.

71. Constable RT, Henkelman RM. Data extrapolation for truncation artifact removal. Magn Reson Med. janv 1991;17(1):108-18.

72. Zaitsev M, Maclaren Julian, Herbst M. Motion Artefacts in MRI: a Complex Problem with Many Partial Solutions. J Magn Reson Imaging. oct 2015;42(4):887-901.

73. Liu J, Kocak M, Supanich M, Deng J. Motion artifacts reduction in brain MRI by means of a deep residual network with densely connected multi-resolution blocks (DRN-DCMB). Magn Reson Imaging. sept 2020;71:69-79.

74. Zhao C, Carass A, Dewey B, Woo J, Oh J, Calabresi P, et al. A Deep Learning Based Anti-aliasing Self Super-Resolution Algorithm for MRI. In: MICCAI. 2018.

75. Stadler A, Schima W, Ba'ssalamah A, Kettenbach J, Eisenhuber E. Artifacts in body MR imaging: Their appearance and how to eliminate them. European radiology. 1 juin 2007;17:1242-55.

76. Hendrix LE, Kneeland JB, Haughton VM, Daniels DL, Szumowski J, Williams AL, et al. MR imaging of optic nerve lesions: value of gadopentetate dimeglumine and fat-suppression technique. American Journal of Roentgenology. 1 oct 1990;155(4):849-54.

77. Mudgal P. Gibbs and truncation artifact | Radiology Case | Radiopaedia.org [Internet]. Radiopaedia. Disponible sur: https://radiopaedia.org/cases/gibbs-and-truncation-artifact

78. Bashir U. Aliasing artifact | Radiology Case | Radiopaedia.org [Internet]. Radiopaedia. Disponible sur: https://radiopaedia.org/cases/aliasing-artifact-1

79. Schmitz B, Aschoff A, Hoffmann M, Grön G. Advantages and pitfalls in 3T MR brain imaging: A pictorial review. AJNR American journal of neuroradiology. 1 nov 2005;26:2229-37.

80. Nyúl LG, Udupa JK, Zhang X. New variants of a method of MRI scale standardization. IEEE Trans Med Imaging. févr 2000;19(2):143-50.

81. Reinhold JC, Dewey BE, Carass A, Prince JL. Evaluating the impact of intensity normalization on MR image synthesis. Medical Imaging 2019: Image Processing. 2019;10949:109493H.

82. Shinohara R, Sweeney E, Goldsmith J, Shiee N, Mateen F, Calabresi P, et al. Statistical normalization techniques for magnetic resonance imaging. Neuroimage Clin. 15 août 2014;6:9-19.

83. Dewey BE, Zhao C, Reinhold JC, Carass A, Fitzgerald KC, Sotirchos ES, et al. DeepHarmony: A deep learning approach to contrast harmonization across scanner changes. Magnetic Resonance Imaging. 1 déc 2019;64:160-70.

84. Maintz JBA, Viergever MA. A survey of medical image registration. Medical Image Analysis. 1 mars 1998;2(1):1-36.

85. Hill DLG, Batchelor PG, Holden M, Hawkes DJ. Medical image registration. Phys Med Biol. févr 2001;46(3):R1-45.

86. Rueckert D, Schnabel JA. Medical Image Registration. In: Deserno TM, éditeur. Biomedical Image Processing [Internet]. Berlin, Heidelberg: Springer Berlin Heidelberg; 2011. p. 131-54. Disponible sur: https://doi.org/10.1007/978-3-642-15816-2_5

87. Penny WD, Friston KJ, Ashburner JT, Kiebel SJ, Nichols TE. Statistical Parametric Mapping: The Analysis of Functional Brain Images. Elsevier; 2011. 689 p.

88. Auzias G, Colliot O, Glaunes JA, Perrot M, Mangin J-F, Trouve A, et al. Diffeomorphic Brain Registration Under Exhaustive Sulcal Constraints. IEEE Transactions on Medical Imaging. juin 2011;30(6):1214-27.

89. Khan AR, Wang L, Beg MF. Multistructure Large Deformation Diffeomorphic Brain Registration. IEEE Transactions on Biomedical Engineering. févr 2013;60(2):544-53.

90. Landberg T, Chavaudra J, Dobbs J, Hanks G, Johansson K-A, Möller T, et al. ICRU Report 50 - Prescribing, Recording, and Reporting Photon Beam Therapy. Journal of the International Commission on Radiation Units and Measurements. 1 sept 1993;os26(1):NP-NP.

91. Landberg T, Chavaudra J, Dobbs J, Gerard J-P, Hanks G, Horiot J-C, et al. ICRU Report 62 - Prescribing, Recording, and Reporting Photon Beam Therapy (Supplement to ICRU Report 50). Journal of the International Commission on Radiation Units and Measurements. 1 nov 1999;os32(1):NP-NP.

92. Hodapp N. ICRU Report 83 : prescribing, recording and reporting photon-beam intensity-modulated radiation therapy (IMRT). Strahlenther Onkol. janv 2012;188(1):97-9.

93. Chavaudra J, Bridier A. Définition des volumes en radiothérapie externe : rapports ICRU 50 et 62. Cancer/Radiothérapie. 1 oct 2001;5(5):472-8.

94. Kantor G, Maingon P, Mornex F, Mazeron JJ. Contours des volumes cibles en radiothérapie. Généralités. Cancer/Radiothérapie. 1 nov 2002;6:56-60.

95. Niyazi M, Brada M, Chalmers AJ, Combs SE, Erridge SC, Fiorentino A, et al. ESTRO-ACROP guideline “target delineation of glioblastomas”. Radiotherapy and Oncology. 1 janv 2016;118(1):35-42.

96. Ajithkumar T, Horan G, Padovani L, Thorp N, Timmermann B, Alapetite C, et al. SIOPE – Brain tumor group consensus guideline on craniospinal target volume delineation for high-precision radiotherapy. Radiotherapy and Oncology. 1 août 2018;128(2):192-7.

97. Charron O, Lallement A, Jarnet D, Noblet V, Clavier J-B, Meyer P. Automatic detection and segmentation of brain metastases on multimodal MR images with a deep convolutional neural network. Comput Biol Med. 1 avr 2018;95:43-54.

98. Sun Y, Shi H, Zhang S, Wang P, Zhao W, Zhou X, et al. Accurate and rapid CT image segmentation of the eyes and surrounding organs for precise radiotherapy. Med Phys. mai 2019;46(5):2214-22.

99. Tang F, Liang S, Zhong T, Huang X, Deng X, Zhang Y, et al. Postoperative glioma segmentation in CT image using deep feature fusion model guided by multi-sequence MRIs. Eur Radiol. févr 2020;30(2):823-32.

100. Hu S-Y, Weng W-H, Lu S-L, Cheng Y-H, Xiao F, Hsu F-M, et al. Multimodal Volume-Aware Detection and Segmentation for Brain Metastases Radiosurgery. ArXiv. 2019;abs/1908.05418.

101. Landberg T, Chavaudra J, Dobbs J, Gerard J-P, Hanks G, Horiot J-C, et al. Report 62. J ICRU. 1 nov 1999;os32(1):NP-NP.

102. Reynaert N, van der Marck S, Schaart D. Monte Carlo Treatment Planning—An Introduction, Nederlandse Commissie Voor Stralingsdosimetrie. 2006. Report No.: 16 of the Netherlands Commission on Radiation Dosimetry Netherlands, Commission on Radiation Dosimetry Subcommission Monte Carlo Treatment Planning.

103. Kyriakou I, Ivanchenko V, Sakata D, Bordage MC, Guatelli S, Incerti S, et al. Influence of track structure and condensed history physics models of Geant4 to nanoscale electron transport in liquid water. Phys Med. févr 2019;58:149-54.

104. Monte Carlo Methods in Practice (Generating Random Numbers) [Internet]. Disponible sur: https://www.scratchapixel.com/lessons/mathematics-physics-for-computer-graphics/monte-carlo-methods-in-practice/generating-random-numbers

105. Legrady D, Halasz M, Kophazi J, Molnar B, Tolnai G. Population-based variance reduction for dynamic Monte Carlo. Annals of Nuclear Energy. 15 déc 2020;149:107752.

106. Kawrakow I, Rogers DWO, Walters BRB. Large efficiency improvements in BEAMnrc using directional bremsstrahlung splitting. Medical Physics. 2004;31(10):2883-98.

107. Zeghari A, Saaidi R, Cherkaoui El Moursli R. Investigation of variance reduction techniques parameters to enhance the efficiency for a 12 MV photon beam. Journal of Radiation Research and Applied Sciences. 2 janv 2019;12(1):192-9.

108. Mohammed M, Chakir E, Boukhal H, Saeed M, El Bardouni T. Evaluation of variance reduction techniques in BEAMnrc Monte Carlo simulation to improve the computing efficiency. Journal of Radiation Research and Applied Sciences. 13 oct 2016;9(4):424-30.

109. Oelkfe U, Scholz C. Dose Calculation Algorithms. In: Schlegel W, Bortfeld T, Grosu A-L, éditeurs. New Technologies in Radiation Oncology [Internet]. Berlin, Heidelberg: Springer; 2006 [cité 3 juill 2021]. p. 187-96. (Medical Radiology). Disponible sur: https://doi.org/10.1007/3-540-29999-8_15

110. Knöös T. 3D dose computation algorithms. J Phys: Conf Ser. mai 2017;847:012037.

111. Yang MC. Optimisation des plans de traitement en radiothérapie grâce aux dernières techniques de calcul de dose rapide [Internet] [phdthesis]. Université Paris Sud - Paris XI; 2014. Disponible sur: https://tel.archives-ouvertes.fr/tel-01011223

112. Ahnesjö A. Collapsed cone convolution of radiant energy for photon dose calculation in heterogeneous media. Med Phys. août 1989;16(4):577-92.

113. Cho W, Suh T-S, Park J-H, Xing L, Lee J-W. Practical implementation of a collapsed cone convolution algorithm for a radiation treatment planning system. Journal of the Korean Physical Society. 1 déc 2012;61(12):2073-83.

114. Whitton K, Hu X, Yu C, Chen D. An FPGA Solution for Radiation Dose Calculation. In 2006. p. 227-36.

115. Fogliata A, Nicolini G, Vanetti E, Clivio A, Winkler P, Cozzi L. The impact of photon dose calculation algorithms on expected dose distributions in lungs under different respiratory phases. Phys Med Biol. 7 mai 2008;53(9):2375-90.

116. Chaikh A, Khamphan C, Kumar T, Garcia R, Balosso J. What should we know about photon dose calculation algorithms used for radiotherapy? Their impact on dose distribution and medical decisions based on TCP/NTCP. Int J Cancer Ther Oncol. 2016;4(4).

117. Sievinen J, Ulmer W, Kaissl W. AAA Photon Dose Calculation Model in Eclipse^{TM}.

118. Blanpain B. Vers un calcul en temps réel de la dose dans un fantôme segmenté en mailles homogènes [phd]. Université de Toulouse, Université Toulouse III - Paul Sabatier; 2009.

119. Oelfke U, Bortfeld T. Inverse planning for photon and proton beams. Medical Dosimetry. 1 juin 2001;26(2):113-24.

120. Semenenko VA, Reitz B, Day E, Qi XS, Miften M, Li XA. Evaluation of a commercial biologically based IMRT treatment planning system. Medical Physics. 2008;35(12):5851-60.

121. Hartmann M, Bogner L. Investigation of intensity-modulated radiotherapy optimization with gEUD-based objectives by means of simulated annealing. Medical Physics. 2008;35(5):2041-9.

122. Mohammadi S, Shang C, Ouhib Z, Leventouri T, Kalantzis G. A computational study on different penalty approaches for constrained optimization in radiation therapy treatment planning with a simulated annealing algorithm. In 2015.

123. Dias J, Rocha H, Ferreira B, Lopes M do C. Simulated annealing applied to IMRT beam angle optimization: A computational study. Physica Medica. 1 nov 2015;31(7):747-56.

124. Menon SV, Paramu R, Bhasi S, Nair RK. Evaluation of Plan Quality Metrics in Stereotactic Radiosurgery/Radiotherapy in the Treatment Plans of Arteriovenous Malformations. J Med Phys. 2018;43(4):214-20.

125. Feuvret L, Noël G, Mazeron J-J, Bey P. Conformity index: a review. Int J Radiat Oncol Biol Phys. 1 févr 2006;64(2):333-42.

126. Kataria T, Sharma K, Subramani V, Karrthick KP, Bisht SS. Homogeneity Index: An objective tool for assessment of conformal radiation treatments. J Med Phys. 2012;37(4):207-13.

127. Trignani M, Genovesi D, Vinciguerra A, Di Pilla A, Augurio A, Di Tommaso M, et al. Parotid glands in whole-brain radiotherapy: 2D versus 3D technique for no sparing or sparing. Radiol Med. mars 2015;120(3):324-8.

128. Convery D, Rosenbloom M. The generation of intensity-modulated fields for conformal radiotherapy by dynamic collimation. 1992;

129. Bortfeld T. IMRT: a review and preview. Phys Med Biol. 7 juill 2006;51(13):R363-379.

130. Rehman J ur, Zahra, Ahmad N, Khalid M, Noor ul Huda Khan Asghar HM, Gilani ZA, et al. Intensity modulated radiation therapy: A review of current practice and future outlooks. Journal of Radiation Research and Applied Sciences. 1 oct 2018;11(4):361-7.

131. Lafond C, Jouyaux F, Bellec J, Henry O, Perdrieux M, Chajon E, et al. [Which IMRT? From « step and shoot » to VMAT: physicist point of view]. Cancer Radiother. oct 2010;14(6-7):539-49.

132. Iqbal K, Isa M, Buzdar SA, Gifford KA, Afzal Muhammad. Treatment planning evaluation of sliding window and multiple static segments technique in intensity modulated radiotherapy. Rep Pract Oncol Radiother. 21 déc 2012;18(2):101-6.

133. Marchesi V. La radiothérapie conformationnelle avec modulation d’intensité (RCMI) - Algorithmes et technique de réalisation. 2015.

134. Kelly M, Amerongen JHM van, Balvert M, Craft D. Dynamic fluence map sequencing using piecewise linear leaf position functions. Biomed Phys Eng Express. févr 2019;5(2):025036.

135. Medeiros H, Goldbarg EFG, Goldbarg MC. On the three-objective static unconstrained leaf sequencing in IMRT. Med Biol Eng Comput. 1 sept 2020;58(9):2025-37.

136. Lust T, Teghem J. Multiobjective decomposition of integer matrices: application to radiotherapy. 5 juin 2010;

137. Elith C, Dempsey SE, Findlay N, Warren-Forward HM. An Introduction to the Intensity-modulated Radiation Therapy (IMRT) Techniques, Tomotherapy, and VMAT. J Med Imaging Radiat Sci. mars 2011;42(1):37-43.

138. Otto K. Volumetric modulated arc therapy: IMRT in a single gantry arc. Med Phys. janv 2008;35(1):310-7.

139. Scaringi C, Agolli L, Minniti G. Technical Advances in Radiation Therapy for Brain Tumors. Anticancer Res. nov 2018;38(11):6041-5.

140. Wagner D, Christiansen H, Wolff H, Vorwerk H. Radiotherapy of malignant gliomas: comparison of volumetric single arc technique (RapidArc), dynamic intensity-modulated technique and 3D conformal technique. Radiother Oncol. déc 2009;93(3):593-6.

141. Shaffer R, Nichol AM, Vollans E, Fong M, Nakano S, Moiseenko V, et al. A comparison of volumetric modulated arc therapy and conventional intensity-modulated radiotherapy for frontal and temporal high-grade gliomas. Int J Radiat Oncol Biol Phys. 15 mars 2010;76(4):1177-84.

142. Rey EMA, Muñoz AR, Jiménez DN, Pardos RG, Truyols MC. EP-1567: Inverse planning versus forward planning for orbital lymphoma. Radiotherapy and Oncology. 1 mai 2017;123:S844.

143. MacDonald RL, Thomas CG, Syme A. Dynamic collimator trajectory algorithm for multiple metastases dynamic conformal arc treatment planning. Med Phys. janv 2018;45(1):5-17.

144. Bokrantz R, Wedenberg M, Sandwall P. Dynamic conformal arcs for lung stereotactic body radiation therapy: A comparison with volumetric-modulated arc therapy. J Appl Clin Med Phys. janv 2020;21(1):103-9.

145. Morales-Paliza MA, Coffey CW, Ding GX. Evaluation of the dynamic conformal arc therapy in comparison to intensity-modulated radiation therapy in prostate, brain, head-and-neck and spine tumors. J Appl Clin Med Phys. 19 déc 2010;12(2):5-19.

146. Uto M, Mizowaki T, Ogura K, Hiraoka M. Non-coplanar volumetric-modulated arc therapy (VMAT) for craniopharyngiomas reduces radiation doses to the bilateral hippocampus: a planning study comparing dynamic conformal arc therapy, coplanar VMAT, and non-coplanar VMAT. Radiation Oncology. 23 juin 2016;11(1):86.

147. Molinier J, Kerr C, Simeon S, Ailleres N, Charissoux M, Azria D, et al. Comparison of volumetric-modulated arc therapy and dynamic conformal arc treatment planning for cranial stereotactic radiosurgery. J Appl Clin Med Phys. 8 janv 2016;17(1):92-101.

148. Halasz LM, Rockhill JK. Stereotactic radiosurgery and stereotactic radiotherapy for brain metastases. Surg Neurol Int. 2 mai 2013;4(Suppl 4):S185-91.

149. Lupattelli M, Alì E, Ingrosso G, Saldi S, Fulcheri C, Borghesi S, et al. Stereotactic Radiotherapy for Brain Metastases: Imaging Tools and Dosimetric Predictive Factors for Radionecrosis. J Pers Med. 4 juill 2020;10(3):E59.

150. Gu L, Qing S, Zhu X, Ju X, Cao Y, Jia Z, et al. Stereotactic Radiation Therapy (SRT) for Brain Metastases of Multiple Primary Tumors: A Single Institution Retrospective Analysis. Front Oncol. 10 déc 2019;9:1352.

151. Fuss M, Salter BJ, Cheek D, Sadeghi A, Hevezi JM, Herman TS. Repositioning accuracy of a commercially available thermoplastic mask system. Radiother Oncol. juin 2004;71(3):339-45.

152. Arino C, Stadelmaier N, Dupin C, Kantor G, Henriques de Figueiredo B. Le masque de contention en radiothérapie : une source d’anxiété pour le patient ? Cancer/Radiothérapie. 1 déc 2014;18(8):753-6.

153. Goyal S, Kataria T. Image Guidance in Radiation Therapy: Techniques and Applications. Radiol Res Pract. 2014;2014:705604.

154. Sorcini B, Tilikidis A. Clinical application of image-guided radiotherapy, IGRT (on the Varian OBI platform). Cancer/Radiothérapie. 1 sept 2006;10(5):252-7.

155. Muralidhar KR, Murthy PN, Kumar R. Commissioning and quality assurance of the X-ray volume Imaging system of an image-guided radiotherapy capable linear accelerator. J Med Phys. 2008;33(2):72-7.

156. Infusino E, Trodella L, Ramella S, D'Angelillo RM, Greco C, Iurato A, et al. Estimation of patient setup uncertainty using BrainLAB Exatrac X-Ray 6D system in image-guided radiotherapy. Journal of Applied Clinical Medical Physics. 2015;16(2):99-107.

157. Jin J-Y, Yin F-F, Tenn SE, Medin PM, Solberg TD. Use of the BrainLAB ExacTrac X-Ray 6D system in image-guided radiotherapy. Med Dosim. 2008;33(2):124-34.

158. Ma J, Chang Z, Wang Z, Jackie Wu Q, Kirkpatrick JP, Yin F-F. ExacTrac X-ray 6 degree-of-freedom image-guidance for intracranial non-invasive stereotactic radiotherapy: comparison with kilo-voltage cone-beam CT. Radiother Oncol. déc 2009;93(3):602-8.

159. Bissonnette J-P, Balter PA, Dong L, Langen KM, Lovelock DM, Miften M, et al. Quality assurance for image-guided radiation therapy utilizing CT-based technologies: A report of the AAPM TG-179. Medical Physics. 2012;39(4):1946-63.

160. Stützel J, Oelfke U, Nill S. A quantitative image quality comparison of four different image guided radiotherapy devices. Radiother Oncol. janv 2008;86(1):20-4.

161. Bidaut LM, Humm JL, Mageras GS, Rothenberg LN. Imaging in Radiation Oncology. W.B. Saunders; 2010.

162. Wong JW. Electronic Portal Imaging Devices (EPID). In: Brady LW, Yaeger TE, éditeurs. Encyclopedia of Radiation Oncology. Berlin, Heidelberg: Springer; 2013. p. 207-13.

163. Liu J, Xu Y, Teymurazyan A, Papandreou Z, Pang G. Development of a novel high quantum efficiency MV x-ray detector for image-guided radiotherapy: A feasibility study. Medical Physics. 2020;47(1):152-63.

164. Langmack KA. Portal imaging. BJR. 1 sept 2001;74(885):789-804.

165. van Herk M, Remeijer P, Rasch C, Lebesque JV. The probability of correct target dosage: dose-population histograms for deriving treatment margins in radiotherapy. Int J Radiat Oncol Biol Phys. 1 juill 2000;47(4):1121-35.

166. Craig T, Battista J, Moiseenko V, Van Dyk J. Considerations for the implementation of target volume protocols in radiation therapy. Int J Radiat Oncol Biol Phys. 1 janv 2001;49(1):241-50.

167. Leong J. Implementation of random positioning error in computerised radiation treatment planning systems as a result of fractionation. Phys Med Biol. mars 1987;32(3):327-34.

168. van Herk M, Witte M, van der Geer J, Schneider C, Lebesque JV. Biologic and physical fractionation effects of random geometric errors. Int J Radiat Oncol Biol Phys. 1 déc 2003;57(5):1460-71.

169. van Herk M. Errors and margins in radiotherapy. Semin Radiat Oncol. janv 2004;14(1):52-64.

170. Abubakar A, Zamri NAM, Shaukat SI, Zin HM. Automated algorithm for calculation of setup corrections and planning target volume margins for offline image-guided radiotherapy protocols. Journal of Applied Clinical Medical Physics. 2021;22(7):137-46.

171. Vinod SK, Jameson MG, Min M, Holloway LC. Uncertainties in volume delineation in radiation oncology: A systematic review and recommendations for future studies. Radiother Oncol. 2016;121(2):169-79.

172. Segedin B, Petric P. Uncertainties in target volume delineation in radiotherapy – are they relevant and what can we do about them? Radiol Oncol. 9 mai 2016;50(3):254-62.

173. Weltens C, Menten J, Feron M, Bellon E, Demaerel P, Maes F, et al. Interobserver variations in gross tumor volume delineation of brain tumors on computed tomography and impact of magnetic resonance imaging. Radiother Oncol. juill 2001;60(1):49-59.

174. Coles CE, Hoole ACF, Harden SV, Burnet NG, Twyman N, Taylor RE, et al. Quantitative assessment of inter-clinician variability of target volume delineation for

medulloblastoma: quality assurance for the SIOP PNET 4 trial protocol. Radiother Oncol. nov 2003;69(2):189-94.

175. Stanley J, Dunscombe P, Lau H, Burns P, Lim G, Liu H-W, et al. The effect of contouring variability on dosimetric parameters for brain metastases treated with stereotactic radiosurgery. Int J Radiat Oncol Biol Phys. 1 déc 2013;87(5):924-31.

176. Cattaneo GM, Reni M, Rizzo G, Castellone P, Ceresoli GL, Cozzarini C, et al. Target delineation in post-operative radiotherapy of brain gliomas: interobserver variability and impact of image registration of MR(pre-operative) images on treatment planning CT scans. Radiother Oncol. mai 2005;75(2):217-23.

177. Sandström H, Jokura H, Chung C, Toma-Dasu I. Multi-institutional study of the variability in target delineation for six targets commonly treated with radiosurgery. Acta Oncol. nov 2018;57(11):1515-20.

178. International Atomic Energy Agency. Accuracy requirements and uncertainties in radiation therapy. Vienna; 2016.

179. Kanakavelu N, Jebaseelan Samuel J. Determination of patient set-up error and optimal treatment margin for intensity modulated radiotherapy using image guidance system. J BUON. avr 2016;21(2):505-11.

180. Tryggestad E, Christian M, Ford E, Kut C, Le Y, Sanguineti G, et al. Inter- and intrafraction patient positioning uncertainties for intracranial radiotherapy: a study of four frameless, thermoplastic mask-based immobilization strategies using daily cone-beam CT. Int J Radiat Oncol Biol Phys. 1 mai 2011;80(1):281-90.

181. Mijnheer BJ, Battermann JJ, Wambersie A. What degree of accuracy is required and can be achieved in photon and neutron therapy? Radiotherapy and Oncology. 1 mars 1987;8(3):237-52.

182. Castro P, García-Vicente F, Mínguez C, Floriano A, Sevillano D, Pérez L, et al. Study of the uncertainty in the determination of the absorbed dose to water during external beam radiotherapy calibration. J Appl Clin Med Phys. 22 janv 2008;9(1):70-86.

183. Whelan B, Oborn B, Liney G, Keall P. MRI Linac Systems. In 2019. p. 155-68.

184. Raaymakers BW, Jürgenliemk-Schulz IM, Bol GH, Glitzner M, Kotte ANTJ, van Asselen B, et al. First patients treated with a 1.5 T MRI-Linac: clinical proof of concept of a high-precision, high-field MRI guided radiotherapy treatment. Phys Med Biol. 14 nov 2017;62(23):L41-50.

185. Otazo R, Lambin P, Pignol J-P, Ladd ME, Schlemmer H-P, Baumann M, et al. MRI-guided Radiation Therapy: An Emerging Paradigm in Adaptive Radiation Oncology. Radiology. févr 2021;298(2):248-60.

186. Dewhirst MW. A potential solution for eliminating hypoxia as a cause for radioresistance. PNAS. 16 oct 2018;115(42):10548-50.

187. Afaq A, Andreou A, Koh DM. Diffusion-weighted magnetic resonance imaging for tumour response assessment: why, when and how? Cancer Imaging. 29 sept 2010;10(1A):S179-88.

188. Borman PTS, Tijssen RHN, Bos C, Moonen CTW, Raaymakers BW, Glitzner M. Characterization of imaging latency for real-time MRI-guided radiotherapy. Phys Med Biol. août 2018;63(15):155023.

189. Wu Q, Adamson J, Wu QJ. Chapter 5 - Treatment Plan Optimization For Adaptive Radiation Therapy. In: Adaptive Radiation Therapy. X. Allen Li. 2011.

190. Green OL, Henke LE, Hugo GD. Practical Clinical Workflows for Online and Offline Adaptive Radiation Therapy. Semin Radiat Oncol. juill 2019;29(3):219-27.

191. Winkel D, Bol GH, Kroon PS, van Asselen B, Hackett SS, Werensteijn-Honingh AM, et al. Adaptive radiotherapy: The Elekta Unity MR-linac concept. Clin Transl Radiat Oncol. sept 2019;18:54-9.

192. Keall PJ, Nguyen DT, O’Brien R, Caillet V, Hewson E, Poulsen PR, et al. The first clinical implementation of real-time image-guided adaptive radiotherapy using a standard linear accelerator. Radiother Oncol. avr 2018;127(1):6-11.

193. Lagendijk JJW, Raaymakers BW, Raaijmakers AJE, Overweg J, Brown KJ, Kerkhof EM, et al. MRI/linac integration. Radiother Oncol. janv 2008;86(1):25-9.

194. Raaijmakers AJE, Raaymakers BW, Lagendijk JJW. Integrating a MRI scanner with a 6 MV radiotherapy accelerator: dose increase at tissue-air interfaces in a lateral magnetic field due to returning electrons. Phys Med Biol. 7 avr 2005;50(7):1363-76.

195. Han E, Wen Z, Lee H, Paulino A, Lee C. Measurement of Electron Return Effect and Skin Dose Reduction by a Bolus in an Anthropomorphic Physical Phantom under a Magnetic Resonance Guided Linear Accelerator (MR-LINAC) System. International Journal of Medical Physics, Clinical Engineering and Radiation Oncology. 1 janv 2018;07:339-46.

196. Chen X, Paulson ES, Ahunbay E, Sanli A, Klawikowski S, Li XA. Measurement validation of treatment planning for a MR-Linac. Journal of Applied Clinical Medical Physics. 2019;20(7):28-38.

197. Shortall J, Osorio EV, Aitkenhead A, Berresford J, Agnew J, Budgell G, et al. Experimental verification the electron return effect around spherical air cavities for the MR-Linac using Monte Carlo calculation. Medical Physics. 2020;47(6):2506-15.

198. Liney GP, Dong B, Weber E, Rai R, Destruel A, Garcia-Alvarez R, et al. Imaging performance of a dedicated radiation transparent RF coil on a 1.0 Tesla inline MRI-linac. Phys Med Biol. 25 juin 2018;63(13):135005.

199. Thorwarth D, Low DA. Technical Challenges of Real-Time Adaptive MR-Guided Radiotherapy. Front Oncol. 2021;11:634507.

200. Smit K, van Asselen B, Kok JGM, Aalbers AHL, Lagendijk JJW, Raaymakers BW. Towards reference dosimetry for the MR-linac: magnetic field correction of the ionization chamber reading. Phys Med Biol. 7 sept 2013;58(17):5945-57.

201. O’Brien DJ, Roberts DA, Ibbott GS, Sawakuchi GO. Reference dosimetry in magnetic fields: formalism and ionization chamber correction factors. Medical Physics. 2016;43(8Part1):4915-27.

202. TURING AM. I.—COMPUTING MACHINERY AND INTELLIGENCE. Mind. 1 oct 1950;LIX(236):433-60.

203. Ertel W. Introduction to Artificial Intelligence. 2^{e} éd. Springer International Publishing; 2017. (Undergraduate Topics in Computer Science).

204. Chowdhury GG. Natural language processing. Annual Review of Information Science and Technology. 2003;37(1):51-89.

205. Hill J, Ford WR, Farreras I. Real conversations with artificial intelligence: A comparison between human-human online conversations and human-chatbot conversations. Comput Hum Behav. 2015;

206. de Salis E, Capallera M, Meteier Q, Angelini L, Abou Khaled O, Mugellini E, et al. Designing an AI-Companion to Support the Driver in Highly Autonomous Cars. In: Kurosu M, éditeur. Human-Computer Interaction Multimodal and Natural Interaction. Cham: Springer International Publishing; 2020. p. 335-49. (Lecture Notes in Computer Science).

207. Ni J, Chen Y, Chen Y, Zhu J, Ali D, Cao W. A Survey on Theories and Applications for Self-Driving Cars Based on Deep Learning Methods. Applied Sciences. janv 2020;10(8):2749.

208. Mitchell TM. Machine Learning. McGraw-Hill; 1997. 414 p.

209. Qin T. Machine Learning Basics. In: Qin T, éditeur. Dual Learning. Singapore: Springer; 2020. p. 11-23.

210. Fang X, Yu F, Yang G, Qu Y. Regression Analysis With Differential Privacy Preserving. IEEE Access. 12 sept 2019;PP:1-1.

211. Hu F, Hao Q. Intelligent Sensor Networks: The Integration of Sensor Networks, Signal Processing and Machine Learning. 1er édition. CRC Press; 2012. 674 p.

212. Battistella E. High Dimensional Graph Theory Approaches for Various Omics Data. Université Paris-Saclay; 2021.

213. Battistella E, Vakalopoulou M, Estienne T, Lerousseau M, Sun R, Robert C, et al. Gene Expression High-Dimensional Clustering towards a Novel, Robust, Clinically Relevant and Highly Compact Cancer Signature. In: IWBBIO 2019 - 7th International Work-Conference on Bioinformatics and Biomedical Engineering. Granada, Spain; 2019.

214. Chapelle O, Schölkopf B, Zien A. Semi-Supervised Learning. London, U.K.: MIT Press. 2006.

215. Zhou X, Zhu F, Liu Q, Fu Y, Huang W. A Sarsa(λ)-Based Control Model for Real-Time Traffic Light Coordination. TheScientificWorldJournal. 23 janv 2014;2014:759097.

216. Arel I, Liu C, Urbanik T, Kohls AG. Reinforcement learning-based multi-agent system for network traffic signal control. IET Intelligent Transport Systems. 1 juin 2010;4(2):128-35.

217. Hester T, Quinlan M, Stone P. Generalized Model Learning for Reinforcement Learning on a Humanoid Robot. In 2010. p. 2369-74.

218. Zhou Z, Li X, Zare RN. Optimizing Chemical Reactions with Deep Reinforcement Learning. ACS Cent Sci. 27 déc 2017;3(12):1337-44.

219. Hu Y-J, Lin S-J. Deep Reinforcement Learning for Optimizing Finance Portfolio Management. In: 2019 Amity International Conference on Artificial Intelligence (AICAI). 2019. p. 14-20.

220. Rosenblatt F. The perceptron: A probabilistic model for information storage and organization in the brain. Psychological Review. 1958;65(6):386-408.

221. Hornik K, Stinchcombe M, White H. Multilayer feedforward networks are universal approximators. Neural Networks. 1 janv 1989;2(5):359-66.

222. Desai V, Rao D. Image Hash using Neural Networks. International Journal of Computer Applications. 15 févr 2013;22.

223. Zamir A, Wu T-L, Sun L, Shen B (William), Shi B, Malik J, et al. Feedback Networks. 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). 2017;

224. Bengio Y, Louradour J, Collobert R, Weston J. Curriculum learning. In 2009. p. 6.

225. Prechelt L. Automatic early stopping using cross validation: quantifying the criteria. Neural Networks. 1 juin 1998;11(4):761-7.

226. Bengio Y. Practical recommendations for gradient-based training of deep architectures. Arxiv. 1 juin 2012;

227. Gavrishchaka V, Senyukova O, Koepke M. Synergy of physics-based reasoning and machine learning in biomedical applications: Towards unlimited deep learning with limited data. Advances in Physics: X. 1 janv 2019;4:1582361.

228. Jiawei Z. Gradient Descent based Optimization Algorithms for Deep Learning Models Training. 2019.

229. Kingma DP, Ba J. Adam: A Method for Stochastic Optimization. ICLR. 2015;

230. Elguindi S, Zelefsky MJ, Jiang J, Veeraraghavan H, Deasy JO, Hunt MA, et al. Deep learning-based auto-segmentation of targets and organs-at-risk for magnetic resonance imaging only planning of prostate radiotherapy. Physics and Imaging in Radiation Oncology. 1 oct 2019;12:80-6.

231. Barkousaraie AS, Ogunmolu O, Jiang S, Nguyen D. A fast deep learning approach for beam orientation optimization for prostate cancer treated with intensity-modulated radiation therapy. Medical Physics. 2020;47(3):880-97.

232. Kazemifar S, McGuire S, Timmerman R, Wardak Z, Nguyen D, Park Y, et al. MRI-only brain radiotherapy: Assessing the dosimetric accuracy of synthetic CT images generated using a deep learning approach. Radiother Oncol. 2019;136:56-63.

233. Ronneberger O, Fischer P, Brox T. U-Net: Convolutional Networks for Biomedical Image Segmentation. Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015 Lecture Notes in Computer Science. 18 mai 2015;9351:233-41.

234. Ibtehaz N, Rahman MS. MultiResUNet : Rethinking the U-Net Architecture for Multimodal Biomedical Image Segmentation. Neural Networks. janv 2020;121:74-87.

235. Khanh TLB, Dao D-P, Ho N-H, Yang H-J, Baek E-T, Lee G, et al. Enhancing U-Net with Spatial-Channel Attention Gate for Abnormal Tissue Segmentation in Medical Imaging. Applied Sciences. janv 2020;10(17):5729.

236. Zhao Y, Bai L, Lyu Y, Huang X. Camera-Based Blind Spot Detection with a General Purpose Lightweight Neural Network. Electronics. 19 févr 2019;8:233.

237. Li W, Wang G, Fidon L, Ourselin S, Cardoso MJ, Vercauteren T. On the Compactness, Efficiency, and Representation of 3D Convolutional Networks: Brain Parcellation as a

Pretext Task. In: Niethammer M, Styner M, Aylward S, Zhu H, Oguz I, Yap P-T, et al., éditeurs. Information Processing in Medical Imaging. Cham: Springer International Publishing; 2017. p. 348-60. (Lecture Notes in Computer Science).

238. Alvarez Andres E, Fidon L, Vakalopoulou M, Lerousseau M, Carré A, Sun R, et al. Dosimetry-Driven Quality Measure of Brain Pseudo Computed Tomography Generated From Deep Learning for MRI-Only Radiation Therapy Treatment Planning. International Journal of Radiation Oncology, Biology, Physics. 1 nov 2020;108(3):813-23.

239. Salvaris M, Dean D, Tok WH. Generative Adversarial Networks. In: Salvaris M, Dean D, Tok WH, éditeurs. Deep Learning with Azure. Berkeley, CA: Apress; 2018. p. 187-208.

240. Mirza M, Osindero S. Conditional Generative Adversarial Nets. ArXiv. 2014;

241. Fetty L, Löfstedt T, Heilemann G, Furtado H, Nesvacil N, Nyholm T, et al. Investigating conditional GAN performance with different generator architectures, an ensemble model, and different MR scanners for MR-sCT conversion. Phys Med Biol. mai 2020;65(10):105004.

242. Xu K, Cao J, Xia K, Yang H, Zhu J, Wu C, et al. Multichannel Residual Conditional GAN-Leveraged Abdominal Pseudo-CT Generation via Dixon MR Images. IEEE Access. 2019;7:163823-30.

243. Singhrao K, Fu J, Parikh NR, Mikaeilian AG, Ruan D, Kishan AU, et al. A generative adversarial network-based (GAN-based) architecture for automatic fiducial marker detection in prostate MRI-only radiotherapy simulation images. Medical Physics. 2020;47(12):6405-13.

244. Hibbard L. Adversarial Prediction of Radiotherapy Treatment Machine Parameters. In: Syeda-Mahmood T, Drechsler K, Greenspan H, Madabhushi A, Karargyris A, Linguraru MG, et al., éditeurs. Multimodal Learning for Clinical Decision Support and Clinical Image-Based Procedures. Cham: Springer International Publishing; 2020. p. 85-94. (Lecture Notes in Computer Science).

245. Zhu J-Y, Park T, Isola P, Efros AA. Unpaired Image-to-Image Translation using Cycle-Consistent Adversarial Networks. ArXiv. 24 août 2020;

246. Kang SK, An HJ, Jin H, Kim J, Chie EK, Park JM, et al. Synthetic CT generation from weakly paired MR images using cycle-consistent GAN for MR-guided radiotherapy. Biomed Eng Lett. 19 juin 2021;

247. Liu Y, Chen A, Shi H, Huang S, Zheng W, Liu Z, et al. CT synthesis from MRI using multi-cycle GAN for head-and-neck radiation therapy. Computerized Medical Imaging and Graphics. 1 juill 2021;91:101953.

248. Boni KNDB, Klein J, Gulyban A, Reynaert N, Pasquier D. Improving generalization in MR-to-CT synthesis in radiotherapy by using an augmented cycle generative adversarial network with unpaired data. Medical Physics. 2021;48(6):3003-10.

249. Koike Y, Anetai Y, Takegawa H, Ohira S, Nakamura S, Tanigawa N. Deep learning-based metal artifact reduction using cycle-consistent adversarial network for intensity-modulated head and neck radiation therapy treatment planning. Physica Medica. 1 oct 2020;78:8-14.

250. Harms J, Lei Y, Wang T, Zhang R, Zhou J, Tang X, et al. Paired cycle-GAN-based image correction for quantitative cone-beam computed tomography. Medical Physics. 2019;46(9):3998-4009.

251. Nie D, Trullo R, Lian J, Petitjean C, Ruan S, Wang Q, et al. Medical Image Synthesis with Context-Aware Generative Adversarial Networks. Med Image Comput Comput Assist Interv. sept 2017;10435:417-25.

252. Mahmood R, Babier A, Mcniven A, Diamant A, Chan T. Automated Treatment Planning in Radiation Therapy using Generative Adversarial Networks. 2018.

253. Maspero M, Savenije MHF, Dinkla AM, Seevinck PR, Intven MPW, Jurgenliemk-Schulz IM, et al. Dose evaluation of fast synthetic-CT generation using a generative adversarial network for general pelvis MR-only radiotherapy. Phys Med Biol. sept 2018;63(18):185001.

254. Peng Y, Chen S, Qin A, Chen M, Gao X, Liu Y, et al. Magnetic resonance-based synthetic computed tomography images generated using generative adversarial networks for nasopharyngeal carcinoma radiotherapy treatment planning. Radiotherapy and Oncology. 1 sept 2020;150:217-24.

255. Koike Y, Akino Y, Sumida I, Shiomi H, Mizuno H, Yagi M, et al. Feasibility of synthetic computed tomography generated with an adversarial network for multi-sequence magnetic resonance-based brain radiotherapy. J Radiat Res. 23 janv 2020;61(1):92-103.

256. Elmahdy MS, Wolterink JM, Sokooti H, Išgum I, Staring M. Adversarial Optimization for Joint Registration and Segmentation in Prostate CT Radiotherapy. In: Shen D, Liu T, Peters TM, Staib LH, Essert C, Zhou S, et al., éditeurs. Medical Image Computing and Computer Assisted Intervention – MICCAI 2019. Cham: Springer International Publishing; 2019. p. 366-74. (Lecture Notes in Computer Science).

257. Huynh E, Hosny A, Guthier C, Bitterman DS, Petit SF, Haas-Kogan DA, et al. Artificial intelligence in radiation oncology. Nat Rev Clin Oncol. déc 2020;17(12):771-81.

258. Siddique S, Chow JCL. Artificial intelligence in radiotherapy. Rep Pract Oncol Radiother. 2020;25(4):656-66.

259. Chan MF, Witztum A, Valdes G. Integration of AI and Machine Learning in Radiotherapy QA. Frontiers in Artificial Intelligence. 2020;3:76.

260. Han X. MR-based synthetic CT generation using a deep convolutional neural network method. Med Phys. avr 2017;44(4):1408-19.

261. Dinkla AM, Wolterink JM, Maspero M, Savenije MHF, Verhoeff JJC, Seravalli E, et al. MR-Only Brain Radiation Therapy: Dosimetric Evaluation of Synthetic CTs Generated by a Dilated Convolutional Neural Network. Int J Radiat Oncol Biol Phys. 15 2018;102(4):801-12.

262. Li W, Li Y, Qin W, Liang X, Xu J, Xiong J, et al. Magnetic resonance image (MRI) synthesis from brain computed tomography (CT) images based on deep learning methods for magnetic resonance (MR)-guided radiotherapy. Quant Imaging Med Surg. juin 2020;10(6):1223-36.

263. Jin C-B, Kim H, Liu M, Jung W, Joo S, Park E, et al. Deep CT to MR Synthesis Using Paired and Unpaired Data. Sensors (Basel). 2019;19(10).

264. Yu L, Zhang Z, Li X, Xing L. Deep Sinogram Completion With Image Prior for Metal Artifact Reduction in CT Images. IEEE Transactions on Medical Imaging. janv 2021;40(1):228-38.

265. Nakamura M, Nakao M, Imanishi K, Hirashima H, Tsuruta Y. Geometric and dosimetric impact of 3D generative adversarial network-based metal artifact reduction algorithm on VMAT and IMPT for the head and neck region. Radiation Oncology. 6 juin 2021;16(1):96.

266. Wu G, Kim M, Wang Q, Gao Y, Liao S, Shen D. Unsupervised deep feature learning for deformable registration of MR brain images. Med Image Comput Comput Assist Interv. 2013;16(Pt 2):649-56.

267. Fan J, Cao X, Yap P-T, Shen D. BIRNet: Brain image registration using dual-supervised fully convolutional networks. Medical Image Analysis. 1 mai 2019;54:193-206.

268. Simonovsky M, Gutiérrez-Becker B, Mateus D, Navab N, Komodakis N. A Deep Metric for Multimodal Registration. International conference on medical image computing and computer-assisted intervention. 17 sept 2016;10-8.

269. Hermessi H, Mourali O, Zagrouba E. Convolutional neural network-based multimodal image fusion via similarity learning in the shearlet domain. Neural Comput & Applic. 1 oct 2018;30(7):2029-45.

270. Islam KT, Wijewickrema S, O’Leary S. A deep learning based framework for the registration of three dimensional multi-modal medical images of the head. Scientific Reports. 21 janv 2021;11(1):1860.

271. Cardenas CE, Yang J, Anderson BM, Court LE, Brock KB. Advances in Auto-Segmentation. Semin Radiat Oncol. juill 2019;29(3):185-97.

272. Xue J, Wang B, Ming Y, Liu X, Jiang Z, Wang C, et al. Deep learning-based detection and segmentation-assisted management of brain metastases. Neuro Oncol. 15 avr 2020;22(4):505-14.

273. Mlynarski P, Delingette H, Alghamdi H, Bondiau P-Y, Ayache N. Anatomically Consistent Segmentation of Organs at Risk in MRI with Convolutional Neural Networks. arXiv:190702003 [cs, eess]. 3 juill 2019;

274. Dolz J, Reyns N, Betrouni N, Kharroubi D, Quidet M, Massoptier L, et al. A deep learning classification scheme based on augmented-enhanced features to segment organs at risk on the optic region in brain cancer patients. arXiv:170310480 [cs]. 5 avr 2017;

275. Estienne T, Lerousseau M, Vakalopoulou M, Alvarez Andres E, Battistella E, Carré A, et al. Deep Learning-Based Concurrent Brain Registration and Tumor Segmentation. Front Comput Neurosci. 2020;14.

276. Chen X, Men K, Li Y, Yi J, Dai J. A feasibility study on an automated method to generate patient-specific dose distributions for radiotherapy using deep learning. Med Phys. janv 2019;46(1):56-64.

277. Kontaxis C, Bol GH, Lagendijk JJW, Raaymakers BW. DeepDose: Towards a fast dose calculation engine for radiation therapy using deep learning. Phys Med Biol. 8 avr 2020;65(7):075013.

278. McIntosh C, Conroy L, Tjong MC, Craig T, Bayley A, Catton C, et al. Clinical integration of machine learning for curative-intent radiation treatment of patients with prostate cancer. Nat Med. juin 2021;27(6):999-1005.

279. van der Merwe D, Van Dyk J, Healy B, Zubizarreta E, Izewska J, Mijnheer B, et al. Accuracy requirements and uncertainties in radiotherapy: a report of the International Atomic Energy Agency. Acta Oncologica. 2 janv 2017;56(1):1-6.

280. Tomori S, Kadoya N, Takayama Y, Kajikawa T, Shima K, Narazaki K, et al. A deep learning-based prediction model for gamma evaluation in patient-specific quality assurance. Med Phys. 31 juill 2018;

281. Kida S, Nakamoto T, Nakano M, Nawa K, Haga A, Kotoku J, et al. Cone Beam Computed Tomography Image Quality Improvement Using a Deep Convolutional Neural Network. Cureus. 29 avr 2018;10(4):e2548.

282. Zhang Y, Yue N, Su M-Y, Liu B, Ding Y, Zhou Y, et al. Improving CBCT quality to CT level using deep learning with generative adversarial network. Med Phys. 1 déc 2020;

283. Hoogcarspel SJ, Van der Velden JM, Lagendijk JJW, van Vulpen M, Raaymakers BW. The feasibility of utilizing pseudo CT-data for online MRI based treatment plan adaptation for a stereotactic radiotherapy treatment of spinal bone metastases. Phys Med Biol. 7 déc 2014;59(23):7383-91.

284. Liu F, Jang H, Kijowski R, Bradshaw T, McMillan AB. Deep Learning MR Imaging-based Attenuation Correction for PET/MR Imaging. Radiology. 2018;286(2):676-84.

285. Zijlstra F, Willemsen K, Florkow MC, Sakkers RJB, Weinans HH, van der Wal BCH, et al. CT synthesis from MR images for orthopedic applications in the lower arm using a conditional generative adversarial network. arXiv:190108449 [cs, eess]. 24 janv 2019;

286. Largent A, Barateau A, Nunes J-C, Lafond C, Greer PB, Dowling JA, et al. Pseudo-CT Generation for MRI-Only Radiation Therapy Treatment Planning: Comparison Among Patch-Based, Atlas-Based, and Bulk Density Methods. International Journal of Radiation Oncology*Biology*Physics. 1 févr 2019;103(2):479-90.

287. Neppl S, Landry G, Kurz C, Hansen DC, Hoyle B, Stöcklein S, et al. Evaluation of proton and photon dose distributions recalculated on 2D and 3D Unet-generated pseudoCTs from T1-weighted MR head scans. Acta Oncol. oct 2019;58(10):1429-34.

288. Spadea MF, Pileggi G, Zaffino P, Salome P, Catana C, Izquierdo-Garcia D, et al. Deep Convolution Neural Network (DCNN) Multiplane Approach to Synthetic CT Generation From MR images-Application in Brain Proton Therapy. Int J Radiat Oncol Biol Phys. 1 nov 2019;105(3):495-503.

289. Lerner M, Medin J, Jamtheim Gustafsson C, Alkner S, Siversson C, Olsson LE. Clinical validation of a commercially available deep learning software for synthetic CT generation for brain. Radiat Oncol. 7 avr 2021;16(1):66.

290. Sara U, Akter M, Uddin MS. Image Quality Assessment through FSIM, SSIM, MSE and PSNR—A Comparative Study. Journal of Computer and Communications. 4 mars 2019;7(3):8-18.

291. Sun H, Fan R, Li C, Lu Z, Xie K, Ni X, et al. Imaging Study of Pseudo-CT Synthesized From Cone-Beam CT Based on 3D CycleGAN in Radiotherapy. Frontiers in Oncology. 2021;11:436.

292. Dhungel A, Weeks M. Performance measurement for a wavelet transform-based video compression. In: Proceedings of the 49th Annual Southeast Regional Conference. New York, NY, USA: Association for Computing Machinery; 2011. p. 216-20. (ACM-SE ’11).

293. Malpica W, Bovik A. SSIM based range image quality assessment. 4th Int Workshop on Video Processing and Quality Metrics for Consumer Electronics. 2009;

294. Pambrun J-F, Noumeir R. Limitations of the SSIM quality metric in the context of diagnostic imaging. In 2015.

295. Maspero M, Bentvelzen LG, Savenije MHF, Guerreiro F, Seravalli E, Janssens GO, et al. Deep learning-based synthetic CT generation for paediatric brain MR-only photon and proton radiotherapy. Radiotherapy and Oncology. 1 déc 2020;153:197-204.

296. Liu F, Yadav P, Baschnagel AM, McMillan AB. MR-based treatment planning in radiation therapy using a deep learning approach. J Appl Clin Med Phys. mars 2019;20(3):105-14.

297. Massa HA, Johnson JM, McMillan AB. Comparison of deep learning synthesis of synthetic CTs using clinical MRI inputs. Phys Med Biol. 23 déc 2020;65(23):23NT03.

298. Liu X, Emami H, Nejad-Davarani SP, Morris E, Schultz L, Dong M, et al. Performance of deep learning synthetic CTs for MR-only brain radiation therapy. Journal of Applied Clinical Medical Physics. 2021;22(1):308-17.

299. Tang B, Wu F, Fu Y, Wang X, Wang P, Orlandini LC, et al. Dosimetric evaluation of synthetic CT image generated using a neural network for MR-only brain radiotherapy. J Appl Clin Med Phys. mars 2021;22(3):55-62.

300. Prabhakar R, Julka PK, Ganesh T, Munshi A, Joshi RC, Rath GK. Feasibility of using MRI alone for 3D radiation treatment planning in brain tumors. Jpn J Clin Oncol. juin 2007;37(6):405-11.

301. Weber DC, Wang H, Albrecht S, Ozsahin M, Tkachuk E, Rouzaud M, et al. Open low-field magnetic resonance imaging for target definition, dose calculations and set-up verification during three-dimensional CRT for glioblastoma multiforme. Clin Oncol (R Coll Radiol). mars 2008;20(2):157-67.

302. Hussein M, Clark CH, Nisbet A. Challenges in calculation of the gamma index in radiotherapy – Towards good practice. Physica Medica. 1 avr 2017;36:1-11.

303. Prabhakar R, Julka PK, Ganesh T, Munshi A, Joshi RC, Rath GK. Feasibility of using MRI alone for 3D radiation treatment planning in brain tumors. Jpn J Clin Oncol. juin 2007;37(6):405-11.

304. Kristensen B, Laursen F, Løgager V, Geertsen P, Krarup-Hansen A. Dosimetric and geometric evaluation of an open low-field magnetic resonance simulator for radiotherapy treatment planning of brain tumours. Radiotherapy and oncology : journal of the European Society for Therapeutic Radiology and Oncology. 1 avr 2008;87:100-9.

305. Stanescu T, Jans H-S, Stavrev P, Fallone B. 3T MR-based treatment planning for radiotherapy of brain lesions. Radiology and Oncology. 1 janv 2006;40:125-32.

306. Stanescu T, Jans H-S, Pervez N, Stavrev P, Fallone BG. A study on the magnetic resonance imaging (MRI)-based radiation treatment planning of intracranial lesions. Phys Med Biol. 7 juill 2008;53(13):3579-93.

307. Fox JC, éditeur. Clinical Emergency Radiology. 2 edition. Cambridge University Press; 2017. 650 p.

308. Robson MD, Bydder GM. Clinical ultrashort echo time imaging of bone and other connective tissues. NMR Biomed. nov 2006;19(7):765-80.

309. Breighner RE, Endo Y, Konin GP, Gulotta LV, Koff MF, Potter HG. Technical Developments: Zero Echo Time Imaging of the Shoulder: Enhanced Osseous Detail by Using MR Imaging. Radiology. 1 mars 2018;286(3):960-6.

310. Bratova I, Paluska P, Grepl J, Sykorova P, Jansa J, Hodek M, et al. Validation of dose distribution computation on sCT images generated from MRI scans by Philips MRCAT. Rep Pract Oncol Radiother. avr 2019;24(2):245-50.

311. Yu V, Keyrilainen J, Suilamo S, Beslimane I, Dresner A, Halkola A, et al. A multi-institutional analysis of a general pelvis continuous Hounsfield unit synthetic CT software for radiotherapy. Journal of Applied Clinical Medical Physics. 1 févr 2021;22.

312. Snyder JE, St-Aubin J, Yaddanapudi S, Boczkowski A, Dunkerley DAP, Graves SA, et al. Commissioning of a 1.5T Elekta Unity MR-linac: A single institution experience. Journal of Applied Clinical Medical Physics. 2020;21(7):160-72.

313. Siemens Healthcare GmbH. MR-only RT planning for brain and pelvis with Synthetic CT · White paper. 2019;

314. Boukellouz W, Moussaoui A, Taleb-Ahmed A. Pseudo-CT Generated Through Multi-Metric Image Registration and Atlas Fusion (Application to T1-Weighted Brain MRI). In: 2017 European Conference on Electrical Engineering and Computer Science (EECS). 2017. p. 230-5.

315. Chegeni N, Birgani MJT, Birgani FF, Fatehi D, Akbarizadeh G, Tahmasbi M. Introduction of a Simple Algorithm to Create Synthetic-computed Tomography of the Head from Magnetic Resonance Imaging. J Med Signals Sens. 2019;9(2):123-9.

316. Uh J, Merchant TE, Li Y, Li X, Hua C. MRI-based treatment planning with pseudo CT generated through atlas registration. Med Phys. mai 2014;41(5):051711.

317. Andreasen D, Leemput KV, Hansen RH, Andersen JAL, Edmund JM. Patch-based generation of a pseudo CT from conventional MRI sequences for MRI-only radiotherapy of the brain. Medical Physics. 2015;42(4):1596-605.

318. Demol B, Boydev C, Korhonen J, Reynaert N. Dosimetric characterization of MRI-only treatment planning for brain tumors in atlas-based pseudo-CT images generated from standard T1-weighted MR images. Med Phys. déc 2016;43(12):6557.

319. Sjölund J, Forsberg D, Andersson M, Knutsson H. Generating patient specific pseudo-CT of the head from MR using atlas-based regression. Phys Med Biol. 21 janv 2015;60(2):825-39.

320. Boukellouz W, Moussaoui A, Taleb-Ahmed A, Boydev C. Multiatlas Fusion with a Hybrid CT Number Correction Technique for Subject-Specific Pseudo-CT Estimation in

the Context of MRI-Only Radiation Therapy. Journal of Medical Imaging and Radiation Sciences. 1 sept 2019;50(3):425-40.

321. Li W, Kazemifar S, Bai T, Nguyen D, Weng Y, Li Y, et al. Synthesizing CT images from MR images with deep learning: model generalization for different datasets through transfer learning. Biomed Phys Eng Express. 5 févr 2021;

322. Cronholm RO, Karlsson A, Siversson C. Whitepaper: MRI only radiotherapy planning using the transfer function estimation algorithm. 2020;

323. Philips. Unleash the real power of MR simulation - MRCAT Brain. 2020.

324. Sled JG, Zijdenbos AP, Evans AC. A nonparametric method for automatic correction of intensity nonuniformity in MRI data. IEEE Trans Med Imaging. févr 1998;17(1):87-97.

325. Cox IJ, Roy S, Hingorani SL. Dynamic histogram warping of image pairs for constant image brightness. Proceedings, International Conference on Image Processing. 1995;

326. Wolterink JM, Dinkla AM, Savenije MHF, Seevinck PR, van den Berg CAT, Išgum I. Deep MR to CT Synthesis Using Unpaired Data. In: Tsaftaris SA, Gooya A, Frangi AF, Prince JL, éditeurs. Simulation and Synthesis in Medical Imaging. Cham: Springer International Publishing; 2017. p. 14-23. (Lecture Notes in Computer Science).

327. Kläser K, Markiewicz P, Ranzini M, Li W, Modat M, Hutton BF, et al. Deep Boosted Regression for MR to CT Synthesis. In: Gooya A, Goksel O, Oguz I, Burgos N, éditeurs. Simulation and Synthesis in Medical Imaging. Cham: Springer International Publishing; 2018. p. 61-70. (Lecture Notes in Computer Science).

328. Emami H, Dong M, Nejad-Davarani SP, Glide-Hurst CK. Generating synthetic CTs from magnetic resonance images using generative adversarial networks. Medical Physics. 2018;45(8):3627-36.

329. Kläser K, Varsavsky T, Markiewicz P, Vercauteren T, Atkinson D, Thielemans K, et al. Improved MR to CT Synthesis for PET/MR Attenuation Correction Using Imitation Learning. In: Burgos N, Gooya A, Svoboda D, éditeurs. Simulation and Synthesis in Medical Imaging. Cham: Springer International Publishing; 2019. p. 13-21. (Lecture Notes in Computer Science).

330. Lei Y, Harms J, Wang T, Liu Y, Shu H-K, Jani AB, et al. MRI-only based synthetic CT generation using dense cycle consistent generative adversarial networks. Med Phys. août 2019;46(8):3565-81.

331. Schmidt MA, Payne GS. Radiotherapy Planning using MRI. Phys Med Biol. 21 nov 2015;60(22):R323-61.

332. Johnstone E, Wyatt JJ, Henry AM, Short SC, Sebag-Montefiore D, Murray L, et al. Systematic Review of Synthetic Computed Tomography Generation Methodologies for Use in Magnetic Resonance Imaging–Only Radiation Therapy. International Journal of Radiation Oncology • Biology • Physics. 1 janv 2018;100(1):199-217.

333. Kang KM, Choi HS, Jeong BK, Song JH, Ha I-B, Lee YH, et al. MRI-based radiotherapy planning method using rigid image registration technique combined with outer body correction scheme: a feasibility study. Oncotarget. 15 août 2017;8(33):54497-505.

334. Wang C, Chao M, Lee L, Xing L. MRI-based treatment planning with electron density information mapped from CT images: a preliminary study. Technol Cancer Res Treat. oct 2008;7(5):341-8.

335. Isensee F, Schell M, Pflueger I, Brugnara G, Bonekamp D, Neuberger U, et al. Automated Brain Extraction of Multisequence MRI Using Artificial Neural Networks. Human Brain Mapping. 2019;40(17):4952-64.

336. Çiçek Ö, Abdulkadir A, Lienkamp SS, Brox T, Ronneberger O. 3D U-Net: Learning Dense Volumetric Segmentation from Sparse Annotation. In: Ourselin S, Joskowicz L, Sabuncu MR, Unal G, Wells W, éditeurs. Medical Image Computing and Computer-Assisted Intervention – MICCAI 2016. Cham: Springer International Publishing; 2016. p. 424-32. (Lecture Notes in Computer Science).

337. Tustison NJ, Avants BB, Cook PA, Zheng Y, Egan A, Yushkevich PA, et al. N4ITK: Improved N3 Bias Correction - IEEE Journals & Magazine. IEEE Transactions on Medical Imaging. juin 2010;29(6):1310-20.

338. Larsen CT, Iglesias JE, Van Leemput K. N3 Bias Field Correction Explained as a Bayesian Modeling Method. In: Cardoso MJ, Simpson I, Arbel T, Precup D, Ribbens A,

éditeurs. Bayesian and grAphical Models for Biomedical Imaging. Cham: Springer International Publishing; 2014. p. 1-12. (Lecture Notes in Computer Science).

339. Otsu N. A Threshold Selection Method from Gray-Level Histograms. IEEE Transactions on Systems, Man, and Cybernetics. janv 1979;9(1):62-6.

340. Mohan R, Chui C, Lidofsky L. Differential pencil beam dose computation model for photons. Med Phys. févr 1986;13(1):64-73.

341. Chu C, Zhmoginov A, Sandler M. CycleGAN, a Master of Steganography. ArXiv. 2017;abs/1712.02950.

342. Fragoso M, Wen N, Kumar S, Liu D, Ryu S, Movsas B, et al. Dosimetric verification and clinical evaluation of a new commercially available Monte Carlo-based dose algorithm for application in stereotactic body radiation therapy (SBRT) treatment planning. Phys Med Biol. 21 août 2010;55(16):4445-64.

343. Petoukhova AL, van Wingerden K, Wiggenraad RGJ, van de Vaart PJM, van Egmond J, Franken EM, et al. Verification measurements and clinical evaluation of the iPlan RT Monte Carlo dose algorithm for 6 MV photon energy. Phys Med Biol. 21 août 2010;55(16):4601-14.

344. Combs SE, Baumert BG, Bendszus M, Bozzao A, Brada M, Fariselli L, et al. ESTRO ACROP guideline for target volume delineation of skull base tumors. Radiotherapy and Oncology. 1 mars 2021;156:80-94.

345. Weygand J, Fuller CD, Ibbott GS, Mohamed ASR, Ding Y, Yang J, et al. Spatial Precision in Magnetic Resonance Imaging-Guided Radiation Therapy: The Role of Geometric Distortion. Int J Radiat Oncol Biol Phys. 15 2016;95(4):1304-16.

346. Xing A, Holloway L, Arumugam S, Walker AR, Rai R, Juresic E, et al. Commissioning and quality control of a dedicated wide bore 3T MRI simulator for radiotherapy planning. In 2016.

347. Sun J, Barnes M, Dowling J, Menk F, Stanwell P, Greer PB. An open source automatic quality assurance (OSAQA) tool for the ACR MRI phantom. Australas Phys Eng Sci Med. mars 2015;38(1):39-46.

348. Jenkinson M, Smith S. A global optimisation method for robust affine registration of brain images. Med Image Anal. juin 2001;5(2):143-56.

349. Jenkinson M, Bannister P, Brady M, Smith S. Improved optimization for the robust and accurate linear registration and motion correction of brain images. Neuroimage. oct 2002;17(2):825-41.

350. Lauritzen AD, Papademetris X, Turovets S, Onofrey JA. Evaluation of CT Image Synthesis Methods:From Atlas-based Registration to Deep Learning. 2019;

351. Yang H, Sun J, Carass A, Zhao C, Lee J, Prince JL, et al. Unsupervised MR-to-CT Synthesis Using Structure-Constrained CycleGAN. IEEE Trans Med Imaging. déc 2020;39(12):4249-61.

352. Yang T, Song J, Li L. A deep learning model integrating SK-TPCNN and random forests for brain tumor segmentation in MRI. Biocybernetics and Biomedical Engineering. 1 juill 2019;39(3):613-23.

353. Pereira S, Pinto A, Alves V, Silva CA. Brain Tumor Segmentation Using Convolutional Neural Networks in MRI Images. IEEE Transactions on Medical Imaging. mai 2016;35(5):1240-51.

354. Peng S, Chen W, Sun J, Liu B. Multi-Scale 3D U-Nets: An approach to automatic segmentation of brain tumor. International Journal of Imaging Systems and Technology. 2020;30(1):5-17.

355. Lee B, Yamanakkanavar N, Choi JY. Automatic segmentation of brain MRI using a novel patch-wise U-net deep architecture. PLOS ONE. août 2020;15(8):e0236493.

356. Keall PJ, Siebers JV, Jeraj R, Mohan R. The effect of dose calculation uncertainty on the evaluation of radiotherapy plans. Medical Physics. 2000;27(3):478-84.

357. Tustison NJ, Avants BB, Cook PA, Zheng Y, Egan A, Yushkevich PA, et al. N4ITK: improved N3 bias correction. IEEE Trans Med Imaging. juin 2010;29(6):1310-20.

358. Isensee F, Schell M, Pflueger I, Brugnara G, Bonekamp D, Neuberger U, et al. Automated brain extraction of multisequence MRI using artificial neural networks. Hum Brain Mapp. 01 2019;40(17):4952-64.

359. Shinohara RT, Sweeney EM, Goldsmith J, Shiee N, Mateen FJ, Calabresi PA, et al. Statistical normalization techniques for magnetic resonance imaging. Neuroimage Clin. 2014;6:9-19.

360. Otsu N. A Threshold Selection Method from Gray-Level Histograms. IEEE Transactions on Systems, Man, and Cybernetics. janv 1979;9(1):62-6.

361. Scoccianti S, Detti B, Gadda D, Greto D, Furfaro I, Meacci F, et al. Organs at risk in the brain and their dose-constraints in adults and in children: a radiation oncologist's guide for delineation in everyday practice. Radiother Oncol. févr 2015;114(2):230-8.

362. Schober P, Boer C, Schwarte LA. Correlation Coefficients: Appropriate Use and Interpretation. Anesth Analg. mai 2018;126(5):1763-8.

363. Lerner M, Medin J, Jamtheim Gustafsson C, Alkner S, Siversson C, Olsson LE. Clinical validation of a commercially available deep learning software for synthetic CT generation for brain. Radiation Oncology. 7 avr 2021;16(1):66.

364. Papanikolaou N, Battista J, Boyer A, Kappas C, Klein E, Mackie T, et al. Tissue inhomogeneity corrections for megavoltage photon beams. AAPM Report No 85; Task Group No. 85; Task Group No. 65. 2004.

365. Karavitaki N, Cudlip S, Adams CBT, Wass JAH. Craniopharyngiomas. Endocr Rev. juin 2006;27(4):371-97.

366. Garrè ML, Cama A. Craniopharyngioma: modern concepts in pathogenesis and treatment. Current Opinion in Pediatrics. août 2007;19(4):471-9.

367. Kalapurakal JA. Radiation therapy in the management of pediatric craniopharyngiomas--a review. Childs Nerv Syst. août 2005;21(8-9):808-16.

368. Sartoretti-Schefer S, Wichmann W, Aguzzi A, Valavanis A. MR differentiation of adamantinous and squamous-papillary craniopharyngiomas. American Journal of Neuroradiology. 1 janv 1997;18(1):77-87.

369. Kortmann R-D. Different Approaches in Radiation Therapy of Craniopharyngioma. Front Endocrinol. 2011;0.

370. Minniti G, Esposito V, Amichetti M, Enrici RM. The role of fractionated radiotherapy and radiosurgery in the management of patients with craniopharyngioma. Neurosurg Rev. avr 2009;32(2):125-32; discussion 132.

371. Combs SE, Thilmann C, Huber PE, Hoess A, Debus J, Schulz-Ertner D. Achievement of long-term local control in patients with craniopharyngiomas using high precision stereotactic radiotherapy. Cancer. 1 juin 2007;109(11):2308-14.

372. Hashizume C, Mori Y, Kobayashi T, Shibamoto Y, Nagai A, Hayashi N. Stereotactic radiotherapy using Novalis for craniopharyngioma adjacent to optic pathways. J Neurooncol. juin 2010;98(2):239-47.

373. Reddy GD, Hansen D, Patel A, Lin Y, Jea A, Lam S. Treatment options for pediatric craniopharyngioma. Surg Neurol Int. 11 mars 2016;7(Suppl 6):S174-8.

374. Beltran C, Roca M, Merchant TE. On the benefits and risks of proton therapy in pediatric craniopharyngioma. Int J Radiat Oncol Biol Phys. 1 févr 2012;82(2):e281-287.

375. Kazemifar S, Barragán Montero AM, Souris K, Rivas ST, Timmerman R, Park YK, et al. Dosimetric evaluation of synthetic CT generated with GANs for MRI-only proton therapy treatment planning of brain tumors. J Appl Clin Med Phys. mai 2020;21(5):76-86.

376. Pileggi G, Speier C, Sharp GC, Izquierdo Garcia D, Catana C, Pursley J, et al. Proton range shift analysis on brain pseudo-CT generated from T1 and T2 MR. Acta Oncol. nov 2018;57(11):1521-31.

377. Klinge B, Lundström M, Rosén M, Bertl K, Klinge A, Stavropoulos A. Dental Implant Quality Register-A possible tool to further improve implant treatment and outcome. Clin Oral Implants Res. oct 2018;29 Suppl 18:145-51.

378. Palmér E, Karlsson A, Nordström F, Petruson K, Siversson C, Ljungberg M, et al. Synthetic computed tomography data allows for accurate absorbed dose calculations in a magnetic resonance imaging only workflow for head and neck radiotherapy. Physics and Imaging in Radiation Oncology. 1 janv 2021;17:36-42.

379. Persson E, Gustafsson C, Nordström F, Sohlin M, Gunnlaugsson A, Petruson K, et al. MR-OPERA: A Multicenter/Multivendor Validation of Magnetic Resonance Imaging-Only Prostate Treatment Planning Using Synthetic Computed Tomography Images. Int J Radiat Oncol Biol Phys. 1 nov 2017;99(3):692-700.

380. Siversson C, Nordström F, Nilsson T, Nyholm T, Jonsson J, Gunnlaugsson A, et al. Technical Note: MRI only prostate radiotherapy planning using the statistical decomposition algorithm. Medical Physics. 2015;42(10):6090-7.

381. Owrangi AM, Greer PB, Glide-Hurst CK. MRI-only treatment planning: benefits and challenges. Phys Med Biol. févr 2018;63(5):05TR01.

382. Han F, Zhou Z, Du D, Gao Y, Rashid S, Cao M, et al. Respiratory motion-resolved, self-gated 4D-MRI using Rotating Cartesian K-space (ROCK): Initial clinical experience on an MRI-guided radiotherapy system. Radiother Oncol. juin 2018;127(3):467-73.

383. Stemkens B, Paulson ES, Tijssen RHN. Nuts and bolts of 4D-MRI for radiotherapy. Phys Med Biol. 23 oct 2018;63(21):21TR01.

384. Persson E, Jamtheim Gustafsson C, Ambolt P, Engelholm S, Ceberg S, Bäck S, et al. MR-PROTECT: Clinical feasibility of a prostate MRI-only radiotherapy treatment workflow and investigation of acceptance criteria. Radiation Oncology. 9 avr 2020;15(1):77.

385. Jans LBO, Chen M, Elewaut D, Van den Bosch F, Carron P, Jacques P, et al. MRI-based Synthetic CT in the Detection of Structural Lesions in Patients with Suspected Sacroiliitis: Comparison with MRI. Radiology. 1 févr 2021;298(2):343-9.

Supplementary Materials

Table A1: Statistical analysis concerning HB, ZMUV, WS and NS cohorts constituted in the MRI standardization experiment. The analysis is based on the MAE, the gamma indices and the DVH differences computed on the PTV.

	p-value HB - ZMUV	**95% Confidence interval HB - ZMUV**	**p-value HB - NS**	**95% Confidence interval HB - NS**	**p-value ZMUV - NS**	**95% Confidence interval ZMUV - NS**
MAE head	**<0.0001**	[6.90, 9.66]	**<0.0001**	[-4.40, -2.75]	**<0.0001**	[-13.19, -10.61]
MAE air	**0.00034**	[5.53, 17.88]	**<0.0001**	[-22.76, -10.07]	**<0.0001**	[-33.81, -21.32]
MAE bone	**<0.0001**	[29.78, 40.44]	0.65	[-3.69, 2.52]	**<0.0001**	[-41.35, -31.79]
MAE water	**<0.0001**	[0.92, 1.85]	**<0.0001**	[-4.20, -3.12]	**<0.0001**	[-5.56, -4.49]
1%/1mm gamma index	0.34	[-0.02, 0.09]	**0.0052**	[0.03, 0.17]	0.16	[-0.02, 0.17]
2%/2mm gamma index	0.19	[-0.01, 0.04]	0.17	[-0.01, 0.05]	0.48	[-0.03, 0.04]

3%/3mm gamma index	**0.011**	[0.00, 0.04]	0.69	[0.00, 0.01]	0.46	[-0.03, 0.01]
Difference PTV $D_{02\%}$	0.88	[-0.04, 0.03]	0.18	[-0.06, 0.01]	0.45	[-0.06, 0.03]
Difference PTV $D_{50\%}$	0.312	[-0.02, 0.05]	0.11	[-0.07, 0.01]	**0.031**	[-0.08, 0.00]
Difference PTV $D_{95\%}$	0.39	[-0.02, 0.06]	**0.0020**	[-0.08, -0.02]	**0.0012**	[-0.11, -0.03]
Difference PTV DVH $D_{98\%}$	0.11	[-0.01, 0.07]	**0.0028**	[-0.10, -0.03]	**0.00074**	[-0.14, -0.04]

	p-value WS - HB	**95% Confidence interval WS - HB**	**p-value WS - ZMUV**	**95% Confidence interval WS - ZMUV**	**p-value WS - NS**	**95% Confidence interval WS - NS**
MAE head	**<0.0001**	[12.61, 15.53]	**<0.0001**	[4.92, 6.57]	**<0.0001**	[16.43, 19.14]

MAE air	**<0.0001**	[36.32, 49.49]	**<0.0001**	[26.94, 37.32]	**<0.0001**	[51.82, 66.13]
MAE bone	**<0.0001**	[45.77, 56.79]	**<0.0001**	[12.63, 17.99]	**<0.0001**	[47.00, 56.94]
MAE water	**<0.0001**	[1.97, 3.29]	**<0.0001**	[0.89, 2.02]	**<0.0001**	[5.74, 7.14]
1%/1mm gamma index	**0.00039**	[-0.20, -0.05]	**0.00056**	[-0.27, -0.05]	**<0.0001**	[-0.32, -0.13]
2%/2mm gamma index	0.073	[-0.04, 0.00]	**0.013**	[-0.05, -0.01]	**0.00094**	[-0.05, -0.01]
3%/3mm gamma index	0.19	[0.00, 0.01]	0.14	[-0.02, 0.00]	0.83	[-0.01, 0.01]
Difference PTV $D_{02\%}$	0.64	[-0.02, 0.04]	0.29	[-0.02, 0.06]	0.28	[-0.02, 0.07]
Difference PTV $D_{50\%}$	0.054	[0.00, 0.06]	0.26	[-0.02, 0.06]	**0.00046**	[0.03, 0.09]
Difference PTV $D_{95\%}$	**0.019**	[0.01, 0.08]	0.092	[0.00, 0.07]	**<0.0001**	[0.06, 0.13]
Difference PTV DVH $D_{98\%}$	**0.0020**	[0.02, 0.11]	0.13	[-0.01, 0.08]	**<0.0001**	[0.07, 0.17]

Table A2: Composition of the training, validation and test sets in terms of MR devices.

MR devices	Training set	Validation set	Testing set
DISCOVERY MR750w (3T)	133	34	57
SIGNA EXCITE (1.5T)	100	46	21
Optima MR450w (1.5T)	7	1	1
Optima MR360 (1.5T)	1	0	0
DISCOVERY MR450 (1.5T)	1	0	0

Table A3: Mean confusion matrix within the air, bone and water areas.

		Model 3			**Model 14**			**Model 48**		
		CT (%)			**CT (%)**			**CT (%)**		
		Air	**Bone**	**Water**	**Air**	**Bone**	**Water**	**Air**	**Bone**	**Water**
pCT (%)	**Air**	32.1 +/- 8.7	3.3 +/- 2.8	1.2 +/- 0.5	36.4 +/- 10.1	1.1 +/- 1.2	0.9 +/- 0.5	61.1 +/- 10.0	0.7 +/- 0.6	1.4 +/- 0.7
	Bone	8.1 +/- 3.4	44.3 +/- 10. 7	2.5 +/- 1.0	4.5 +/- 1.6	62.5 +/- 9.2	3.2 +/- 1.0	3.2 +/- 2.1	85.8 +/- 6.7	4.9 +/- 1.6
	Water	59.8 +/- 8.7	52.46 +/- 10.67	96.3 +/- 1.3	59.1 +/- 10.2	36.4 +/- 8.9	95.9 +/- 1.3	35.7 +/- 9.8	13.6 +/- 6.4	93.6 +/- 2.1

List A1: Publications.

Publications in peer-reviewed journals:

Alvarez Andres E, Fidon L, Vakalopoulou M, Lerousseau M, Carré A, Sun R, et al. Dosimetry-Driven Quality Measure of Brain Pseudo Computed Tomography Generated From Deep Learning for MRI-Only Radiation Therapy Treatment Planning. Int J Radiat Oncol Biol Phys. 1 nov 2020;108(3):813-23.

Oral communications in national congresses:

Alvarez Andres E, Fidon L, Vakalopoulou M, Noël G, Beaudre A, Niyoteka S et al. Assessing the impact of key preprocessing concepts on the pseudo CT generation. Société Française de Physique Médicale. 2019.

Posters:

Alvarez Andres E, Gasnier A, Veres C, Dhermain F, Corbin S, Auville F et al. Synthetic CT from MRI with deep learning: Assessing the clinical impact of generated errors. European Society for Radiotherapy and Oncology (ESTRO). 2021.

Alvarez Andres E, Fidon L, Vakalopoulou M, Lerousseau M, Carré A, Sun R et al. Optimizing the generation of Pseudo CT from MRI based on a highly efficient 3D neural network. ESTRO. 2020.

Alvarez Andres E, Caussé M, Fidon L, Ermeneux L, Bolle S, Martin V et al. Assessment of the generalizability to pediatric protontherapy of a 3D network generating Pseudo CT. ESTRO. 2020.

Alvarez Andres E, Fidon L, Vakalopoulou M, Noël G, Beaudre A, Niyoteka S, et al. 3D Deep Learning a s a powerful brain Pseudo Computed Tomography generator. ESTRO. 2019.

université PARIS-SACLAY

ÉCOLE DOCTORALE
Cancérologie : biologie - médecine - santé (CBMS)

Title: Characterization of pseudo CT from MRI with deep learning - Application to brain tumors treated with radiotherapy

Keywords: pseudo CT, MR, Deep learning, Brain tumors, Dosimetry

Abstract: Current brain tumor radiotherapy treatments require the acquisition of a Computed Tomography (CT) used during the segmentation and dosimetry steps, and a Magnetic Resonance Imaging (MRI) being mostly important for the target volumes delineation. Yet, dealing with multiple modalities implies to spatially register them, which has been shown to include 2mm errors, currently considered with a margin increase. Thus, generating pseudo Computed Tomography (pCT) from MRI appears to be an appealing approach to reduce margins and surrounding healthy tissues irradiation. The first step of the thesis aimed at characterizing parameters playing a key role in the Deep Learning (DL)-derived pCT quality, namely the training set size, the MRI sequence used as network input, the MRI standardization approach, the bias field correction filter and the network architecture. To do so, a large cohort composed of more than 400 patients was constituted, gathering images from multiple MR devices and tumor locations, to ensure the model robustness. The obtained pCT were first evaluated via the mean absolute error, based on intensities. Further dosimetry analyses were performed. Except for the training set size, all the studied approaches led to equivalent dosimetry performances.

With the goal to introduce pCT-based dosimetry in clinical practice and to deploy a methodology to validate the non-inferiority of MRI-only based-radiotherapy treatment planning compared to conventional CT-based radiotherapy treatment planning, a second study aiming at assessing the best-suited dosimetry criteria for an unbiased pCT evaluation was performed. Their correlations with intensity-based metrics were also calculated. Lastly, the impact of scenarios simulating extreme errors pCT was quantified, based on the previously defined metrics. A new test set of 71 brain patients was constituted reflecting tumor locations encountered in clinics and treatment modalities used in our center. Global and local 1%/1mm gamma indices with non-extreme dose thresholds were proved to be relevant for the brain pCT task evaluation. Additionally, dose volume histograms-based metrics differences for target and organs at risk volumes should also be computed since they reflect delineated structure-wise pCT dosimetry performance.

Lastly, pCT generation was applied to proton-based radiotherapy, via a collaboration with the Centre de Protonthérapie d'Orsay. The previously developed adults DL-model was tested on paediatrics to assess its generalizability. Satisfying clinical performances were reached, except for a few patients, potentially suggesting the transferability of the evaluated model. Future work consists in a dosimetry analysis in a larger scale, with the composition of a cohort of 198 children representing 4 different histologies. The goal is to assess the most efficient network training and validation approach, with stopping power maps to ensure the non-influence of CT devices parameters on images.

Thus, a solid understanding of key points for pCT generation and a methodology for pCT characterization have been achieved. The resulting recommendations have the key role to facilitate the quantification and interpretation of pCT quality evaluation criteria in the context of clinical trials set up, such as the ongoing observational GliopCT.

Titre : Caractérisation de pseudo CT générés à partir d'images IRM à l'aide de méthodes deep learning - Application aux tumeurs cérébrales traitées par radiothérapie

Mots clés : pseudo CT, IRM, Deep learning, Tumeurs cérébrales, Dosimetrie

Résumé : De nos jours, les traitements de tumeurs cérébrales par radiothérapie nécessitent l'acquisition d'un scanner utilisé pendant les étapes de segmentation et de dosimétrie, ainsi que d'une Imagerie par Résonance Magnétique (IRM) jouant un rôle important durant l'étape de segmentation des volumes cibles en particulier. Cependant, cette double modalité implique de recaler spatialement les images, processus qui induit des erreurs de 2mm, actuellement prises en compte par une augmentation de marges. Ainsi, générer des pseudo scanners (pCT) à partir d'images IRM apparaît comme étant une solution attractive pour diminuer les marges et réduire l'irradiation des tissus sains périphériques.

La première étape de cette thèse avait pour but de caractériser les paramètres jouant un rôle clé dans la qualité de pCT générés par Deep Learning (DL), à savoir la taille de la cohorte d'entrainement, la séquence IRM utilisée en entrée du réseau, la technique de standardisation des images IRM, le filtre de correction d'inhomogénéités de champ et l'architecture du réseau. Pour ce faire, une large cohorte composée de plus de 400 patients a été constituée, rassemblant des images de multiples appareils d'IRM et localisations tumorales, afin d'assurer la robustesse du modèle. Les pCT obtenus ont tout d'abord été évalués à l'aide de l'erreur absolue moyenne, basée sur les intensités. Des analyses dosimétriques ont ensuite été menées. Toutes les approches étudiées ont atteint des performances dosimétriques équivalentes, excepté pour la taille du jeu d'entrainement.

Pour introduire une dosimétrie basée sur les pCT en pratique clinique et déployer une méthodologie de validation de la non-infériorité de la planification de traitement dans le cas d'une radiothérapie basée seulement sur IRM comparée à celle conventionnellement basée sur scanner, une deuxième étude visant à définir les métriques dosimétriques les plus adaptées à une évaluation de pCT sans biais a été réalisée. Leurs corrélations avec des métriques basées sur les intensités ont été calculées. Enfin, l'impact de scenarios simulant des erreurs extrêmes de pCT a été quantifié, basé sur les métriques optimales préalablement définies. Un nouveau jeu de test de 71 patients a été constitué, reflétant les localisations tumorales rencontrées en clinique et les modalités de traitement propres à notre centre. Les gamma index globaux et locaux pour le critère 1%/1mm associés à des seuils de dose non-extrêmes ont montré leur pertinence pour la tâche d'évaluation de pCT cérébraux. De plus, les différences de métriques issues des histogrammes dose/volume des volumes cibles et organes à risque doivent aussi être calculées car elles reflètent la performance dosimétrique du pCT pour chaque structure segmentée.

Enfin, la génération de pCT a été appliquée à la radiothérapie basée sur protons, grâce à une collaboration avec le Centre de Protonthérapie d'Orsay. Le modèle de DL préalablement développé sur des patients adultes a été testé sur des patients pédiatriques, afin d'évaluer sa généralisabilité. Des performances cliniques satisfaisantes ont été atteintes, excepté pour quelques patients, pouvant potentiellement prouver la transférabilité du modèle évalué. Les travaux futurs comprennent une évaluation dosimétrique à plus grande échelle, avec la composition d'une cohorte de 198 enfants représentant 4 histologies différentes. Le but est de déterminer l'approche d'entrainement et de validation du réseau la plus efficace à l'aide de cartes de pouvoirs d'arrêt pour s'affranchir de l'influence des paramètres des scanners sur les images.

Ainsi, une solide compréhension des points clés de la génération de pCT ainsi qu'une méthodologie de leur caractérisation ont été réalisées. Les recommandations en découlant ont le rôle clé de faciliter la quantification et l'interprétation de critères d'évaluation de la qualité de pCT dans le contexte de mise en place d'essais cliniques, tel que l'essai observationnel en cours GliopCT.