Radiomics and Artificial Intelligence for personalized treatment of HPV-induced cancers

Stéphane Niyoteka

université PARIS-SACLAY

Radiomics and Artificial Intelligence for personalized treatment of HPV-induced cancers

Radiomique et Intelligence Artificielle pour un traitement personnalisé des cancers HPV-induits

Thèse de doctorat de l'Université Paris-Saclay

École doctorale n°582 : cancérologie : biologie - médecine - santé (CBMS)
Spécialité de doctorat : Science du cancer
Graduate School : Life Sciences and Health.
Référent: Faculté de médecine

Thèse préparée dans l'unité de recherche **Radiothérapie Moléculaire et Innovation Thérapeutique (Université Paris-Saclay, Institut Gustave Roussy, Inserm)**, sous la direction de **Charlotte ROBERT**, MCU, et la co-direction de **Eric DEUTSCH**, PU-PH

Thèse soutenue à Paris-Saclay, le 08 Juin 2023, par

Stéphane NIYOTEKA

Composition du jury

Membres du jury avec voix délibérative

Corinne BALLEYGUIER Professeur Associée, Université Paris-Saclay, UMR 1030, Villejuif	Présidente du jury
Issam El NAQA Prof, H. LEE MOFFITT CANCER CENTER AND RESEARCH INSTITUTE, INC	Rapporteur et Examinateur
Mathieu HA Directeur de recherche (DR2) INSERM, UMR 1101, Brest	Rapporteur et Examinateur
Benjamin LEMASSON Chargé de recherche, Université Grenoble Alpes, INSERM	Examinateur

THESE DE DOCTORAT
NNT : 2023UPASL056

université PARIS-SACLAY

ÉCOLE DOCTORALE
Cancérologie : biologie - médecine - santé (CBMS)

Titre: Radiomique et Intelligence Artificielle pour un traitement personnalisé des cancers HPV-induits

Mots clés: Intelligence Artificielle, Radiomique, Radiothérapie, Human Papillomavirus, IRM, TEP

Résumé: Les cancers HPV induits sont responsables d'une forte proportion des cancers du col de l'utérus localement avancés (LACC), des cancers de l'anus et de l'oropharynx. Afin d'améliorer la prise en charge et le traitement de ces cancers, le développement de schémas thérapeutiques combinant des modalités de traitement complémentaires telles que la chimioradiothérapie (CRT) et l'immunothérapie semble être une option thérapeutique prometteuse. Dans ce contexte, la recherche de biomarqueurs spécifiques à la réponse au traitement reste un enjeu majeur. Les récentes innovations technologiques en informatique, i.e. le Machine Learning (ML) et le Deep-Learning (DL) basé sur l'analyse d'images médicales, la radiomique, pourraient fournir des outils complémentaires et puissants pour une meilleure compréhension du cancer. La radiomique consiste à extraire des caractéristiques quantitatives d'images médicales, caractérisant la forme de la tumeur, ses intensités sur l'image et son hétérogénéité spatiale. Les algorithmes de ML basés sur les caractéristiques radiomiques ont déjà démontré leur capacité à identifier des motifs complexes permettant l'élaboration de modèles phénotypiques, pronostiques et prédictifs robustes.

Dans ce travail, le ML appliqué à l'imagerie en oncologie, à savoir la tomographie par émission de positons (TEP) avec 18F-fluorodésoxyglucose ([18F]-FDG TEP) et l'imagerie par résonance magnétique (IRM), a été exploré comme outil potentiel pour l'optimisation du traitement des cancers HPV-induits. Dans un premier temps, des modèles ML basés sur les variables cliniques, biologiques et radiomiques extraites des images TEP au [18F]-FDG ont été développés pour prédire la survie des patients. Deuxièmement, étant donné que les caractéristiques radiomiques dépendent fortement des paramètres d'acquisition et de reconstruction, un réseau conditionnel antagoniste génératif (cGAN) a été entraîné pour générer des images IRM synthétiques robustes aux paramètres d'acquisition. Dans le but de transposer ces modèles dans la pratique clinique, une méthode de standardisation basée sur un CycleGAN a été entraînée sur une cohorte rétrospective de patients traités pour un LACC et sa contribution a été évaluée sur deux tâches cliniques : la classification du stade tumoral et la prédiction de rechute.

D'après nos résultats, les modèles de prédiction de survie basés sur les caractéristiques radiomiques extraites des images TEP au [18F]-FDG pourraient prédire la survie avec une signature commune aux LACC, et aux cancers du canal anal et de l'oropharynx. Les résultats ont été validés sur des jeux de données indépendants, mais la valeur ajoutée par rapport aux grandeurs conventionnelles telles que le volume tumoral métabolique reste à démontrer. Nous avons montré la supériorité des réseaux de neurones pour la standardisation des images IRM en comparaison à des méthodes plus conventionnelles. Dans cette thèse, nous avons démontré que les modèles de ML pourraient ouvrir la voie à une radiomique reproductible en routine clinique, pour conduire à un traitement personnalisé des cancers HPV-induits.

Title: Radiomics and Artificial Intelligence for personalized treatment of HPV-induced cancers
Keywords: Artificial Intelligence, Radiomics, Radiotherapy, Human Papillomavirus, MRI, PET

Abstract: HPV-induced cancers account for a high proportion of locally advanced cervical cancers (LACC), anal cancers, and oropharyngeal cancers. In order to improve the management and treatment of these cancers, the development of therapeutic regimens combining complementary treatment modalities such as chemoradiotherapy (CRT) and immunotherapy appears to be a promising therapeutic option. In this context, the search for specific biomarkers of treatment response remains a major challenge. Recent technological innovations in computer science, i.e. Machine Learning (ML) and Deep-Learning (DL) based on medical image analysis, radiomics, could provide complementary and powerful tools for a better understanding of cancer. Radiomics consists in extracting quantitative features from medical images, characterizing the tumor shape, its intensities on the image and its spatial heterogeneity. ML algorithms based on radiomic features have already demonstrated their ability to identify complex patterns allowing the development of robust phenotypic, prognostic and predictive models.

In this work, ML applied to oncology imaging, namely 18F-fluorodeoxyglucose positron emission tomography (PET) ([18F]-FDG PET) and magnetic resonance imaging (MRI), was explored as a potential tool for optimizing treatment of HPV-induced cancers. First, ML models based on clinical, biological and radiomic variables extracted from [18F]-FDG PET images were developed to predict patient survival. Second, because radiomic features are highly dependent on acquisition and reconstruction parameters, a conditional generative adversarial network (cGAN) was trained to generate synthetic MR images robust to acquisition parameters. With the aim of translating these models into clinical practice, a CycleGAN-based standardization method was trained on a retrospective cohort of patients treated for LACC and its contribution was evaluated on two clinical tasks: tumor stage classification and relapse prediction.

Based on our results, survival prediction models using radiomic features extracted from [18F]-FDG PET images could predict survival outcomes using a common HPV signature in LACC, anal cancers, and oropharyngeal cancers. Results have been validated on external independent data sets, but the added value compared to conventional quantities such as metabolic tumor volume (MTV) remains to be demonstrated. We have shown the superiority of neural network-based MR images standardization compared to more conventional methods. In this thesis, we demonstrated that ML models could pave the way for reproducible radiomics in clinical routine that would in turn lead to personalized treatment of HPV-induced cancers.

Acknowlegments

I would like to thank first of all my thesis supervisors Charlotte Robert and Eric Deutsch. Thank you Charlotte for your advice, mentoring and invaluable contributions along the way. Eric, you have been a never-ending source of knowledge. I am very glad to have completed this work under your guidance.

Thanks to all the members of the INSERM U1030 laboratory. Special thanks to Emilie Alvarez-Andres, Angela Rouyar-Nicolas, and Rahimeh Rouhi. In addition, the "experts in everything", Nathan Benzazon, François de Kermenguy, and Cathyanne Schott provided just the right amount of madness to ensure that this project was not just science-oriented. Thank you for these sparkling conversations about communities, chatGPT, and our societies in general.

I would also like to thank Thomas Berger for guiding my first steps into the world of research, for our conversations along the Kattegat Sea area, and for his friendship.

This thesis would not have been possible without the support of my friends: Michael, Elvis, Lynca, Thierry, Fleury, Maso, Steve, Raïssa, Peace, Alexis, Bella, Francis, and Audrey, to name but a few.

To my parents, whose unwavering love, support, and sacrifices have been the cornerstone of my success. You have instilled in me values of hard work, determination, and resilience, which have propelled me forward. Thank you for gifting me with extraordinary siblings: Inès, Audrey, Joseph, and Bertrand who stood by my side all the long.

Finally, I would like to thank my wife Audrey whose affection and support could only be described in pale, hollow words.

Contents

Introduction

Cancer is a complex disease affecting millions of people worldwide with almost one in six deaths, rendering cancer the second overall cause of death in the world according to the World Health Organisation. In addition to the human toll, cancer carries a heavy economic burden estimated at $25.2 trillion from 2020 to 2050[1]. The most common cancers are breast, lung, colorectal, and prostate cancer. Among the diversity of cancer-related risk factors, chronic infections with bacteria and viruses are common risk factors for cancer. Human papillomavirus (HPV), a mainly sexually transmitted infection, accounts for 31.1% of cancers worldwide, especially in low- and middle-income countries. HPV is responsible for most cervical cancers, oropharyngeal cancers (and a substantial part in head and neck cancers), anal cancers, penile cancers, vaginal cancers, and vulvar cancers. Between 30 and 50% of cancers can currently be prevented by avoiding risk factors and implementing existing evidence-based prevention strategies. The cancer burden can also be reduced through early detection of cancer and appropriate treatment and care of patients who develop cancer.

Medical imaging has proved to be an essential component in cancer management. It is now mandatory to proceed to multiple modality acquisitions from the diagnosis of the disease to the follow-up of the patient. It has been shown that medical images are more than tools for a one-time inspection: images are data[2]. Artificial Intelligence (AI), especially machine learning (ML) thanks to various pattern recognition tools has taken advantage of medical images, via the field of radiomics. Radiomics is a field of medical imaging analysis that uses high-dimensional, mineable features to build image-based models for better patient management. In HPV-induced cancers, studies have reported that CT-based and MRI-based radiomics could predict HPV status in head and neck cancers[3, 4, 5]. In addition to prediction of HPV status, radiomic studies have shown promising results in assessing treatment response. Several studies have assessed the power of radiomic features in the prediction of lymph node metastasis in cervical cancer[6, 7, 8]. Radiomic prognostic performances have also been studied. Multiple sources have reported that image-based models could be used for personalized treatment after promising results in prediction of progression, recurrence, or survival in HPV related cancers[9, 10, 11, 12, 13, 14].

Despite excellent results and flourishing literature, some limitations still hinder the clinical implementation of developed radiomic-based models. Radiomic-features are indeed highly dependent on the scanning devices used during image acquisition. Given that most studies are not multi-institutional (images are acquired with the same acquisition parameters), the generalizability of developed models will be negatively affected rendering such models obsolete. ML models also suffer from inadequate practices in the field, such as inadequate feature selection or absence of an independent validation cohort, which do not allow us to conclude on a realistic implementation in clinical routine. The purpose of this thesis was to develop radiomic-oriented ML algorithms for personalized treatment in HPV-related cancers, namely locally advanced cervical cancer (LACC), anal squamous cell carcinomas (ASCC), and oropharyngeal squamous cell carcinomas (OSCC). In this thesis, some of the limitations encountered in radiomic studies were addressed.

In Chapter 1, a brief summary of the genesis of cancer is presented. The role of AI based on medical imaging in cancer management is discussed.

In Chapter 2, HPV cancers in general are described. We focused on LACC, ASCC, and OPSCC as these

types of cancers account for up to 90% of most HPV-induced cancers. The role of MRI and FDG-PET in the management of these cancers is also discussed as an introduction to the work done in this thesis.

In Chapter 3, ML in medical imaging analysis is described with an emphasis on radiomic modeling, from image acquisition to ML algorithms building. General concepts, good practices, and practices to avoid are detailed.

In Chapter 4, efforts were made to collect a large dataset of FDG-PET images from five institutions of patients treated for HPV-related cancers. A radiomic signature based on PET imaging to predict survival and the evaluation of its generalization properties on multiple HPV-induced cancers, was developed.

Signal intensities in MRI are highly dependent on imaging parameters, which will in turn impact radiomic features. Strategies to mitigate the effects of acquisition and reconstruction imaging parameters for strong reproducibility and robustness of radiomic features must be established. In Chapter 5, the lack of robust standardization methods for MRI-based radiomic studies in LACC was addressed. A Deep-Learning (DL) based standardization method was developed and compared to other conventional methods.

In Chapter 6, we discussed tumor automatic segmentation in LACC and the potential use of the developed standardization method coupled to radiomics to assess the impact of combining immunotherapy with chemoradiation (CRT) in the framework of the AtezoLACC clinical trial.

In Chapter 7, conclusions and general perspectives are discussed.

Chapter 1

Cancer: Evolution towards Precision Medicine

1.1 Genesis of Cancer

Fossilized bone tumors, human mummies in ancient Egypt, and ancient manuscripts have shown that mankind has had cancer throughout recorded history. The oldest description of the disease can be retraced back to about 3000 BC in ancient Egypt on a manuscript called the Edwin Smith Papyrus. This document is a copy of an ancient medical text on trauma surgery describing 8 cases of tumors or ulcers of the breast that were removed by cauterization. The oldest known scientific treatise about cancer explicitly said that there was no treatment. Fortunately, knowledge of the matter has improved significantly over the years. The Greek physician Hippocrates (460-370 BC), used the word "carcinos", meaning crab, to describe ulcers evolving into tumors. The Roman physician, Celsus (25 BC - 50 AD), later translated the Greek term into cancer, the Latin word for crab. The highly spreading nature of tumor cells was clearly captured by the crab analogy. It was thought up to several hundred years AD that an imbalance in body fluids was the cause of cancer. This theory is known as the humoral theory of cancer in reference to the belief that the human body was made of 4 humors or body fluids. Later on in the 1700s, a theory based on lymph nodes gained rapid success. Cancer was then thought to be caused by another body fluid called lymph and tumors were believed to grow from lymph constantly thrown out by the blood. It was not until 1775 that the first link of environmental exposure to the development of cancer was shown by Percival Pott. By 1838, Johannes Muller demonstrated that tumor cells were not made of lymph but developed from normal cells. Rudolph Virchow correctly linked the origin of cancers to normal cells, believing that cancer was caused by severe irritation in the tissues (the 'chronic irritation theory') [15]. Virchow's theory was then improved in the 1860s by Karl Thiersch, showing that cancers metastasized through the spread of malignant cells, and not through some unidentified fluid [1]. Throughout the 17th and 18th centuries, cancer was also believed contagious after members of the same household experienced breast cancer. In the 1900s, several studies showed that external causes such as viruses, tobacco, and coal tar, were directly linked to the development of cancers[16, 17, 18, 19], even though earlier knowledge on the subject existed. The discovery of

[1] https://www.cancer.org/treatment/understanding-your-diagnosis/history-of-cancer/modern-knowledge-and-cancer-causes.html

DNA structure by Crick and Watson [20] led to the missing evidence for a genetic component in some cancer growth. It is now believed that all causes of cancer, carcinogens, cause gene damages which will in turn evolve into cancer. Twelve key mechanisms are commonly accepted as hallmarks of cancer [21] namely:

Sustaining Proliferative Signaling Unlike non-pathological cells, cancer cells can acquire the ability to maintain proliferative signals. The enabling signals are conveyed in large part by growth factors that bind cell-surface receptors signaling pathways that regulate progression through the cell cycle as well as cell growth (that is, increases in cell size); often these signals influence yet other cell-biological properties, such as cell survival and energy metabolism.

Evading Growth Suppressors Cancer cells can also circumvent powerful programs that negatively regulate cell proliferation. The two main inhibitory proteins of cell proliferation are the RB protein, associated with retinoblastoma, and the p53 protein, both involved in the phenomena of proliferation, senescence, and apoptosis.

Resisting Cell Death Tumor cells develop a variety of strategies to limit or circumvent apoptosis. The most common is the loss of TP53 tumor suppressor function, but tumors may achieve similar results by increasing expression of antiapoptotic regulators (Bcl-2, Bcl-xL) or of survival signals (Igf1/2).

Enabling Replicative Immortality Normal cells are able to pass through only a limited number of successive cell growth-and-division cycles in contrast to cancer cells which possess an enzyme very rare in normal cells, telomerase, allowing them unlimited replication.

Inducing Angiogenesis The tumor has the ability to create a vascular system necessary for its growth: this is an angiogenesis induction phenomenon. This tumor-associated neovasculature will address its needs in the form of nutrients and oxygen as well as metabolic wastes and carbon dioxide evacuation.

Activating Invasion and Metastasis By a succession of cellular biological changes, beginning with local invasion, then intravasation into nearby blood and lymphatic vessels, transit through the lymphatic and hematogenous systems, followed by escape of cancer cells from the lumina of such vessels into the parenchyma of distant tissues (extravasation), the formation of small nodules of cancer cells (micrometastases), and finally the growth of micrometastatic lesions into macroscopic tumors, cancer cells invade and colonize normal organs.

Genome Instability and Mutation Non-pathological cells have a DNA repair system that limits the rate of spontaneous mutations. Mutations in cancer cells are related to altered genome maintenance processes and surveillance systems.

Tumor-promoting inflammation The presence of infiltrating immune cells in the tumor can induce positive effects on tumor progression. Inflammatory cells produce active molecules in the microenvironment and can release substances such as free radicals, which are highly mutagenic to the cells.

Reprogramming Energy Metabolism The chronic and often uncontrolled cell proliferation that represents the essence of neoplastic disease involves not only deregulated control of cell proliferation but also

corresponding adjustments of energy metabolism in order to fuel cell growth and division. By limiting their energy metabolism largely to glycolysis, cancer cells can reprogram their energy production.

Evading Immune Destruction By inhibiting the action of the molecules produced by the immune system to destroy cancer cells, they are able to escape eradication.

Phenotypic Plasticity Cancer cells can unlock the normally restricted capability for phenotypic plasticity in order to evade or escape from the state of terminal differentiation.

Nonmutational Epigenetic Reprogramming Genome (DNA) instability and mutation is a fundamental component of cancer formation but mutation-free cancer evolution and purely nonmutational epigenetic programming can cause cancer growth. Epigenetic reprogramming is the process by which an organism's genotype interacts with the environment to produce its phenotype and provides a framework for explaining individual variations and the uniqueness of cells, tissues, or organs despite identical genetic information[22].

Polymorphic Microbiomes The diversity and variability of the plethora of microorganisms, collectively termed the microbiota, have profound impact on health and disease. Polymorphic variability in the microbiomes between individuals in a population can have a profound impact on cancer phenotypes.

Senescent Cells Cellular senescence is a typically irreversible form of proliferative arrest, likely evolved as a protective mechanism for maintaining tissue homeostasis, as a complementary mechanism to programmed cell death that serves to inactivate and in due course remove diseased, dysfunctional, or otherwise unnecessary cells. By the activation of a senescence-associated secretory phenotype (SASP) involving the release of a plethora of bioactive proteins, including chemokines, cytokines, and proteases, senescent cells have been shown to contribute to proliferative signaling, avoiding apoptosis, inducing angiogenesis, stimulating invasion and metastasis, and suppressing tumor immunity[23, 24]

1.2 Cancer Treatment

Cancer treatment has kept improving with the understanding of the intrinsic nature of the disease while benefiting from technological advances in the field. Physicians, for a long time, believed that no treatment existed, with surgery more or less localized to the tumor area being the only available treatment. Since then, the scientific community has highlighted various mechanisms leading to the development of cancer, which has allowed the development of treatments, in addition to surgery, to stop the proliferation of tumor cells. Although treatment is not always curative for these cancers, there has been significant improvement in progression-free survival (PFS) and overall survival (OS).

Nevertheless, surgery remains today the main course of treatment for the majority of cancers. Alongside surgery in the case of solid tumors, are radiation therapy and antitumor drugs, chemotherapy and immunotherapy. The first use of chemotherapy in treating cancer was in the 1930s. Chemotherapy is a drug treatment using chemicals to kill fast-growing cells. Used as primary treatment, it can be curative in some types of

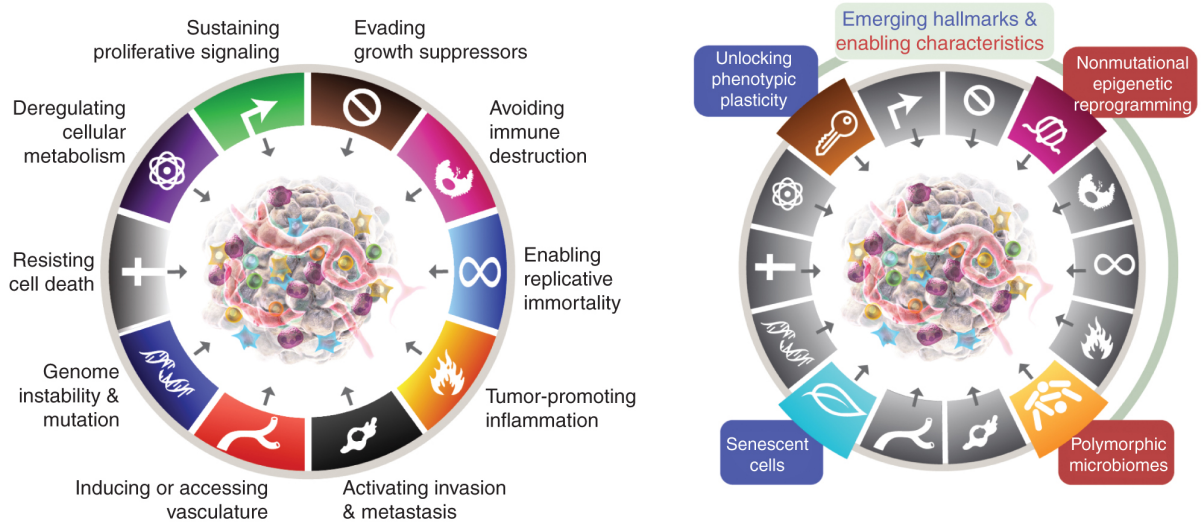


Figure 1.1: The Hallmarks of Cancer. Left, from the Hallmarks of Cancer proposed in 2000 and in 2011 [25, 26] Hallmarks of Cancer embody eight hallmark capabilities and two enabling characteristics. Right, the figure incorporates additional proposed emerging hallmarks and enabling characteristics involving "unlocking phenotypic plasticity", "nonmutational epigenetic reprogramming," "polymorphic microbiomes," and "senescent cells." *Reprinted from [21] under the Creative Commons Attribution-NonCommercial 4.0 International.*

advanced cancers, including acute lymphoblastic and acute myelogenous leukemia, germ cell cancer, small cell lung cancer, ovarian cancer, and choriocarcinoma. Chemotherapy can also be used concomitantly with RT or as neoadjuvant therapy. Neoadjuvant chemotharapy aims at reducing the size of the primary tumor and preventing micrometastases. This type of treatment improves on conservative surgical techniques in preserving the functionality of important organs. Neoadjuvant chemotherapy is indicated for anal, breast, lung, gastroesophageal, rectal, bladder, and head and neck cancer (HNC), as well as some types of sarcoma[27]. Radiation therapy (RT) is the use of high doses of ionizing radiation, targeted to tumor sites, and administered mostly to cause DNA damage that results in the arrest or slowing of tumor cell growth. Cancer cells exposed to cytotoxic doses of radiotherapy also undergo a potently immunostimulatory cell death variant that has been named immunogenic cell death[28]. It has been shown that ionizing radiation could also reduce tumor growth outside the field of radiation, known as the abscopal effect[29]. The earliest radiation treatments were delivered as single large exposures using low-energy cathode-ray tubes or radium-filled glass tubes close to tumors. Unfortunately, this type of treatment implied very high doses which were often associated with extensive normal tissue damage[30]. Since then, machines for treatment delivery have been constantly improved with devices producing higher energy X-rays that penetrate deeper into tissue allowing for treatment of internal tumors and causing lower skin burns. In the earliest 1950s, RT devices, such as compact linear accelerators installed on gantries capable of rotating 360 degrees around patients or teletherapy units using cobalt-60, had custom-designed metal blocks placed in the head of the treatment machines to shape radiation beams. Recent advances from the past decades have allowed the use of multi-leaves collimators (MLCs) which consist of small metallic leaves located in the head of the treatment unit. Each leaf within an MLC moves independently of the others to create beams with specific characteristics of shape. This technique, known as three-dimensional conformal radiation therapy, allows both conformal (that is, high doses "wrap" closely around the target volume) and homogeneous (with little variability

of dose within the target) delivery[30]. This allowed for the development of volumetric modulated arc therapy (VMAT), which from an arc trajectory, delivers doses dynamically during rotation. By moving the MLCs during the beam-on time, the field aperture is split into smaller segments, and a checkerboard pattern across the length and width of each beam is achieved such that each small segment delivers a different intensity. This is called intensity-modulated radiation therapy (IMRT) and this technique offers the potential for improved dose distributions in many clinical situations[31]. Unlike IMRT, VMAT do not support beam modulation by the MLCs. In the case of very small tumors, high-precision devices can deliver a high radiation dose under stereotactic conditions (Cyber knife or Gamma knife). In addition, new treatment techniques are being assessed. This is particularly the case for FLASH radiotherapy. Re-discovered in 2014 in mice, ultrahigh dose rate irradiation increases differential responses between normal and tumor tissues. This therapy delivers a single high dose of irradiation in a very short time (less than 200 milliseconds) with an equal probability of tumor control and little or no normal tissue damage compared to current radiotherapy[32, 33]. FLASH therapy opens up new and very promising treatment possibilities and could represent a major breakthrough not only in cancer management but also in its understanding. In addition to photons and electrons being widely used for cancer treatment, particle beams have also been developed. Due to the unique energy absorption profile of protons in tissues, proton beam therapy is among the most highly conformal radiation modalities[34, 35]. The main limitation of proton therapy is due to the large financial investment required partly for building cyclotrons used to generate therapeutic proton beams. Smaller and more affordable cyclotrons are now in development, and if successful, this may result in the more widespread availability of proton-based radiotherapy in the coming decades. In the future, it is expected that various techniques of radiotherapy, such as proton therapy, heavy particle therapy, and FLASH therapy, will be widely or specifically used.

Surgery, chemotherapy, and RT can be used as primary or sole treatment but in most cases are used in combination. Since the 1990s, several clinical trials have shown that the combination of radiotherapy with cytotoxic chemotherapy can result in improved local tumor control, organ preservation, and patient survival in some oncological settings[36]. Over the past few years, a new form of therapy that uses drugs that block immune checkpoints, with the goal to boost the natural ability of lymphocyte cells to fight cancer and enhance the body's immune response against cancer, called immunotherapy, has become an important therapeutic alternative. The immune system was shown to play a major role in therapeutic responses to radiotherapy[29]. Remarkable systemic effects have been observed in patients receiving radiotherapy to control tumors that were progressing during immune checkpoint blockade, stimulating interest in using radiation to overcome primary and acquired cancer resistance to immunotherapy[37]. This has motivated several clinical trials combining multiple variants of immunotherapy with different forms of radiotherapy. As a recent example in this context, chemoradiotherapy plus the immune checkpoint inhibitor durvalumab for the treatment of unresectable stage III non-small-cell lung cancer (NSCLC)[38] was approved. More than 150 clinical trials are exploring the combination of standard, full-dose (chemo)radiotherapy plus immunotherapy[39, 40].

1.3 Medical imaging in radiotherapy

Before the age of medical imaging, cancer diagnosis required in the majority of cases exploratory surgery and up to now, in most situations, biopsy is the only way to definitively diagnose cancer. With the progress of medical imaging, namely ultrasound (US), computed tomography (CT), magnetic resonance imaging (MRI), single-photon emission computed tomography (SPECT), and positron emission tomography (PET), biopsies procedures were coupled with medical images. These diagnostic images allow both, anatomical and functional imaging. Anatomical imaging enables detection of structural details that are sometimes, but not invariably, associated with cancer mass whereas functional imaging allows detection of molecular signals that indicate the presence of biochemical activity and changes (changes in metabolism, blood flow, etc.)[41].

Anatomic images in daily clinical routine are based on structural features such as location, size, morphology, and structural changes to adjacent tissues. Anatomical imaging is an essential tool used in radiotherapy treatment planning to accurately localize tumors and normal structures within a patient's body. This imaging is used to create a 3D image of the patient's anatomy, which helps radiation oncologists and medical physicists to design a customized treatment plan that delivers radiation precisely to the tumor while minimizing exposure to surrounding healthy tissues. Anatomical imaging is also used during the course of radiotherapy treatment to monitor tumor response and assess the need for any adjustments in the treatment plan. For example, repeat imaging may be performed to assess tumor shrinkage or growth, which may necessitate changes in the radiation dose or treatment fields. CT is the modality of choice for initial cancer evaluation because of its widespread availability, short scan times, excellent multiplanar reconstruction capabilities, and familiarity among clinicians. T1-weighted (T1w) or T2-weighted (T2w) MRI are particularly well suited to specific anatomic locations such as prostate cancer, uterine cancer, the paranasal sinuses, salivary glands, oral cavity, and certain liver cancers but suffer from a poor contrast in soft tissue. Structural imaging alone may not provide the clinician with all of the information that is necessary to fully characterize or monitor lesions in those with cancer or at risk for cancer[42]. With anatomic imaging, it is often not clear whether masses are the result of malignant or benign etiologies, as in solitary pulmonary nodules or borderline-size lymph nodes. Similarly, small cancers are hardly detectable with traditional anatomic methods, because they have not yet formed a mass[43]. Moreover, assessment of lymph nodes based on structural imaging features alone is limited in sensitivity and specificity[42].

Functional imaging (mostly PET, SPECT and some sequences of MRI) helps to address many of the limitations of anatomic imaging during radiotherapy. Functional imaging techniques in MRI can be used to study tumor physiology through perfusion imaging that studies the vascularity of malignancies (dynamic contrast-enhanced MRI), diffusion imaging that measures the diffusion of water molecules (diffusion-weighted MRI, or diffusion tensor MRI), and lymph node imaging for tumor staging (ferumoxtran-10-enhanced MRI). Functional images are wildly used to assess tumor molecular processes through metabolic imaging (such as glucose metabolism, amino-acid metabolism) using PET with different metabolic radiotracer, cell proliferation imaging, oxygenation imaging, or molecular imaging with targeted imaging agents. These additional information can be acquired before treatment delivery (e.g. tumor staging) to improve treatment planning or after radiotherapy to assess the tumor response. Functional imaging can also be implemented through the use of CT, and US, as well as through PET, SPECT, and MRI, and when combined with anatomic images (such as PET/CT or PET-MRI),

is emerging as a particularly valuable tool, providing both anatomic precision and functional information in a single image set[43].

Much of the progress in radiotherapy in recent decades has been due to technological advances in medical imaging. Medical imaging has improved invasive procedures like biopsies through image-guided biopsy with CT, real-time X-ray (fluoroscopy), US, MRI, or a combination of these modalities. In addition to improving surgical practices, early diagnosis, and follow-up of cancer patients, advances in medical imaging have also allowed for the optimization of cancer treatment. This has been particularly significant in RT with the use of increasingly precise techniques guided by imaging. In particular, axial imaging methods and three-dimensional treatment planning systems have significantly improved our ability to deliver homogenous radiation doses, especially in difficult anatomic locations and where there are unusual shapes. CT–based imaging before treatment planning is now a standard procedure for radiation oncology centers. The use of functional and anatomical imaging such as MRI and PET scans has also become common for tumor volumes and normal organ anatomy characterization. In prostate cancer, for example, MRI can better differentiate between the gland and the surrounding muscle. In most cancers, PET scans can better discriminate between hypermetabolic tumor and adjacent normal tissue[30]. Innovations in imaging modalities have led to significant improvement to position the patient and precisely deliver radiation to a tumor. One of the most prominent imaging techniques in image-guided radiotherapy (IGRT) is the acquisition of cone beam computed tomographies (CBCTs), through X-rays tubes integrated into linear accelerators (linac), with the patient in treatment position. CBCTs provide accurate 3-dimensional (3D) knowledge about the patient's anatomy for every treatment fraction[44, 45]. The MR-Linac combines an MV linac with onboard MR imaging. The MR unit enables monitoring of the patient throughout the treatment, before, during, and after dose delivery. MR monitoring allows for real-time target and organs at risk (OARs) visualization, enabling a fully integrated on-table adaptive workflow[46, 47]. In particular, MR-linac will enable daily online adaptive RT, especially in cancer where inter- and intra-fractional changes of the target could lead to suboptimal tumor coverage and/or high exposure of OARs.

Medical imaging has therefore helped in improving detection and characterization of tumors, optimizing dose delivery during RT, monitoring tumor response to therapeutic intervention, improving delineation of residual or recurrent tumors, and assessing patient prognosis before and after treatment. It is clear that the role of imaging in cancer treatment is crucial and the scientific community agrees that its potential has not been fully explored.

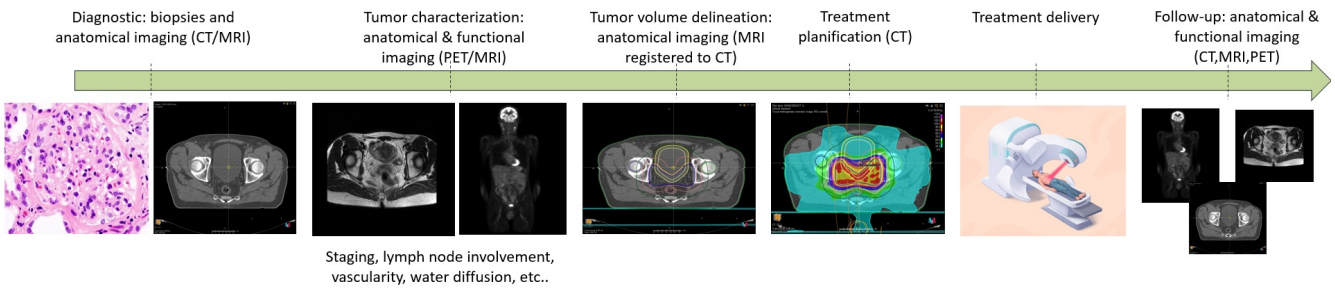


Figure 1.2: The radiotherapy workflow and the place of medical imaging at different steps.

1.4 Cancer in the era of Big Data

1.4.1 Big Data and Artificial Intelligence

Cancer is daunting in the breadth and scope of its diversity, spanning genetics, cell and tissue biology, pathology, and response to therapy[21]. The integrative concept embodied in the hallmarks of cancer is helping to distill this complexity into an increasingly logical science, and the provisional new dimensions presented in this perspective may add value to that endeavor, to more fully understand mechanisms of cancer development and malignant progression, and apply that knowledge to cancer medicine. Various studies use different approaches to understand cancer in its complexity from the molecular to the macroscopic level. Ever more powerful experimental and computational tools and technologies are providing an avalanche of "big data" about the myriad manifestations of the diseases that cancer encompasses[21]. It's essential to remember, however, that data by themselves are useless but must be analyzed, interpreted, and acted on. Thus, it is essential to develop algorithms, as analytical tools of these accumulated data, that will prove transformative in regard to cancer understanding[48]. Many genius individuals have participated in the development of these powerful algorithms since the ground-breaking work in the field of artificial intelligence (AI) in the mid-20th century by the British logician and computer pioneer Alan Mathison Turing. Artificial intelligence (AI) is a broadly encompassing term, coined by McCarthy et al.[49] in the 1950s, referring to the branch of computer science in which algorithms are developed to emulate what an intelligent human might do in the same situation[50]. Arthur Samuel considered a pioneer in AI and computer gaming, developed the first accepted today as a self-learning computer program in 1952, the Samuel Checker Playing Program[51]. The more its program played the game, the more it learned from its experience, by using a minimax algorithm that analyzed moves to find winning strategies. Since then, there have been tremendous efforts to improve AI performances, and in 1997, the IBM super-computer Deep-Blue defeated the world chess champion, Garry Kasparov, proving that machines were indeed capable of human-like intelligence. AI technology is indeed capable of extracting valuable data from huge sources of complex information, without having to rely on humans in some cases. Where traditional analytical tools are often limited by the maximum amount of data that can be analyzed, AI reveals its full potential when data sources are growing, allowing it to learn and refine insights with ever-increasing accuracy, which makes it suitable for "Big Data". Clearly, the more data there is, the more powerful AI computers are and the more efficiently they can discover patterns buried in the data. AI, similarly to a doctor progressing through residency, learns rules from data. Algorithms navigate through vast numbers of variables, including patient-level observations, looking for combinations that reliably predict outcomes. AI is a general term that encompasses various fields such as ML, robotics, natural language processing (NLP), and expert systems. Deep learning (DL) is a particular subset of ML that structures algorithms in layers to create an "artificial neural network". Machine Learning (ML) usually provides computer systems with the ability to learn and enhance from experience automatically without being specifically programmed[52]. The main difference between ML and traditional regression models (which are now often considered as ML), is that ML handles a large number of variables — sometimes, remarkably, more predictors than observations — and combines them in nonlinear and highly interactive ways[**obermeyer_predicting_2016**]. Further details on ML are given in Chapter 3

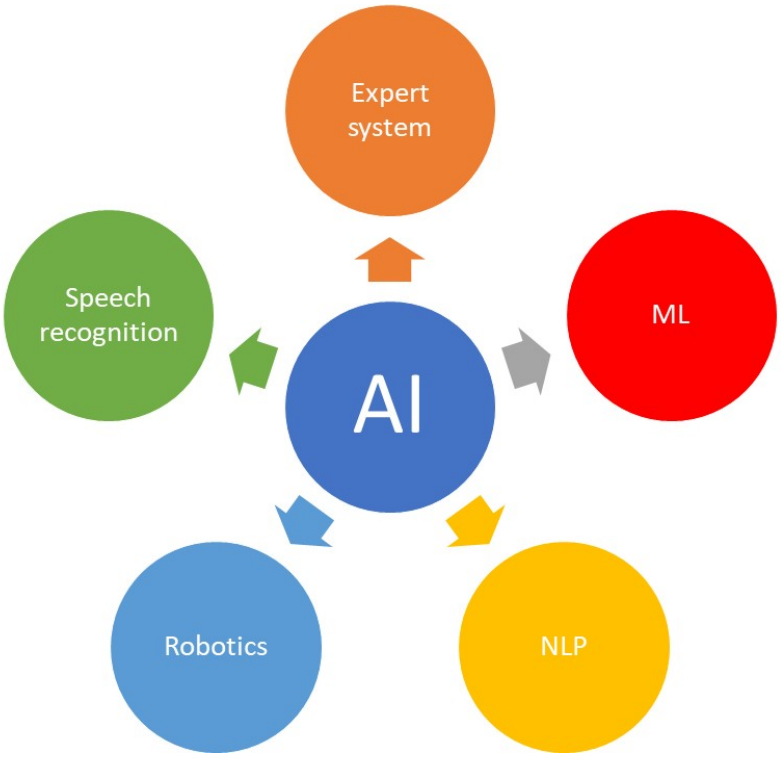


Figure 1.3: Subsets of Artificial Intelligence. ML; Machine Learning; NLP; Natural Language Processing

1.4.2 Artificial Intelligence for cancer management

AI techniques open the possibility for new and innovative approaches in the treatment of diseases, in particular cancer, as they allow the processing of large and complex amounts of data on individual variability. The application of these techniques could help explain why patients with the same clinical diagnosis or symptoms develop different responses to the same treatment. In other words, this would lead to precision medicine, as the ability to classify individuals into groups differing in their susceptibility, prognosis, or response to a particular disease and treatment[53]. This could be particularly applied to immunotherapy as it has been seen that only a small fraction of treated patients respond to this particular and promising treatment[54]. AI is therefore a natural tool to explore the potential of all kinds of medical data, including biological, clinical, genomic, proteomic, histological data, and in particular medical images as seen in Figure 1.4. Given that medical imaging is crucial and contains multiparametric information (structural, functional, molecular, and metabolic), for precision medicine to reach its highest potential, medical imaging must be a necessary and integral part in ML research[55]. A field that shows great promise in this context is radiomics, a new "data-driven" approach to the analysis of medical images. Radiomics is defined as "the high-throughput mining of quantitative image features from standard-of-care medical imaging that enables data to be extracted and applied within clinical-decision support systems to improve diagnostic, prognostic, and predictive accuracy"[56].

Whenever possible, radiomic models are combined with uncorrelated data such as those provided by clinical variables such as demographics, patient risk factors, biological analyses, or genomics. Such models that include both radiomics and clinical covariates generally improve predictive capabilities[14, 11].

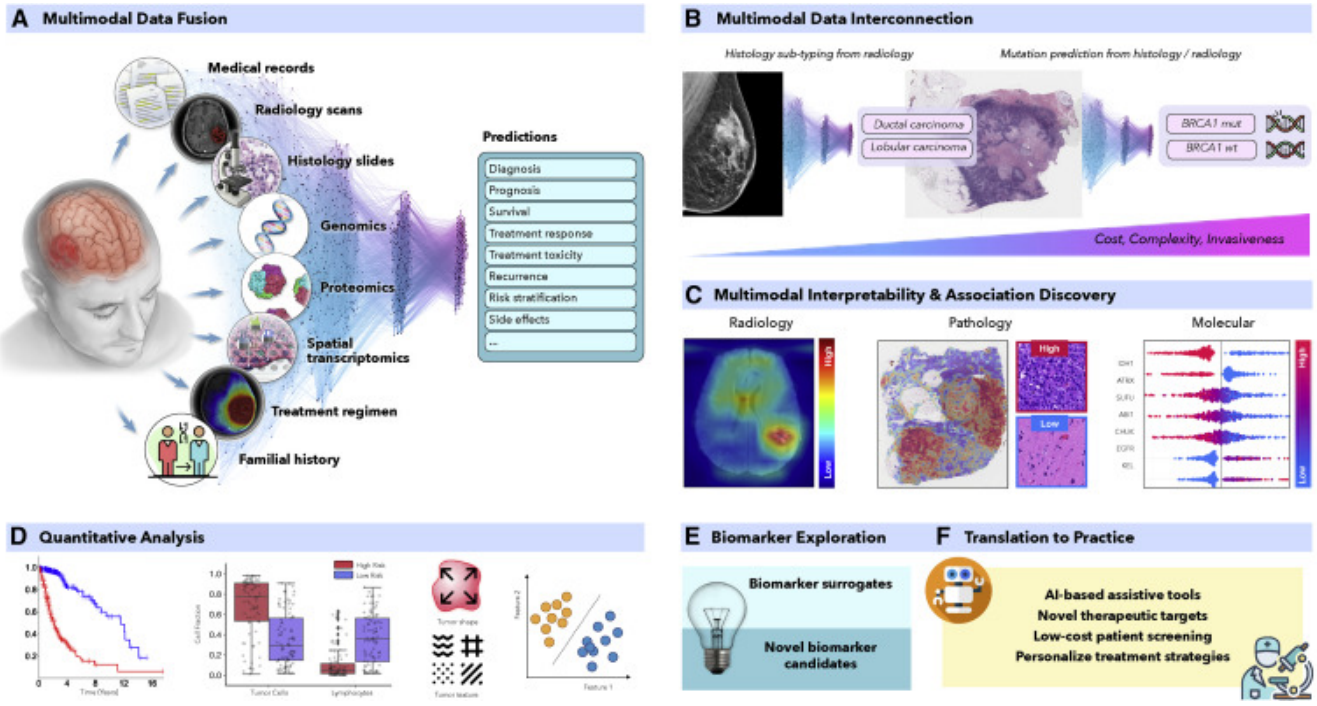


Figure 1.4: AI-driven multimodal data integration (A and C–F) (A) AI models can integrate complementary information and clinical context from diverse data sources to provide more accurate outcome predictions. The clinical insights identified by such models can be further elucidated through (C) interpretability methods and (D) quantitative analysis to guide and accelerate the discovery of new biomarkers or therapeutic targets (E and F). (B) AI can reveal novel multimodal interconnections, such as relations between certain mutations and changes in cellular morphology or associations between radiology findings and histology tumor subtypes or molecular features. Such associations can serve as non-invasive or cost-efficient alternatives to existing biomarkers to support large-scale patient screening (E and F). *Reprinted from [57] under the terms of the Creative Commons Attribution-NonCommercial-No Derivatives License (CC BY NC ND).*

1.4.3 AI tools in oncology clinical routine

Multiple clinical trials are being run to demonstrate the benefit of AI-based models for better cancer care. Table 1.1 (non-exhaustive list) presents a brief overview of ongoing clinical trials. These clinical trials cover the different stages of cancer treatment and aim to improve different tools. Some of them aim at improving patient screening for better cancer detection[58] (NCT04874064) while others aim at improving RT treatment planning tools[59] (NCT05653063). There are also clinical trials involved in assessing the power of radiomic models in predicting the efficacy of immunotherapy in patients with stage III non-small cell lung cancer[60] (NCT04984148), or prediction of survival in glioblastoma[61] (NCT02666066).

The evident success of AI in the medical field has been accompanied by the validation by competent authorities (CE marking, FDA approval) of various tools for routine clinical practice. Mostly, softwares to speed treatment planning in RT have been implemented. As an example, ART plan (Therapanacea, France), a software designed to assist the contouring process of the organs at risk (OARs) on 3D images of cancer patients for whom radiotherapy treatment has been planned, has been CE marked and FDA-/TGA cleared. It can also be used for image registration and generation of synthetic CT from MR images. Similarly, Mirada Medical^{Ltd} (United Kingdom) and LimbusAI (Canada) developed a tool based on AI for automation and optimization of the RT workflow, encompassing automatic OARs multi-modality contouring. DL Precise, a one-click software tool developed by DeepLook Medical (United States) for automatic segmentation and measurement of suspicious objects across all medical imaging modalities, has also been FDA-cleared. Ethos^{TM}, developed by Varian Medical Systems has also been deployed after FDA approval, for daily on-the-fly adaptive radiotherapy (DART).

1.4.4 Ethics in AI

In contrast to promising results in cancer care, AI poses ethical issues that have to be addressed. Safeguards must be clearly established to implement the responsible research paradigm into their current practice. The use of AI comes with certain risks such as data bias (incomplete data, data selection, effect of garbage in garbage out), data security and storage, training bias, violation of privacy, lack of transparency and interpretability, harm to patients, etc. Therefore the practice in clinical routine and medical research must be regulated and have to follow specific principles: the autonomy, the beneficence, and the nonmaleficence principles. The autonomy principle is usually expressed as the right of competent adults to make informed decisions about their own medical care. It underlies the right to refuse treatment, the right to participate in research or refuse it. To exercise this right would require Informed Consent. Under the beneficence principle, AI must be used to provide treatment in an ethical manner not only by respecting patients decisions and protecting them from harm but also by making efforts to secure their well-being. The principle of nonmaleficence assures that any AI model has an obligation not to inflict harm on patients. Ethics guidelines for trustworthy AI have been issued by the High-Level Expert Group on Artificial Intelligence (AI HLEG), an independent commission led by the European Union[62]. AI HLEG identified seven conditions:

- Human agency and oversight: an AI system should be aimed to supporting human agency and fundamental rights, rather than reducing, limiting, or undermining human independence.

Study Title	Type of Cancer	Hospitals
The Application Value of Artificial Intelligence in MRI Precision Diagnosis and Treatment of Bladder Cancer	Bladder Cancer	The first affiliated hospital of Nanjing Medical University Nanjing, Jiangsu, China
Artificial Intelligence in Breast Cancer Screening Programs in Córdoba (AITIC)	Breast Cancer	Hospital Universitario Reina Sofia Córdoba, Spain
TRIple Negative Breast Cancer Markers In Liquid Biopsies Using Artificial Intelligence	Breast Cancer	St. Joseph's Health Care London London, Ontario, Canada The Ottawa Hospital Cancer Center, Ottawa, Ontario, Canada Sunnybrook Health Sciences Centre Toronto, Ontario, Canada Jewish General Hospital, Montreal, Quebec, Canada
ARCHERY - Artificial Intelligence-Based Radiotherapy Treatment Planning for Cervical, Head and Neck and Prostate Cancer	Cervical, Head and Neck, and Prostate Cancer	University College, London, United Kingdom
Radiomics and Metabolomics in the Follow-up of CAR T-cells for Refractory or Relapsed Non-Hodgkin's Lymphoma	Refractory or Relapsed Non-Hodgkin's Lymphoma	CHU de Nice, Nice, France
Construction of CT Radiomics Model for Predicting the Efficacy of Immunotherapy in Patients With Stage III Non-small Cell Lung Cancer	Stage III Non-small Cell Lung Cancer	Guangdong Provincial People's Hospital Guangzhou, Guangdong, China
Radiomics for Prediction of Survival in Glioblastoma	Glioblastoma	Maastricht Radiation Oncology (MAASTRO clinic), Maastricht, Limburg, Netherlands
Radiomics-based Prediction Model of Tumor Spread Through Air Space in Lung Adenocarcinoma	Lung Adenocarcinoma	Departement of general surgery "Paride Stefanini", Roma, Italy
Evaluation of the Use of Radiomics in 18F-FDOPA PET Examinations for the Characterization of Gliomas	Gliomas	CHRU of Nancy, Vandœuvre-lès-Nancy, France
Can MRI of the Prostate Combined With a Radiomics Evaluation Determine the Invasive Capacity of a Tumour	Prostate Cancer	Victoria General Hospital, Halifax, Nova Scotia, Canada

Table 1.1: Brief overview of clinical trials on AI into clinical routine. An electronic search was performed on *https://www.clinicaltrials.gov/* with the keywords "Artificial Intelligence in cancer" or "radiomics in cancer".

- Technical robustness and safety: trustworthy AI requires algorithms that are sufficiently secure, reliable, and robust to deal with errors or inconsistencies in all phases of the AI system's work.
- Privacy and data governance: citizens should have full control over their personal data. This data must not be used to harm or disadvantage them.
- Transparency: emphasizes the AI system's traceability, explainability, and communication.
- Diversity, non-discrimination, and fairness: AI systems should consider all degrees of people's talents, skills, and needs, as well as guarantee user accessibility
- Societal and environmental well-being: AI systems should be utilized to reinforce positive social change and increase sustainability and environmental responsibility.
- Accountability: mechanisms should be put in place to ensure responsibility and accountability for AI systems and the results of their processes, including the opportunity to review and report negative consequences.

Furthermore, a concise set of principles for better scientific data management and stewardship has been defined, the "FAIR Guiding Principles"[63], stating that all research objects should be: i) Findable; ii) Accessible; iii) Interoperable; and iv) Reusable. Implementation of the FAIR principles within the radiomics field, in particular, can facilitate its faster clinical translation.

Several initiatives have been taken to guarantee "FAIR radiomics". To name a few, the Image biomarker standardization initiative (IBSI) (see Chapter 3) and the radiomics ontology provided guidelines for standardized practices. The radiomics ontology frequently updates a repository on the National Center for Biomedical Ontology BioPortal [2] in order to improve the interoperability of radiomics analyses via consistent tagging of radiomic features, segmentation algorithms, and imaging filters. This ontology could provide a standardized way of reporting on radiomic data and methods and would summarize the implementation details of a given radiomics workflow.

It is important to note that these guidelines apply not only to "data" in the conventional sense, but also to the algorithms, tools, and workflows that led to that data.

[2] bioportal.bioontology.org/ontologies/RO

Bibliography

[1] Simiao Chen et al. "Estimates and Projections of the Global Economic Cost of 29 Cancers in 204 Countries and Territories From 2020 to 2050". In: *JAMA Oncology* (Feb. 2023). ISSN: 2374-2437. DOI: 10.1001/jamaoncol.2022.7826. URL: https://doi.org/10.1001/jamaoncol.2022.7826 (visited on 04/05/2023).

[2] Robert J. Gillies, Paul E. Kinahan, and Hedvig Hricak. "Radiomics: Images Are More than Pictures, They Are Data". eng. In: *Radiology* 278.2 (Feb. 2016), pp. 563–577. ISSN: 1527-1315. DOI: 10.1148/radiol.2015151169.

[3] Marta Bogowicz et al. "Comparison of PET and CT radiomics for prediction of local tumor control in head and neck squamous cell carcinoma". eng. In: *Acta Oncol* 56.11 (Nov. 2017), pp. 1531–1536. ISSN: 1651-226X. DOI: 10.1080/0284186X.2017.1346382.

[4] Karen Buch et al. "Quantitative variations in texture analysis features dependent on MRI scanning parameters: A phantom model". In: *J Appl Clin Med Phys* 19.6 (Oct. 2018), pp. 253–264. ISSN: 1526-9914. DOI: 10.1002/acm2.12482. URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6236836/ (visited on 10/13/2022).

[5] Beomseok Sohn et al. "Machine Learning Based Radiomic HPV Phenotyping of Oropharyngeal SCC: A Feasibility Study Using MRI". en. In: *The Laryngoscope* 131.3 (2021). _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/lary.28889, E851–E856. ISSN: 1531-4995. DOI: 10.1002/lary.28889. URL: https://onlinelibrary.wiley.com/doi/abs/10.1002/lary.28889 (visited on 04/05/2023).

[6] Meiling Xiao et al. "Multiparametric MRI-Based Radiomics Nomogram for Predicting Lymph Node Metastasis in Early-Stage Cervical Cancer". en. In: *J Magn Reson Imaging* (Feb. 2020), jmri.27101. ISSN: 1053-1807, 1522-2586. DOI: 10.1002/jmri.27101. URL: https://onlinelibrary.wiley.com/doi/abs/10.1002/jmri.27101 (visited on 03/19/2020).

[7] Taotao Dong et al. "Development and Validation of a Deep Learning Radiomics Model Predicting Lymph Node Status in Operable Cervical Cancer". In: *Frontiers in Oncology* 10 (2020). ISSN: 2234-943X. URL: https://www.frontiersin.org/articles/10.3389/fonc.2020.00464 (visited on 01/10/2023).

[8] Xueming Xia et al. “Radiomics Based on Nomogram Predict Pelvic Lymphnode Metastasis in Early-Stage Cervical Cancer”. en. In: *Diagnostics* 12.10 (Oct. 2022). Number: 10 Publisher: Multidisciplinary Digital Publishing Institute, p. 2446. ISSN: 2075-4418. DOI: 10.3390/diagnostics12102446. URL: https://www.mdpi.com/2075-4418/12/10/2446 (visited on 01/05/2023).

[9] Reinhard R. Beichel et al. “FDG PET based prediction of response in head and neck cancer treatment: Assessment of new quantitative imaging features”. In: *PLoS One* 14.4 (Apr. 2019), e0215465. ISSN: 1932-6203. DOI: 10.1371/journal.pone.0215465. URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6474600/ (visited on 08/04/2022).

[10] Thomas Berger et al. “Predicting radiotherapy-induced xerostomia in head and neck cancer patients using day-to-day kinetics of radiomics features”. eng. In: *Phys Imaging Radiat Oncol* 24 (Oct. 2022), pp. 95–101. ISSN: 2405-6316. DOI: 10.1016/j.phro.2022.10.004.

[11] P. J. Brown et al. “Prediction of outcome in anal squamous cell carcinoma using radiomic feature analysis of pre-treatment FDG PET-CT”. eng. In: *Eur J Nucl Med Mol Imaging* 46.13 (Dec. 2019), pp. 2790–2799. ISSN: 1619-7089. DOI: 10.1007/s00259-019-04495-1.

[12] Nicolas Giraud et al. “MRI-Based Radiomics Input for Prediction of 2-Year Disease Recurrence in Anal Squamous Cell Carcinoma”. In: *Cancers (Basel)* 13.2 (Jan. 2021), p. 193. ISSN: 2072-6694. DOI: 10.3390/cancers13020193. URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7827348/ (visited on 04/22/2022).

[13] Sylvain Reuzé et al. “Prediction of cervical cancer recurrence using textural features extracted from 18 F-FDG PET images acquired with different scanners”. In: *Oncotarget* 8.26 (May 2017). Publisher: Impact Journals, pp. 43169–43179. ISSN: 1949-2553. DOI: 10.18632/oncotarget.17856. URL: https://www.oncotarget.com/article/17856/text/ (visited on 01/26/2021).

[14] R. Da-Ano et al. “Performance comparison of modified ComBat for harmonization of radiomic features for multicenter studies”. eng. In: *Sci Rep* 10.1 (June 2020), p. 10248. ISSN: 2045-2322. DOI: 10.1038/s41598-020-66110-w.

[15] Edward Walter and Mike Scott. “The life and work of Rudolf Virchow 1821–1902: “Cell theory, thrombosis and the sausage duel””. In: *J Intensive Care Soc* 18.3 (Aug. 2017), pp. 234–235. ISSN: 1751-1437. DOI: 10.1177/1751143716663967. URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5665122/ (visited on 12/28/2022).

[16] Yamagiwa K. “Experimentelle studie uber die pathogenese der epithelialgeschwulste”. In: *Mitt. Med. Fad. Tokio* 15 (1915), pp. 295–344. URL: https://cir.nii.ac.jp/crid/1571980075393572992 (visited on 12/28/2022).

[17] P. Rous. “A SARCOMA OF THE FOWL TRANSMISSIBLE BY AN AGENT SEPARABLE FROM THE TUMOR CELLS”. eng. In: *J Exp Med* 13.4 (Apr. 1911), pp. 397–411. ISSN: 0022-1007. DOI: 10.1084/jem.13.4.397.

[18] Franz Hermann Müller. “Tabakmißbrauch und Lungencarcinom”. de. In: *Z Krebs-forsch* 49.1 (Jan. 1940), pp. 57–85. ISSN: 1432-1335. DOI: 10.1007/BF01633114. URL: https://doi.org/10.1007/BF01633114 (visited on 12/28/2022).

[19] Lutz Gissmann and Harald Zur Hausen. “Partial characterization of viral DNA from human genital warts (condylomata acuminata)”. en. In: *International Journal of Cancer* 25.5 (1980). _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/ijc.2910250509, pp. 605–609. ISSN: 1097-0215. DOI: 10.1002/ijc.2910250509. URL: http://onlinelibrary.wiley.com/doi/abs/10.1002/ijc.2910250509 (visited on 12/28/2022).

[20] J. D. Watson and F. H. C. Crick. “Molecular Structure of Nucleic Acids: A Structure for Deoxyribose Nucleic Acid”. en. In: *Nature* 171.4356 (Apr. 1953). Number: 4356 Publisher: Nature Publishing Group, pp. 737–738. ISSN: 1476-4687. DOI: 10.1038/171737a0. URL: http://www.nature.com/articles/171737a0 (visited on 12/28/2022).

[21] Douglas Hanahan. “Hallmarks of Cancer: New Dimensions”. en. In: *Cancer Discov* 12.1 (Jan. 2022). Publisher: American Association for Cancer Research Section: Review, pp. 31–46. ISSN: 2159-8274, 2159-8290. DOI: 10.1158/2159-8290.CD-21-1059. URL: https://cancerdiscovery.aacrjournals.org/content/12/1/31 (visited on 01/13/2022).

[22] Wan-yee Tang and Shuk-mei Ho. “Epigenetic reprogramming and imprinting in origins of disease”. In: *Rev Endocr Metab Disord* 8.2 (June 2007), pp. 173–182. ISSN: 1389-9155. DOI: 10.1007/s11154-007-9042-4. URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4056338/ (visited on 01/06/2023).

[23] Douglas V. Faget, Qihao Ren, and Sheila A. Stewart. “Unmasking senescence: context-dependent effects of SASP in cancer”. eng. In: *Nat Rev Cancer* 19.8 (Aug. 2019), pp. 439–453. ISSN: 1474-1768. DOI: 10.1038/s41568-019-0156-2.

[24] Shenghui He and Norman E. Sharpless. “Senescence in Health and Disease”. eng. In: *Cell* 169.6 (June 2017), pp. 1000–1011. ISSN: 1097-4172. DOI: 10.1016/j.cell.2017.05.015.

[25] Douglas Hanahan and Robert A Weinberg. “The Hallmarks of Cancer”. en. In: *Cell* 100.1 (Jan. 2000), pp. 57–70. ISSN: 0092-8674. DOI: 10.1016/S0092-8674(00)81683-9. URL: https://www.sciencedirect.com/science/article/pii/S0092867400816839 (visited on 03/01/2023).

[26] Douglas Hanahan and Robert A. Weinberg. “Hallmarks of cancer: the next generation”. eng. In: *Cell* 144.5 (Mar. 2011), pp. 646–674. ISSN: 1097-4172. DOI: 10.1016/j.cell.2011.02.013.

[27] Manuel Arruebo et al. “Assessment of the Evolution of Cancer Treatment Therapies”. In: *Cancers (Basel)* 3.3 (Aug. 2011), pp. 3279–3330. ISSN: 2072-6694. DOI: 10.3390/cancers3033279. URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3759197/ (visited on 01/05/2023).

[28] Eric Deutsch et al. “Optimising efficacy and reducing toxicity of anticancer radioimmunotherapy”. eng. In: *Lancet Oncol* 20.8 (Aug. 2019), e452–e463. ISSN: 1474-5488. DOI: 10.1016/S1470-2045(19)30171-8.

[29] Sandra Demaria et al. “Ionizing radiation inhibition of distant untreated tumors (abscopal effect) is immune mediated”. en. In: *International Journal of Radiation Oncology*Biology*Physics* 58.3 (Mar. 2004), pp. 862–870. ISSN: 0360-3016. DOI: 10.1016/j.ijrobp.2003.09.012. URL: https://www.sciencedirect.com/science/article/pii/S0360301603019382 (visited on 03/01/2023).

[30] Philip P. Connell and Samuel Hellman. “Advances in Radiotherapy and Implications for the Next Century: A Historical Perspective”. In: *Cancer Research* 69.2 (Jan. 2009), pp. 383–392. ISSN: 0008-5472. DOI: 10.1158/0008-5472.CAN-07-6871. URL: https://doi.org/10.1158/0008-5472.CAN-07-6871 (visited on 03/02/2023).

[31] Romuald Gajewski and Gary Arthur. “Intensity-modulated radiation therapy: the state of the art Edited by Jatinder R. Palta and T. Rockwell Mackie. AAPM Medical Physics Monograph No.29.” en. In: *Australas. Phys. Eng. Sci. Med.* 26.3 (Sept. 2003), pp. 145–146. ISSN: 1879-5447. DOI: 10.1007/BF03178785. URL: https://doi.org/10.1007/BF03178785 (visited on 03/02/2023).

[32] Vincent Favaudon et al. “Ultrahigh dose-rate FLASH irradiation increases the differential response between normal and tumor tissue in mice”. eng. In: *Sci Transl Med* 6.245 (July 2014), 245ra93. ISSN: 1946-6242. DOI: 10.1126/scitranslmed.3008973.

[33] Marie-Catherine Vozenin et al. “FLASH radiotherapy International Workshop”. en. In: *Radiotherapy and Oncology*. FLASH radiotherapy International Workshop 139 (Oct. 2019), pp. 1–3. ISSN: 0167-8140. DOI: 10.1016/j.radonc.2019.07.020. URL: https://www.sciencedirect.com/science/article/pii/S0167814019330130 (visited on 03/02/2023).

[34] Catherine T. Lee et al. “Treatment planning with protons for pediatric retinoblastoma, medulloblastoma, and pelvic sarcoma: how do protons compare with other conformal techniques?” eng. In: *Int J Radiat Oncol Biol Phys* 63.2 (Oct. 2005), pp. 362–372. ISSN: 0360-3016. DOI: 10.1016/j.ijrobp.2005.01.060.

[35] Agustinus J. A. J. van de Schoot et al. “Dosimetric advantages of proton therapy compared with photon therapy using an adaptive strategy in cervical cancer”. eng. In: *Acta Oncol* 55.7 (July 2016), pp. 892–899. ISSN: 1651-226X. DOI: 10.3109/0284186X.2016.1139179.

[36] C. Chargari, J. -C. Soria, and E. Deutsch. “Controversies and challenges regarding the impact of radiation therapy on survival”. en. In: *Annals of Oncology* 24.1 (Jan. 2013), pp. 38–46. ISSN: 0923-7534. DOI: 10.1093/annonc/mds217. URL: https://www.sciencedirect.com/science/article/pii/S0923753419370309 (visited on 03/01/2023).

[37] María E. Rodríguez-Ruiz et al. "Immunological Mechanisms Responsible for Radiation-Induced Abscopal Effect". en. In: *Trends in Immunology* 39.8 (Aug. 2018), pp. 644–655. ISSN: 1471-4906. DOI: 10.1016/j.it.2018.06.001. URL: https://www.sciencedirect.com/science/article/pii/S1471490618301091 (visited on 04/20/2023).

[38] Scott J. Antonia et al. "Overall Survival with Durvalumab after Chemoradiotherapy in Stage III NSCLC". In: *New England Journal of Medicine* 379.24 (Dec. 2018). Publisher: Massachusetts Medical Society _eprint: https://doi.org/10.1056/NEJMoa1809697, pp. 2342–2350. ISSN: 0028-4793. DOI: 10.1056/NEJMoa1809697. URL: https://doi.org/10.1056/NEJMoa1809697 (visited on 12/13/2022).

[39] Josephine Kang, Sandra Demaria, and Silvia Formenti. "Current clinical trials testing the combination of immunotherapy with radiotherapy". en. In: *J Immunother Cancer* 4.1 (Dec. 2016). Publisher: BMJ Specialist Journals Section: Clinical Trials Monitor, p. 51. ISSN: 2051-1426. DOI: 10.1186/s40425-016-0156-7. URL: https://jitc.bmj.com/content/4/1/51 (visited on 04/19/2023).

[40] Gustave Roussy, Cancer Campus, Grand Paris. *Randomized Phase II Trial Assessing the Inhibitor of Programmed Cell Death Ligand 1 (PD-L1) Immune Checkpoint Atezolizumab in Locally Advanced Cervical Cancer*. Clinical trial registration NCT03612791. submitted: July 20, 2018. clinicaltrials.gov, Feb. 2022. URL: https://clinicaltrials.gov/ct2/show/NCT03612791 (visited on 03/21/2023).

[41] "Functional Imaging". en. In: *Encyclopedia of Pain*. Ed. by Robert F. Schmidt and William D. Willis. Berlin, Heidelberg: Springer, 2007, pp. 806–806. ISBN: 978-3-540-29805-2. DOI: 10.1007/978-3-540-29805-2_1571. URL: https://doi.org/10.1007/978-3-540-29805-2_1571 (visited on 03/02/2023).

[42] Drew A. Torigian et al. "Functional Imaging of Cancer with Emphasis on Molecular Techniques". en. In: *CA: A Cancer Journal for Clinicians* 57.4 (2007). _eprint: https://onlinelibrary.wiley.co pp. 206–224. ISSN: 1542-4863. DOI: 10.3322/canjclin.57.4.206. URL: https://onlinelibrary.wiley.com/doi/abs/10.3322/canjclin.57.4.206 (visited on 03/02/2023).

[43] John E. Niederhuber et al. *Abeloff's Clinical Oncology*. Elsevier, Jan. 2019. DOI: 10.1016/B978-0-323-47674-4.00124-9. URL: http://www.scopus.com/inward/record.url?scp=85123663709&partnerID=8YFLogxK (visited on 03/02/2023).

[44] Uwe Oelfke et al. "Linac-integrated kV-cone beam CT: Technical features and first applications". en. In: *Medical Dosimetry*. Image-Guided Radiation Therapy, Part 1: Target Delineation and Localization 31.1 (Mar. 2006), pp. 62–70. ISSN: 0958-3947. DOI: 10.1016/j.meddos.2005.12.008. URL: https://www.sciencedirect.com/science/article/pii/S0958394705001950 (visited on 03/02/2023).

[45] Kavitha Srinivasan, Mohammad Mohammadi, and Justin Shepherd. "Applications of linac-mounted kilovoltage Cone-beam Computed Tomography in modern radiation therapy: A review". In: *Pol J Radiol* 79 (July 2014), pp. 181–193. ISSN: 1733-134X. DOI: 10.12659/PJR.890745. URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4085117/ (visited on 03/02/2023).

[46] Sebastian Klüter. "Technical design and concept of a 0.35 T MR-Linac". en. In: *Clinical and Translational Radiation Oncology* 18 (Sept. 2019), pp. 98–101. ISSN: 2405-6308. DOI: 10.1016/j.ctro.2019.04.007. URL: https://www.sciencedirect.com/science/article/pii/S2405630819300692 (visited on 03/02/2023).

[47] Olga L. Green et al. "First clinical implementation of real-time, real anatomy tracking and radiation beam control". en. In: *Medical Physics* 45.8 (2018). _eprint: https://onlinelibrary.wiley.com/d pp. 3728–3740. ISSN: 2473-4209. DOI: 10.1002/mp.13002. URL: https://onlinelibrary.wiley.com/doi/abs/10.1002/mp.13002 (visited on 03/02/2023).

[48] Ziad Obermeyer and Ezekiel J. Emanuel. "Predicting the Future — Big Data, Machine Learning, and Clinical Medicine". In: *N Engl J Med* 375.13 (Sept. 2016). Publisher: Massachusetts Medical Society, pp. 1216–1219. ISSN: 0028-4793. DOI: 10.1056/NEJMp1606181. URL: http://www.nejm.org/doi/10.1056/NEJMp1606181 (visited on 12/28/2022).

[49] John McCarthy et al. "A Proposal for the Dartmouth Summer Research Project on Artificial Intelligence, August 31, 1955". en. In: *AI Magazine* 27.4 (Dec. 2006). Number: 4, pp. 12–12. ISSN: 2371-9621. DOI: 10.1609/aimag.v27i4.1904. URL: https://ojs.aaai.org/aimagazine/index.php/aimagazine/article/view/1904 (visited on 03/07/2023).

[50] Kaustav Bera et al. "Artificial intelligence in digital pathology - new tools for diagnosis and precision oncology". eng. In: *Nat Rev Clin Oncol* 16.11 (Nov. 2019), pp. 703–715. ISSN: 1759-4782. DOI: 10.1038/s41571-019-0252-y.

[51] E.A. Weiss. "Biographies: Eloge: Arthur Lee Samuel (1901-90)". In: *IEEE Annals of the History of Computing* 14.3 (1992). Conference Name: IEEE Annals of the History of Computing, pp. 55–69. ISSN: 1934-1547. DOI: 10.1109/85.150082.

[52] Iqbal H. Sarker. "Machine Learning: Algorithms, Real-World Applications and Research Directions". en. In: *SN COMPUT. SCI.* 2.3 (Mar. 2021), p. 160. ISSN: 2661-8907. DOI: 10.1007/s42979-021-00592-x. URL: https://doi.org/10.1007/s42979-021-00592-x (visited on 03/03/2023).

[53] National Research Council (US) Committee on A Framework for Developing a New Taxonomy of Disease. *Toward Precision Medicine: Building a Knowledge Network for Biomedical Research and a New Taxonomy of Disease*. eng. The National Academies Collection: Reports funded by National Institutes of Health. Washington (DC): National Academies Press (US), 2011. ISBN: 978-0-309-22222-8. URL: http://www.ncbi.nlm.nih.gov/books/NBK91503/ (visited on 01/03/2023).

[54] Daniel S. Chen and Ira Mellman. "Elements of cancer immunity and the cancer–immune set point". en. In: *Nature* 541.7637 (Jan. 2017). Number: 7637 Publisher: Nature Publishing Group, pp. 321–330. ISSN: 1476-4687. DOI: 10.1038/nature21349. URL: http://www.nature.com/articles/nature21349 (visited on 12/13/2022).

[55] European Society of Radiology (ESR). "Medical imaging in personalised medicine: a white paper of the research committee of the European Society of Radiology (ESR)". eng. In: *Insights Imaging* 6.2 (Apr. 2015), pp. 141–155. ISSN: 1869-4101. DOI: 10.1007/s13244-015-0394-0.

[56] Philippe Lambin et al. "Radiomics: the bridge between medical imaging and personalized medicine". en. In: *Nat Rev Clin Oncol* 14.12 (Dec. 2017). Bandiera_abtest: a Cg_type: Nature Research Journals Number: 12 Primary_atype: Reviews Publisher: Nature Publishing Group Subject_term: Cancer genomics;Cancer imaging;Medical imaging;Personalized medicine Subject_term_id: cancer-genomics;cancer-imaging;medical-imaging;personalized-medicine, pp. 749–762. ISSN: 1759-4782. DOI: 10.1038/nrclinonc.2017.141. URL: https://www.nature.com/articles/nrclinonc.2017.141 (visited on 09/20/2021).

[57] Jana Lipkova et al. "Artificial intelligence for multimodal data integration in oncology". en. In: *Cancer Cell* 40.10 (Oct. 2022), pp. 1095–1110. ISSN: 1535-6108. DOI: 10.1016/j.ccell.2022.09.012. URL: https://www.sciencedirect.com/science/article/pii/S153561082200441X (visited on 12/07/2022).

[58] Mark Basik. *Triple Negative Breast Cancer Markers in Liquid Biopsies Using Artificial Intelligence (TRICIA Study)*. Clinical trial registration NCT04874064. submitted: April 30, 2021. clinicaltrials.gov, May 2021. URL: https://clinicaltrials.gov/ct2/show/NCT04874064 (visited on 04/19/2023).

[59] University College, London. *ARCHERY - Artificial Intelligence Based Radiotherapy Treatment Planning for Cervical, Head and Neck and Prostate Cancer*. Clinical trial registration NCT05653063. submitted: November 3, 2022. clinicaltrials.gov, Dec. 2022. URL: https://clinicaltrials.gov/ct2/show/NCT05653063 (visited on 04/19/2023).

[60] YI PAN. *Construction of CT Radiomics Model to Assess PD-L1 Status and Predict the Efficacy of Chemoradiotherapy Combined With Immunotherapy in Unresectable Locally Advanced Non-small Cell Lung Cancer*. Clinical trial registration NCT04984148. submitted: July 25, 2021. clinicaltrials.gov, Nov. 2021. URL: https://clinicaltrials.gov/ct2/show/NCT04984148 (visited on 04/19/2023).

[61] Maastricht Radiation Oncology. *Radiomics for the Prediction of Survival in GBM After Radiotherapy With/Without Temozolomide*. Clinical trial registration NCT02666066. submitted: January 24, 2016. clinicaltrials.gov, Aug. 2018. URL: https://clinicaltrials.gov/ct2/show/NCT02666066 (visited on 04/19/2023).

[62] Nathalie A. Smuha. "The EU Approach to Ethics Guidelines for Trustworthy Artificial Intelligence". en. In: *Computer Law Review International* 20.4 (Aug. 2019). Publisher: Verlag Dr. Otto Schmidt, pp. 97–106. ISSN: 2194-4164. DOI: 10.9785/cri-2019-200402. URL: https://www.degruyter.com/document/doi/10.9785/cri-2019-200402/html (visited on 03/23/2023).

[63] Mark D. Wilkinson et al. "The FAIR Guiding Principles for scientific data management and stewardship". en. In: *Sci Data* 3.1 (Mar. 2016). Number: 1 Publisher: Nature Publishing Group, p. 160018. ISSN: 2052-4463. DOI: 10.1038/sdata.2016.18. URL: https://www.nature.com/articles/sdata201618 (visited on 03/24/2023).

Chapter 2

Human papillomavirus in Oncology

2.1 Overview

Papillomaviruses (PVs) are a large and diverse family of double-stranded circular DNA viruses that by and large exhibit rigid species and tissue tropism[1] containing eight genes. The L1 gene, one of these eight genes, encodes the principal capsid protein and is used for classification and construction of phylogenetic trees, as it can be found in the genome of all known PVs. These viruses are traditionally referred to as types, a type being a cloned full-length PV genome, whose L1 nucleotide sequence is at least 10% dissimilar from that of any other PV types[2]. The classification in types is therefore based on genome sequence homology, biological function, and pathological effect. The consensus within the community of papillomavirus researchers established that the name of an animal PV should be based on the scientific name of the host, using the host genus and species designation. Types of these viruses that infect humans are called human papillomaviruses (HPV). HPVs are one of the most common sexually transmitted infections. There are more than 220 known types of HPV[2, 3]. HPV, and PVs in general, are known to infect the vertebrate epithelia through the oral cavity or upper respiratory tract, the anogenital tract, or the skin. The infection can lead to a large range of pathologies, from warts (papillomas) to dysplasia and cancer but can also persist asymptomatically. HPV are classified into two groups, low and high risk, and only a few are at high risk of leading to cancer. Low-risk HPV infections have low chances of leading to tumor growth but can cause warts on or around the genitals, anus, mouth, or throat. High-risk HPVs can cause several types of cancer. There are about 14 high-risk HPV (hrHPV) types namely HPV 16, 18, 31, 33, 35, 39, 45, 51, 52, 56, 58, 59, 66, and 68[4]. Two of these, HPV16 and HPV18, are responsible for most HPV-related cancers namely cervical cancer, oropharyngeal cancer, anal cancer, penile cancer, vaginal cancer, and vulvar cancer. Since the squamous cells on the inner surfaces of these organs are mostly infected, most HPV-related cancers represent a type of cancer called squamous cell carcinoma. Other cancers are due to infection in the glandular regions of epithelial tissue and are called adenocarcinomas. Today 5% of all cancers are considered HPV-related[5] and HPV is the most common pathogen responsible for female cancers[6], with cervical cancer being one of the most common cancers and a leading cause of death especially in low- and middle-income countries. In this chapter, we will focus only on the types of HPV that are known to cause cancer.

2.2 HPV induced cancers

2.2.1 Cervical cancer

Cervical cancer remains the fourth most common cancer worldwide among females after breast, colorectal, and lung cancer with approximately 604 000 new cases and 342 000 deaths annually[7]. The cervix is the lowermost part of the uterus and is a cylindrical structure composed of stroma and epithelium. The ectocervix, which projects into the vagina, is lined by squamous epithelium. The endocervical canal, which extends from the opening into the uterus (internal os) to the opening into the vagina (external os), is lined by columnar epithelium. Almost all cases of cervical carcinoma originate from the ecto- or endocervical mucosa in the transformation zone, the area of the cervix between the old and new squamocolumnar junction[8]. Cervical cancers are considered to be related to HPV infections in 99% of cases[9]. Most women will encounter at least one hrHPV infection but only one-tenth of all infections could lead to the development of precancerous lesions. Cervical cancer occurs as a long-term outcome of these persistent infections of the lowermost part of the uterus by one of the hrHPV types. Of the estimated 604 000 new cervical cancer cases annually worldwide, HPV 16 and HPV 18 account for 71% of cases; while HPV types 31, 33, 45, 52, and 58 account for another 19% of cervical cancer cases[10]. The United Nations 2030 Agenda with The World Health Organization (WHO) have called for a global initiative for the elimination of cervical cancer as a public health problem by implementing the following 90%–70%–90% triple pillar intervention strategy before the year 2030[11]:

- 90% of girls fully vaccinated with two doses of HPV vaccine by the age of 15 years;
- 70% of women screened using a high-performance screening test at the age of 35 and 45 years;
- 90% of women detected with cervical lesions to receive treatment and care.

Studies have reported evidence of effectiveness of HPV vaccination (a single dose can prevent hrHPV infection as well as three or two doses) on significantly reducing the risk of invasive cervical cancer[12, 13]. It has been estimated that worldwide HPV vaccination with high coverage could prevent about 8.7 million cases by 2094[14].

Histopatology

The diagnosis of cervical cancer is made by histological analysis on biopsy. HPV dependency is mostly determined using the $P16^{INK4a}$ immunohistochemistry (IHC) by looking for a diffuse overexpression of the p16 molecule, which is a test for the presence of potentially transforming HPV infection. Cervical cancers are almost totally caused by HPV infection in epithelial tissue with a majority of squamous carcinomas (80-90%) and adenocarcinomas (10-20%). Macroscopically, carcinomas appear as irregular, ulcerated, exophytic masses. There are several histological variants but the most common histological form is non-keratinized squamous cell carcinoma. These different histological forms do not alter the therapeutic management or prognosis. Non-HPV squamous cell carcinomas have similar morphologic features to HPV-related carcinomas with keratinizing histology. Adenocarcinomas are a more heterogeneous category. Macroscopically, they also present as an exophytic,

polyoid, and irregular mass. Purely infiltrative forms, which are more common, are more difficult to diagnose. HPV-independent adenocarcinomas account for approximately 20% of cervical adenocarcinomas. We also distinguish adenosquamous carcinomas which are tumors with both squamous and glandular differentiation. Other histological types of cervical cancer are much rare. Basal adenoid carcinoma is a tumor with an excellent prognosis after surgical excision without metastatic spread. It should not be confused with squamous cell carcinoma in a basaloid variant. Neuroendocrine carcinomas of the cervix (small cell carcinoma and large cell carcinoma) are very aggressive tumors. They are HPV-induced and overexpress p16. Their morphological characteristics are identical to those of endocrine tumors of other organs.

Staging and treatment of cervical cancer

Staging in cervical cancer is based on tumor size, lymph node involvement, and the presence of metastases. According to the FIGO (Fédération Internationale de Gynécologie et d'Obstétrique) classification of cervical cancer stage[15] and the TNM (Tumor-node-metastasis) classification of malignant tumours[16], the different stages are presented in Table 2.1.

Treatment of cervical cancer is mainly surgery or chemoradiation depending on cancer stage as seen in Figure 2.1. Surgery is suitable for early stages (IB1, IB2, IIA1), where cervical conization (IA1), simple hysterectomy, or radical hysterectomy (IA2) may be selected according to the stage of disease[8]. In FIGO 2018 stages IB2 and IIA1, surgery or radiation therapy may be chosen as the primary treatment on the basis of other patient factors and local resources, as both have similar outcomes. If not operable, patients are treated with chemoradiotherapy (CRT) (weekly intravenous cisplatin 40 mg/m^2, 5–6 cycles, 1 day per cycle, plus 45–50 Gy external-beam radiotherapy delivered in 1.8–2 Gy per fraction) $\pm$ simultaneously integrated boost to metastatic lymph nodes, followed by brachytherapy (BT) based on recommendations published by the Groupe Européen de Curiethérapie - European Society for Radiation Oncology (GEC-ESTRO)[17]. More detailed information on cancer staging and radiation management can be found in [8].

Follow-up should be performed every 3-4 months in the first 2 years and every 6-12 months in years 3-5. More than 75% of recurrences occur within 2-3 years in LACC [19], therefore, annual pelvic examinations should be rescheduled after 5 years of recurrence-free follow-up.

2.2.2 Anal cancer

The anal canal is a tube at the end of the rectum that measures about 4 cm in length. Anatomically, the anal canal extends from the dentate line to the anal verge. The dentate line demarcates a transition from a glandular mucosa to the squamous epithelium and is located 2–3 cm proximal to the anal verge. Functionally, the accepted boundaries of the anal canal span from the proximal aspect of the internal anal sphincter (IAS) to the anal verge, with a length of approximately 3–6 cm[20]. An estimated 29 000 persons, predominantly women, are diagnosed with anal cancers every year, for which HPV infection is considered the main cause[21, 22]. Anal cancer is a rare disease that accounts for $< 1\%$ and $< 3\%$ of all new cancer diagnoses and gastrointestinal tumors, respectively[23]. Two different categories of tumors can be found in the anal canal region. We distinguish tumors that develop from the mucosa, called the anal canal tumors, and tumors that arise within the skin at or

Stage grouping			
FIGO	TNM		
Stage 0	Tis	N0	M0
Stage IA	T1a	N0	M0
Stage IA1	T1a1	N0	M0
Stage IA2	T1a2	N0	M0
Stage IB	T1b	N0	M0
Stage IB1	T1b1	N0	M0
Stage IB2	T1b2	N0	M0
Stage IIA	T2a	N0	M0
Stage IIB	T2b	N0	M0
Stage IIIA	T3a	N0	M0
Stage IIIB	T1, T2, T3a, N1	M0	
	T3b	Any N	M0
Stage IVA	T4	Any N	M0
Stage IVB	Any T	Any N	M1

Table 2.1: LACC TNM classification. Tis: Tumor in situ, T1: tumor confined to the cervix, T1a: stromal invasion with a maximal depth of 5 mm, T1a1: Measured depth of stromal invasion 3 mm or less in depth, T1a2: measured depth of stromal invasion more than 3 mm and not more than 5 mm, T1b: Lesion confined to the cervix with depth of invasion greater than 5mm, T1b1: tumor of 2 cm or less in greatest dimension, T1b2: lesion more than 2 cm in greatest dimension but no more than 4cm in greatest dimension. T2: tumor of 4 cm but not >5 cm in greatest dimension, T3: tumor of >5 cm in greatest dimension, and T4: tumor of any size invades adjacent organ(s). N refers to regional lymph nodes. NX: Regional lymph nodes cannot be assessed, N0: No regional lymph node metastasis, N1: Regional lymph node metastasis. The regional lymph nodes are the paracervical, parametrial, hypogastric (internal ilial, obturator), common and external iliac, presacral, and lateral sacral nodes. M refers to distant, metastasis. MX: Distant metastasis cannot be assessed, M0: No distant metastasis, M1: Distant metastasis.

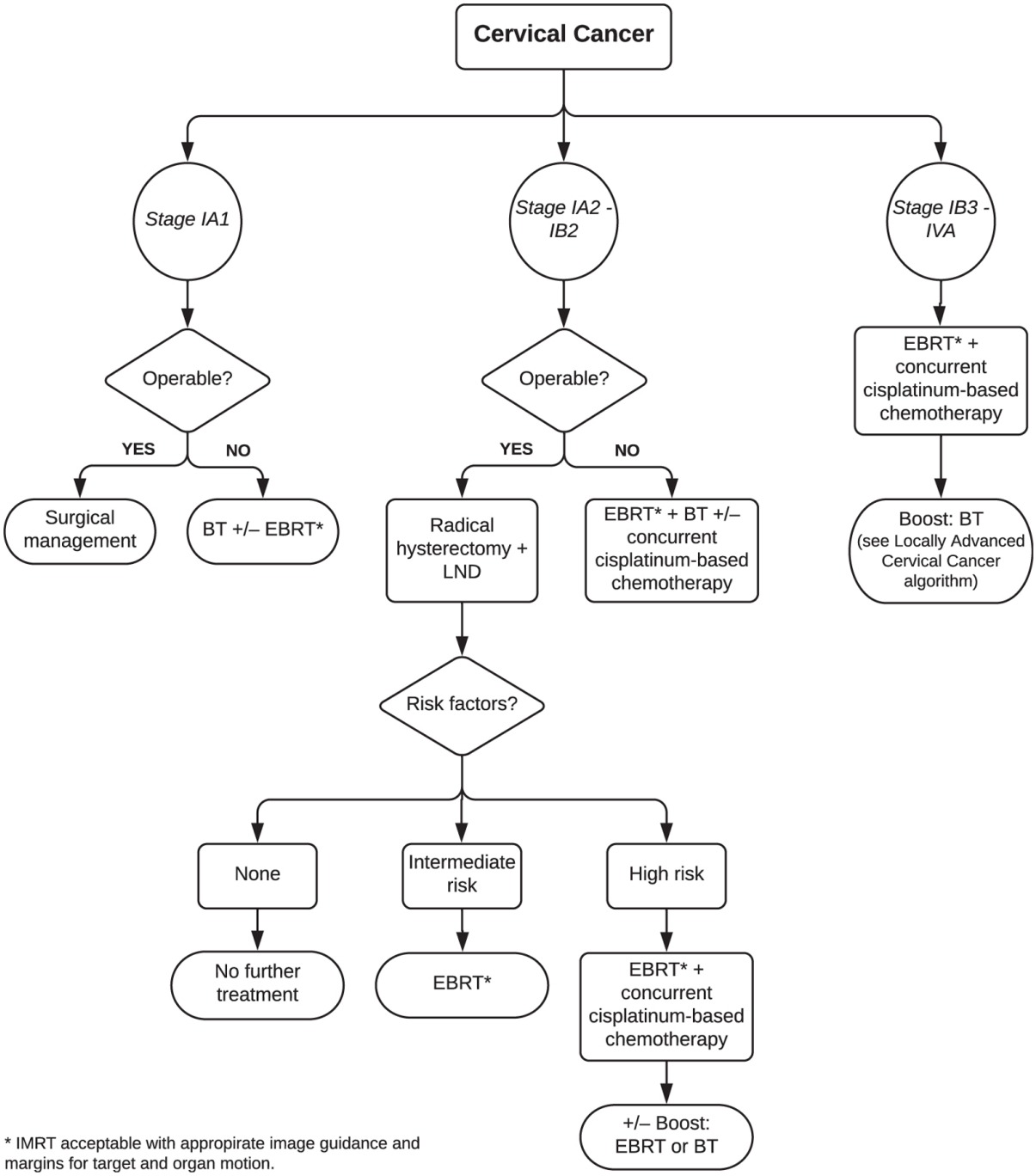


Figure 2.1: Overview of the management and treatment of cervical cancer based on stage of the disease. BT = brachytherapy; EBRT = external beam radiation therapy; IMRT = intensity modulated radiation therapy; LND = lymph node dissection; RT = radiation therapy. *Reprinted from [18] under the Creative Commons Attribution-NonCommercial 4.0 International.*

distal to the squamous mucocutaneous junction, termed as perianal cancers[24]. Multiple risk factors have been identified such as HPV infection, anoreceptive intercourse, cigarette smoking, and immunosuppression. HIV infection is also associated with anal cancer; there is a higher incidence in HIV-positive patients but the direct relationship between HIV and anal cancer has been difficult to separate from the prevalence of HPV in this population [25]. At least 23 HPV subtypes have been shown to infect the anogenital cells and similar to cervical cancer, HPV subtypes 16 and 18, are mostly attributable to progression of premalignant anal intraepithelial lesions into invasive anal canal cancer.

Histopathology

The diagnosis of anal cancer is made by biopsy-proven histology analysis. Similarly to LACC, anal cancer induced by HPV leads to over-expression of p16 protein and is diagnosed through the $P16^{INK4a}$ IHC. The workflow for anal cancer diagnosis is shown in Figure 2.2. Almost all anal cancers are squamous cell carcinoma (SCC). HPV infections are considered to be the primary cause of SCC being associated with 88% of anal cancers (AC)[26]. SCC can display various patterns which can lead to interobserver variability in diagnosis. Histological sub-types of basaloid, transitional, spheroidal, and cloacogenic cell cancers have no impact on management and are all included under the same heading of SCC[27]. The other histological types, although rare, are adenocarcinoma, malignant melanoma, gastrointestinal stromal tumors, poorly differentiated neuroendocrine tumors, and lymphoma. A variety of other tumor types may arise less frequently in the anal canal and make recognition more difficult. These histopathological types other than SCC are frequently diagnosed incidentally and often at an advanced stage, rendering poor overall prognoses. Therefore, histological confirmation is mandatory. Adenocarcinoma of the anal canal is a rare entity, accounting for less than 5%-10% of all anal cancers, and develops from the glandular cells that produce mucus in the anal canal. HPV16 is the type most commonly detected in SCC, whereas HPV18 predominates in adenocarcinomas.

Screening programs using anal cytology and high-resolution anoscopy have been proposed for high-risk populations (Gay, Bisexual, and HIV-negative women with a history of anal intercourse or other HPV-related anogenital malignancies) based on achievements obtained in cervical cytology screening. However, no randomized controlled study has yet demonstrated a preventive effect of screening in these high-risk populations and thus it cannot be routinely advocated at present[23].

Staging and treatment of anal cancer

Staging in anal cancer follows the recommendations of the Union for International Cancer Control (UICC) TNM 8th edition staging system classification of malignant tumors which are based on clinical assessment of the diameter of the tumor in anal cancer. Classification into different stages is detailed in Table 2.2.

Similar to cervical cancer, stage I is treated with local excision surgery followed by low-dose CRT for patients with histological margins of $\leq 1mm$. While there is no definitive consensus on treatment, recommendations have been made to treat anal cancer patients of localized stage I-III by CRT with a dose of at least >50 Gy to the clinical target volume (CTV) plus concomitant fluorouracil (5-FU) plus mitomycin C (MMC) during the first and the fifth week of RT, followed by surgery if possible. Capecitabine can be possibly used as an alternative to 5-FU in combination with MMC and RT.

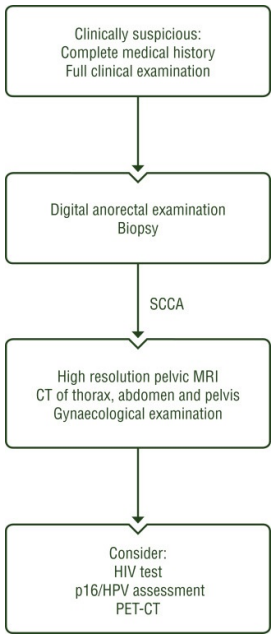


Figure 2.2: Diagnostic algorithm for anal cancer. CT: computed tomography; HIV: human immunodeficiency virus; HPV: human papillomavirus; MRI: magnetic resonance imaging; PET: positron emission tomography; SCCA: squamous-cell carcinoma of the anus. *Reprinted from [23] under the Creative Commons Attribution-NonCommercial 4.0 International.*

Stage grouping			
Stage I	T1	N0	M0
Stage IIA	T2	N0	M0
Stage IIB	T3	N0	M0
Stage IIIA	T1, T2	N1	M0
Stage IIIB	T4	N0	M0
Stage IIIC	T3, T4	N1	M0
Stage IV	Any T	Any N	M1

Table 2.2: Anal cancer TNM classification. T1: tumor of 2 cm or less in greatest dimension, T2: tumor of >2 cm but not >5 cm in greatest dimension, T3: tumor of >5 cm in greatest dimension, and T4: tumor of any size invades adjacent organ(s). N refers to regional lymph nodes. NX: Regional lymph nodes cannot be assessed, N0: No regional lymph node metastasis, N1: Regional lymph node metastasis. The regional lymph nodes are the perirectal, internal iliac, inguinal, and external iliac lymph nodes. M refers to distant, metastasis. MX: Distant metastasis cannot be assessed, M0: No distant metastasis, M1: Distant metastasis

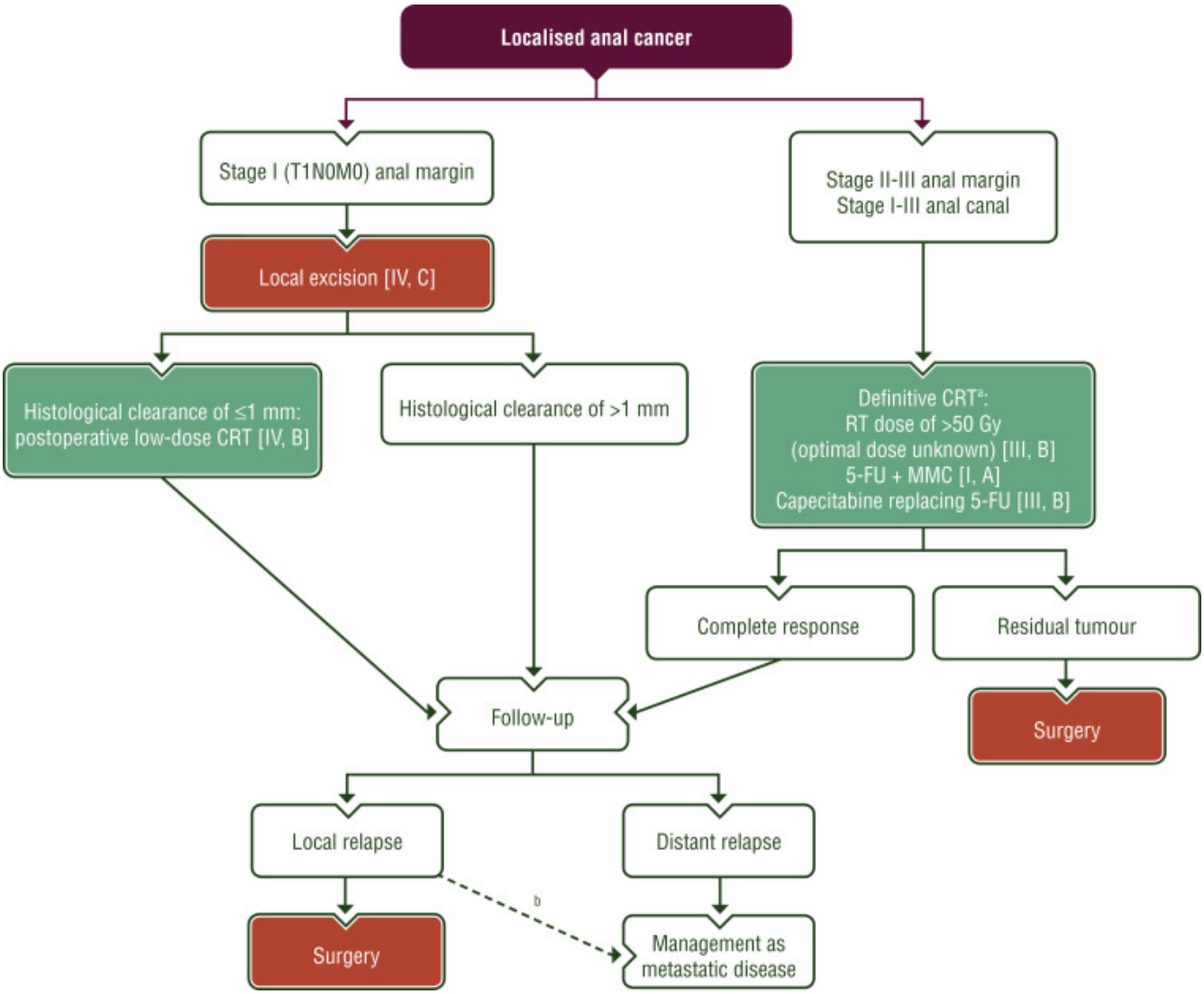


Figure 2.3: Treatment workflow for anal cancer. Purple: general categories or stratification; red: surgery; green: combination of treatments or other systemic treatments; white: other aspects of management. 5-FU, 5-fluorouracil; CRT, chemoradiotherapy; M, metastasis; N, node; MMC, mitomycin C; RT, radiotherapy; T, tumour. [a] Optimum timepoint to assess clinical tumor response after CRT is 26 weeks [II, B]. [b] In cases where surgery cannot be carried out. *Reprinted from [23] under the Creative Commons Attribution-NonCommercial 4.0 International.*

Patients in complete remission should be evaluated every 3-6 months for a period of 2 years, and every 6-12 months until 5 years, with a clinical examination including digital rectal exam (DRE) and palpation of the inguinal lymph nodes. Most relapses occur before 3 years (only <1% after 3 years)[28] so extended imaging surveillance after this time is not recommended.

2.2.3 Head and Neck cancer

Head and Neck cancer (HNC) is responsible for approximately 900,000 cases and over 400,000 deaths annually and ranks as the seventh most common cancer worldwide[29, 30]. HNCs are usually located in the squamous cells along the mucosal epithelium surfaces in the oral cavity (lips, buccal mucosa, hard palate, anterior tongue, floor of mouth, and retromolar trigone), pharynx (palatine tonsils, lingual tonsils, base of tongue, soft palate, uvula, posterior pharyngeal wall, bottom part of the throat, extending from the hyoid bone to the cricoid cartilage) and larynx. These cancers are collectively called HNC squamous cell carcinoma (HNSCC) and consist of 90% of all HNCs. Other types of HNCs such as adenoid cystic carcinoma (AdCC) are very rare. HNSCCs have been associated with well-established risk factors like tobacco smoking, alcohol use, and viral infections, such as HPV, referred to as mucosal hrHPV types that have been clearly associated with anogenital cancers. Oral cavity and larynx cancers are generally associated with tobacco consumption, alcohol abuse, or both, whereas pharynx cancers are increasingly attributed to infection with HPV, primarily HPV 16 and, to a lesser extent, HPV 18[31, 32, 33]. HPV-positive HNSCC outside of the oropharynx is rare (<6%). HPV-negative HNCs have worse prognostic than HPV-positive and while the incidence of non-viral HNCs is decreasing, HPV-positive HNCs incidence has dramatically increased in the last few decades. No screening strategy has proved to be effective, and careful physical examination remains the primary approach for early detection. It is feasible that HPV-positive HNSCC could be prevented by successful vaccination campaigns worldwide.

Histology

Squamous cells abnormal development of the surface epithelium prior to invasion of the subepithelial connective tissues include abnormal cellular organization, increased mitotic activity, and nuclear enlargement with pleomorphism. With progression, the carcinoma in situ breaks through the basement membrane and infiltrates the subepithelial connective tissue as cohesive nests and cords. With advanced tumor growth, nests of invasive cells invade skeletal muscle, craniofacial bones, and facial skin. Invasion may be associated with tumor extension along nerves (i.e., perineural invasion) and involvement of lymphatic spaces. The microscopic appearance may vary as a function of tumor differentiation, but the prototypic HNSCC is moderately differentiated. Subtypes of HNSCC include the basaloid variant, the spindle-cell variant, and the papillary variant. The basaloid squamous variant (solid lobules of cells with peripheral palisading, scant cytoplasm, and dark nuclei) is a rapidly growing tumor associated with poor patient outcomes. The spindle-cell variant is characterized by the proliferation of noncohesive spindle cells. Its microscopic appearance more closely resembles a sarcoma than a carcinoma. The papillary variant is characterized by a prominent exophytic component of papillary growth. In contrast to the benign squamous papilloma, the papillary fronds are lined by overtly malignant squamous cells [34]. HPV-positive HNSCC presents a basaloid, nonkeratinizing, poorly differentiated histology, while HPV-negative

HNSCC is keratinized and moderately differentiated. HPV-induced HNSCC, similar to other HPV-related cancers, is characterized by p16 overexpression.

Staging and treatment of HNSCC

Staging in HNC follows the recommendations of the UICC TNM 8th edition staging system classification of malignant tumours[16]. Classification into different stages is detailed in Table 2.3. Table 2.3 refers to the stage grouping of oropharyngeal squamous cell carcinoma (OPSCC), given that, in this work, data of OPSCC patients were used.

Stage grouping			
Stage 0	Tis	N0	M0
Stage I	T1,T2	N0, N1	M0
Stage II	T1,T2	N2	M0
	T3	N0, N1, N2	M0
Stage III	T1, T2, T3	N3	M0
	T4	Any N	M0
Stage IV	Any T	Any N	M1

Table 2.3: OPSCC TNM classification. TX: Primary tumor cannot be assessed. T0: No evidence of primary tumor. Tis: Carcinoma in situ. T1: Tumor with a maximum size of 2 cm or less. T2: Tumor with a maximum size of more than 2 cm but not more than 4 cm. T3: Tumor with a maximum size of more than 4 cm or extension to the lingual surface of the epiglottis. T4: tumor of any size invades adjacent organ(s). N refers to regional lymph nodes. NX: Regional lymph nodes cannot be assessed. N0: No regional lymph node metastasis, N1: Unilateral metastasis in lymph node(s), all with a maximum size of 6 cm or less. N2: Contralateral or bilateral metastasis in lymph node(s), all with a maximum size of 6 cm or less. N3: Metastasis in lymph node(s) with a maximum size of greater than 6 cm. M refers to distant, metastasis. MX: Distant metastasis cannot be assessed, M0: No distant metastasis, M1: Distant metastasis

The standard care for oropharyngeal cancers is RT or surgery (followed by RT or CRT if indicated) for stage I disease, and definitive (C)RT or adjuvant (C)RT for locally advanced tumors (Figure 2.4). In regard to RT, all patients with HNSCC should be treated by Intensity-Modulated RT (IMRT) or Volumetric Modulated Arc Therapy (VMAT) regardless of their HPV status. The standard chemotherapy treatment includes cisplatin at a dose of 100 mg/m^2 given on days 1, 22, and 43 of concomitant RT (70 Gy), or weekly cisplatine at a dose of 40 mg/m^2. However, if patients are unfit for cisplatin, carboplatin combined with 5-FU or cetuximab concomitant to RT as well as hyperfractionated or accelerated RT without chemotherapy can be treatment alternatives. Immunotherapy was shown effective for patients with recurrent/metastatic HNSCC expressing PD-L1, using Pembrolizumab in combination with platinum/5-FU or pembrolizumab monotherapy. For recurrent/metastatic patients with HNSCC not expressing PD-L1, platinum/5-FU/cetuximab remains the standard therapy. Nivolumab is both FDA- and EMA-approved for recurrent/metastatic patients who progress within 6 months of platinum therapy. Guidelines details can be found in [35, 36].

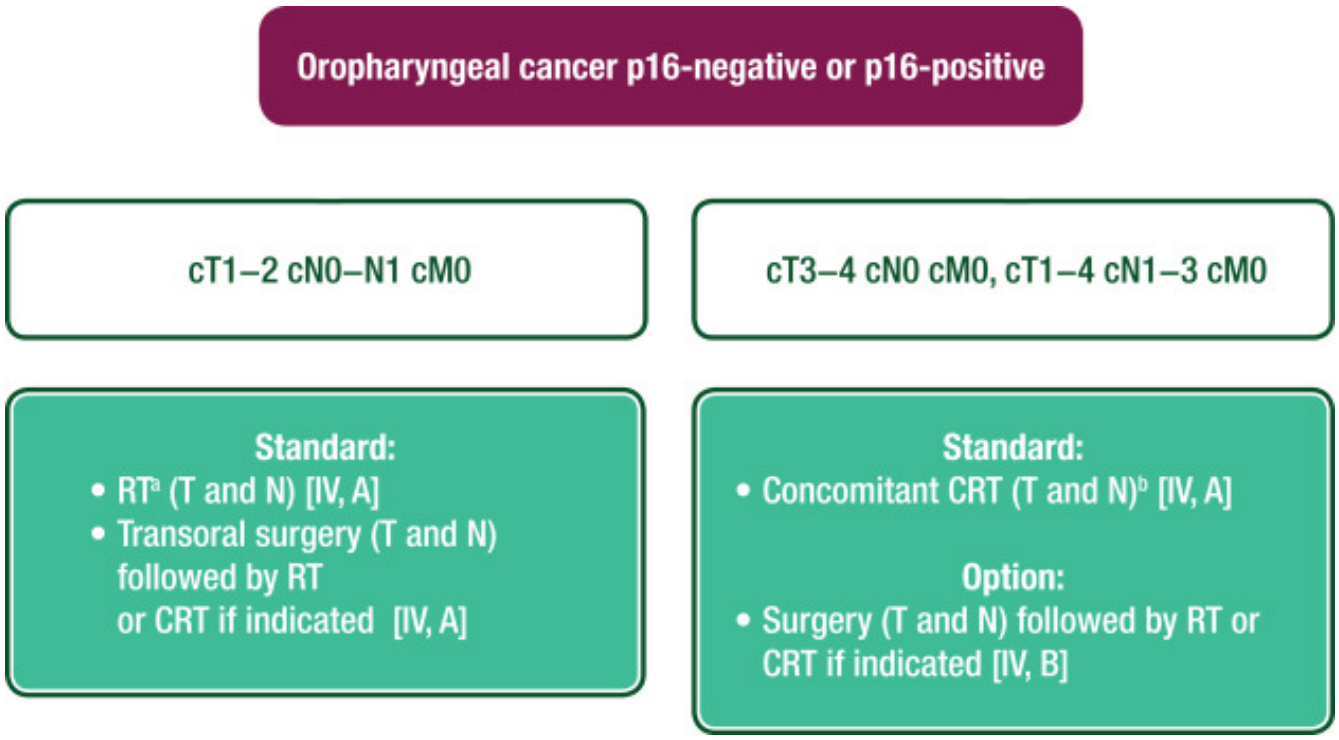


Figure 2.4: Management of oropharyngeal cancer (p16-negative stage I–IVB; p16-positive stage I–III). CRT, chemoradiotherapy; M, metastasis; N, node; RT, radiotherapy; T, tumour. [a] Altered fractionation (accelerated or hyperfractionated) RT is a valid option for T1–N1, T2–N0 or T2–N1. [b] Altered fractionation (accelerated or hyperfractionated) RT is a valid option for T1–N1 or T2–N1. *Reprinted from [36] under the Creative Commons Attribution-NonCommercial 4.0 International.*

2.3 Role of imaging in HPV induced cancers

In clinical routine, CT scans, PET and MRI are the most used imaging techniques. CT (which can be enhanced with a contrast agent) because of its high availability, is almost always the first step in diagnosis of patients and occupies an important place in RT, where it is mandatory for treatment planning. The diagnosis can then be refined by the other modalities. PET and MRI provide additional insights and can be critical in solving clinical differential diagnoses. Many types of solid tumors can be better characterized by PET/CT and PET-MRI scans, including: brain, breast, cervical, colorectal, esophageal, HNC, lung, lymphatic system, pancreatic, prostate, skin, thyroid.

In this section, the role of PET and MRI in HPV-induced cancers is discussed.

2.3.1 Positron Emission Tomography

PET and SPECT are image modalities that exploit the nuclear properties of radiotracers, i.e, radioactive isotopes chemically bound to a biological complex normally used by the body such as glucose (or glucose analogs), water, or ammonia, or into molecules that bind to specific receptors or sites within the body. The radiotracer is usually injected intravenously but can also be swallowed, or inhaled, depending on what area of the body is being examined. The biological complex acts characteristically and is concentrated into areas of higher levels of metabolic or biochemical activity. Mostly, in oncology, the biological complex used is glucose analog, and gets metabolized by tumor cells since these cells have increased glycolysis compared to normal cells. In PET images, the positron emitters fluorine-18 (18F) coupled with glucose analog compound to form [18F]Fluorodeoxyglucose ([18F]-FDG) is the most used radiopharmaceuticals. The radionuclide is continuously emitting positrons that collide with medium electrons and produce two photons of 511 keV. These two photons are emitted in approximately opposite directions and are finally detected in the scintillators arranged in a ring, which are one of the elements of the PET detector blocks, and form the basis for locating the annihilation site in the reconstructed PET image[37]. PET enables functional imaging (visualizes the expression of receptors, enzymes, and other molecular target structures) and can be coupled with computed tomography (CT) or magnetic resonance imaging (MRI) as a hybrid technique. The working principles of PET are shown in figure 2.5.

In [18F]-FDG PET, FDG uptake in tumors is usually quantified by measuring the standardized uptake value (SUV). The SUV is a simplified measure, and is now the most widely used method for the quantification of 18F-FDG PET studies, although other measures have been developed as well[39, 40]. The SUV is defined as [18F]-FDG retention normalized to injected activity whole body concentration :

$$SUV = \frac{A[\frac{kBq}{mL}]}{\frac{D_i[kBq]}{w[g]}} \tag{2.1}$$

where A is the tumor activity concentration in the considered voxel, D_i is the decay-corrected activity of injected radiolabeled FDG, and w is the weight of the patient [g][41]. Mean and Maximal Standard Uptake Value (SUVmean and SUVmax) are the most commonly used imaging biomarkers derived from the analysis of 18F-FDG PET images used to guide clinical decision in differential diagnosis, treatment response prediction,

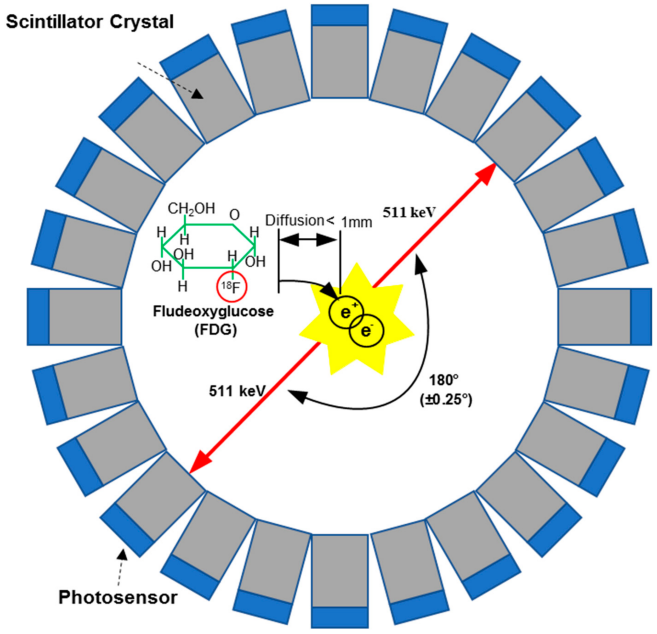


Figure 2.5: The basic principle of a positron emission tomography (PET) system: A PET detector ring detects a pair of gamma photons with an energy of 511 keV (red arrows) which results from the annihilation of an electron with a positron emitted by the radiotracer (here, FDG). *Reprinted from [38] under the Creative Commons Attribution-NonCommercial 4.0 International.*

and prognosis[42, 43]. In addition to SUVmax, metabolic tumor volume and total lesion glycolysis have been extensively investigated in recent studies, and many lines of evidence support that these parameters are related to patient outcome[44]. The reproducibility of these biomarkers is hampered by several factors, which can be either physiologic (i.e., blood glucose concentration) or technical (e.g. type of detectors and associated electronics, reconstruction algorithms)[45].

[18F]-FDG PET/CT is particularly useful in the detection of the primary tumor, of pelvic lymph nodes, and in the initial staging of patients with LACC and anal cancers. In both cancers, its sensitivity was found superior for the detection of regional nodal involvement, and distant metastatic disease compared to conventional imaging[46, 47, 48]. In HNCs, physical examination and endoscopy followed by imaging modalities such as neck ultrasound, neck MRI and neck-chest CT, have been shown to be less accurate than [18F]-FDG PET pre-treatment staging. For primary tumor assessment, the extent of the tumor may be underestimated especially if the exam is performed with low-dose enhanced CT. However, the main indication of [18F]-FDG PET/CT in newly diagnosed HNCs remains the detection of cervical lymph node involvement, which is one of the most important prognostic factors[49]. [18F]-FDG PET is therefore considered mandatory in clinical routine[50]. In addition, there is growing interest in PET-guided RT for more precise target volume segmentation and the possibility of implementing functional signal-guided dose escalation[51, 52]. PET/CT can also detect and differentiate radiation and surgical changes from residual or recurrent tumors because cancer cells retain more FDG for longer periods of time than inflammatory tissues. Studies have shown that PET/CT had a high sensitivity (of more than 90%) for localization of recurrent disease[53, 54, 55]. As well, [18F]-FDG PET is used for restaging in case of tumor recurrence whether local or distant [48, 23, 47, 49].

Despite clear benefits of this modality in multiple situations, misinterpretation of tracer uptake may be

unavoidable in some cases where high activity in normal tissues may lead to confusion when distinguishing pathologic activity from physiologic activity. Infections and inflammations related to cancer therapy or foreign bodies are also associated with increased FDG uptake. This is due to high glycolysis of macrophages in inflammatory and neoplastic lesions. Therefore, caution should be paid to timing in tumor follow-up. Finally, Some tumors such as salivary gland tumors (mucoepidermoid and adenoid cystic carcinoma), cystic lesions, and necrotic lesions have inherently low FDG uptake[53].

2.3.2 Magnetic Resonance Imaging

MRI is a medical imaging technique based on nuclear magnetic resonance (NMR). When nuclei with an intrinsic magnetic moment (odd numbers of nucleons) are placed in a strong magnetic field $\vec{B_0}$, the macroscopic magnetization vector $\vec{M}$ resulting from all the microscopic magnetization vectors, oriented in the same direction as the applied magnetic field, becomes non-null. By applying a resonating radio frequency (RF) pulse, at a specific frequency called Larmor frequency, often perpendicular to $\vec{B_0}$, the nuclei magnetic vector is deflected by an angle alpha, which will decrease the longitudinal component of $\vec{M}$ and increase its transverse component. The RF will influence the longitudinal and transverse components, through simultaneous but different mechanisms: by inducing the transition on the high energy level and by the phasing of protons. When the RF is stopped, a radio wave signal is emitted because the magnetic vector $\vec{M}$ is returning to its resting state. This is called relaxation. We distinguish longitudinal and transversal relaxations. The process of relaxation is described in Figure 2.6.

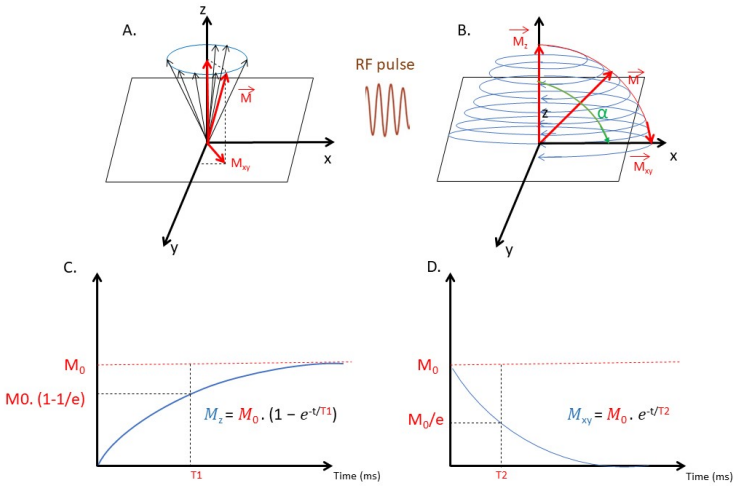


Figure 2.6: Relaxation phenomena in MRI. (A) Macroscopic magnetization vector $\vec{M}$ resulting from the presence of a strong magnetic field $\vec{B_0}$ and its components. (B) Excitation phase: the resonating RF pulse deflects $\vec{M}$ by an angle alpha. Here alpha is equal to 90°, the longitudinal component is nullified. (C) Longitudinal relaxation characterized by a time constant T1, (D) Transversal relaxation characterized by a time constant T2.

Longitudinal relaxation is characterized by the return of the longitudinal component of $\vec{M}$ to its maximum. This relaxation is accompanied by the emission of the energy absorbed during the excitation phase. The energy emission with the surrounding medium, called lattice is done by thermal exchange (spin-lattice or thermal

relaxation). Longitudinal relaxation is called T1 relaxation because it follows an exponential process with T1 as a first-order time constant characterizing the tissue. T1 can be viewed as the time required for the longitudinal component of $\vec{M}$ to reach $(1-1/e)$ or about 63% of its maximum value. When the 90°RF emission is stopped, in addition to inducing the disappearance of the longitudinal component of $\vec{M}$, this leads to a dephasing of the spins of protons: the transversal component of $\vec{M}$ decreases rapidly. This phenomenon is characterized by interactions of protons between them. Transversal relaxation is called T2 relaxation or spin-spin relaxation and follows an exponential decay with time constant T2, where T2 can be viewed as the time required for the transversal component of $\vec{M}$ to reach $(1/e)$ or lose about 63% of its maximum value. T2 is intrinsically related to the molecular medium, while T1 is dependent on the intensity of the magnetic field $\vec{B_0}$. Furthermore, T2 can be influenced by local static field disturbance in addition to $\vec{B_0}$.

The emitted signal after NMR is detected with a coil antenna by free induction decay (FID). However, this signal is affected by both the inhomogeneities of molecular origin, to which are added the proper inhomogeneities of the magnetic field$\vec{B_0}$. To get rid of these inhomogeneities, and since the signal is decaying, a 180°RF is emitted after a time TE/2 after the 90°pulse. This will allow the spins to rephase and an echo will be emitted at a time called echo time (TE). At this point, the signal is measured to collect the maximum intensity signal, this is the echo spin sequence. RF and echo emission can be repeated periodically. The amount of time between two consecutive series of RF and echo pulses is called repetition time (TR). Given that different tissues relax at different rates when the transmitted RF pulse is switched off, modulation of TE and TR will highlight different emphases on tissues. Therefore, TE will define T2 weighting, while TR will define T1 weighting.

- a short TR and a short TE give a T1-weighted image
- a long TR and a long TE give a T2 weighted image
- a long TR and a short TE give a proton density-weighted image

The quality of an MR image is described using conventional quality metrics, including signal-to-noise ratio (SNR), spatial resolution, and image contrast. The signal emitted during an MRI acquisition depends on intrinsic tissue properties but also on machine parameters such as the magnetic field $\vec{B_0}$, RF pulses, TR, TE, voxel volume, number of excitations (NEX), bandwidth, etc. Spatial resolution is directly dependent on voxel volume, itself associated with the selected slice thickness and field of view, while TR and TE affect image contrast. In T1-weighted MR image, optimal TR is around tissue T1 and TE should be minimal. For T2-weighted images, TR and TE should be long enough at the expense of optimal SNR. NEX also impacts the signal in MRI as SNR is proportional to the square root of NEX. Therefore as NEX increases, noise begins to be canceled out with the square root of the NEX. In MRI, bandwidth refers to the receiver (or acquisition) bandwidth and the transmit bandwidth. The receiver bandwidth impacts the MR image quality as a narrow bandwidth will increase SNR but might add spatial distortion, while a larger bandwidth reduces SNR but allows for faster imaging. The transmit bandwidth is proportional to the slice thickness of the acquired MR image.

Originally, MRI was anatomical but sequences have been developed over the years to broaden the scope of this modality. Advanced techniques have allowed for functional MR images such as Diffusion-weighted imaging (DWI), perfusion-weighted imaging (PWI), diffusion tensor imaging (DTI), and magnetic resonance spectroscopy (MRS).

MRI displays high soft tissue contrast which allows for accurate delineation between the different tissues. The use of T2-weighted MRI is needed for optimal assessment of primary tumor and lymph nodes both in LACC, anal cancers, and HNCC[56, 23, 35]. Therefore MRI is an indispensable image modality in the planning and delivery of radiotherapy (RT) and brachytherapy (BT)[57, 58]. Lymph node assessment is notoriously difficult to predict using imaging modalities, and validation of nodal assessment criteria has not been possible due to the scarcity of surgical specimens for histopathology correlation. However, node involvement is best assessed using high-resolution T2-weighted MRI[59]. Moreover, MRI plays also a key role in post-treatment response assessment and surveillance. In anal cancer, patients with locally advanced anal cancer may benefit from intensive MRI surveillance in the first 12 months[23]. For all these reasons, MRI is now a widely used modality, and its role in RT and cancer management, in general, will only increase over the years, especially with the event of synthetically generated CT with AI.

Bibliography

[1] Megan E. Spurgeon and Paul F. Lambert. "Mus musculus Papillomavirus 1: a New Frontier in Animal Models of Papillomavirus Pathogenesis". In: *J Virol* 94.9 (Apr. 2020), e00002–20. ISSN: 0022-538X. DOI: 10.1128/JVI.00002-20. URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7163119/ (visited on 12/14/2022).

[2] Hans-Ulrich Bernard et al. "Classification of papillomaviruses (PVs) based on 189 PV types and proposal of taxonomic amendments". eng. In: *Virology* 401.1 (May 2010), pp. 70–79. ISSN: 1096-0341. DOI: 10.1016/j.virol.2010.02.002.

[3] Koenraad Van Doorslaer et al. "ICTV Virus Taxonomy Profile: Papillomaviridae". eng. In: *J Gen Virol* 99.8 (Aug. 2018), pp. 989–990. ISSN: 1465-2099. DOI: 10.1099/jgv.0.001105.

[4] Emma J Crosbie et al. "Human papillomavirus and cervical cancer". en. In: *The Lancet* 382.9895 (Sept. 2013), pp. 889–899. ISSN: 0140-6736. DOI: 10.1016/S0140-6736(13)60022-7. URL: https://www.sciencedirect.com/science/article/pii/S0140673613600227 (visited on 12/14/2022).

[5] Catherine de Martel et al. "Worldwide burden of cancer attributable to HPV by site, country and HPV type". eng. In: *Int J Cancer* 141.4 (Aug. 2017), pp. 664–670. ISSN: 1097-0215. DOI: 10.1002/ijc.30716.

[6] Arnaud John Kombe Kombe et al. "Epidemiology and Burden of Human Papillomavirus and Related Diseases, Molecular Pathogenesis, and Vaccine Evaluation". In: *Frontiers in Public Health* 8 (2021). ISSN: 2296-2565. URL: https://www.frontiersin.org/articles/10.3389/fpubh.2020.552028 (visited on 12/14/2022).

[7] Hyuna Sung et al. "Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries". en. In: *CA: A Cancer Journal for Clinicians* 71.3 (2021). _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.3322/caac.21660, pp. 209–249. ISSN: 1542-4863. DOI: 10.3322/caac.21660. URL: https://onlinelibrary.wiley.com/doi/abs/10.3322/caac.21660 (visited on 11/02/2022).

[8] Neerja Bhatla et al. "Cancer of the cervix uteri: 2021 update". en. In: *International Journal of Gynecology & Obstetrics* 155.S1 (2021). _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/ijg pp. 28–44. ISSN: 1879-3479. DOI: 10.1002/ijgo.13865. URL: https://onlinelibrary.wiley.com/doi/abs/10.1002/ijgo.13865 (visited on 12/14/2022).

[9] Rodrigo Pinheiro Araldi et al. “The human papillomavirus (HPV)-related cancer biology: An overview”. eng. In: *Biomed Pharmacother* 106 (Oct. 2018), pp. 1537–1556. ISSN: 1950-6007. DOI: 10.1016/j.biopha.2018.06.149.

[10] F. X. Bosch et al. “The causal relation between human papillomavirus and cervical cancer”. eng. In: *J Clin Pathol* 55.4 (Apr. 2002), pp. 244–265. ISSN: 0021-9746. DOI: 10.1136/jcp.55.4.244.

[11] *Global burden of cancer attributable to infections in 2018: a worldwide incidence analysis - PubMed*. URL: https://pubmed.ncbi.nlm.nih.gov/31862245/ (visited on 02/14/2022).

[12] Mélanie Drolet et al. “Population-level impact and herd effects following the introduction of human papillomavirus vaccination programmes: updated systematic review and meta-analysis”. eng. In: *Lancet* 394.10197 (Aug. 2019), pp. 497–509. ISSN: 1474-547X. DOI: 10.1016/S0140-6736(19)30298-3.

[13] Jiayao Lei et al. “HPV Vaccination and the Risk of Invasive Cervical Cancer”. eng. In: *N Engl J Med* 383.14 (Oct. 2020), pp. 1340–1348. ISSN: 1533-4406. DOI: 10.1056/NEJMoa1917338.

[14] Maxime Bonjour et al. “Global estimates of expected and preventable cervical cancers among girls born between 2005 and 2014: a birth cohort analysis”. eng. In: *Lancet Public Health* 6.7 (July 2021), e510–e521. ISSN: 2468-2667. DOI: 10.1016/S2468-2667(21)00046-3.

[15] Sergio Pecorelli. “Revised FIGO staging for carcinoma of the vulva, cervix, and endometrium”. en. In: *International Journal of Gynecology & Obstetrics* 105.2 (2009). _eprint: https://onlinelibrary.wil pp. 103–104. ISSN: 1879-3479. DOI: 10.1016/j.ijgo.2009.02.012. URL: https://onlinelibrary.wiley.com/doi/abs/10.1016/j.ijgo.2009.02.012 (visited on 04/22/2022).

[16] James D. Brierley, Mary K. Gospodarowicz, and Christian Wittekind. *TNM Classification of Malignant Tumours*. en. Google-Books-ID: 642GDQAAQBAJ. John Wiley & Sons, Jan. 2017. ISBN: 978-1-119-26357-9.

[17] Richard Pötter et al. “Recommendations from gynaecological (GYN) GEC ESTRO working group (II): Concepts and terms in 3D image-based treatment planning in cervix cancer brachytherapy—3D dose volume parameters and aspects of 3D image-based anatomy, radiation physics, radiobiology”. en. In: *Radiotherapy and Oncology* 78.1 (Jan. 2006), pp. 67–77. ISSN: 0167-8140. DOI: 10.1016/j.radonc.2005.11.014. URL: https://www.sciencedirect.com/science/article/pii/S0167814005005463 (visited on 03/16/2022).

[18] Junzo Chino et al. “Radiation Therapy for Cervical Cancer: Executive Summary of an ASTRO Clinical Practice Guideline”. In: *Pract Radiat Oncol* 10.4 (2020), pp. 220–234. ISSN: 1879-8500. DOI: 10.1016/j.prro.2020.04.002. URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC8802172/ (visited on 12/14/2022).

[19] Ritu Salani et al. “An update on post-treatment surveillance and diagnosis of recurrence in women with gynecologic malignancies: Society of Gynecologic Oncology (SGO) recommendations”. en. In: *Gynecologic Oncology* 146.1 (July 2017), pp. 3–10. ISSN: 0090-8258. DOI: 10.1016/j.ygyno.2017.03.022. URL: https://www.sciencedirect.com/science/article/pii/S009082581730238X (visited on 03/14/2022).

[20] David B. Erlichman et al. “MRI anatomy and pathology of the anal canal”. eng. In: *J Magn Reson Imaging* 50.4 (Oct. 2019), pp. 1018–1032. ISSN: 1522-2586. DOI: 10.1002/jmri.26776.

[21] Catherine de Martel et al. “Global burden of cancer attributable to infections in 2018: a worldwide incidence analysis”. eng. In: *Lancet Glob Health* 8.2 (Feb. 2020), e180–e190. ISSN: 2214-109X. DOI: 10.1016/S2214-109X(19)30488-7.

[22] Gary M. Clifford et al. “A meta-analysis of anal cancer incidence by risk group: Toward a unified anal cancer risk scale”. In: *Int J Cancer* 148.1 (Jan. 2021), pp. 38–47. ISSN: 0020-7136. DOI: 10.1002/ijc.33185. URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7689909/ (visited on 12/14/2022).

[23] S. Rao et al. “Anal cancer: ESMO Clinical Practice Guidelines for diagnosis, treatment and follow-up”. English. In: *Annals of Oncology* 32.9 (Sept. 2021). Publisher: Elsevier, pp. 1087–1100. ISSN: 0923-7534, 1569-8041. DOI: 10.1016/j.annonc.2021.06.015. URL: https://www.annalsofoncology.org/article/S0923-7534(21)02064-0/fulltext#secsectitle0095 (visited on 11/25/2022).

[24] Paulo M. Hoff, Renata Coudry, and Camila Motta Venchiarutti Moniz. “Pathology of Anal Cancer”. eng. In: *Surg Oncol Clin N Am* 26.1 (Jan. 2017), pp. 57–71. ISSN: 1558-5042. DOI: 10.1016/j.soc.2016.07.013.

[25] Hope E. Uronis and Johanna C. Bendell. “Anal cancer: an overview”. eng. In: *Oncologist* 12.5 (May 2007), pp. 524–534. ISSN: 1083-7159. DOI: 10.1634/theoncologist.12-5-524.

[26] Ulrike Wieland and Alexander Kreuter. “Anal cancer risk: HPV-based cervical screening programmes”. English. In: *The Lancet Infectious Diseases* 19.8 (Aug. 2019). Publisher: Elsevier, pp. 799–800. ISSN: 1473-3099, 1474-4457. DOI: 10.1016/S1473-3099(19)30296-8. URL: https://www.thelancet.com/journals/laninf/article/PIIS1473-3099(19)30296-8/fulltext (visited on 11/26/2021).

[27] F. T. Bosman et al. “WHO classification of tumours of the digestive system.” English. In: *WHO classification of tumours of the digestive system.* Ed. 4 (2010). Publisher: World Health Organization. URL: https://www.cabdirect.org/cabdirect/abstract/20113051318 (visited on 12/14/2022).

[28] Kathinka S. Slørdahl et al. “Treatment outcomes and prognostic factors after chemoradiotherapy for anal cancer”. eng. In: *Acta Oncol* 60.7 (July 2021), pp. 921–930. ISSN: 1651-226X. DOI: 10.1080/0284186X.2021.1918763.

[29] Jacques Ferlay et al. “Cancer incidence and mortality worldwide: sources, methods and major patterns in GLOBOCAN 2012”. eng. In: *Int J Cancer* 136.5 (Mar. 2015), E359–386. ISSN: 1097-0215. DOI: 10.1002/ijc.29210.

[30] Mayur D. Mody et al. “Head and neck cancer”. eng. In: *Lancet* 398.10318 (Dec. 2021), pp. 2289–2299. ISSN: 1474-547X. DOI: 10.1016/S0140-6736(21)01550-6.

[31] Andrew P. Stein et al. “Prevalence of Human Papillomavirus in Oropharyngeal Cancer: A Systematic Review”. eng. In: *Cancer J* 21.3 (2015), pp. 138–146. ISSN: 1540-336X. DOI: 10.1097/PPO.0000000000000115.

[32] Tatyana Isayeva et al. “Human papillomavirus in non-oropharyngeal head and neck cancers: a systematic literature review”. eng. In: *Head Neck Pathol* 6 Suppl 1.Suppl 1 (July 2012), S104–120. ISSN: 1936-0568. DOI: 10.1007/s12105-012-0368-1.

[33] Dominique S. Michaud et al. “High-risk HPV types and head and neck cancer”. eng. In: *Int J Cancer* 135.7 (Oct. 2014), pp. 1653–1661. ISSN: 1097-0215. DOI: 10.1002/ijc.28811.

[34] Sara I. Pai and William H. Westra. “Molecular Pathology of Head and Neck Cancer: Implications for Diagnosis, Prognosis, and Treatment”. In: *Annu Rev Pathol* 4 (2009), pp. 49–70. ISSN: 1553-4006. DOI: 10.1146/annurev.pathol.4.110807.092158. URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3703474/ (visited on 12/15/2022).

[35] Vincent Grégoire et al. “Delineation of the primary tumour Clinical Target Volumes (CTV-P) in laryngeal, hypopharyngeal, oropharyngeal and oral cavity squamous cell carcinoma: AIRO, CACA, DAHANCA, EORTC, GEORCC, GORTEC, HKNPCSG, HNCIG, IAG-KHT, LPRHHT, NCIC CTG, NCRI, NRG Oncology, PHNS, SBRT, SOMERA, SRO, SSHNO, TROG consensus guidelines”. eng. In: *Radiother Oncol* 126.1 (Jan. 2018), pp. 3–24. ISSN: 1879-0887. DOI: 10.1016/j.radonc.2017.10.016.

[36] J. -P. Machiels et al. “Squamous cell carcinoma of the oral cavity, larynx, oropharynx and hypopharynx: EHNS–ESMO–ESTRO Clinical Practice Guidelines for diagnosis, treatment and follow-up††These Guidelines were developed by the European Head and Neck Society (EHNS), the European Society for Medical Oncology (ESMO) and the European SocieTy for Radiotherapy & Oncology (ESTRO). The three societies nominated authors to write the guidelines as well as reviewers to comment on them. These guidelines were approved by the EHNS Executive Board, the ESMO Guidelines Committee and the ESTRO Executive Board in June 2020.” en. In: *Annals of Oncology* 31.11 (Nov. 2020), pp. 1462–1475. ISSN: 0923-7534. DOI: 10.1016/j.annonc.2020.07.011. URL: https://www.sciencedirect.com/science/article/pii/S092375342039949X (visited on 12/15/2022).

[37] Thorsten Derlin et al. “Molecular Imaging in Oncology Using Positron Emission Tomography”. In: *Dtsch Arztebl Int* 115.11 (Mar. 2018), pp. 175–181. ISSN: 1866-0452. DOI: 10.3238/arztebl.2018.0175. URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5913576/ (visited on 01/11/2023).

[38] Wei Jiang, Yamn Chalich, and M. Jamal Deen. "Sensors for Positron Emission Tomography Applications". en. In: *Sensors* 19.22 (Jan. 2019). Number: 22 Publisher: Multidisciplinary Digital Publishing Institute, p. 5019. ISSN: 1424-8220. DOI: 10.3390/s19225019. URL: https://www.mdpi.com/1424-8220/19/22/5019 (visited on 01/11/2023).

[39] G. J. Hunter et al. "Simplified measurement of deoxyglucose utilization rate". eng. In: *J Nucl Med* 37.6 (June 1996), pp. 950–955. ISSN: 0161-5505.

[40] M. M Graham, L. M Peterson, and R. M Hayward. "Comparison of simplified quantitative analyses of FDG uptake". en. In: *Nuclear Medicine and Biology* 27.7 (Oct. 2000), pp. 647–655. ISSN: 0969-8051. DOI: 10.1016/S0969-8051(00)00143-8. URL: https://www.sciencedirect.com/science/article/pii/S0969805100001438 (visited on 01/11/2023).

[41] Paul E. Kinahan and James W. Fletcher. "PET/CT Standardized Uptake Values (SUVs) in Clinical Practice and Assessing Response to Therapy". In: *Semin Ultrasound CT MR* 31.6 (Dec. 2010), pp. 496–505. ISSN: 0887-2171. DOI: 10.1053/j.sult.2010.10.001. URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3026294/ (visited on 01/11/2023).

[42] Nanda C. Krak et al. "Effects of ROI definition and reconstruction method on quantitative outcome and applicability in a response monitoring trial". eng. In: *Eur J Nucl Med Mol Imaging* 32.3 (Mar. 2005), pp. 294–301. ISSN: 1619-7070. DOI: 10.1007/s00259-004-1566-1.

[43] Sylvain Reuzé et al. "Prediction of cervical cancer recurrence using textural features extracted from 18 F-FDG PET images acquired with different scanners". In: *Oncotarget* 8.26 (May 2017). Publisher: Impact Journals, pp. 43169–43179. ISSN: 1949-2553. DOI: 10.18632/oncotarget.17856. URL: https://www.oncotarget.com/article/17856/text/ (visited on 01/26/2021).

[44] Sylvain Reuzé et al. "Radiomics in Nuclear Medicine Applied to Radiation Therapy: Methods, Pitfalls, and Challenges". en. In: *International Journal of Radiation Oncology*Biology*Physics.* Imaging in Radiation Oncology 102.4 (Nov. 2018), pp. 1117–1142. ISSN: 0360-3016. DOI: 10.1016/j.ijrobp.2018.05.022. URL: https://www.sciencedirect.com/science/article/pii/S0360301618308150 (visited on 01/06/2023).

[45] Julian M. M. Rogasch et al. "Influences on PET Quantification and Interpretation". In: *Diagnostics (Basel)* 12.2 (Feb. 2022), p. 451. ISSN: 2075-4418. DOI: 10.3390/diagnostics12020451. URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC8871060/ (visited on 04/26/2023).

[46] E. de Winton et al. "The impact of 18-fluorodeoxyglucose positron emission tomography on the staging, management and outcome of anal cancer". eng. In: *Br J Cancer* 100.5 (Mar. 2009), pp. 693–700. ISSN: 1532-1827. DOI: 10.1038/sj.bjc.6604897.

[47] Aamer Mahmud, Raymond Poon, and Derek Jonker. "PET imaging in anal canal cancer: a systematic review and meta-analysis". In: *Br J Radiol* 90.1080 (Dec. 2017), p. 20170370. ISSN: 0007-1285. DOI: 10.1259/bjr.20170370. URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6047643/ (visited on 12/15/2022).

[48] Chitra Viswanathan et al. "FDG-PET assessment of cervical cancer". In: *PET Clin* 13.2 (Apr. 2018), pp. 165–177. ISSN: 1556-8598. DOI: 10.1016/j.cpet.2017.11.004. URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5996999/ (visited on 12/15/2022).

[49] P. CASTALDI et al. "Role of 18F-FDG PET-CT in head and neck squamous cell carcinoma". In: *Acta Otorhinolaryngol Ital* 33.1 (Feb. 2013), pp. 1–8. ISSN: 0392-100X. URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3631810/ (visited on 08/04/2022).

[50] D. Mak et al. "Role of FDG-PET/CT in staging and follow-up of head and neck squamous cell carcinoma". eng. In: *Q J Nucl Med Mol Imaging* 55.5 (Oct. 2011), pp. 487–499. ISSN: 1824-4785.

[51] Saskia A. Cooke et al. "18F-FDG-PET guided vs whole tumour radiotherapy dose escalation in patients with locally advanced non-small cell lung cancer (PET-Boost): Results from a randomised clinical trial". eng. In: *Radiother Oncol* 181 (Apr. 2023), p. 109492. ISSN: 1879-0887. DOI: 10.1016/j.radonc.2023.109492.

[52] Olga Hamming-Vrieze et al. "Biological PET-guided adaptive radiotherapy for dose escalation in head and neck cancer: a systematic review". eng. In: *Q J Nucl Med Mol Imaging* 62.4 (Dec. 2018), pp. 349–368. ISSN: 1827-1936. DOI: 10.23736/S1824-4785.18.03087-X.

[53] Galal Omami, Dania Tamimi, and Barton F. Branstetter. "Basic principles and applications of 18F-FDG-PET/CT in oral and maxillofacial imaging: A pictorial essay". In: *Imaging Sci Dent* 44.4 (Dec. 2014), pp. 325–332. ISSN: 2233-7822. DOI: 10.5624/isd.2014.44.4.325. URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4245476/ (visited on 01/11/2023).

[54] G. W. Goerres et al. "Impact of whole body positron emission tomography on initial staging and therapy in patients with squamous cell carcinoma of the oral cavity". eng. In: *Oral Oncol* 39.6 (Sept. 2003), pp. 547–551. ISSN: 1368-8375. DOI: 10.1016/s1368-8375(03)00016-2.

[55] Lee A. Zimmer et al. "The use of combined PET/CT for localizing recurrent head and neck cancer: the Pittsburgh experience". eng. In: *Ear Nose Throat J* 84.2 (Feb. 2005), pp. 104, 106, 108–110. ISSN: 0145-5613.

[56] Emma C. Fields and Elisabeth Weiss. "A practical review of magnetic resonance imaging for the evaluation and management of cervical cancer". In: *Radiat Oncol* 11 (Feb. 2016), p. 15. ISSN: 1748-717X. DOI: 10.1186/s13014-016-0591-0. URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4736634/ (visited on 10/10/2022).

[57] Kari Tanderup et al. “From point A to the sculpted pear: MR image guidance significantly improves tumour dose and sparing of organs at risk in brachytherapy of cervical cancer”. eng. In: *Radiother Oncol* 94.2 (Feb. 2010), pp. 173–180. ISSN: 1879-0887. DOI: 10.1016/j.radonc.2010.01.001.

[58] Richard Pötter et al. “The EMBRACE II study: The outcome and prospect of two decades of evolution within the GEC-ESTRO GYN working group and the EMBRACE studies”. eng. In: *Clin Transl Radiat Oncol* 9 (Feb. 2018), pp. 48–60. ISSN: 2405-6308. DOI: 10.1016/j.ctro.2018.01.001.

[59] D. M. Koh et al. “Pelvic phased-array MR imaging of anal carcinoma before and after chemoradiation”. eng. In: *Br J Radiol* 81.962 (Feb. 2008), pp. 91–98. ISSN: 1748-880X. DOI: 10.1259/bjr/96187638.

Chapter 3

From image acquisition to precision medicine

Radiomics is defined as "the high-throughput mining of quantitative image features from standard-of-care medical imaging that enables data to be extracted and applied within clinical-decision support systems to improve diagnostic, prognostic, and predictive accuracy"[1]. Radiomics aims to identify and combine new imaging biomarkers, not assessable to the human eye, in complex models, providing predictive or prognostic information about patients and their pathologies based on statistical analyses. These biomarkers, i.e. image descriptors, reflect tissue intrinsic properties and heterogeneity and, indirectly, molecular and genetic substrates.

Predictions have been made that radiologists will beneficially incorporate AI methods into their daily practice[2]. Success of AI, especially ML in imaging offers a new and promising set of methods for analyzing image data and will most likely increase diagnostic certainty, and lead to faster turnaround, better outcomes for patients, and better quality of work life for radiologists. In this section, we are going to discuss the radiomics workflow from data acquisition to ML model development. Radiomic models developed in this thesis were based on MR images and PET scans, therefore, this section focuses on these modalities.

3.1 Image acquisition

As imaging data are collected during routine clinical practice, large data sets are — in principle — readily available, thus offering an incredibly rich resource for scientific and medical discovery. Image data are acquired at various stages, starting with diagnostic stages, both prior to and during treatment, but also during the follow-up of patients to assess response to treatment. Worldwide, the images are stored in DICOM (Digital Image and Communication in Medicine) format into the institutional archiving systems (PACS, Picture Archiving and Communication System). MR images and PET scans are mostly used for diagnosis, staging, and to help monitor the patient during and after therapy. For practical reasons, DICOM images are often converted to the Neuroimaging Informatics Technology Initiative (NIfTI) format for AI analyses[3].

3.2 Segmentation

Segmentation is a critical step in the radiomic process because features are extracted from the segmented volumes in radiomic studies. Most of the time, the region of interest (ROI) is the tumor itself. Radiomic features can be extracted from bidimensional ROI or from the whole 3D volume. Tumor segmentation is challenging because many tumors show unclear borders. It is contentious because there are ongoing debates over whether to seek ground truth or reproducibility and how much to rely on manual or automatic segmentation.[4]. To delineate as accurately as possible the tumor volume, taking into consideration real tumor borders, information accessible only through histology have to be combined with 3D images. Studies[5, 6] have explored this successfully. Manual segmentation by expert readers has been shown to be subject to high inter-reader variability. This method is also labor intensive and not always feasible for radiomic analysis based on very large data sets[7]. Nevertheless, manual segmentation remains an acceptable option if inter-expert variability is negligible, mostly in monocentric studies where images are delineated by the same expert or by few experts. In large multi-institutional studies, manual segmentation would be flawed even when performed by experienced specialists. Semi-automatic or fully automatic segmentation should be therefore preferred. Semi-automatic segmentation refers to automatic segmentation followed by manual checking and editing of the segment boundaries. Several methods for automatic segmentation methods have been proposed i.e., level-set-based active contour model[8, 9], deformable models like the active contour algorithm[10, 11], region growing technique using several constraints[12], K-means clustering[13], and fuzzy C-means clustering[14, 15]. One of the major challenges in radiomic studies and in radiotherapy in general concerns the implementation of fully automated delineation methods to reduce inter-reader variability. DL-based automatic segmentation methods are showing promising results in obtaining the most robust and reproducible contours[16, 17, 18].

3.3 Image pre-processing

Before image intensity standardization, some limitations inherent to each imaging modality must be addressed by means of image pre-processing. Image pre-processing refers to various techniques used in order to reduce image artifacts, increase comparability, and therefore the developed model's accuracy and generalizability. Some artifacts on MR images and PET scans are due to bad image acquisitions. If present prior to radiomic features calculation, those images should be excluded from the study. Here, pre-processing steps prior to radiomics extraction on MR and PET images are discussed.

In MR images, the signal returned is negatively impacted by poor radio frequency (RF) coil uniformity, static field inhomogeneity, RF penetration, gradient-driven eddy currents, and patient anatomy[19, 20]. In addition to these artifacts, the images acquired do not always have the same voxel size which is a source of variability in MRI-based radiomics[21]. These sources of varaibilities lead to the need to apply at least a bias field inhomogeneity correction coupled to a voxel size resampling on MR images.

Apart from biological factors and technical errors, the quality of PET images is influenced by a wide variety of factors, some inherent to the modality itself (partial volume effects, choice and arrangement of detectors, acquisition time, etc...) and others inherent to the selected reconstruction parameters. Several

techniques for denoising and limiting the partial volume effects have been investigated to improve SNR and spatial resolution and SUV quantification of images beyond what the reconstruction algorithm initially produced. Ideally, reconstruction parameters should be kept the same for all imaging data in a radiomics study. As for voxel size, considered the most important source of variability in radiomic features from PET images[22], there are two ways to reduce this effect, namely through reconstruction to the same voxel dimensions or by image interpolation[23]. Differences in SUV units would impact not only histogram-based features but also textural features. Due to these variations, most of the previously published radiomic studies have been performed using small, retrospective, and monocentric cohorts[24]. This is the main limitation of radiomic studies based on PET images.

In the following sections, we are discussing bias field correction in MR images and voxel resampling.

3.3.1 Bias field inhomogeneity

The static polarising magnetic field in MRI is supposed to be spatially invariant. Any deviation from this principle affects, among other things, the uniformity of the image, its spatial integrity, and blurring, and can lead to image artifacts. Bias field corresponds to the low-frequency background nonuniformities (inhomogeneities) of the image domain that exist in MR images due to inhomogeneity in the magnetic field. Although the bias field can not be always observed by human eyes, it can significantly affect the performance of many image processing, segmentation, and analysis techniques[25]. Nonparametric intensity nonuniformity normalization (N3) and its improved version (N4ITK)[26] are reference algorithms for intensity inhomogeneity correction. In this algorithm, the corrupted image is considered a multiplicative component of the uncorrupted image corrupted by a smooth polynomial bias field and an additive component as shown in the equation below:

$$v(x) = u(x)f(x) + n(x) \tag{3.1}$$

where v is the observed image, u is the uncorrupted image, f is the bias field, n is the noise (assumed to be independent and Gaussian), and x designates the spatial position or voxel. Assuming a noise-free scenario, the image model becomes:

$$\hat{v(x)} = \hat{u}(x) + \hat{f}(x) \tag{3.2}$$

where $\hat{u} = log(u)$, $\hat{f} = log(f)$. The particularity of N4ITK is that the B-spline fitting is improved compared to N3. Thus for N4ITK, from equation (3.1), the iterative component of the algorithm is optimized as described by the following equation:

$$\hat{u}^n = \hat{u}^{n-1} - \hat{f}_r^n = \hat{u}^{n-1} - \{S^*\hat{u}^{n-1} - E[\hat{u}|\hat{u}^{n-1}]\} \tag{3.3}$$

where S^* is a modified B-spline estimator, and $\hat{f}_r^n$ is the estimated residual bias field at the nth iteration.

3.3.2 Image denoising and partial volume correction

Noise reduction techniques in medical images can be broadly divided into two categories: techniques applied during the acquisition and techniques used after the acquisition of the image. Here, we are referring to techniques

used after image acquisition. Several digital processing techniques have been developed in medical imaging and applied to MR and PET images[27]. Basic techniques for image denoising are digital filters typically applying a low pass filter to remove peaks and replace certain values with a specific measure such as local average. Commonly digital filters are Gaussian averaging, mean, median, and Diffusion filters[28, 29]. However, these filters tend to remove small and lower contrast lesions along with the noise. Therefore more complex and computationally expensive techniques have been assessed: adaptive filters such as median adaptive filter [30] and bilateral filters[31], multi-scale methods, deep-learning (DL) based techniques, Genetic Algorithms (GA) and Fuzzy Logic (FL) based on multi-scale analysis have been implemented.

On MR images, the noise corrupting the images is Rician. State-of-the-art MR image denoising techniques include filtering methods, transform domain approaches, statistical techniques, algorithms utilizing sparsity and self-similarity, low-rank approximation[32, 33, 34]. Machine learning (ML) methods, including DL-based methods[35], have also been utilized for this task.

In PET imaging, denoising is mainly done by Gaussian smoothing and locally adaptive filtering[27]. The spatial width of the Gaussian smoothing filter must be selected carefully to provide a compromise between SNR and spatial resolution. Otherwise, it might result in excessive smoothing which in turn would result in loss of resolution and blurring of adjacent objects. Multi-scale transform coupled to non-local means filtering achieved also good results while preserving structures and details[36]. Bilateral filtering and multi-scale techniques have also been successfully applied. Similarly to denoising in MRI, ML methods, especially DL based, are a new trend in PET image denoising[37, 38, 39]. More details on image denoising for both modalities can be found in [27].

In PET images, the partial volume effects (PVEs) is a phenomenon encompassing two effects: image blurring and image sampling The first effect, 3D image blurring, is due to the limited spatial resolution of the imaging system (the detector design and the reconstruction process). This results in loss of apparent activity as part of the signal from the source "spills out" and hence is seen outside the actual sources. The second effect of PVEs, image sampling (also called the tissue fraction effect), is due to the fact that most voxels include signals coming from different type of tissues. The signal intensity in each voxel is the mean of the signal intensities of the underlying tissues included in that voxel. It is interesting to note that, even if the imaging system had perfect spatial resolution, PVE would still affect the image due to the image sampling effect. This explains why high-resolution imaging such as MRI and CT are still affected by PVE. Technically, the spatial resolution is regarded as the point-spread function (PSF), which essentially corresponds to the image of a point source[40]. The reconstructed PET image can be described in the frequency domain as the product of the Fourier transform of the true activity distribution and the modulation transfer function (MTF) of the system. The ultimate goal of partial volume correction (PVC) is to reverse the effect of the system PSF in a PET (or SPECT) image and thereby restore the true activity distribution, qualitatively and quantitatively[41].

PVC can be performed using two main approaches: image reconstruction with resolution modeling $\pm$ introduction of anatomical priors, and post-reconstruction image restoration. The first approach is mostly done by deconvolution during iterative image reconstruction by incorporating the PSF in the system matrix[42, 43, 44]. This is done to enhance spatial resolution during reconstruction. Several methods have been assessed to improve this approach: anatomically based PVC methods using structural information from other imaging modalities[45, 46], penalized iterative reconstruction algorithms[47, 48, 49], minimization of cross-entropy

method[50], etc. Unfortunately, each method has its limitation and no consensus has been reached on the best method. The second approach consists of performing image enhancement by post-processing the image reconstructed without resolution modeling (or in an attempt to reduce residual partial volume effects after reconstruction with resolution modeling)[41]. This can be performed either using a deconvolution operation on a reconstructed image or by incorporating into the reconstructed image some high-frequency information taken from a structural image[51, 52, 53, 54, 55, 56]. Similarly to the first approach, these methods suffer from several limitations and there is no consensus on which method is most appropriate.

To date, the available literature suggests that the denoising models in MRI and PET struggle to preserve small or low-contrast structures. Similarly, PVC in PET images did not show a significant impact on standard PET metrics and there is no evidence that pre-processing images with denoising and/or partial volume effects provide significant improvement for radiomics application. Therefore, for these reasons and given that no consensus exist on the best denoising or PVC method[24], image denoising and PVC were not applied in this thesis.

3.3.3 Voxel size harmonization

Radiomic features are supposed to be extracted from images whose voxels have a common voxel spacing, i.e, are isotropic. This is particularly relevant for textural features computed by taking into account the relationship among values of neighboring voxels[57]. Spatial resampling is thus indispensable to ensure comparability and rotational invariance of extracted radiomic features. Resampling should be done using high-order interpolator (cubic or b-spline for example) to avoid smoothing texture in the image. The masks should be resampled to isometric voxels also, using nearest-neighbor interpolation, or other similar methods. Voxel size harmonization covers both upsampling (decreasing spacing or increasing the number of voxels) and downsampling (increasing spacing or decreasing the number of voxels) depending on the given data.

3.4 Image intensities standardization

After image pre-processing, given that MR images are non-quantitative contrast images acquired on multiple devices using different acquisition and reconstruction parameters, and within a long time frame, before extracting features, MR images need to be standardized, to mitigate the so-called 'center-effect'. This problem can however be treated by a step of binning of the intensities which can be an alternative method to image standardization (relative binning) in addition to the need to group the close intensities to reduce the sensitivity to the noise and to decrease the computing time.

In PET images, only the intensity binning step is necessary because image intensities are converted into SUV units and considered normalized. Nevertheless, differences in SUV values can differ from 15% or 30% [58] due to acquisition or reconstruction parameters, respectively. Specifically, matrix size of the images and variation in voxel size were found to impact substantially the image intensity thus the SUV normalization while the choice of reconstruction method, number of iterations and iterations, and width of the Gaussian filter used for post-reconstruction smoothing might also increase differences in SUV values but to a lesser extent[59].

Several methods have been proposed in order to harmonize image intensities and allow reproducible radiomic studies[60, 25, 61, 62]. Among these methods, statistical methods based mainly on first-order statistics have been widely used.

3.4.1 Image normalization principles

For any voxel v from an MR image expressed in arbitrary units, measured for a subject i at time j, normalization can be described as any transformation of the type:

$$Y_{ij}(v) \rightarrow N_{ij}(Y_{ij}(v)) \tag{3.4}$$

In 2014, Shinohara et el.[61] introduced a set of 7 principles, which we refer to as the statistical principles of image normalization (SPIN) as an attempt to provide guidelines/definitions for better practices in regard to image normalization. They stipulated that any normalization process should produce units that:

- have a common interpretation across locations within the same tissue type
- are replicable
- preserve the rank of intensities
- have similar distributions for the same tissues of interest within and across patients
- are not influenced by biological abnormality or population heterogeneity
- are minimally sensitive to noise and artifacts and
- do not result in loss of information associated with pathology or other phenomena.

3.4.2 First-order normalization techniques

These techniques refer to normalization processes based on first-order statistics like minimum, maximum, median, mean intensity, or standard deviation of the intensities in the images or a specified ROI. The most popular are the Max-Min normalization, z-score, and histogram-based normalization.

Max-Min or Linear normalization

Max-min normalization (usually called feature scaling) performs a linear transformation on the original intensities and consists of rescaling all intensities in the range [0, 1]. Max-min normalization preserves the relationships among the original data values and consists of subtracting I_{min} corresponding to the minimum intensity of the image from each voxel intensity $I(x)$ and dividing the result by the difference between the image maximum intensity I_{max} and I_{min}:

$$I_{max-min}(x) = \frac{I(x) - I_{min}}{I_{max} - I_{min}} \tag{3.5}$$

The cost of having this bounded range is that we will end up with smaller standard deviations, which can suppress the effect of outliers. The max-min technique has been applied for MR image normalization in several radiomic studies[63, 64].

Z-score normalization

The Z-Score method normalizes image histograms by subtracting μ, corresponding to the mean intensity value of the image, from each voxel intensity I(x) and dividing the result by the standard deviation of the image intensity valuesσ:

$$I_{zscore}(x) = \frac{I(x) - \mu}{\sigma} \tag{3.6}$$

Z-score normalization sets mean intensity and standard deviation to 0 and 1 respectively. It is a straightforward, easy-to-implement technique and by far the most used for MRI normalization in radiomics[65, 66, 67, 68, 69, 70].

Histogram matching method

A histogram-matching method was detailed for the first time by Nyul et al.[60] and modified later by Nyul et al.[71] and Shah et al.[72]. The main idea underlying this method is two-fold: the first stage creates a template histogram, with landmarks of interest usually defined through averaging histograms in a reference population[71]. Then, for each subject in the study, the histogram is mapped via a piecewise linear transformation to the template defined using quantiles as knots[72]. The working principles is illustrated on Figure3.1:

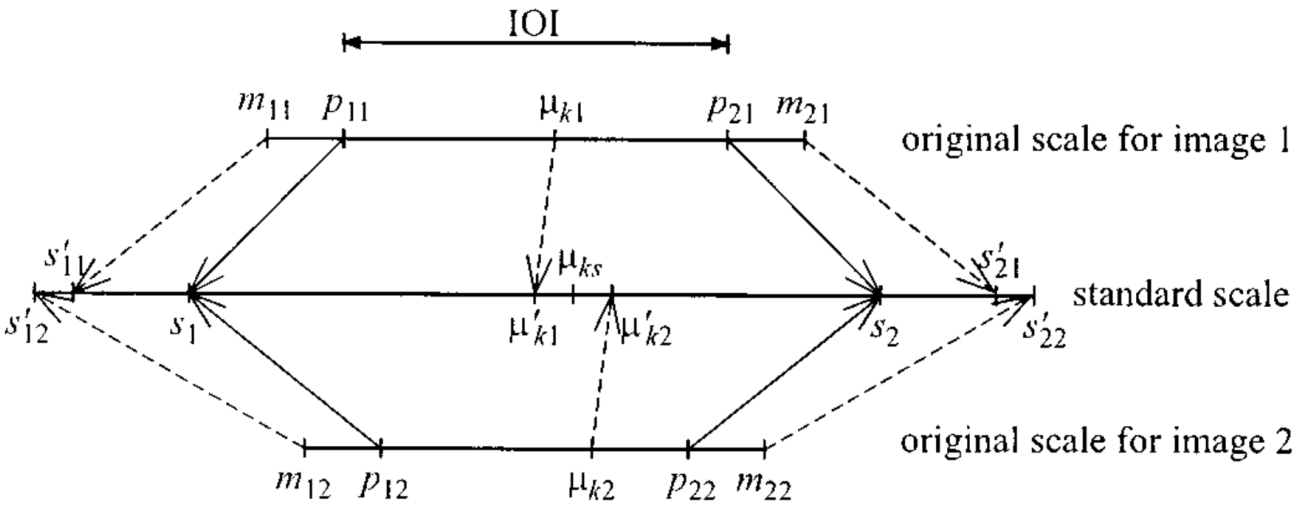


Figure 3.1: Histogram matching method as proposed by Nyul et al.[71]. Only two input images are shown. For $j = 1,2$, m_{1j}, m_{2j} are the minimum and maximum intensities in the image j, p_{1j} and p_{2j} are the minimum and maximum percentile intensities, μ_{kj} is one of the landmarks of the histogram, μ'_{kj} is the mapped value of μ_{kj}, and μ_{ks} is the mean of the μ'_{kj}: the actual parameter we are looking for on the standard scale. *Reprinted from[71].*

In Shah et al.[72], the authors demonstrate good results with this method by defining landmarks as intensity percentiles at 1,10,20,...,90,99 percents (where the intensity values below 1% and above 99% are extrapolated from the [1,10] and [90,99] percent intervals). This method has been applied to harmonize MR images in different radiomic studies showing good results[66, 73, 74, 67]. In the following chapters, this method will be referred to as 'Nyul'.

Pros and cons of cited methods Max-min normalization does not handle well outliers but z-score normalization suffers less from this limitation. However, z-score assumes that the intensities are normally distributed and does not consider the skewness and kurtosis of the intensities distribution which can lead to false conclusion

if the distribution is not symmetrical. Despite producing replicable results, being easy to implement, and being computationally fast, Nyul method presents some limitations. This method is based on assumptions that might be questionable: the distribution of tissue type is supposedly the same across subjects, abnormal pathologies are not considered among patients, and technical artifacts (for example, from patient motion and residual spatial inhomogeneity after correction) are not taken into account. To overcome these limitations, statistical normalization methods based on the intensity values of biological tissue were assessed.

3.4.3 Tissue specific standardization methods

Madhabushi et al.[75] developed statistical methods capable of automatically partition the image into homogeneous regions, corresponding to different tissue regions. Unlike Nyul method, which utilized the median and other quartile locations from the whole image histogram, this method used landmarks derived from the individual scale regions in the image to perform the nonlinear mapping of intensities. Similarly, Shinohara et al.[61] selected the normal-appearing white matter (NAWM) in the brain as reference tissue, adjusting the intensity distribution in other tissues accordingly. This method is called the WhiteStripe method. The image intensity in each voxel x after normalization using WhiteStripe is:

$$I_{WhiteStripe}(x) = \frac{I(x) - \mu_{ws}}{\sigma_{ws}} \tag{3.7}$$

where $I(x)$ is the voxel intensity before normalization, μ_{ws} the mean value intensity value of NAWM, σ_{ws} the standard deviation of the NAWM, and $I_{WhiteStripe}(x)$ the voxel intensity after WhiteStripe normalization. Many studies in brain cancers have applied the WhiteStripe technique to improve robustness of radiomic feeatures[66, 67]. Carré et al.[66] compared different normalization techniques on a tumor-grade classification task and recommended z-score as the best technique compared to Nyul and WhiteStripe. Li et al.[67] implemented WhiteStripe and three additional normalization techniques based on different values of intensity of the white matter. They concluded that Nyul, WhiteStripe, z-score, and techniques based on white matter tissue achieved similar results in radiomic features robustness enhancement.

Similar normalization methods based on other anatomical tissues have been developed. In HNC, Wahid et al.[63] developed a method based on cheek fat. This method standardized the image with respect to left and right cheek fat (healthy tissue) and was adapted from van Dijk et al.[76]. The intensity of each voxel was divided by the mean intensity of the cheek fats and multiplied by an arbitrary scaling value of 350 using the equation:

$$I_{cheekfat}(x) = \frac{I(x)}{\mu_{fat}} * 350 \tag{3.8}$$

where $I(x)$ and $I_{cheekfat}$ were the original and standardized intensities, respectively, and μ_{fat} was the mean intensity of both cheek fat ROIs per patient. Issakson et al.[74] implemented a prostate-specific normalization based on healthy prostate tissue intensities on MR images of prostate cancer patients. They demonstrated that normalization based on healthy prostate tissue intensities as a landmark had better results compared to Nyul. In a similar study on MR images of patients with prostate cancer, Scalco et al.[77] implemented MRI normalization based on the mean value of the urine in the bladder and found that this method had lower performances compared to Nyul.

3.4.4 Where do we stand on MR images standardization?

Despite the clearly promising results of different normalization methods in MRI-based radiomics, whether for reducing feature variability or improving the performance of the developed radiomic signatures, it is not clear which method should be preferred. All methods have pros and cons, and their implementation should be task motivated and accurately described. Here, it is important to report that standardization methods can also be applied a ***posteriori*** on extracted features. The commonly used method is the ComBat harmonization[78, 79, 80, 81] and will be discussed in subsection3.5.3. Many studies have shown the benefits of combining image normalization with feature harmonization using the ComBat method[82, 67, 83]. Orlhac et al. demonstrated that multicenter harmonization using WhiteStripe and ComBat could remove protocol-based variations in MR images of brain tumor patients (T1w and FLAIR images) acquired in two different centers[82]. Li et al.[67] also showed that image normalization techniques in MRI might not be enough and could benefit from the addition of ComBat.

3.5 Radiomic features extraction

3.5.1 Gray-level discretization

Intensity discretization is another required step in the workflow of radiomic studies. This step has to be applied right before features extraction as it consists of converting continuous values into discretized levels in order to group close gray levels together and reduce the impact of noise[57]. This step facilitates the calculation and comparison of texture features[84] among patients. Two methods are usually used for gray-level discretization: relative discretization also known as fixed bin number (FBN) or absolute discretization also known as fixed bin size (FBS).

Relative discretization consists of grouping the gray levels in a fixed number of bins between the minimum and maximum intensity of each ROI:

$$R(x) = (N_b * \frac{I(x) - I_{ROI,min}}{I_{ROI,max} - I_{ROI,min}}) + 1$$

where $R(x)$ is the resampled value in voxel x, $I(x)$ is the value in voxel x in the original image, N_b corresponds to the fixed number of bins, $I_{ROI,min}$ and $I_{ROI,max}$ are the minimum and the maximum intensity in the ROI respectively.

Absolute discretization consists of grouping the gray levels in bins of a fixed size between two bounds (I_{min} and I_{max}). It is described as follows:

$$R(x) = (N_b * \frac{I(x) - I_{min}}{I_{max} - I_{min}}) + 1$$

where $R(x)$ is the resampled value in voxel x, $I(x)$ is the value in voxel x in the original image, N_b corresponds to the fixed number of bins. I_{min} and I_{max} are mostly defined as the minimum and maximum voxel intensity across patients respectively.

The relation between the number of bins N_b and the bin width W_b is given by the following equation:

$$W_b = \frac{I_{max} - I_{min}}{N_b}$$

In PET, Leijnaar et al.[85] showed that FBS yields features with lower correlation with the corresponding tumor volume than with FBN. However, Orlhac et al.[57] found that FBS introduced spurious correlation with SUV. Goya-Outi et al.[86], in a study on multimodal MRI, showed that the choice of discretization was crucial in regard to modeling performance. Since there are no clear indications of which method is the best, gray-level discretization should be applied carefully in regard to image modality. For comparison purposes, it may be useful to systematically implement and report both[23, 24].

3.5.2 Features extraction

For a long time, radiomic features referred to handcrafted features, but they also include deep features extracted by layers of deep neural networks. Radiomic features should be extracted through a high-quality pipeline in order to be reproducible. The Image Biomarker Standardization Initiative (IBSI) has issued common definitions of radiomic features and guidelines to ensure calibration and verification of radiomics software[23].

Handcrafted features are usually categorized into the following subgroups:

Shape features : describe the shape of the selected ROI and its geometric properties such as volume, maximum diameter along different orthogonal directions, maximum surface, tumor compactness, and sphericity. For example, the surface-to-volume ratio of a spiculated tumor will show higher values than that of a round tumor of similar volume.

First-order statistics : these features describe the distribution of individual voxel values without concern for spatial relationships. These are histogram-based properties reporting the mean, median, maximum, and minimum values of the voxel intensities on the image, as well as their skewness (asymmetry), kurtosis (flatness), uniformity, and randomness (entropy).

Second-order statistics : these features refer to textural features, which are obtained by calculating the statistical inter-relationships between neighboring voxels [87]. They provide a measure of the spatial arrangement of the voxel intensities, and hence of intra-lesion heterogeneity. Such features can be derived from the gray-level co-occurrence matrix (GLCM), quantifying the incidence of voxels with the same intensities at a predetermined distance along a fixed direction, or from the Gray-level run-length matrix (GLRLM), quantifying consecutive voxels with the same intensity along fixed directions. They can also be derived from the gray-level size zone matrix (GLSZM) measuring the number of neighboring voxels with the same value for any voxel value, or from the neighborhood gray-level difference matrix (NGDLM) calculating the difference between neighboring voxels based on voxel values. Textural features can be computed either in two dimensions (2D), three dimensions (3D), or in 2.5 dimensions, meaning that values are calculated in 2D for each slice and then averaged over all slices in the volume[88].

Higher-order statistics : these features are obtained by statistical methods after applying filters or mathematical transforms to the images; for example, with the aim of identifying repetitive or non-repetitive

patterns, suppressing noise, or highlighting details. These include fractal analysis, Minkowski functionals, wavelet transform, and Laplacian transforms of Gaussian-filtered images, which can extract areas with increasingly coarse texture patterns.

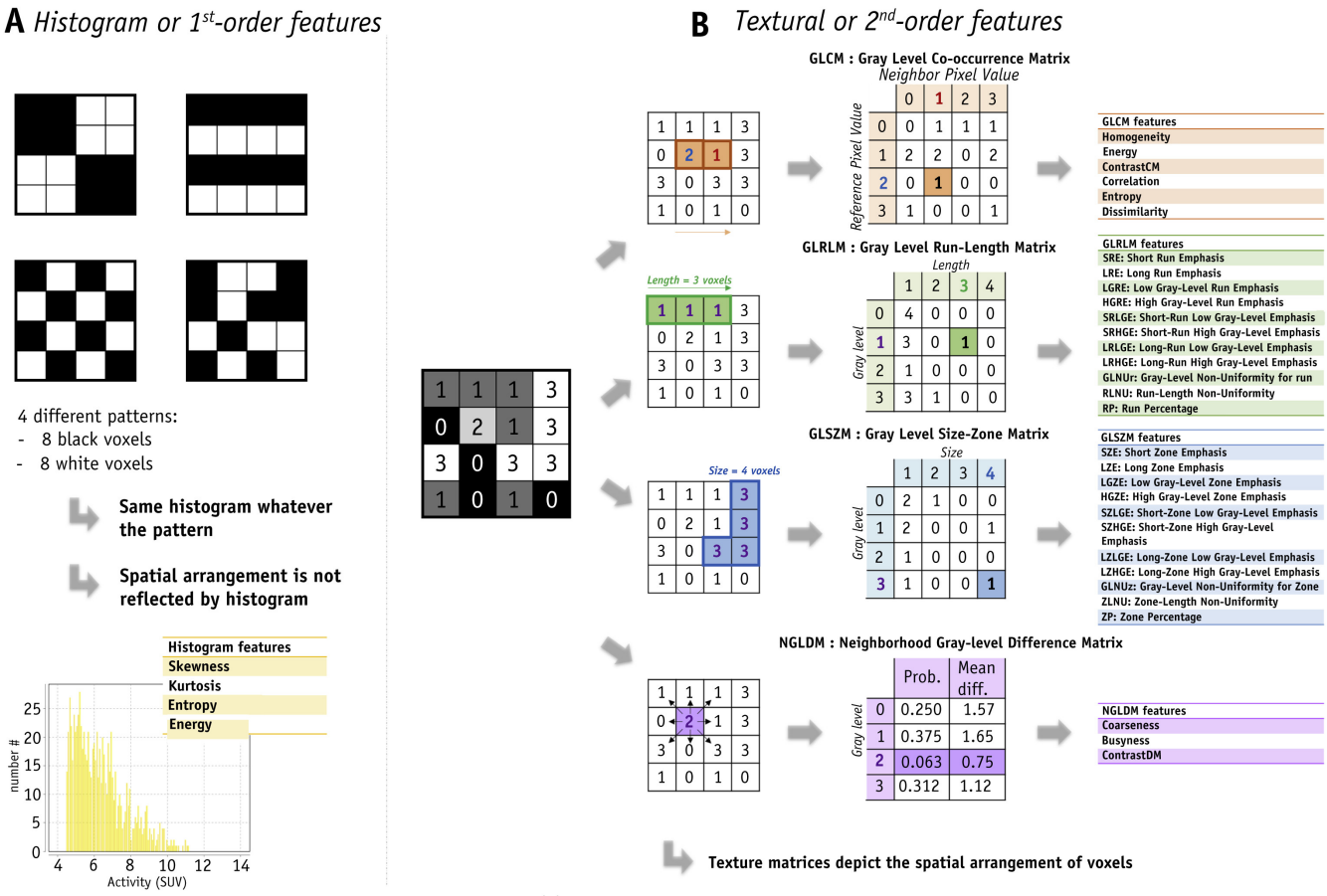


Figure 3.2: Radiomic features. (A) Histogram or first-order features reflect voxel intensity distribution only. (B) Textural or second-order features derived from texture matrices reflect the complex and unique spatial arrangement of voxels. *Reprinted from [89] under the terms and conditions provided by Elsevier and Copyright Clearance Center.*

Deep features are features extracted from an image by a neural network, relevant to a specific task without being predefined by the user. These features are self-designed in contrast to handcrafted features. Multiple studies have highlighted that models using deep features outperformed those using handcrafted features[90, 91] but have also demonstrated the benefits of integrating both handcrafted and deep features[92, 93].

3.5.3 ComBat Harmonizaztion

ComBat is a technique commonly used for harmonization of features extracted from multiple centers, i.e devices with different acquisition and reconstruction parameters to mitigate the 'center effect' in radiomics. Originally developed for combating the batch effect in microarray genomic analysis [78], it was later demonstrated that the ComBat harmonization technique successfully removes inter-site technical variability, while preserving inter-patient biological variability on MR images[79, 80]. ComBat harmonization implements the location (mean) and

scale (variance) (L/S) method. The main goal of the L/S method is to set the same mean and/or variance on all batches. Correcting for the batch effect consists of estimating additive and multiplicative parameters accounting for batch effects for each radiomic feature in a given batch while taking into account present covariates. To improve the estimations of these parameters, ComBat harmonization technique can be applied with an empirical Bayes (EB) method. When ComBat is applied with the EB method, instead of running L/S on each feature separately, the L/S model parameters are estimated by pooling features in each batch to "borrow information" across features to "shrink" the batch effect parameter estimates toward the overall mean of the batch effect estimates. The EB method can be used with a parametric or a non-parametric approach. Here we applied ComBat harmonization with EB estimates and non-parametric prior method. The L/S adjustment method considers the value of a feature f as:

$$Y_{ijf} = \alpha_f + X\beta_f + \gamma_{if} + \delta_{if}\epsilon_{ijf}, \tag{3.9}$$

where Y_{ijf} represents the value corresponding to feature f for sample j from batch i, α_f is the overall feature expression, X is the covariates matrix, and β_f is the vector of regression coefficients corresponding to X. The error terms, ϵ_{ijf}, can be assumed to follow a Normal distribution with expected value of zero and variance σ_f^2. The γ_{if} and δ_{if} represent the additive and multiplicative batch effects of batch i for feature f, respectively. Correcting for the batch effect consists then of estimating the parameters γ_{if} and δ_{if} using the following equation:

$$Y^*_{ijf} = \frac{Y_{ijf} - \hat{\alpha}_f - X\hat{\beta}_f - \hat{\gamma}_{if}}{\hat{\delta_{if}}} + \hat{\alpha_f} + X\hat{\beta}_f, \tag{3.10}$$

where $\hat{\alpha}_f$, $\hat{\beta}_f$, $\hat{\gamma}_{if}$ and $\hat{\delta_{if}}$ are estimators of parameters α_f,β_f, γ_{if} and δ_{if}.

As stipulated in section 3.4, the benefit of using ComBat, combined with or without other normalization techniques, in multi-institutional studies has been demonstrated[82, 67, 83, 94]. However, studies have also shown mixed results with ComBat showing no improvement compared to raw features[95] or underperforming in increasing robustness of radiomic features compared to a quantile normalization technique[96]. Despite these results, the benefit of ComBat is now widely accepted and efforts have been made to improve this method[81, 97, 98, 99, 100]. An electronic search was performed on PubMed with the keywords "ComBat" and "radiomics" (80 articles were found). Table 3.1 presents some of the recent studies using ComBat harmonization in the radiomics field for cancer management.

Author (year of publication)	Modality - number of scanners	Standardization	Task (patients]	Results
Leithner et al.[101](2023)	MRI (T1w, T1c, T2w, and FLAIR) - 2	ComBat-B and ComBat-NB	Impact of ComBat harmonization on nonbinary tissue classification: liver, spleen, and paraspinal muscle (80)	ComBat-B and ComBat-NB harmonized features achieved higher accuracy than unharmonized radiomic features ($P = 0.005$)
Lucia et al. [102](2023)	18F-FDG PET/CT and MRI -	ComBat	Prediction of PALN involvement in LACC (239)	18F-FDG PET/CT radiomic features harmonized with ComBat outperformed clinical variables in prediction of PALN involvement. C-statistic was in the range of 0.88- 0.96 and 0.57-0.67 for radiomic features and clinical variables, respectively
Salome et al.[103] (2023)	MRI (T1w, contrast-enhanced T1w , T2w, and T2w-FLAIR) - 29	Fuzzy C-Means, kernel density estimation, Gaussian mixture models, Nyul, WhiteStripe, z-score, Combat	Impact of MR intensity normalization methods on overall survival prediction in primary and recurrent high-grade glioma (338)	ComBat and Nyul methods were found to be the top performing for both MRI sequences with C-ondex values ranging from 0.66 to 0.68

Continue on the next page

Table 3.1: My caption (cont.)

Author (year of publication)	Modality - number of scanners	Standardization	Task (patients]	Results
Zounek et al. [104] (2023)	[18F]FET PET and [18F]GE-180 PET	ComBat	Effect of ComBat harmonization on radiomic feature extracted of pooled [18F]FET or [18F]GE-180 PET images of gliomas acquired with different reconstruction settings (19)	ComBat resulted in almost complete harmonization (87%) according to Friedman test and little to no improvement according to the coefficient of variation and inter-rater reliability measures
Xu et al.[105](2023)	18F-FDG PET/CT	8 different strategies for ComBat harmonization: before or after feature selection, in feature groups separately or all features directly, and with center or clustering-determined labels.	Assess the impact of several ComBat harmonization strategies, intra-tumoral sub-volume characterization, and automatic segmentations for PFS prediction through radiomics modeling for patients with HNC (325)	The best performances were obtained when ComBat was applied after feature selection, with C-index values ranging from 0.65-0.66. The best model was clinical variables with radiomic features harmonized by ComBat applied after FS, for feature groups separately and relying on clustering-determined labels with C-index of 0.71. The use of radiomics from tumor sub-volume characterization or automatically segmented achieved similar results

Continue on the next page

Table 3.1: My caption (cont.)

Author (year of publication)	Modality - number of scanners	Standardization	Task (patients]	Results
Song et al.[106] (2023)	MRI (T2w, DWI, and fat-saturation contrast-enhanced T1w) - 5	ComBat	Assessment of microsatellite instability (MSI) status in (EC) (338)	A higher MSI prediction performance was achieved after ComBat harmonization method. The SVM algorithm had the best performance, with AUCs of 0.921, 0.903, and 0.937 in the training, internal validation, and external validation cohorts, respectively
Castaldo et al.[70] (2022)	MRI (T2w)	ComBat, ARSyNseq, SVA, mixed effect - 2	Comparison of tree harmonization methods of radiomic features extracted from MRI of prostate cancers from different centers (210)	The ComBat method outperformed all other methods by achieving 70% accuracy and 78% AUC with the random forest method in classification of patients affected by prostate cancer
Bos et al.[96] (2023)	MRI (T1w)	ComBat, quantile normalization - 2	Validation of an MR-based radiomics model predictive of LRC in OPSCC patients	Quantile normalization was found more robust than ComBat harmonization. With appropriate harmonization, radiomics can be applied on data acquired from different vendors and protocols

Continue on the next page

Table 3.1: My caption (cont.)

Author (year of publication)	Modality - number of scanners	Standardization	Task (patients]	Results
Du et al.[107] (2023)	CT	deep-learning based harmonization method, ComBat	Comparison of a deep-learning-based harmonization method with other well-known harmonization methods for compensating the kernel effect in CT images on LNM prediction (85) and for differential diagnosis between LC and TB (164)	The convolutional neural network (CNN) harmonization methods outperformed all other harmonization methods. For LC vs TB diagnosis: best AUCs of CNN vs. others were 0.87 vs. 0.54-0.86 (p = 0.0001-0.55). For LNM prediction, best AUCs of CNN vs. others were 0.78 vs.0.70-0.73 (p = 0.07-0.40)
Horng et al. (2022)[99]	CT	Nested ComBat, GMM ComBat - 48	Generalized ComBat harmonization methods for radiomic features with multi-modal distributions and multiple batch effects in NSCLC(281)	GMM ComBat was seen to be the best model for radiomic features robustness. Nested ComBat and GMM ComBat achieved similar results on the whole dataset (C-index of 0.63). Both methods had higher performances compared to ComBat

Continue on the next page

Table 3.1: My caption (cont.)

Author (year of publication)	Modality - number of scanners	Standardization	Task (patients]	Results
Horng et al. (2022)[98]	CT	OPNested ComBat and OPNested +/− GMM ComBat - 48	Improvement of GMM ComBat harmonization method for multiple batch effects (281)	OPNested + GMM ComBat and OPNested - GMM ComBat exhibited the best harmonization performance, and the best C-statistics for survival prediction. However, for survival prediction ComBat achieved similar results as OPNested - GMM ComBat
Paquier et al.[108] (2022)	DW-MRI	ComBat - 6	Role of ComBat harmonisation in DW-MRI for radiomic analysis	ComBat method could improve significantly the within-subject coefficients of variation compared to reproducibility without ComBat in radiomic features (p-value $<$ 0.001)
Carré et al. [100] (2022)	MRI (T1w, T1w-Gd, T2w-flair)	ComBat, AutoComBat harmonization after z-score image normalization	Comparisosn of AutoComBat and ComBat harmonization techniques on MRI of patients with low-grade glioma and glioblastoma (243)	On T2w-flair, AutoComBat, using either metadata plus quality metrics or metadata alone as inputs, performs better than the conventional ComBat, highlighting its potential for data harmonization

Continue on the next page

Table 3.1: My caption (cont.)

Author (year of publication)	Modality - number of scanners	Standardization	Task (patients]	Results
Frood at al.[94] (2022)	18F-FDG PET/CT	ComBat	Assessment of Pre-treatment FDG PET/CT-derived radiomic features for outcome prediction in classical Hodgkin lymphoma (289)	Pre-treatment FDG PET/CT-derived radiomic features achieved accuracy of 0.81 after ComBat harmonization in predicting 2-year event-free survival in classical Hodgkin lymphoma
Chang et al.[109] (2022)	MRI (T2w)	modified cycle GAN coupled to histogram matching, ComBat, CLAHE	CycleGAN harmonization method for image standardization	The proposed method overperformed ComBat and CLAHE in reducing variabilities in radiomic features and in predicting LNM in LACC (accuracy of 0.89)
Refaee et al.[110]	Phantom CT scans - 5	Reconstruction Kernel Normalization (RKN) and ComBat harmonization	Effect of Reconstruction Kernel Normalization (RKN) technique, ComBat harmonization in CT Reconstruction Kernels (28)	ComBat improves the number of robust features in 54% and RKN in 28% of scenarios. Combined, ComBat and RKN achieved features robustness statistical improvement in 79% of scenarios
Acquitter et al.[83] (2022)	MRI (T1w, contrast-enhancement T1w , T2w, and FLAIR	WhiteStripe coupled with ComBat - 7	Detection of radionecrosis after treatment of Glioblastoma using harmonized features extracted from multiparametric MRI (28)	Harmonized radiomic models achieved better predictive. Balanced accuracy of 0.61 with the model based on raw data and 0.72 with ComBat harmonization

Continue on the next page

Table 3.1: My caption (cont.)

Author (year of publication)	Modality - number of scanners	Standardization	Task (patients]	Results
Whitney et al.[95] (2021)	DCE-MRI	ComBat - 5	Classification of breast cancer (benign vs malignant) after features harmonization	Pre- and post-harmonized radiomic features achieved similar results
Li et al.[67] (2021)	MRI (T1w)	z-score, WhiteStripe, Nyul, FCM-based, GMM-based, and KDE-based normalization with or without ComBat	The scanner effects at the radiomic feature level still exist after applying these intensity normalization methods. ComBat can be applied in combination with image normalization to remove the scanner effect	
Da-ano et al.[81] (2020)	MRI and 18F-FDG PET/CT	M-ComBat, B-ComBat, BM-ComBat, ComBat	Evaluation of three modified versions of ComBat in predicting patients with local failure in LACC (197) and lack of response in LALC (98)	Scanner effect was successfully removed with all ComBat versions. The predictive performance of the radiomic models was always improved with harmonization and the improved ComBat versions provided the best results
Da-ano et al.[111] (2021)	MRI and 18F-FDG PET/CT	Transfer learning technique applied to M-ComBat, B-ComBat, ComBat	Integrate transfer learning technique into ComBat and its improved versions	The proposed TL approach enables applying a previously determined ComBat transform to new, previously unseen data

Continue on the next page

Table 3.1: My caption (cont.)

Author (year of publication)	Modality - number of scanners	Standardization	Task (patients]	Results
Badic et al.[97]	CT	B-ComBat	Prediction of recurrence after surgery in colorectal cancer patients (193)	After B-ComBat harmonization the prediction AUC increased from [0.52-0.63] to [0.64-0.79]
Ferreira et al.[112] (2021)	18F-FDG PET/CT	ComBat - 3	Predict disease-free survival in cervical cancer (158)	ComBat did not improve the predictive performance of the best models
Saint Martin et al.[113] (2021)	MRI (T1, T2, and T1-DCE) phantoms	Z-score or Nyul, coupled to ComBat	Pipeline to increase reproducibility in breast MRI radiomic studies	Harmonisation by ComBat lowered the percentage of radiomic features significantly different between the three coils from 87% after bias field correction and MR normalisation to 3% in the gel, while preserving or improving performance of lesion classification in the phantoms
Garau et al.[114] (2020)	Low-dose CT	ComBat	Impact of the Combat harmonization method to compensate for multicenter datasets variabilities in early lung cancer diagnosis (182)	No significant improvements were observed when applying the Combat harmonization method to the classification of lung nodules from multicenter data

Continue on the next page

Table 3.1: My caption (cont.)

Author (year of publication)	Modality - number of scanners	Standardization	Task (patients]	Results

Table 3.1: Radiomic studies assessing ComBat harmonization performance in radiomics. AUC; Area under the Receiving Operating Curve LDA; Linear Discriminant Analysis. ComBat-B; ComBat harmonization with empirical Bayes estimation. ComBat-NB; ComBat harmonization without empirical Bayes estimation. LACC; Locally Advanced Cervical Cancer. LALC; Locally Advanced Laryngeal Cancer PALN; Para-Aortic Lymph Node. GLCM; Gray-Level Co-occurrence Matrix. GLZLM; Gray-Level size Zone Matrix. LZLGE; Large Zone Low Gray-Level Emphasis. PFS; Progression-Free Survival. DWI; diffusion-weighted imaging. DCE; Dynamic Constrast-Enhanced. EC; Endometrial Cancer. SVM; Support Vector Machine. ARSyNseq; Anova-simultaneous component analysis (ASCA) Removal of SYstematic Noise SVA; Surrogate Variable Analysis. LRC; Loco-Regional Control. OPSCC; Oropharyngeal Squamous Cell Carcinoma. LNM; lymph node metastasis. LC; Lung Cancer. TB; Pulmonary Tuberculosis. GMM; Gaussian Mixture Model. FCM; Fuzzy C-Means. KDE; Kernel Density Estimate. NSCLC; non-small cell lung cancer. OP; Optimized order. GAN; Generative Adversarial Network. CLAHE; Contrast Limited Adaptive Histogram Equalization. M-ComBat allows transforming all features distributions to a chosen reference, instead of the overall mean, providing more flexibility. B-ComBat adds bootstrap and Monte Carlo for improved robustness in the estimation. BM-ComBat combines both modifications. SVD; Singular value decomposition

3.6 Radiomics-based AI models

AI is broadly a set of systems that can accurately perform inferences from a large amount of data, based on advanced computational algorithms[2]. The relationship between radiomics and AI is mutual. Due to its ever-growing high-dimensional nature, the field of radiomics needs much more powerful analytic tools, and AI appears to be a potential candidate for this purpose, with its extreme capabilities. On the other hand, in medical image analysis, AI applications inevitably need the radiomics because the features that are used to train and build the AI models are delivered through radiomic approaches, specifically, feature extraction and feature engineering techniques[115]. It has become clear that AI methods will have a strong impact in oncology. It is a priority to establish good practices in AI applied to medical imaging for successful clinical implementation.

3.6.1 Data preparation

Data preparation is by far the most important and most difficult step in any AI project. It is a crucial step for accurate and robust ML models building. Given that we have standard implementations of highly parameterized machine learning algorithms in open source libraries[116, 117], fitting models has become routine. As such, the most challenging part of each predictive modeling project is how to prepare the one thing that is unique to the project: the data. Any AI model is only as good as the quality of the training data. This is often summarized as garbage in, garbage out[118]. Also, data can be very complex, containing compressed complex nonlinear relationships that may need to be exposed, and messy with statistical noise, errors, missing values, and conflicting examples. Messy data is often dealt with using deep domain expertise. Thus, incorrect values, such as being mistyped, corrupted, duplicated, missing, and other obviously erroneous observations, are identified and removed. In medical images, this means checking for corrupted images, artifacts, poor quality, etc.

Before applying any data preparation technique, one must clearly define the problem at hand, gather all available information and discuss the project with subject matter experts. At this step, several input variables might be included or disregarded. The outputs have also to be correctly selected. Visualizing selected inputs and using statistical methods can, at this level, provide a first glance at the quality of selected inputs. Outliers for a variable or patient outliers can be identified by plots. This can help with data cleaning. It may also provide insight into the probability distribution that underlies the data. This may help in determining whether data transforms that change a variable's probability distribution would be appropriate. Statistical hypothesis tests can be used to determine whether a variable matches a given probability distribution. There may also be interplay between the data preparation step and the evaluation of models. The choice of algorithm is based on information on the underlying mechanisms which can in turn inform the selection and configuration of data preparation methods. For example, input variables might be required to be presented in a specific way or have a particular probability distribution given the algorithm we desire to train. In some cases, variables must be encoded (into binary for instance) or transformed before we can apply a machine learning algorithm, such as converting strings to numbers. We might need to apply different methods on different input variables or different subsets of input variables may require different sequences of data preparation methods. Choosing which methods to apply to the variables can be overwhelming, given the large number of methods, each of which may have its own configuration and requirements. Performance metrics may also influence the preparation of

the target variable such as scoring regression models based on prediction error using a specific unit of measure. This highlights the highly iterative nature of any predictive modeling project. Minimal data preparation should be targeted but this should be done in a proper way, reducing errors of any kind (statistical, noise, incorporating missing values and conflicting examples) but in such a manner it does not erase useful patterns. The broader philosophy of data preparation is to discover how to best expose the underlying structure of the problem to the learning algorithms[119].

In radiomics studies, raw images typically cannot be used directly, the raw data must be pre-processed prior to being used to train and evaluate an ML model. Data preparation includes data acquisition, image pre-processing, and feature extraction (see the previous section from section3.3 to section3.5.2. This step is not only image modality specific but also depends on the goals of the project, and the algorithms that will be used to model your data. For example, bias field correction only concerns MR images but voxel resampling prior to radiomic extraction is for all modalities.

3.6.2 Data splitting

After data preparation, the data set is conventionally split into training, validation, and test sets. In AI in general, when training a model, poor generalization on unseen data has to be avoided. To ensure that the trained model is robust and its performance is not just limited on the training set, the model has to be tested on another set of unseen data. Data splitting refers to dividing the dataset into two or more subsets: the training, validation, and test sets. The training set should be used to train the model, the validation set used for model optimization, and the final model performance should be evaluated on the test set. Data splitting is data dependent and different strategies can be preferred as illustrated in Figure 3.3. To this day, there is no consensus on data splitting but common split percentages include: 80%-20%, 67%-33%, 50%-50% for train (training and validation) and test sets respectively[119]. Several techniques for data splitting in presence of limited data have been developed to avoid overly optimistic results in addition to the classical train-test splits:

Cross-validation(CV) : CV is a statistical method of evaluating and comparing learning algorithms by dividing the training set into two segments: one used to learn or train a model and the other used to validate the model or hold-out validation. The training and validation sets must cross over in successive rounds such that each data point has a chance of being in both sets. The basic form of CV is k-fold CV. In k-fold CV, the data is first partitioned into k equally (or nearly equally) sized segments or folds. Subsequently, k iterations of training and validation are performed such that within each iteration a different fold of the data is held-out for validation while the remaining $k-1$ folds are used for learning. Variations of the CV have been assessed such as the leave-one-out CV (LOOCV), which is a special case of k-fold CV where k equals the number of instances in the data or the repeated K-Fold CV, which is a k-fold CV run several times. Further details can be found in [120]. CV procedure can be used both when optimizing the hyperparameters of a model on a dataset, and when comparing and selecting a model for the dataset. Using the same CV procedure and dataset is likely to lead to an optimistically biased evaluation of the model performance. To overcome this bias, one can use a nested CV. In ML, classes are often imbalanced. It is therefore recommended to apply CV with stratified sampling, i.e, stratified

CV. Stratified CV will return folds by preserving the percentage of samples for each class.

Nested Cross-validation (nCV) : In nCV, data is split into k outer folds and then inner folds are created in each outer training set to select features, train models, and tune parameters. An outer cross-validation procedure is performed to provide a performance estimate used to select the optimal model. In each fold of the outer cross-validation, the hyperparameters of the model are tuned independently to minimize an inner cross-validation estimate of generalization performance. nCV should also be applied with stratified sampling.

Data preparation methods that may induce data leakage, where knowledge of the hold-out test set leaks into the dataset used to train the model, must not be applied to the whole dataset as this could result in an incorrect estimate of model performance. Methods such as normalization of the input variables, or feature selection should not be applied to the whole dataset.

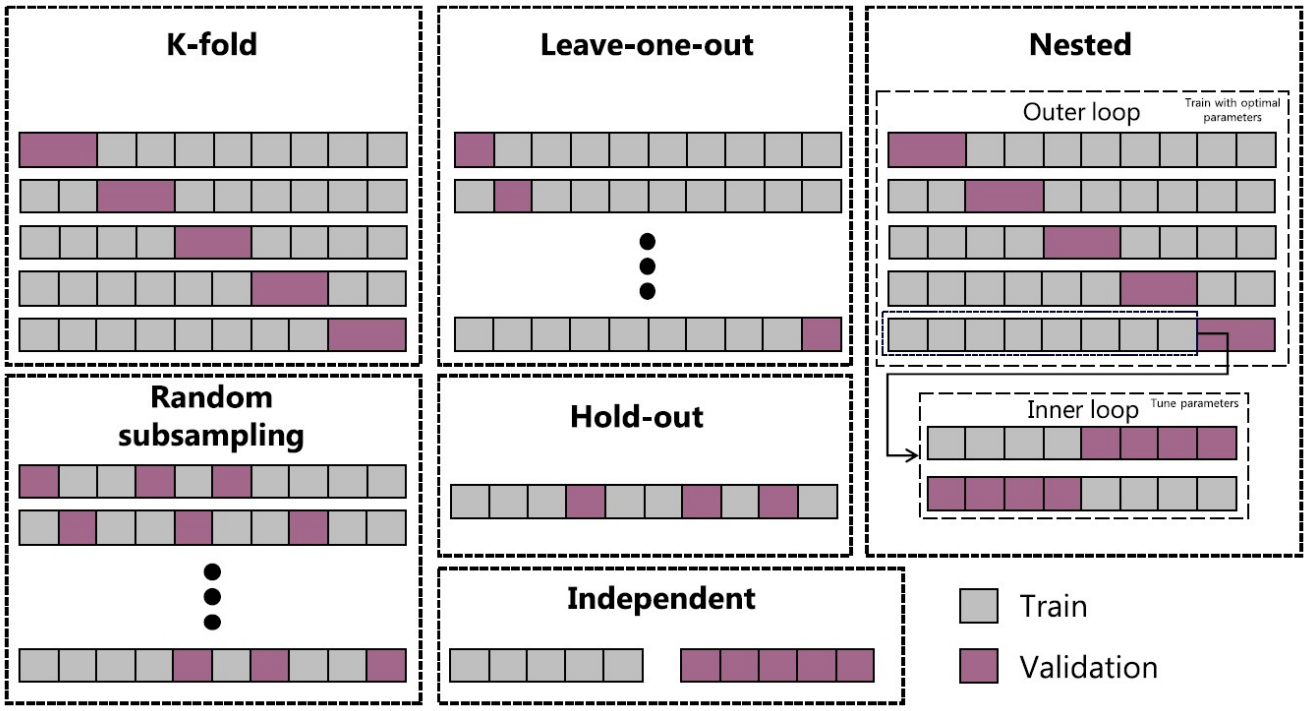


Figure 3.3: Simplistic representation of validation schemes. K-fold CV splits all samples (n) into equal-sized groups known as folds (if K = n, this is identical to the Leave-One-Out method). K-1 folds are used to learn the prediction function, and the fold that is left out is utilized for validation. A nCV consists of an inner loop CV enclosed inside an outer CV. The inner loop controls model selection and hyperparameter tuning (similar to the validation set), while the outer loop is in charge of error estimates (similar to test set). Random subsampling randomly splits the data set into training and validation subsets. Unlike k-fold cross-validation, where the data set is divided into folds, the split here is random. A single split is formed using random sampling in the hold-out approach. Independent validation uses a new data set, preferably an external data set, for the validation part. CV; Cross-validation, nCV; nested cross validation *Reprinted from[121].*

3.7 Model building

3.7.1 Feature selection

Feature selection refers to techniques for selecting a subset of input features that are most relevant to the target variable that is being predicted. This is important as irrelevant and redundant input variables can distract or mislead learning algorithms possibly resulting in lower predictive performance. Furthermore, in most radiomic studies, the number of patients is lower than the number of features, which generates a high-dimensional data problem, leading to overly optimistic results and preventing the generalizability of the trained models. Additionally, it is desirable to develop models only using the data that is required to make a prediction, e.g. to favor the simplest possible well-performing model[119]. There are two main types of feature selection techniques: supervised and unsupervised. When the outcome is ignored during the elimination of predictors, the technique is unsupervised[122].

Supervised techniques for feature selection may be divided into wrapper, filter, and intrinsic:

Wrapper methods : These methods evaluate multiple models using procedures that add and/or remove predictors to find the optimal combination that maximizes model performance[122]. Model performance maximization can be based on p-values, R-squared, Adjusted R-squared (in case of regression models), or accuracy, precision, and recall (for classification models). Although effective, these approaches can be computationally very expensive, especially for large training datasets and sophisticated models. Some of the most wrapper methods are recursive feature elimination (RFE), and Bi-directional elimination (Stepwise Selection)[123].

Filter methods : these are statistical methods that evaluate the relevance of the predictors outside of the predictive models and subsequently select only the predictors that pass some criterion[122]. Most of these techniques are univariate, meaning that they evaluate each predictor in isolation. In this case, the existence of correlated predictors makes it possible to select important, but redundant, predictors. The obvious consequences of this issue are that too many predictors are chosen and, as a result, collinearity problems arise. Different statistical tests are used: Pearson's test, Spearman's test, T-test, ANOVA, Chi-squared test, etc.

Intrinsic methods : These methods refer to ML algorithms' built-in feature selection, meaning that the model will only include predictors that help to maximize accuracy. In these cases, the model can pick and choose which representation of the data is best[122].

Feature selection is also related to dimensionality reduction techniques in that both methods seek fewer input variables to a predictive model. The difference is that feature selection selects features to keep or remove from the dataset, whereas dimensionality reduction creates a projection of the data resulting in entirely new input features. Dimensionality reduction is often used for data visualization. Dimensionality reduction yields a more compact, more easily interpretable representation of the target concept, focusing the user's attention on the most relevant variables[124] by projecting the data to a lower dimensional subspace which captures the "essence" of the data. The most common approach to dimensionality reduction is called principal components analysis

or PCA[125]. There are other techniques for dimensionality reduction such as manifold learning techniques or auto-encoders.

Feature selection should be applied with great caution and should never be applied on the whole data set or hold-out folds in cross-validation. Otherwise, it will lead to a selection bias due to the leak of information by the pre-filtering of the features[126]. There is no best method for feature selection; feature selection techniques must be carefully chosen depending on the inputs and the output response variables.

3.7.2 Machine Learning

Machine learning (ML) is a branch of AI and computer science that uses statistical and optimization methods, to train algorithms to uncover key insights in given data. Unlike, expert systems that are made for problem-solving, ML purpose is learning. An expert system requires a strict set of rules to mimic experts' decision-making process. Three main types of ML models can be distinguished based on the presence or absence of human influence on raw data, specific feedback or reward, or based on predefined labels: i) Supervised learning where all the datasets being used has been pre-labeled by users to allow the algorithm to see how accurate its predictions are. Some of the most used algorithms for supervised learning are polynomial regression, random forest, linear regression, logistic regression, decision trees, K-nearest neighbors, and convolutional neural networks. Typical supervised tasks involve function approximation, like regression and classification. Classification can be binary, or involve multiple classes, as in determining a particular pathology among several labels, or concern not the whole image but each pixel, as done for image segmentation[127]. Regression can also be done in a pixel-wise way, image enhancement (e.g. improving a low-quality image, the input, by mapping it to its higher quality counterpart, the output label or annotation) or image-to-image mapping (e.g. mapping a CT image, the input, to the corresponding dose distribution, the output). ii) Unsupervised learning where the models learn hidden patterns on the dataset without pre-existing labels. These models are mostly used for probability density estimation, like clustering (finding separated groups of similar data items), outlier or anomaly detection (isolated items), or even manifold learning and dimensionality reduction (subspaces on which data concentrate). Unsupervised learning has been, so far, less used than its supervised counterpart, although useful applications for medical imaging exist, such as domain adaptation (e.g., adapting a segmentation model trained on an image modality to work on a different image modality) [128], data generation (e.g., generate artificial realistic images) [129, 130] or even image segmentation[131]. The most used algorithms are partial least squares, fuzzy means, singular value decomposition, K-means clustering, apriori, hierarchical clustering, principal component analysis, and convolutional neural networks. A hybrid framework halfway between supervised and unsupervised, called semi-supervised learning, uses both labeled and unlabeled data in a training dataset. Groups identified as clusters by unsupervised learning can be used as possible class labels. Semi-supervised learning has been applied for the generation or translation of images from a specific class to another in a semi-supervised setting (e.g., generation of synthetic CTs from MR images)[132], and segmentation or classification of images with partially labeled data[133]. Generative Adversarial Networks have been widely used in semi-supervised learning mostly for synthetic image generation[134, 135] but also for segmentation[136] and recently for image normalization to improve reproducibility and discriminative power of radiomic features[137, 138, 139]. iii)Reinforcement learning which gives feedback to the algorithm to learn from its own experiences by trial and error using a

"rewards/punishments" system. Reinforcement learning for medical imaging is still scarcely used but promising applications mimicking physician behavior for typical tasks have been assessed[140, 141].

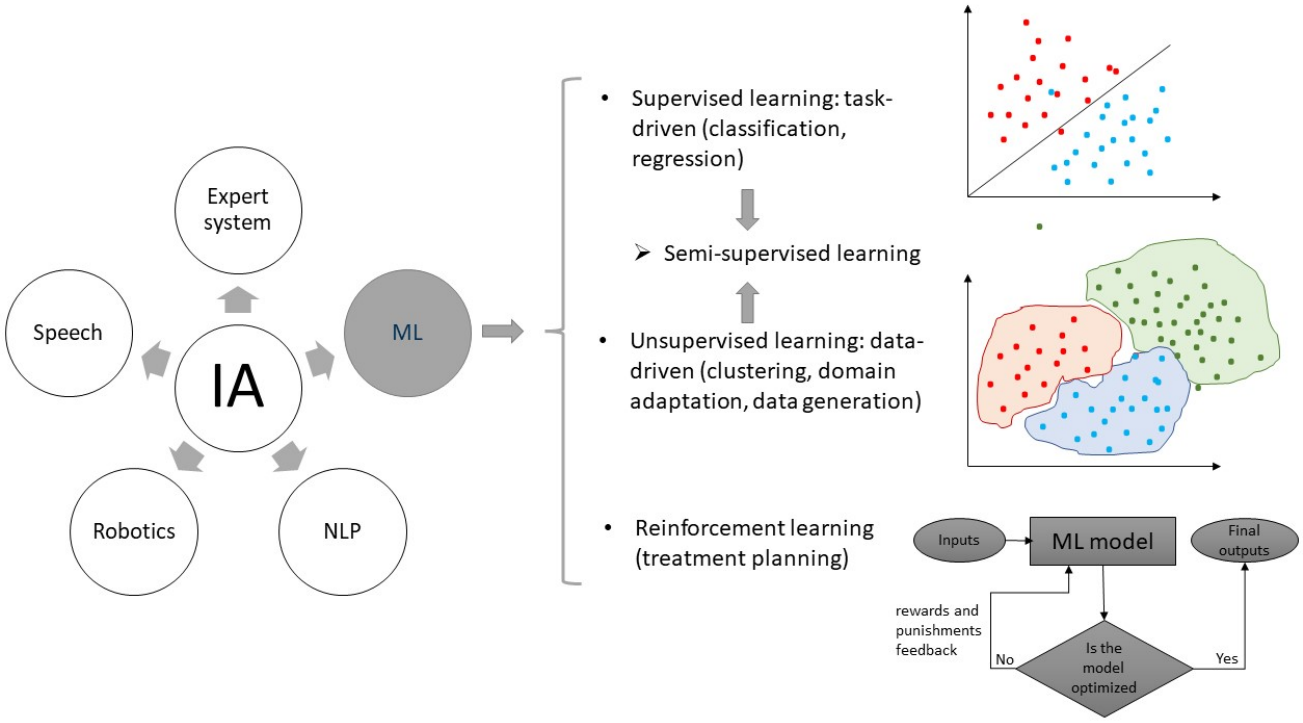


Figure 3.4: AI and its different subsets. Three types of ML can be distinguished: supervised learning (requires annotated data), unsupervised learning (the model learns hidden patterns on the dataset without pre-existing labels) and reinforcement learning (learns in trial and error process). Semi-supervised learning refers to models using both labeled and unlabeled data in the training dataset.

Based on the provided input data, which can be labeled or unlabeled, the trained model will produce an estimate of a pattern in the dataset. This model learns this pattern by optimizing an error function called the loss function. The loss function characterizes the difference between the output of an ML model and the expected result. The loss function has to be optimized during learning by finding a set of weights and biases that have low loss, on average, across all examples. If this function is a distance, for example, this distance will be minimized numerically.

In addition to model parameters (weights and biases), ML models must use specific hyperparameters for optimal results. Hyperparameters can be set manually before training (default hyperparameters). For example, the number of trees in a random forest, and the learning rate in a neural network are hyperparameters while weights are learned during training. ML model hyperparameter optimization includes finding a combination of hyperparameters that reduces a predefined loss function and in turn increases the accuracy on a given independent data. Different strategies can be used for hyperparameter optimization. Traditionally, hyperparameters were selected manually by trial and error. This is still done but better, faster, and automatic methods have been developed:

- Random Search: a random combination of hyperparameters over a pre-defined grid of hyperparameters is tested.

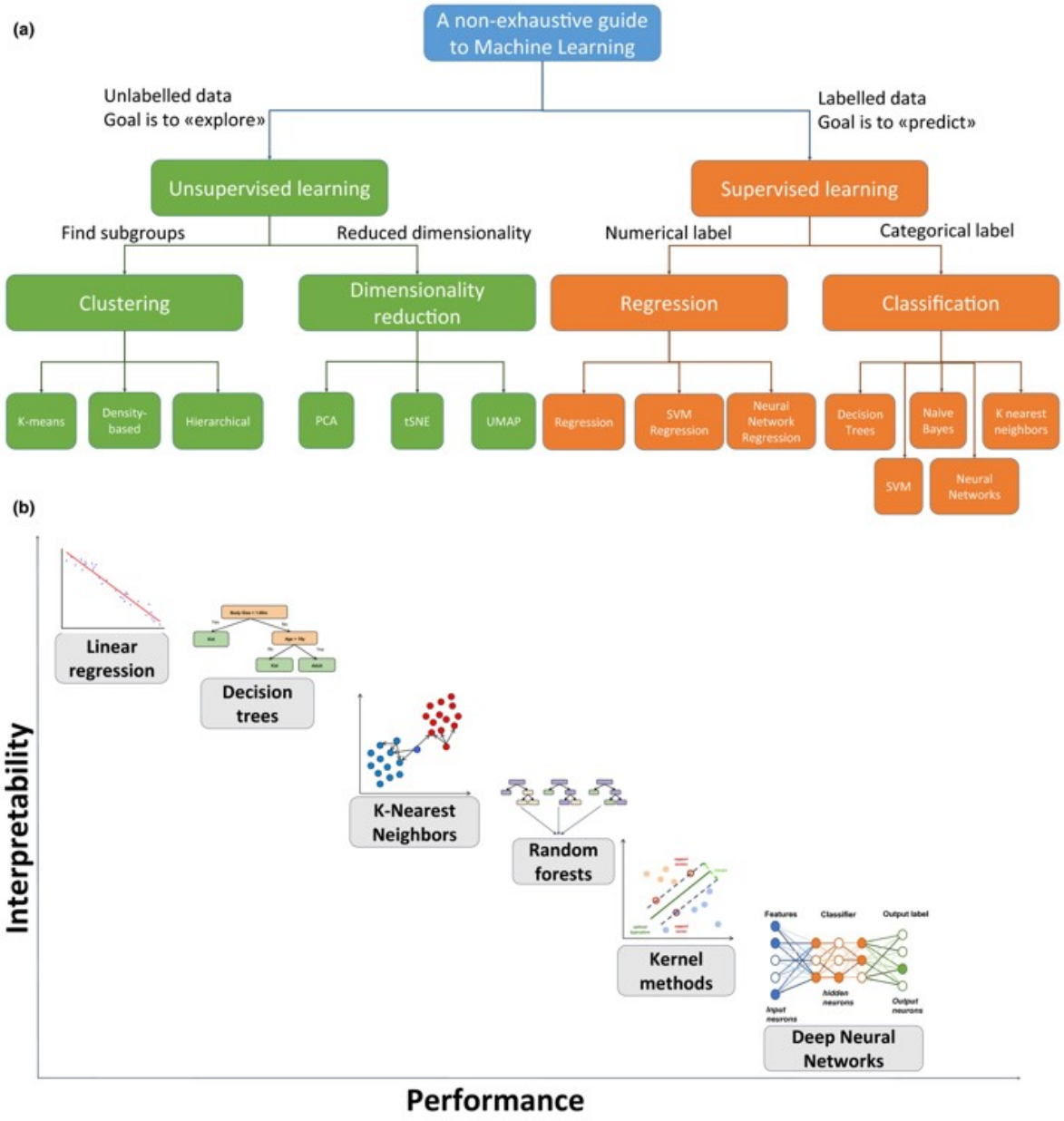


Figure 3.5: Overview of the main machine learning (ML) algorithms. (a) Supervised and Unsupervised learning. (b) Overview of ML methods from simpler and more interpretable to more advanced algorithms with potentially higher performance at the expense of less interpretability. Position of methods on the figure is qualitative and in practice depends on the number of free parameters, model complexity, data type, and the exact definition of interpretability used. PCA, principal component analysis; SVM, support vector machine; tSNE, t-distributed stochastic neighbor embedding; UMAP, uniform manifold approximation and projection. *Readapted from [142] under Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International.*

- Grid Search: each possible combination of all of the hyperparameters is evaluated. This can result in a very high number of combinations which can be computationally expensive and time-consuming depending on the trained model. Given its random search pattern, random search is usually preferred in the first time to form a grid of the most relevant hyperparameters[143]. Grid Search can then be applied on this grid to select to best hyperparameters.

- Bayesian Optimization: is sequential model-based optimization (SMBO) algorithm that allows the use of the results of a previous iteration to improve the sampling method of the next iteration. In this manner, only settings closer which bring the model closer to the desired output are considered. Bayesian optimization reduces the loss function using a posterior distribution of functions (Gaussian process).

- Genetic Algorithms: inspired by the biological concept of evolution, use stochastic global search procedure to efficiently discover the top-performing model[144]. This method uses a dynamic process where from N models with predefined hyperparameters, the best performing are kept. From these models, N offsprings models with similar hyperparameters are again defined and tested. This cycle is repeated until hyperparameters yielding the best performances are found.

In brief, through an accurate combination of parameters and hyperparameters, the goal of an ML model is to provide predicted outputs close to the input in the training set, otherwise, it is said that it underfits and has a high bias. Models should also generalize beyond the training set. The opposite is called overfitting. Overfitting is an undesirable machine learning behavior referring to an ML model that gives accurate predictions for training data but underperforms on unseen data. Underfitting can occur when the model is too simple or when input variables are not informative enough. Overfitting often occurs when the model is too complex or there are too many features over a small set of training examples.

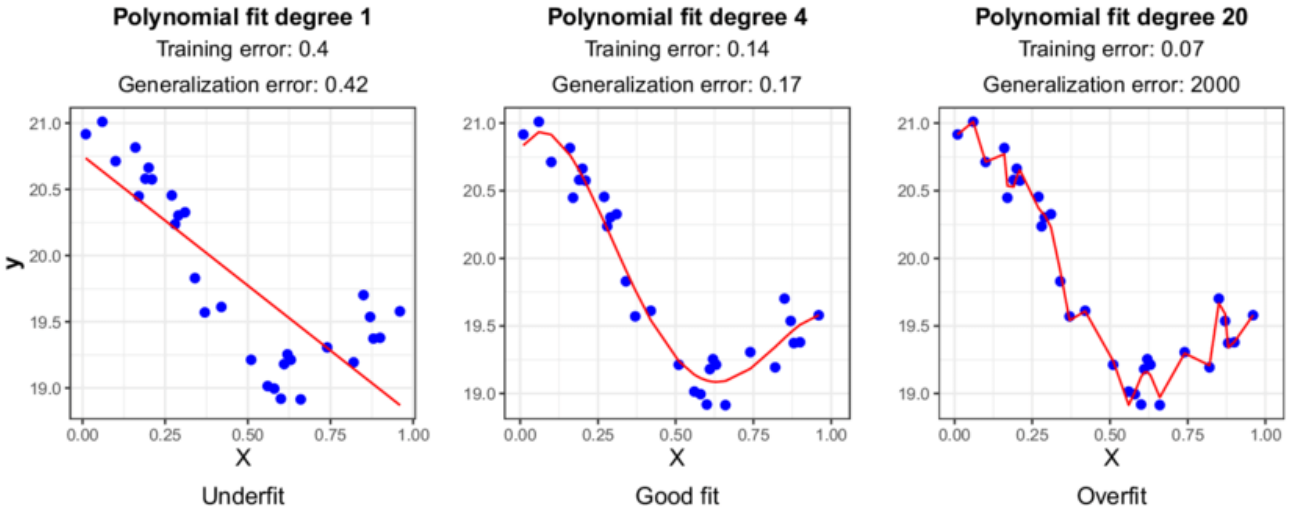


Figure 3.6: Illustration of the underfitting/overfitting issue on a simple regression case. Data points are shown as blue dots and the model fits as red lines. Underfitting occurs with a linear model (left panel), a good fit with a polynomial of degree 4 (center panel), and overfitting with a polynomial of degree 20 (right panel). Root mean square error is chosen as the objective function for evaluating the training error and the generalization error, assessed by using 10-fold cross-validation. *Reprinted from [142] under Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International.*

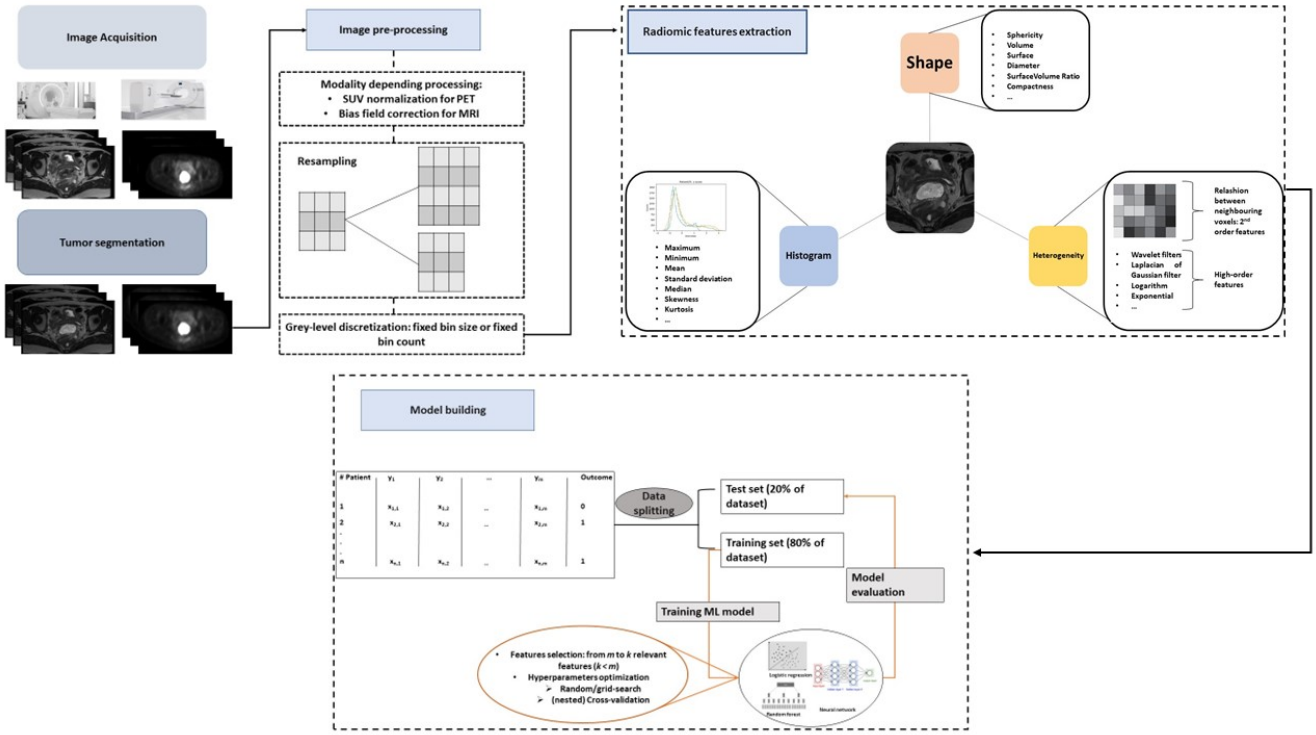


Figure 3.7: Workflow for radiomic studies. Acquired medical images are segmented by a physician (manually or using automatic tools). The images and the masks of the ROI are then pre-processed depending on the modality. After, gray-level discretization, radiomic features are extracted to be fed to ML algorithms for a specific task. ROI; Region of Interest.

3.7.3 Deep Learning

Artificial neural networks (ANN), deep learning, and machine learning tend to be used interchangeably. However, ANN is a sub-field of ML and DL is a sub-field of neural network. ANN constitute a collection of neurons and edges, drawing their origins from circuit analysis. Different weights can be applied to each edge connecting the neurons. Each neural node is connected to another and is characterized by its weight and a threshold called bias. Bias is a constant value (or constant vector) that enables the training process not to be limited to certain values only. Biases are added to the product of inputs and weights. At each neuron, an activation function is applied to a weighted input signal to generate an output signal. A sigmoidal function is often used, consisting of a first-order lowpass filter of a unit step function[142]. Neural networks are subdivided into an input layer, hidden layer(s), and output layer, as shown in 3.9. The hidden layers perform the layer of abstraction needed to go from the input layer to the output layer. The number of hidden layers defines whether the system is a shallow learning system (with one or a few hidden layers) or deep learning (with many hidden layers). The first theoretical model of an artificial neuron was proposed in 1943[145] and in the late 1950s, the first ANN was implemented[146]. It was a single-layer ANN combined with a learning ability for binary classification. The main limitation of the perceptron and single-layer networks in general was their inability to handle linear inseparable problems. Further improvements to address the limitations of single-layer networks have been made with multilayer ANN, which can become in turn very time-consuming and computationally expensive. Therefore, there is an inherent trade-off between the number of hidden layers and time required to train the model[142].

At the beginning of the training of a DL model, similarly to any ML model, randomly generated weights are initialized and biases are set to zero. The input data is then transferred (forward pass) to the input layer which will distribute it to the hidden layers. The hidden layers will then make decisions from the previous layer and weigh up how a stochastic change within itself, detriments or improves the final output, and this is referred to as the process of learning. Neural Networks learn through iterative tuning of parameters (weights and biases) during the training stage, commonly referred to as backpropagation in supervised learning. The model training process typically entails several iterations of a forward pass, backpropagation, and parameters update.

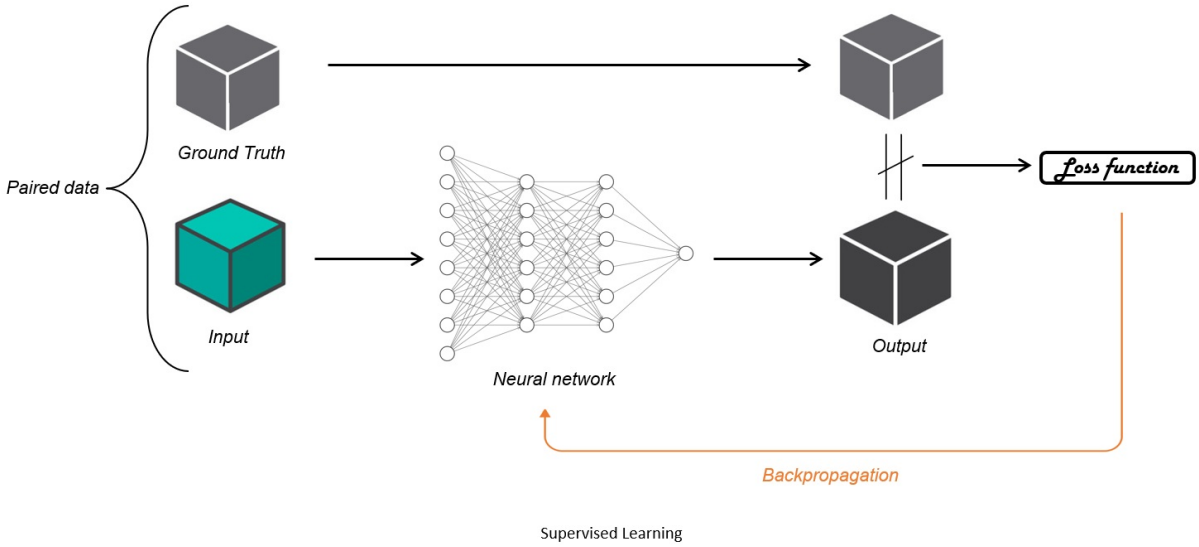


Figure 3.8: Illustration of the backpropagation process in neural network. The input is forward passed through the DL model that generates an output. The output is compared to the ground truth and an error rate via the loss function is backpropagated through the neural networks to adjust the weights. Until the output and the ground truth are similar, the loss function is optimized and weights are tuned iteratively via this process.

Different DL architectures have been used in various studies based on medical images with performances that outperform human experts[147, 148, 149]. Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs) are the most popular discriminative architectures in radiomics[150, 151, 152, 153] but other networks have also been utilized[154, 155, 156]. LeCun et al. constructed a convolutional neural network for handwritten zip code recognition in 1989 and used the term "convolution" first, which is the original version of LeNet[157]. CNN is a kind of feedforward neural network that is able to extract features from data with convolution structures. CNNs are a multilayer network made of convolutional, pooling, and fully-connected layers, overlapping one another. Each neuron is only connected to a small number of neurons, which is effective in reducing parameters and speed up convergence. A group of connections can share the same weights, which reduces parameters further. Pooling layers purpose is to subsample an input representation (image, hidden layer output matrix, etc.) by reducing its dimension reducing computational cost, trivial features and the number of parameters to be considered by the next layer. For all these reasons CNN is particularly adapted for image-based DL[158].

RNN have been an interesting and important part of neural network research and have already been applied to

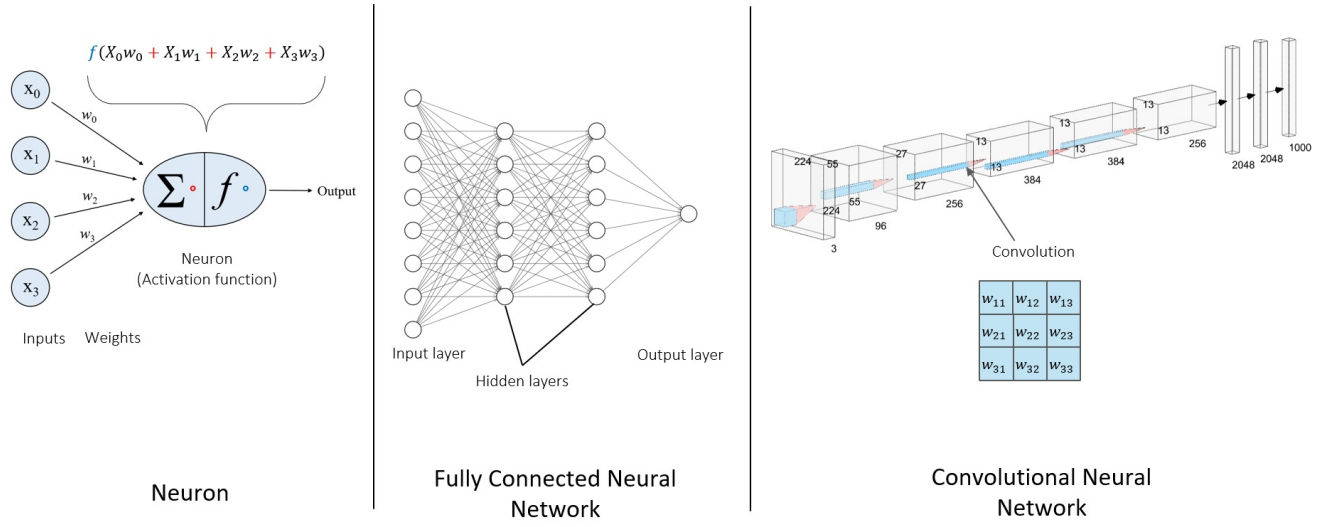


Figure 3.9: Illustration of a neuron, a fully connected neural network and a convolutional neural network.

a wide variety of problems mostly time sequences of events and ordered data such as characters in words. When input data are not independent, ordinary feedforward neural networks cannot be used, the ANN architecture needs to incorporate dependencies between these inputs. RNNs are type of networks with loops within layers that allow information of previous inputs to generate the next output of the sequence but also take the previous output as input. RNNs allow training with inputs of varying lengths. However, they are computationally expensive and can induce vanishing gradient difficulties during the training. Since its first publication by Goodfellow et al.[159], generative adversarial networks and cycle GAN are increasingly being used for image generation[160, 161, 162, 163]. GAN is an unsupervised model consisting of two CNNs, a generative model G and a discriminative model D, capable of generating realistic output (image or text for example) via an adversarial relationship between the generative network and the discriminative network. Model G with random noise z generates a sample $G(z)$ that subjects to the data distribution P_{data} learned by G. The model D can determine whether the input sample is real data x or generated data $G(z)$, it outputs a single scalar representing the probability that x came from the data rather than from $G(z)$.

The training process follows a two-player minimax game with loss function $\mathbb{L}_{GAN}(G, D)$:

$$\min_G \max_D \mathbb{L}_{GAN}(G, D) = \mathbb{E}_x[logD(x)] + \mathbb{E}_z[log(1 - D(G(z)))] \tag{3.11}$$

G and D are both trained simultaneously: we adjust parameters for G to minimize $log(1 - D(G(z))$ and adjust parameters for D to minimize$logD(X)$.

Since the original GAN allows no explicit control on the actual data generation, Mirza et al.[164] proposed the conditional GAN (cGAN) to incorporate additional information such as class labels in the synthesis process. cGAN is an extension of the GAN model in which both the generator and the discriminator are conditioned on some additional information. In the context of sCT generation from MR images for example, the sCT image output is conditioned on the MR image input.

GAN training requires a large amount of paired images. This means that to train successfully a GAN to

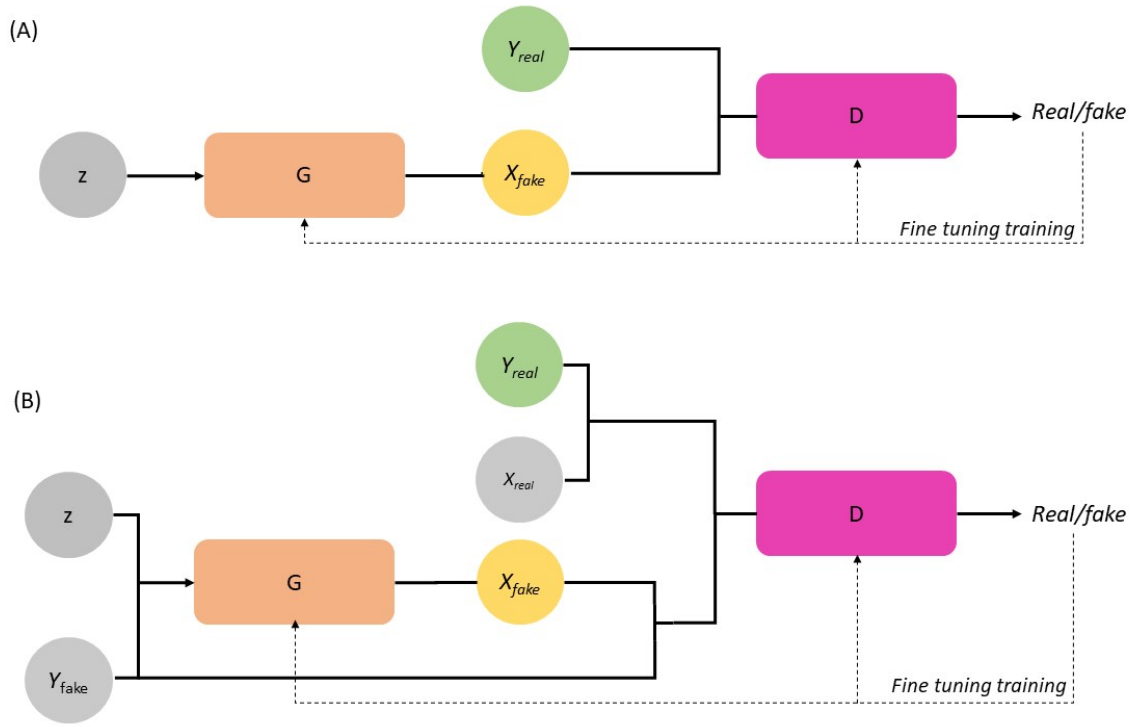


Figure 3.10: GAN vs cGAN architecture. (A) Diagram of a conventional GAN, (B) diagram of a conditional GAN. GAN: Generative adversarial network; cGAN: conditional Generative Adversarial Network; G: Generator. D: Discriminator; z: noise; X_{real}: image source from Domain X X_{fake}: generated image; Y_real: target image from domain Y; Y_{fake}: fake target.

generate output from domain X to domain Y, one has to gather paired data from domain X and domain Y. Therefore, to overcome this limitation, unsupervised cycle generative adversarial networks (CycleGAN) to translate an image from a source domain X to a target domain Y in the absence of paired data were proposed[165]. Mathematically, CycleGAN consists of training a translator $G : X ß Y$ and another translator $F : Y ß X$, with G and F being inverses of each, both mappings should be bijections. X and Y are the domain source and domain target respectively.

The GAN training process described above becomes in CycleGAN:

$$\min_{G,F} \max_{D_X,D_Y} \mathbb{L}_{CycleGAN}(G,F,D_X,D_Y) = \min_{G} \max_{D_Y} \mathbb{L}_{GAN}(G,D_Y,X,Y) + \min_{F} \max_{D_X} \mathbb{L}_{GAN}(F,D_X,Y,X) + \lambda \min_{G,F} \mathbb{L}_{cyc}(G,F)$$

, where λ controls the relative importance of the two objectives and

$$\min_{G,F} \mathbb{L}_{cyc}(G,F) = \mathbb{E}_{x\sim p_{data}}(x)[||F(G(x)) - x||_1] + \mathbb{E}_{z\sim p_z}(z)[||G(F(y)) - y||_1] \tag{3.12}$$

Radiomic features, despite promising results, suffer from several limitations inherent to image modalities and to their design process. DL technologies based on multilayer neural networks may offer realistic solutions to address some of these limitations. Tumor segmentation is a source of variability in radiomics whereas

robust registration might provide a tool for implementing multimodal radiomic studies. Estienne et al.[166] proposed 3D deep CNN for both image registration and tumor segmentation. Primakov et al.[18] developed and validated a modified 2D CNN for automatic tumor detection and segmentation. Moreover, thanks to DL, data representation (feature extraction) and prediction (e.g, classification or regression) can be performed jointly. Therefore, selected deep features would be not impacted by statistical bias due to the feature selection process. DL could also be used for image standardization. In this thesis, MR images standardization will be explored through GAN and compared to traditional models.

3.7.4 Performance metrics for model assessment

Performance metrics are target-dependent. Usually, supervised models are trying to tackle classification or regression problems. For binary classification tasks, accuracy and area under the receiving operator characteristic curve (ROC AUC) are mostly used as performance metrics. These performance measures are derived from the "confusion matrix" shown in Table 3.11. This matrix allows evaluating the efficiency of a model to predict correctly the different classes with different metrics such as precision, recall, and specificity. Precision corresponds to the ratio of correctly predicted positive values to the total number of predicted positive values. Recall or sensitivity also called true positive rate (TPR) is the ratio of correctly predicted positive values to the total number of positive values in the dataset. False Positive Rate (FPR), also known as specificity, corresponds to the proportion of negative values predicted incorrectly. Accuracy corresponds to the number of correctly predicted values divided by the total number of predicted values. Accuracy is the most used evaluation metric but can be misleading when the dataset is imbalanced. For example, if one class accounts for 80% of the dataset, a model predicting only the minority class 1 in 10 times will have an accuracy greater than 0.8. This model would clearly be inadequate despite its high accuracy. ROC AUC shows the TPR (recall) and FPR dependence. In binary classification, each point on the ROC curve is located by choosing different thresholds for classification of a subject in the dataset in the positive or the negative class. The top left corner of a ROC curve is the ideal case with 100% of positive values correctly classified (TPR = 1) and 0% of negative values incorrectly predicted at 1 (FPR = 0). As it is ideal to maximize the TPR while minimizing the FPR, a larger area under the ROC curve (AUC) shows a high predictive capability of a model[142]. However, AUC does not place more emphasis on one class over the other. Given that precision is directly influenced by class imbalance, precision-recall curves have been proposed as they are more informative than ROC curves in highly imbalanced datasets[167]. One should, in addition to accuracy, compute the sensitivity, specificity, and F-score of the model. The F-Measure conveys the balance between precision and sensitivity. The measure is 0 when either the precision or the sensitivity is 0. The Youden index is another commonly used metric to evaluate the ability of a classifier to avoid misclassifications. It puts equal weights on a classifier's performance in both the positive and negative cases.

$$\textit{Youden Index}(\gamma) = sensitivity - (1 - specificity) \tag{3.13}$$

The balanced accuracy is the average between the sensitivity and the specificity, which measures the average accuracy obtained from both the minority and majority classes. The balanced accuracy is a better characterization

of a classification model. It is given by:

$$\textit{Balanced accuracy} = \frac{1}{2}(sensitivity \times specificity) \tag{3.14}$$

For this reason, it is recommended to always report multiple metrics and the degree of imbalance in the data. Several techniques exist to handle imbalanced data but three major approaches are mainly used: algorithm modifications, cost-sensitive learning, and data sampling. The most common approach is data sampling. Data sampling refers to techniques designed to change the class distribution in the training dataset in an attempt to balance the class frequencies. Representative work in this area includes random oversampling, random undersampling, synthetic sampling with data generation, cluster-based sampling methods, and integration of sampling and boosting[168]. Data sampling can be divided into oversampling and undersampling. Oversampling techniques reproduce or synthesize new examples from the examples in the minority class while undersampling techniques delete or select a subset of examples from the majority class. Oversampling and undersampling methods can also be combined.

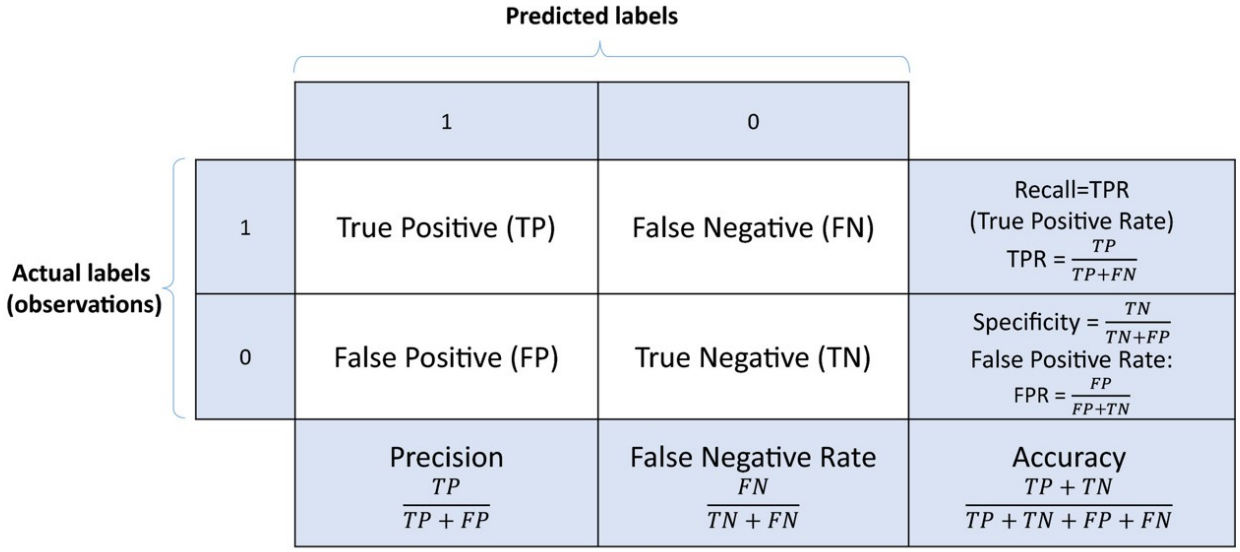


Figure 3.11: Confusion matrix for a classification task with two classes. The confusion matrix indicates how successful the algorithm was at predicting labels in a binary classification problem where labels take values 0 (called "negative") or 1 (called "positive") by evaluating the predicted vs. the real labels. Every data point in the test set belongs to one of the four categories and different measures can be derived from these numbers

There are three metrics that are commonly used for evaluating and reporting the performance of a regression model; they are Mean Squared Error (MSE), Root Mean Squared Error (RMSE), and Mean Absolute Error (MAE). The purpose of ML models in a regression task is to find a pattern that minimizes these distances.

For survival analysis, the most frequently used evaluation metric is the concordance index (C-index, C-statistic). C-index is a measure of the model's ability to correctly provide a trustworthy ranking of survival times based on individual risk scores. It is the fraction of pairs in a cohort where the observation with the higher survival time has the higher probability of survival predicted by the model.

3.8 ML and radiomics for cancer management optimization

Several ML/DL models have been applied successfully in the field of oncology, at all stages of radiotherapy, from tumor early detection[169, 170, 171], to treatment planning[172, 173, 130, 18] to treatment response prediction[174, 175, 176, 177, 178]. Here, we are going to focus on studies related to HPV-induced cancers, namely LACC, ASCC, and OPSCC.

3.8.1 Radiomics for early detection and characterization of lesions

Screening policies have been promoted by the World Health Organization (WHO) to prevent cancer mortality, particularly in patients at risk. Cancer characterization at an early stage increases the chance for early diagnosis and effective treatment for most cancers. ML, through radiomics, might improve screening and replace classic tumor characterization based on macroscopic and invasive samples (biopsies, blood test) [179]. In other words, radiomics could be used for differential diagnosis (malignant or benign lesiosn), or for differential prognosis (good or bad prognosis).

In HNC, HPV status has been shown to be a prognosis factor and if supported by sufficient evidence, radiomics-based biomarkers could be used in the future as a viable alternative to confirm HPV status after or without positive p16 immunohistochemical tests[180]. Song et al.[170] demonstrated that radiomic features extracted from 562 CT scans could potentially be used for identifying HPV-positive patients with OPSCC and that HPV status was associated with disease-free survival (DFS). Similarly, Bagher-Ebadian et al.[169, 181] showed that radiomic-based classifier outperforms clinical factors in characterizing HPV for patients with OPSCC. Apart from OPSCC, in HNC it has been shown that in sinonasal squamous cell carcinoma (SCC), inverted papilloma (IP) is an uncommon phenotype with aggressive behavior and a propensity for postoperative recurrence. The characterization of IP is based on endoscopic incisional biopsy but its sensitivity has shown sampling errors. Ramkumar et al.[182] found that MRI-based textural analysis had the potential to differentiate sinonasal SCC from IP (accuracy 89.1%) with results comparable to manual assessment by neuroradiologists ($P = 0.0004$). The use of radiomics biomarkers has also been proposed to identify molecular subtypes in HNC[183, 184, 185]

For patients with early-stage cervical cancer, pelvic lymph node metastasis (PLNM) and clinical stage are the two primary factors that affect disease-free survival (DFS) and overall survival (OS) in patients with cervical cancer[186]. Mu et al.[187] studied the value of texture features and other commonly used semi-quantitative indices extracted from 18F-FDG PET images in predicting the stage of patients with cervical cancer. Xia et al. [171] developed and validated a noninvasive individualized radiomic-based nomogram integrating a radiomic signature and clinicopathologic factors for PLNM in patients with early-stage cervical cancer before surgery. This radiomic nomogram also showed higher accuracy than the clinical diagnostic criteria of PLNM (stromal invasion depth, FIGO stage, and maximal tumor diameter). Li et al. conducted a meta-analysis of MRI-based radiomic features for predicting PLNM in patients with cervical cancer including 12 studies comprising 793 female patients. Li et al. concluded that using MRI-based radiomic features can help improve the accuracy

of predicting lymph node metastasis in cervical cancer patients[188]. Hystological type of LACC has also been identified using radiomics features extracted from 18F-FDG PET/CT images[189, 190]

Moreover, studies have successfully shown that image textural features could be useful to identify the immune phenotype of tumors before immunotherapy alone or combined with CRT. Chen et al.[191] demonstrated that several 18F-FDG PET/CT-derived textural features could provide additional information to determine tumor PD-L1 expression which correlates with response to PD-1 blockers, such as nivolumab or pembrolizumab. Sun et al.[192, 193] developed and validated a CT-based radiomic signature to quantify CD8 cell count and predict clinical outcomes of patients treated with immunotherapy in multiple solid tumors.

3.8.2 Image-based DL for tumor detection and segmentation

Detection and segmentation of abnormalities on medical images are highly important in patient care. Medical image segmentation has become more and more automated in the field of RT well before the arrival of DL. For example, an automated method, atlas-based segmentation (ABS)[194], is used in several commercially available software solutions. ABS algorithms use multiple libraries of predefined, expert-delineated structures, that vary in size and shape to cover anatomical variations[195]. These predefined structures can be transferred to a new image with the help of image registration methods using a single atlas or multiple ones. One disadvantage of ABS-based methods is that all data used to generate one or several atlases must be available during matching, making huge atlases time-consuming and not feasible in clinical workflows. Furthermore, ABS-based methods have been shown to perform poorly on cancerous lesions. As an alternative to ABS, tumor segmentation based on deep neural networks has established itself over the past ten years as the state of art technique and has outperformed classical automatic segmentation methods[196, 197], even though segmentation based on conventional ML methods has also been implemented in several studies, albeit with less success[198, 199, 200].

Automatic segmentation might be able to reduce or achieve similar inter-observer variability associated with organ or lesion manual segmentation[7, 201, 16]. Deng et al.[202] proposed an automatic segmentation method using traditional ML techniques for evaluation of HNC lesions demonstrating superior segmentation performance (area overlap measure of 0.76 ± 0.08 and accuracy of $86 \pm 8\%$) compared to other methods in the literature at that time. Despite high variability in regard to Dice Similarity Coefficient (DSC) between DL segmentation and ground-truth segmentation derived by radiologists, automatic segmentation has shown promising results in LACC. Torheim et al.[199] developed a fully automatic method for tumor delineation based on voxel classification using Fisher's Linear Discriminant Analysis (LDA) on multi-sequence MR images and achieved mean sensitivity and specificity of 85% and 93% respectively. Several DL algorithms applied to different MRI modalities have been assessed for LACC detection and segmentation. Lin et al.[172] implemented a U-Net on Diffusion-Weighted MR images and apparent diffusion coefficient (ADC) images. Bnouni et al.[136] developed a multi-stage deep learning architecture based on cycle-GAN to segment pelvic multi-organs using complementary multi-view MRI. In a multi-institutional study by Rouhi et al.(cite the paper in review), different deep neural network models were trained on multi-vendors T2-weighted MRI for both 2D and 3D segmentation, to present a comprehensive comparison for tumor segmentation in LACC. The best model achieved DSC of 0.72 ± 0.16. Rahimeh et al. also proposed a failure detection method based on radiomic features to identify cases with sub-optimal DSC. Huang et al[153] proposed an end-to-end modified U-Net to segment clinical target volume (CTV) for cervical cancer

brachytherapy and achieved a median accuracy of 0.81 in terms of average surface distance. In addition to gross tumor, other organs have also been segmented with DL models. Sartor et al.[173] have investigated automatic delineation of CTV lymph nodes, a key structure during RT delivery, and their model yielded a median DSC of 0.82.

Notably, automated DL-based segmentation methods have rarely been applied to anorectal cancers and to complex pelvic OAR structures, like small and large bowels. Nevertheless, few studies have tackled the task of tumor segmentation using DL in anorectal (most of them in rectal cancer) cancers[203, 204, 205, 173], most of which are on CT. Men et al. proposed a CT-based deep dilated convolutional neural network (DDCNN)-based method[203], for fast and consistent auto-segmentation of CTV and OARS of patients with rectum cancer. The mean DSC value was 87.7%. Using a ResNet-101 [205], Men at al. showed that increasing positions-related information improved the trained model accuracy for CTV segmentation. A 2D U-net network was developed on T2-weighted MR images for rectal tumor detection and segmentation by Wang et al.[204]. The DSC was 0.71 ± 0.13 and no significant difference has been observed between automated segmentation and manual segmentation considering DSC.

These studies have shown that automatic tumor segmentation in HPV-induced cancers could be applied with accuracy comparable to manual tumor delineation and potentially less inter-observer variability.

3.8.3 Response prediction

Another application of radiomics is the development and refinement of signatures that can improve upon prognostic and/or predictive models for specific cancers[206]. Several studies have focused on the predictive ability of radiomics models regarding radiotherapy and chemotherapy outcomes and showed that radiomics could predict outcomes not only on recurrence and metastatic status, survival, and treatment response but also in adjacent non-cancerous tissues, such as glandular tissues (parotid and major salivary glands). As an example, several side effects like acute toxicity due to ionizing radiation have been assessed. Multiple studies have assessed the ability of radiomic features to predict xerostomia, a radiation common side effect causing dry mouth in HNC[207, 208, 209] at different timepoints.

Several studies have shown promising results to predict survival in patients with HNC, analyzing various cancer anatomic subsites, tumor segmentation methods, and textural indices[210, 211, 212]. Vallierres et al.[213] built multivariate models based on five features extracted from FDG-PET images that could reliably classify HPV status and predict treatment failure in HNC. Zhuo et al.[214] extracted radiomics from multi-modalities MR sequences and developed a model predicting survival in different phenotypes of patients with non-metastatic nasopharyngeal carcinoma (NPC). Their model achieved a C-index of 0.81 and was seen to be superior to and more stable than conventional TNM staging system with a C-index of 0.76. Another study by Starke et al.[215] combining CT and PET scans from multiple timepoints during treatment showed the predictive ability of radiomics features in HNC. CT-based features allowed for more accurate risk prediction, while FDG-PET-based models are able to stratify patients into low- and high-risk groups more reliably. Similarly, multiple studies have demonstrated the added-value of radiomic features for response prediction in LACC whether on CT, PET, or MR images[175, 216, 81]. Combining radiomic features extracted from PET and MRI images, Lucia et al.[216] validated a radiomic signature for prediction of disease-free survival (DFS) and locoregional

control (LRC) in LACC with an accuracy ≥ 0.90 for both outcomes. Additionally to LACC, and HNC cancers, radiomic features extracted from [18F]-FDG PET and MRI have also been successfully assessed for recurrence and survival prediction in anal squamous cell carcinoma (ASCC) [217, 218, 219, 176]. Brown et al.[176] showed that radiomic features extracted from [18F]-FDG PET/CT could improve prediction of progression-free survival (PFS) in ASCC compared to conventional clinical variables. Giraud et al.[219] showed that pre-therapeutic MRI radiomics mixed with two clinical variables could predict 2-year disease control after CRT and could contribute to identifying high-risk patients.

3.8.4 Radiomics and clinical implementation

Despite a flourishing literature on radiomics, most studies encountered radiomics intrinsic limitations and methodological issues underlined years ago that need to be addressed before their application into clinical routine.

Firstly, most published radiomic studies would be difficult to reproduce and validate due to the lack of standardized definitions and validated reference values. The Image Biomarker Standardization Initiative (IBSI) was thus formed with the following objectives: (a) establish nomenclature and definitions for commonly used radiomic features; (b) establish a general image processing scheme for the calculation of features from imaging; (c) provide data sets and associated reference values for verification and calibration of software implementations for image processing and feature computation; and (d) provide a set of reporting guidelines for studies involving radiomic analyses[23]. In this thesis, radiomic feature calculations were compliant with the IBSI guidelines.

Secondly, the main limitation for clinical implementation is the data itself. This limitation affects the field of ML in general and consists of the difficulty of constituting large, high-quality datasets that cover the complexity of the task at hand. The majority of early research on radiomics had a retrospective, monocentric design, which advocates for caution in interpreting the reported findings[220]. Particularly, most of these studies display small data size which may lead to a patient selection bias, not accurately depicting a real-life scenario. A direct consequence of this is the lack of an independent validation dataset which may cause the predictive model to overfit. This also results in studies involving few device manufacturers and data acquisition techniques. Given that radiomic features are strongly impacted by acquisition and reconstruction parameters[66, 221, 222, 223, 224, 225, 84, 226], it is, therefore, crucial to develop robust image standardization methods for generalizable radiomic studies. Otherwise, this could lead to biased models that cannot duplicate their performances in new research trials. Nevertheless, efforts have been made to collect large and multi-institutional high-quality open-source datasets to overcome the lack of data. This has been done through different data challenges [227, 228]. BReast tumor Image classification on Gigapixel HisTopathological images (BRIGHT) is an open-source data challenge (only available through registration) [1] for the development, testing, and evaluation of models for automatic breast tumor subtyping of frequent lesions along with rare pathologies, by using clinical Hematoxylin & Eosin (H&E) stained annotated gigapixel Whole-Slide Images (WSIs). BRIGHT was the first breast tumor subtyping challenge that included atypical lesions. More than 550 annotated WSIs across a wide spectrum of tumor subtypes are available. WSI classification into three classes as per cancer risk, and WSI classification into six fine-grained lesion subtypes are the two tasks proposed in the challenge. HEad and neCK

[1] https://research.ibm.com/haifa/Workshops/BRIGHT/

TumOR (HECKTOR) challenge[228] was organized for automatic segmentation of HNC primary Gross Tumor Volume (GTV), the automatic prediction of PFS from the same FDG-PET/CT. A third task of predicting PFS with ground truth GTVt annotations provided to the participants was also included. The data were collected from six centers for a total of 325 images, split into 224 training and 101 testing cases. The Brain Tumor segmentation (BraTS) challenge aims at the development of state-of-the-art methods for the segmentation of brain tumors in multi-parametric MR images by gathering a publicly available dataset and a community benchmark[229, 227]. Furthermore, to pinpoint the clinical relevance of this segmentation task, BraTS also proposed a task of prediction of patient OS, via integrative analyses of radiomic features and machine learning (ML) algorithms. BraTS made available data consisting for each patient of a native T1-weighted scan (T1), post-contrast T1-weighted scan (T1Gd), a native T2-weighted scan (T2), and a T2 Fluid Attenuated Inversion Recovery (T2-FLAIR) scan. These initiatives were found to be key to ensuring generalization performance and developing trusted and ethical models. To achieve AI models robust enough to be implemented in clinical routine, such actions should be strongly supported.

Although ML has shown tremendous progress which in turn has resulted in promising applications for better cancer management, current practices are still flawed and need to be harmonized before clinical implementation. Guidelines have been formulated for good practices in the fields of ML, especially to ensure standardized image acquisition, data preparation, and model building. Radiomic-based algorithm is based on a multi-step workflow, each step must be performed carefully and described thoroughly to avoid bias and overly optimistic not reproducible results. Studies should be conducted on multi-institutions datasets. Therefore, robust standardization methods must be applied to mitigate the so-called 'center effect'. Segmentation of the ROI should be done by the same experts or by the same automatic or semi-automatic method. Feature selection and model building should be conducted precisely to avoid selection biases and reported accurately. Furthermore, studies should always include an independent cohort of unseen data to evaluate the generalizability of the constructed model. ML and especially DL models are highly complex and have a large number of parameters. This renders the interpretation of these models quite complicated but indispensable[230, 231].

Bibliography

[1] Philippe Lambin et al. "Radiomics: the bridge between medical imaging and personalized medicine". en. In: *Nat Rev Clin Oncol* 14.12 (Dec. 2017). Bandiera_abtest: a Cg_type: Nature Research Journals Number: 12 Primary_atype: Reviews Publisher: Nature Publishing Group Subject_term: Cancer genomics;Cancer imaging;Medical imaging;Personalized medicine Subject_term_id: cancer-genomics;cancer-imaging;medical-imaging;personalized-medicine, pp. 749–762. ISSN: 1759-4782. DOI: 10.1038/nrclinonc.2017.141. URL: https://www.nature.com/articles/nrclinonc.2017.141 (visited on 09/20/2021).

[2] James H. Thrall et al. "Artificial Intelligence and Machine Learning in Radiology: Opportunities, Challenges, Pitfalls, and Criteria for Success". eng. In: *J Am Coll Radiol* 15.3 Pt B (Mar. 2018), pp. 504–508. ISSN: 1558-349X. DOI: 10.1016/j.jacr.2017.12.026.

[3] Michele Larobina and Loredana Murino. "Medical Image File Formats". en. In: *J Digit Imaging* 27.2 (Apr. 2014), pp. 200–206. ISSN: 1618-727X. DOI: 10.1007/s10278-013-9657-9. URL: https://doi.org/10.1007/s10278-013-9657-9 (visited on 01/06/2023).

[4] Robert J. Gillies, Paul E. Kinahan, and Hedvig Hricak. "Radiomics: Images Are More than Pictures, They Are Data". eng. In: *Radiology* 278.2 (Feb. 2016), pp. 563–577. ISSN: 1527-1315. DOI: 10.1148/radiol.2015151169.

[5] Amaury Leroy et al. "Magnetic Resonance Imaging Virtual Histopathology from Weakly Paired Data". en. In: *Proceedings of the MICCAI Workshop on Computational Pathology*. ISSN: 2640-3498. PMLR, Sept. 2021, pp. 140–150. URL: https://proceedings.mlr.press/v156/leroy21a.html (visited on 03/13/2023).

[6] Wei Shao et al. "ProsRegNet: A deep learning framework for registration of MRI and histopathology images of the prostate". In: *Med Image Anal* 68 (Feb. 2021), p. 101919. ISSN: 1361-8415. DOI: 10.1016/j.media.2020.101919. URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7856244/ (visited on 03/13/2023).

[7] Charlotte L. Brouwer et al. "3D Variation in delineation of head and neck organs at risk". eng. In: *Radiat Oncol* 7 (Mar. 2012), p. 32. ISSN: 1748-717X. DOI: 10.1186/1748-717X-7-32.

[8] Stanley Osher and Ronald Fedkiw. *Level Set Methods and Dynamic Implicit Surfaces*. English. 2003rd edition. New York Berlin Heidelberg: Springer, Oct. 2002. ISBN: 978-0-387-95482-0.

[9] Dirk Smeets et al. “Semi-automatic level set segmentation of liver tumors combining a spiral-scanning technique with supervised fuzzy pixel classification”. en. In: *Medical Image Analysis* 14.1 (Feb. 2010), pp. 13–20. ISSN: 1361-8415. DOI: 10.1016/j.media.2009.09.002. URL: https://www.sciencedirect.com/science/article/pii/S136184150900084X (visited on 01/06/2023).

[10] P.J. Yim and D.J. Foran. “Volumetry of hepatic metastases in computed tomography using the watershed and active contour algorithms”. In: *16th IEEE Symposium Computer-Based Medical Systems, 2003. Proceedings.* ISSN: 1063-71258. June 2003, pp. 329–335. DOI: 10.1109/CBMS.2003.1212810.

[11] Rui Lu, Pina Marziliano, and Choon Hua Thng. “Liver tumor volume estimation by semi-automatic segmentation method”. eng. In: *Conf Proc IEEE Eng Med Biol Soc* 2005 (2005), pp. 3296–3299. ISSN: 1557-170X. DOI: 10.1109/IEMBS.2005.1617181.

[12] B. Zhao et al. “Two-dimensional multi-criterion segmentation of pulmonary nodules on helical CT images”. eng. In: *Med Phys* 26.6 (June 1999), pp. 889–895. ISSN: 0094-2405. DOI: 10.1118/1.598605.

[13] Amjad Rehman Khan et al. “Brain tumor segmentation using K-means clustering and deep learning with synthetic data augmentation for classification”. eng. In: *Microsc Res Tech* 84.7 (July 2021), pp. 1389–1399. ISSN: 1097-0029. DOI: 10.1002/jemt.23694.

[14] Jayaram K. Udupa and Supun Samarasekera. “Fuzzy Connectedness and Object Definition: Theory, Algorithms, and Applications in Image Segmentation”. en. In: *Graphical Models and Image Processing* 58.3 (May 1996), pp. 246–261. ISSN: 1077-3169. DOI: 10.1006/gmip.1996.0021. URL: https://www.sciencedirect.com/science/article/pii/S1077316996900210 (visited on 01/06/2023).

[15] László G. Nyúl, Alexandre X. Falcão, and Jayaram K. Udupa. “Fuzzy-connected 3D image segmentation at interactive speeds”. en. In: *Graphical Models* 64.5 (Sept. 2002), pp. 259–281. ISSN: 1524-0703. DOI: 10.1016/S1077-3169(02)00005-9. URL: https://www.sciencedirect.com/science/article/pii/S1077316902000059 (visited on 01/06/2023).

[16] J. van der Veen et al. “Interobserver variability in organ at risk delineation in head and neck cancer”. In: *Radiation Oncology* 16.1 (June 2021), p. 120. ISSN: 1748-717X. DOI: 10.1186/s13014-020-01677-2. URL: https://doi.org/10.1186/s13014-020-01677-2 (visited on 01/03/2023).

[17] Diana Veiga-Canuto et al. “Comparative Multicentric Evaluation of Inter-Observer Variability in Manual and Automatic Segmentation of Neuroblastic Tumors in Magnetic Resonance Images”. In: *Cancers (Basel)* 14.15 (July 2022), p. 3648. ISSN: 2072-6694. DOI: 10.3390/cancers14153648. URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC9367307/ (visited on 01/03/2023).

[18] Sergey P. Primakov et al. “Automated detection and segmentation of non-small cell lung cancer computed tomography images”. eng. In: *Nat Commun* 13.1 (June 2022), p. 3423. ISSN: 2041-1723. DOI: 10.1038/s41467-022-30841-3.

[19] Em Bellon et al. “MR artifacts: a review”. In: *American Journal of Roentgenology* 147.6 (Dec. 1986). Publisher: American Roentgen Ray Society, pp. 1271–1281. ISSN: 0361-803X. DOI: 10.2214/ajr.147.6.1271. URL: https://www.ajronline.org/doi/10.2214/ajr.147.6.1271 (visited on 10/13/2022).

[20] Boubakeur Belaroussi et al. “Intensity non-uniformity correction in MRI: Existing methods and their validation”. en. In: *Medical Image Analysis* 10.2 (Apr. 2006), pp. 234–246. ISSN: 1361-8415. DOI: 10.1016/j.media.2005.09.004. URL: https://www.sciencedirect.com/science/article/pii/S1361841505000976 (visited on 10/11/2022).

[21] Shin-Hyung Park et al. “Robustness of magnetic resonance radiomic features to pixel size resampling and interpolation in patients with cervical cancer”. eng. In: *Cancer Imaging* 21.1 (Feb. 2021), p. 19. ISSN: 1470-7330. DOI: 10.1186/s40644-021-00388-5.

[22] Stephen Yip et al. “Comparison of texture features derived from static and respiratory-gated PET images in non-small cell lung cancer”. eng. In: *PLoS One* 9.12 (2014), e115510. ISSN: 1932-6203. DOI: 10.1371/journal.pone.0115510.

[23] Alex Zwanenburg et al. “The Image Biomarker Standardization Initiative: Standardized Quantitative Radiomics for High-Throughput Image-based Phenotyping”. eng. In: *Radiology* 295.2 (May 2020), pp. 328–338. ISSN: 1527-1315. DOI: 10.1148/radiol.2020191145.

[24] M. Hatt et al. “Joint EANM/SNMMI guideline on radiomics in nuclear medicine”. en. In: *Eur J Nucl Med Mol Imaging* 50.2 (Jan. 2023), pp. 352–375. ISSN: 1619-7089. DOI: 10.1007/s00259-022-06001-6. URL: https://doi.org/10.1007/s00259-022-06001-6 (visited on 01/10/2023).

[25] A. Madabhushi and J.K. Udupa. “Interplay between intensity standardization and inhomogeneity correction in MR image processing”. In: *IEEE Transactions on Medical Imaging* 24.5 (May 2005). Conference Name: IEEE Transactions on Medical Imaging, pp. 561–576. ISSN: 1558-254X. DOI: 10.1109/TMI.2004.843256.

[26] Nicholas J. Tustison et al. “N4ITK: Improved N3 Bias Correction”. In: *IEEE Trans Med Imaging* 29.6 (June 2010), pp. 1310–1320. ISSN: 0278-0062. DOI: 10.1109/TMI.2010.2046908. URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3071855/ (visited on 10/13/2022).

[27] Sameera V. Mohd Sagheer and Sudhish N. George. “A review on medical image denoising algorithms”. en. In: *Biomedical Signal Processing and Control* 61 (Aug. 2020), p. 102036. ISSN: 1746-8094. DOI: 10.1016/j.bspc.2020.102036. URL: https://www.sciencedirect.com/science/article/pii/S1746809420301920 (visited on 03/13/2023).

[28] Jong-Sen Lee. “Digital Image Enhancement and Noise Filtering by Use of Local Statistics”. In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* PAMI-2.2 (Mar. 1980). Conference Name: IEEE Transactions on Pattern Analysis and Machine Intelligence, pp. 165–168. ISSN: 1939-3539. DOI: 10.1109/TPAMI.1980.4766994.

[29] Yongjian Yu and S.T. Acton. “Speckle reducing anisotropic diffusion”. In: *IEEE Transactions on Image Processing* 11.11 (Nov. 2002). Conference Name: IEEE Transactions on Image Processing, pp. 1260–1270. ISSN: 1941-0042. DOI: 10.1109/TIP.2002.804276.

[30] T. Loupas, W.N. McDicken, and P.L. Allan. “An adaptive weighted median filter for speckle suppression in medical ultrasonic images”. In: *IEEE Transactions on Circuits and Systems* 36.1 (Jan. 1989). Conference Name: IEEE Transactions on Circuits and Systems, pp. 129–135. ISSN: 1558-1276. DOI: 10.1109/31.16577.

[31] C. Tomasi and R. Manduchi. “Bilateral filtering for gray and color images”. In: *Sixth International Conference on Computer Vision (IEEE Cat. No.98CH36271)*. Jan. 1998, pp. 839–846. DOI: 10.1109/ICCV.1998.710815.

[32] José V. Manjón et al. “Adaptive non-local means denoising of MR images with spatially varying noise levels”. en. In: *Journal of Magnetic Resonance Imaging* 31.1 (2010). _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/jmri.22003, pp. 192–203. ISSN: 1522-2586. DOI: 10.1002/jmri.22003. URL: https://onlinelibrary.wiley.com/doi/abs/10.1002/jmri.22003 (visited on 03/13/2023).

[33] P. V. Sudeep et al. “Nonlocal linear minimum mean square error methods for denoising MRI”. en. In: *Biomedical Signal Processing and Control* 20 (July 2015), pp. 125–134. ISSN: 1746-8094. DOI: 10.1016/j.bspc.2015.04.015. URL: https://www.sciencedirect.com/science/article/pii/S1746809415000786 (visited on 03/13/2023).

[34] José V. Manjón, Pierrick Coupé, and Antonio Buades. “MRI noise estimation and denoising using non-local PCA”. en. In: *Medical Image Analysis* 22.1 (May 2015), pp. 35–47. ISSN: 1361-8415. DOI: 10.1016/j.media.2015.01.004. URL: https://www.sciencedirect.com/science/article/pii/S1361841515000171 (visited on 03/13/2023).

[35] He Wang et al. “High-field mr diffusion-weighted image denoising using a joint denoising convolutional neural network”. en. In: *Journal of Magnetic Resonance Imaging* 50.6 (2019). _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/jmri.26761, pp. 1937–1947. ISSN: 1522-2586. DOI: 10.1002/jmri.26761. URL: https://onlinelibrary.wiley.com/doi/abs/10.1002/jmri.26761 (visited on 03/14/2023).

[36] Hajer Jomaa et al. “Denoising of dynamic PET images using a multi-scale transform and non-local means filter”. en. In: *Biomedical Signal Processing and Control* 41 (Mar. 2018), pp. 69–80. ISSN: 1746-8094. DOI: 10.1016/j.bspc.2017.11.002. URL: https://www.sciencedirect.com/science/article/pii/S1746809417302550 (visited on 03/13/2023).

[37] Fumio Hashimoto et al. “Dynamic PET Image Denoising Using Deep Convolutional Neural Networks Without Prior Training Datasets”. In: *IEEE Access* 7 (2019). Conference Name: IEEE Access, pp. 96594–96603. ISSN: 2169-3536. DOI: 10.1109/ACCESS.2019.2929230.

[38] Jianan Cui et al. “Population and individual information guided PET image denoising using deep neural network”. In: May 2019, p. 99. DOI: 10.1117/12.2534901.

[39] Scott Wollenweber and Tyler Bradshaw. “Assessment of Machine Learning Techniques for PET Image De-Noising”. en. In: *Journal of Nuclear Medicine* 60.supplement 1 (May 2019). Publisher: Society of Nuclear Medicine Section: Physics, Instrumentation & Data Sciences, pp. 571–571. ISSN: 0161-5505, 2159-662X. URL: https://jnm.snmjournals.org/content/60/supplement_1/571 (visited on 03/13/2023).

[40] Steven Smith. *The Scientist & Engineer's Guide to Digital Signal Processing: Smith, Steven W.: 9780966017632: Amazon.com: Books*. URL: https://www.amazon.com/Scientist-Engineers-Digital-Signal-Processing/dp/0966017633 (visited on 03/14/2023).

[41] Kjell Erlandsson et al. “A review of partial volume correction techniques for emission tomography and their applications in neurology, cardiology and oncology”. en. In: *Phys. Med. Biol.* 57.21 (Oct. 2012). Publisher: IOP Publishing, R119. ISSN: 0031-9155. DOI: 10.1088/0031-9155/57/21/R119. URL: https://dx.doi.org/10.1088/0031-9155/57/21/R119 (visited on 03/14/2023).

[42] A.J. Reader et al. “EM algorithm system modeling by image-space techniques for PET reconstruction”. In: *IEEE Transactions on Nuclear Science* 50.5 (Oct. 2003). Conference Name: IEEE Transactions on Nuclear Science, pp. 1392–1397. ISSN: 1558-1578. DOI: 10.1109/TNS.2003.817327.

[43] Brian F. Hutton and Yiu H. Lau. “Application of distance-dependent resolution compensation and post-reconstruction filtering for myocardial SPECT”. en. In: *Phys. Med. Biol.* 43.6 (June 1998), p. 1679. ISSN: 0031-9155. DOI: 10.1088/0031-9155/43/6/022. URL: https://dx.doi.org/10.1088/0031-9155/43/6/022 (visited on 03/14/2023).

[44] P. H. Pretorius et al. “Reducing the influence of the partial volume effect on SPECT activity quantitation with 3D modelling of spatial resolution in iterative reconstruction”. en. In: *Phys. Med. Biol.* 43.2 (Feb. 1998), p. 407. ISSN: 0031-9155. DOI: 10.1088/0031-9155/43/2/014. URL: https://dx.doi.org/10.1088/0031-9155/43/2/014 (visited on 03/14/2023).

[45] Shyam M. Srinivas et al. “A recovery coefficient method for partial volume correction of PET images”. en. In: *Ann Nucl Med* 23.4 (June 2009), pp. 341–348. ISSN: 1864-6433. DOI: 10.1007/s12149-009-0241-9. URL: https://doi.org/10.1007/s12149-009-0241-9 (visited on 03/14/2023).

[46] Francesca Gallivanone et al. “PVE Correction in PET-CT Whole-Body Oncological Studies From PVE-Affected Images”. In: *IEEE Transactions on Nuclear Science* 58.3 (June 2011). Conference Name: IEEE Transactions on Nuclear Science, pp. 736–747. ISSN: 1558-1578. DOI: 10.1109/TNS.2011.2108316.

[47] J.A. Fessler, N.H. Clinthorne, and W.L. Rogers. “Regularized emission image reconstruction using imperfect side information”. In: *IEEE Transactions on Nuclear Science* 39.5 (Oct. 1992). Conference Name: IEEE Transactions on Nuclear Science, pp. 1464–1471. ISSN: 1558-1578. DOI: 10.1109/23.173225.

[48] G. Gindi et al. “Bayesian reconstruction of functional images using anatomical information as priors”. In: *IEEE Transactions on Medical Imaging* 12.4 (Dec. 1993). Conference Name: IEEE Transactions on Medical Imaging, pp. 670–680. ISSN: 1558-254X. DOI: 10.1109/42.251117.

[49] Xiaolong Ouyang et al. “Incorporation of correlated structural images in PET image reconstruction”. In: *IEEE Transactions on Medical Imaging* 13.4 (Dec. 1994). Conference Name: IEEE Transactions on Medical Imaging, pp. 627–640. ISSN: 1558-254X. DOI: 10.1109/42.363105.

[50] Babak A. Ardekani et al. “Minimum cross-entropy reconstruction of PET images using prior anatomical information”. en. In: *Phys. Med. Biol.* 41.11 (Nov. 1996), p. 2497. ISSN: 0031-9155. DOI: 10.1088/0031-9155/41/11/018. URL: https://dx.doi.org/10.1088/0031-9155/41/11/018 (visited on 03/14/2023).

[51] Boon-Keng Teo et al. “Partial-Volume Correction in PET: Validation of an Iterative Postreconstruction Method with Phantom and Patient Data”. en. In: *Journal of Nuclear Medicine* 48.5 (May 2007). Publisher: Society of Nuclear Medicine Section: BASIC SCIENCE INVESTIGATIONS, pp. 802–810. ISSN: 0161-5505, 2159-662X. DOI: 10.2967/jnumed.106.035576. URL: https://jnm.snmjournals.org/content/48/5/802 (visited on 03/14/2023).

[52] Anthonin Reilhac et al. “Iterative-based Partial Volume Effects correction with wavelet-based regularization for quantitative PET imaging”. In: *2011 IEEE Nuclear Science Symposium Conference Record*. ISSN: 1082-3654. Oct. 2011, pp. 3788–3791. DOI: 10.1109/NSSMIC.2011.6153717.

[53] A. S. Kirov, J. Z. Piao, and C. R. Schmidtlein. “Partial volume effect correction in PET using regularized iterative deconvolution with variance control based on local topology”. en. In: *Phys. Med. Biol.* 53.10 (Apr. 2008), p. 2577. ISSN: 0031-9155. DOI: 10.1088/0031-9155/53/10/009. URL: https://dx.doi.org/10.1088/0031-9155/53/10/009 (visited on 03/14/2023).

[54] David L. Barbee et al. “A method for partial volume correction of PET-imaged tumor heterogeneity using expectation maximization with a spatially varying point spread function”. en. In: *Phys. Med. Biol.* 55.1 (Dec. 2009), p. 221. ISSN: 0031-9155. DOI: 10.1088/0031-

9155/55/1/013. URL: https://dx.doi.org/10.1088/0031-9155/55/1/013 (visited on 03/14/2023).

[55] P. Calvini et al. "Fusion of the MR image to SPECT with possible correction for partial volume effects". In: *IEEE Transactions on Nuclear Science* 53.1 (Feb. 2006). Conference Name: IEEE Transactions on Nuclear Science, pp. 189–197. ISSN: 1558-1578. DOI: 10.1109/TNS.2005.862959.

[56] Miho Shidahara et al. "Functional and structural synergy for resolution recovery and partial volume correction in brain PET". en. In: *NeuroImage* 44.2 (Jan. 2009), pp. 340–348. ISSN: 1053-8119. DOI: 10.1016/j.neuroimage.2008.09.012. URL: https://www.sciencedirect.com/science/article/pii/S105381190800983X (visited on 03/14/2023).

[57] Fanny Orlhac et al. "Radiomics in PET Imaging: A Practical Guide for Newcomers". English. In: *PET Clinics* 16.4 (Oct. 2021). Publisher: Elsevier, pp. 597–612. ISSN: 1556-8598, 1879-9809. DOI: 10.1016/j.cpet.2021.06.007. URL: https://www.pet.theclinics.com/article/S1556-8598(21)00046-8/fulltext (visited on 09/13/2022).

[58] Marinke Westerterp et al. "Quantification of FDG PET studies using standardised uptake values in multi-centre trials: effects of image reconstruction, resolution and ROI definition parameters". eng. In: *Eur J Nucl Med Mol Imaging* 34.3 (Mar. 2007), pp. 392–404. ISSN: 1619-7070. DOI: 10.1007/s00259-006-0224-1.

[59] Alex Zwanenburg. "Radiomics in nuclear medicine: robustness, reproducibility, standardization, and how to avoid data analysis traps and replication crisis". en. In: *Eur J Nucl Med Mol Imaging* 46.13 (Dec. 2019), pp. 2638–2655. ISSN: 1619-7089. DOI: 10.1007/s00259-019-04391-8. URL: https://doi.org/10.1007/s00259-019-04391-8 (visited on 03/13/2023).

[60] L. G. Nyúl and J. K. Udupa. "On standardizing the MR image intensity scale". eng. In: *Magn Reson Med* 42.6 (Dec. 1999), pp. 1072–1081. ISSN: 0740-3194. DOI: 10.1002/(sici)1522-2594(199912)42:6<1072::aid-mrm11>3.0.co;2-m.

[61] Russell T. Shinohara et al. "Statistical normalization techniques for magnetic resonance imaging". en. In: *NeuroImage: Clinical* 6 (Jan. 2014), pp. 9–19. ISSN: 2213-1582. DOI: guo. URL: https://www.sciencedirect.com/science/article/pii/S221315821400117X (visited on 10/13/2022).

[62] Y. Ge et al. "Numerical tissue characterization in MS via standardization of the MR image intensity scale". eng. In: *J Magn Reson Imaging* 12.5 (Nov. 2000), pp. 715–721. ISSN: 1053-1807. DOI: 10.1002/1522-2586(200011)12:5<715::aid-jmri8>3.0.co;2-d.

[63] Kareem A. Wahid et al. "Intensity standardization methods in magnetic resonance imaging of head and neck cancer". en. In: *Physics and Imaging in Radiation Oncology* 20 (Oct. 2021), pp. 88–93. ISSN: 2405-6316. DOI: 10.1016/j.phro.2021.11.001. URL: https:

//www.sciencedirect.com/science/article/pii/S2405631621000658 (visited on 10/13/2022).

[64] Shunchao Guo et al. "Multimodal MRI Image Decision Fusion-Based Network for Glioma Classification". In: *Frontiers in Oncology* 12 (2022). ISSN: 2234-943X. URL: https://www.frontiersin.org/articles/10.3389/fonc.2022.819673 (visited on 03/14/2023).

[65] Jin-Ming Cao et al. "A radiomics model of liver CT to predict risk of hepatic encephalopathy secondary to hepatitis B related cirrhosis". eng. In: *Eur J Radiol* 130 (Sept. 2020), p. 109201. ISSN: 1872-7727. DOI: 10.1016/j.ejrad.2020.109201.

[66] Alexandre Carré et al. "Standardization of brain MR images across machines and protocols: bridging the gap for MRI-based radiomics". en. In: *Sci Rep* 10.1 (July 2020). Bandiera_abtest: a Cc_license_type: cc_by Cg_type: Nature Research Journals Number: 1 Primary_atype: Research Publisher: Nature Publishing Group Subject_term: Cancer imaging;CNS cancer;Computational science;Diagnostic markers Subject_term_id: cancer-imaging;cns-cancer;computational-science;diagnostic-markers, p. 12340. ISSN: 2045-2322. DOI: 10.1038/s41598-020-69298-z. URL: https://www.nature.com/articles/s41598-020-69298-z (visited on 09/20/2021).

[67] Yingping Li et al. "Impact of Preprocessing and Harmonization Methods on the Removal of Scanner Effects in Brain MRI Radiomic Features". en. In: *Cancers* 13.12 (Jan. 2021). Number: 12 Publisher: Multidisciplinary Digital Publishing Institute, p. 3000. ISSN: 2072-6694. DOI: 10.3390/cancers13123000. URL: https://www.mdpi.com/2072-6694/13/12/3000 (visited on 01/06/2023).

[68] Hubert Beaumont et al. "Harmonization of radiomic feature distributions: impact on classification of hepatic tissue in CT imaging". eng. In: *Eur Radiol* 31.8 (Aug. 2021), pp. 6059–6068. ISSN: 1432-1084. DOI: 10.1007/s00330-020-07641-8.

[69] R. W. Y. Granzier et al. "Test-Retest Data for the Assessment of Breast MRI Radiomic Feature Repeatability". eng. In: *J Magn Reson Imaging* 56.2 (Aug. 2022), pp. 592–604. ISSN: 1522-2586. DOI: 10.1002/jmri.28027.

[70] Rossana Castaldo et al. "A Complex Radiomic Signature in Luminal Breast Cancer from a Weighted Statistical Framework: A Pilot Study". eng. In: *Diagnostics (Basel)* 12.2 (Feb. 2022), p. 499. ISSN: 2075-4418. DOI: 10.3390/diagnostics12020499.

[71] L. G. Nyúl, J. K. Udupa, and X. Zhang. "New variants of a method of MRI scale standardization". eng. In: *IEEE Trans Med Imaging* 19.2 (Feb. 2000), pp. 143–150. ISSN: 0278-0062. DOI: 10.1109/42.836373.

[72] Mohak Shah et al. "Evaluating intensity normalization on MRIs of human brain with multiple sclerosis". eng. In: *Med Image Anal* 15.2 (Apr. 2011), pp. 267–282. ISSN: 1361-8423. DOI: 10.1016/j.media.2010.12.003.

[73] Amandine Crombé et al. “Intensity harmonization techniques influence radiomics features and radiomics-based predictions in sarcoma patients”. en. In: *Sci Rep* 10.1 (Sept. 2020). Number: 1 Publisher: Nature Publishing Group, p. 15496. ISSN: 2045-2322. DOI: 10.1038/s41598-020-72535-0. URL: https://www.nature.com/articles/s41598-020-72535-0 (visited on 03/14/2023).

[74] Lars J. Isaksson et al. “Effects of MRI image normalization techniques in prostate cancer radiomics”. eng. In: *Phys Med* 71 (Feb. 2020), pp. 7–13. ISSN: 1724-191X. DOI: 10.1016/j.ejmp.2020.02.007.

[75] Anant Madabhushi and Jayaram K. Udupa. “New methods of MR image intensity standardization via generalized scale”. en. In: *Medical Physics* 33.9 (2006). _eprint: https://aapm.onlinelibrary.w pp. 3426–3434. ISSN: 2473-4209. DOI: 10.1118/1.2335487. URL: https://onlinelibrary.wiley.com/doi/abs/10.1118/1.2335487 (visited on 03/14/2023).

[76] Lisanne V. van Dijk et al. “Parotid gland fat related Magnetic Resonance image biomarkers improve prediction of late radiation-induced xerostomia”. en. In: *Radiotherapy and Oncology* 128.3 (Sept. 2018), pp. 459–466. ISSN: 0167-8140. DOI: 10.1016/j.radonc.2018.06.012. URL: https://www.sciencedirect.com/science/article/pii/S0167814018333097 (visited on 03/15/2023).

[77] Elisa Scalco et al. “T2w-MRI signal normalization affects radiomics features reproducibility”. eng. In: *Med Phys* 47.4 (Apr. 2020), pp. 1680–1691. ISSN: 2473-4209. DOI: 10.1002/mp.14038.

[78] W. Evan Johnson, Cheng Li, and Ariel Rabinovic. “Adjusting batch effects in microarray expression data using empirical Bayes methods”. In: *Biostatistics* 8.1 (Jan. 2007), pp. 118–127. ISSN: 1465-4644. DOI: 10.1093/biostatistics/kxj037. URL: https://doi.org/10.1093/biostatistics/kxj037 (visited on 09/20/2021).

[79] Jean-Philippe Fortin et al. “Harmonization of multi-site diffusion tensor imaging data”. eng. In: *Neuroimage* 161 (Nov. 2017), pp. 149–170. ISSN: 1095-9572. DOI: 10.1016/j.neuroimage.2017.08.047.

[80] Jean-Philippe Fortin et al. “Harmonization of cortical thickness measurements across scanners and sites”. en. In: *NeuroImage* 167 (Feb. 2018), pp. 104–120. ISSN: 1053-8119. DOI: 10.1016/j.neuroimage.2017.11.024. URL: https://www.sciencedirect.com/science/article/pii/S105381191730931X (visited on 03/22/2022).

[81] R. Da-Ano et al. “Performance comparison of modified ComBat for harmonization of radiomic features for multicenter studies”. eng. In: *Sci Rep* 10.1 (June 2020), p. 10248. ISSN: 2045-2322. DOI: 10.1038/s41598-020-66110-w.

[82] Fanny Orlhac et al. “How can we combat multicenter variability in MR radiomics? Validation of a correction procedure”. eng. In: *Eur Radiol* 31.4 (Apr. 2021), pp. 2272–2280. ISSN: 1432-1084. DOI: 10.1007/s00330-020-07284-9.

[83] Clément Acquitter et al. “Radiomics-Based Detection of Radionecrosis Using Harmonized Multiparametric MRI”. en. In: *Cancers* 14.2 (Jan. 2022). Number: 2 Publisher: Multidisciplinary Digital Publishing Institute, p. 286. ISSN: 2072-6694. DOI: 10.3390/cancers14020286. URL: https://www.mdpi.com/2072-6694/14/2/286 (visited on 03/15/2023).

[84] Baderaldeen A. Altazi et al. “Reproducibility of F18-FDG PET radiomic features for different cervical tumor segmentation methods, gray-level discretization, and reconstruction algorithms”. en. In: *Journal of Applied Clinical Medical Physics* 18.6 (2017). _eprint: https://aapm.onlinelibrary.wiley.com/doi/pdf/10.1002/acm2.12170, pp. 32–48. ISSN: 1526-9914. DOI: 10.1002/acm2.12170. URL: http://onlinelibrary.wiley.com/doi/abs/10.1002/acm2.12170 (visited on 10/14/2022).

[85] Ralph T. H. Leijenaar et al. “The effect of SUV discretization in quantitative FDG-PET Radiomics: the need for standardized methodology in tumor texture analysis”. eng. In: *Sci Rep* 5 (Aug. 2015), p. 11075. ISSN: 2045-2322. DOI: 10.1038/srep11075.

[86] Jessica Goya-Outi et al. “Computation of reliable textural indices from multimodal brain MRI: suggestions based on a study of patients with diffuse intrinsic pontine glioma”. en. In: *Phys. Med. Biol.* 63.10 (May 2018). Publisher: IOP Publishing, p. 105003. ISSN: 0031-9155. DOI: 10.1088/1361-6560/aabd21. URL: https://dx.doi.org/10.1088/1361-6560/aabd21 (visited on 03/19/2023).

[87] Philippe Lambin et al. “Radiomics: extracting more information from medical images using advanced feature analysis”. eng. In: *Eur J Cancer* 48.4 (Mar. 2012), pp. 441–446. ISSN: 1879-0852. DOI: 10.1016/j.ejca.2011.11.036.

[88] Joost J. M. van Griethuysen et al. “Computational Radiomics System to Decode the Radiographic Phenotype”. eng. In: *Cancer Res* 77.21 (Nov. 2017), e104–e107. ISSN: 1538-7445. DOI: 10.1158/0008-5472.CAN-17-0339.

[89] Sylvain Reuzé et al. “Radiomics in Nuclear Medicine Applied to Radiation Therapy: Methods, Pitfalls, and Challenges”. en. In: *International Journal of Radiation Oncology*Biology*Physics.* Imaging in Radiation Oncology 102.4 (Nov. 2018), pp. 1117–1142. ISSN: 0360-3016. DOI: 10.1016/j.ijrobp.2018.05.022. URL: https://www.sciencedirect.com/science/article/pii/S0360301618308150 (visited on 01/06/2023).

[90] Petros-Pavlos Ypsilantis et al. “Predicting Response to Neoadjuvant Chemotherapy with PET Imaging Using Convolutional Neural Networks”. eng. In: *PLoS One* 10.9 (2015), e0137036. ISSN: 1932-6203. DOI: 10.1371/journal.pone.0137036.

[91] Parnian Afshar et al. “From Handcrafted to Deep-Learning-Based Cancer Radiomics: Challenges and Opportunities”. In: *IEEE Signal Processing Magazine* 36.4 (July 2019). Conference Name: IEEE Signal Processing Magazine, pp. 132–160. ISSN: 1558-0792. DOI: 10.1109/MSP.2019.2900993.

[92] A. Bizzego et al. *Integrating deep and radiomics features in cancer bioimaging*. en. Pages: 568170 Section: New Results. Mar. 2019. DOI: 10.1101/568170. URL: https://www.biorxiv.org/content/10.1101/568170v1 (visited on 01/06/2023).

[93] Zhenyuan Ning et al. "Multi-modal magnetic resonance imaging-based grading analysis for gliomas by integrating radiomics and deep features". In: *Ann Transl Med* 9.4 (Feb. 2021), p. 298. ISSN: 2305-5839. DOI: 10.21037/atm-20-4076. URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7944310/ (visited on 01/06/2023).

[94] Russell Frood et al. "Utility of pre-treatment FDG PET/CT-derived machine learning models for outcome prediction in classical Hodgkin lymphoma". eng. In: *Eur Radiol* 32.10 (Oct. 2022), pp. 7237–7247. ISSN: 1432-1084. DOI: 10.1007/s00330-022-09039-0.

[95] Heather M. Whitney et al. "Multi-Stage Harmonization for Robust AI across Breast MR Databases". eng. In: *Cancers (Basel)* 13.19 (Sept. 2021), p. 4809. ISSN: 2072-6694. DOI: 10.3390/cancers13194809.

[96] Paula Bos et al. "External validation of an MR-based radiomic model predictive of locoregional control in oropharyngeal cancer". eng. In: *Eur Radiol* 33.4 (Apr. 2023), pp. 2850–2860. ISSN: 1432-1084. DOI: 10.1007/s00330-022-09255-8.

[97] Bogdan Badic et al. "Prediction of recurrence after surgery in colorectal cancer patients using radiomics from diagnostic contrast-enhanced computed tomography: a two-center study". eng. In: *Eur Radiol* 32.1 (Jan. 2022), pp. 405–414. ISSN: 1432-1084. DOI: 10.1007/s00330-021-08104-4.

[98] Hannah Horng et al. "Improved generalized ComBat methods for harmonization of radiomic features". eng. In: *Sci Rep* 12.1 (Nov. 2022), p. 19009. ISSN: 2045-2322. DOI: 10.1038/s41598-022-23328-0.

[99] Hannah Horng et al. "Generalized ComBat harmonization methods for radiomic features with multi-modal distributions and multiple batch effects". en. In: *Sci Rep* 12.1 (Mar. 2022). Number: 1 Publisher: Nature Publishing Group, p. 4493. ISSN: 2045-2322. DOI: 10.1038/s41598-022-08412-9. URL: https://www.nature.com/articles/s41598-022-08412-9 (visited on 03/16/2023).

[100] Alexandre Carré et al. "AutoComBat: a generic method for harmonizing MRI-based radiomic features". en. In: *Sci Rep* 12.1 (July 2022). Number: 1 Publisher: Nature Publishing Group, p. 12762. ISSN: 2045-2322. DOI: 10.1038/s41598-022-16609-1. URL: https://www.nature.com/articles/s41598-022-16609-1 (visited on 01/25/2023).

[101] Doris Leithner et al. "Impact of ComBat harmonization on PET radiomics-based tissue classification: a dual-center PET/MR and PET/CT study". en. In: *Journal of Nuclear Medicine* (Feb. 2022). Publisher: Society of Nuclear Medicine Section: AI/Advanced Image Analysis. ISSN: 0161-5505, 2159-662X. DOI: 10.2967/jnumed.121.263102. URL: https:

//jnm.snmjournals.org/content/early/2022/02/24/jnumed.121.263102 (visited on 04/28/2022).

[102] François Lucia et al. "Multicentric development and evaluation of 18F-FDG PET/CT and MRI radiomics models to predict para-aortic lymph node involvement in locally advanced cervical cancer". eng. In: *Eur J Nucl Med Mol Imaging* (Mar. 2023). ISSN: 1619-7089. DOI: 10.1007/s00259-023-06180-w.

[103] Patrick Salome et al. "MR Intensity Normalization Methods Impact Sequence Specific Radiomics Prognostic Model Performance in Primary and Recurrent High-Grade Glioma". eng. In: *Cancers (Basel)* 15.3 (Feb. 2023), p. 965. ISSN: 2072-6694. DOI: 10.3390/cancers15030965.

[104] Adrian Jun Zounek et al. "Feasibility of radiomic feature harmonization for pooling of [18F]FET or [18F]GE-180 PET images of gliomas". eng. In: *Z Med Phys* (Jan. 2023), S0939–3889(22)00138–6. ISSN: 1876-4436. DOI: 10.1016/j.zemedi.2022.12.005.

[105] Hui Xu et al. "Radiomics prognostic analysis of PET/CT images in a multicenter head and neck cancer cohort: investigating ComBat strategies, sub-volume characterization, and automatic segmentation". eng. In: *Eur J Nucl Med Mol Imaging* (Jan. 2023). ISSN: 1619-7089. DOI: 10.1007/s00259-023-06118-2.

[106] Xiao-Li Song et al. "Multisequence magnetic resonance imaging-based radiomics models for the prediction of microsatellite instability in endometrial cancer". eng. In: *Radiol Med* 128.2 (Feb. 2023), pp. 242–251. ISSN: 1826-6983. DOI: 10.1007/s11547-023-01590-0.

[107] Dongyang Du et al. "Deep learning-based harmonization of CT reconstruction kernels towards improved clinical task performance". eng. In: *Eur Radiol* 33.4 (Apr. 2023), pp. 2426–2438. ISSN: 1432-1084. DOI: 10.1007/s00330-022-09229-w.

[108] Zelda Paquier et al. "Pre-trial quality assurance of diffusion-weighted MRI for radiomic analysis and the role of harmonisation". eng. In: *Phys Med* 103 (Nov. 2022), pp. 138–146. ISSN: 1724-191X. DOI: 10.1016/j.ejmp.2022.10.009.

[109] Xiao Chang et al. "Self-supervised learning for multi-center magnetic resonance imaging harmonization without traveling phantoms". eng. In: *Phys Med Biol* 67.14 (July 2022). ISSN: 1361-6560. DOI: 10.1088/1361-6560/ac7b66.

[110] Turkey Refaee et al. "CT Reconstruction Kernels and the Effect of Pre- and Post-Processing on the Reproducibility of Handcrafted Radiomic Features". eng. In: *J Pers Med* 12.4 (Mar. 2022), p. 553. ISSN: 2075-4426. DOI: 10.3390/jpm12040553.

[111] Ronrick Da-Ano et al. "A transfer learning approach to facilitate ComBat-based harmonization of multicentre radiomic features in new datasets". eng. In: *PLoS One* 16.7 (2021), e0253653. ISSN: 1932-6203. DOI: 10.1371/journal.pone.0253653.

[112] Marta Ferreira et al. "[18F]FDG PET radiomics to predict disease-free survival in cervical cancer: a multi-scanner/center study with external validation". en. In: *Eur J Nucl Med Mol Imaging* 48.11 (Oct. 2021), pp. 3432–3443. ISSN: 1619-7089. DOI: 10.1007/s00259-021-05303-5. URL: https://doi.org/10.1007/s00259-021-05303-5 (visited on 04/22/2022).

[113] Marie-Judith Saint Martin et al. "A radiomics pipeline dedicated to Breast MRI: validation on a multi-scanner phantom study". eng. In: *MAGMA* 34.3 (June 2021), pp. 355–366. ISSN: 1352-8661. DOI: 10.1007/s10334-020-00892-y.

[114] Noemi Garau et al. "External validation of radiomics-based predictive models in low-dose CT screening for early lung cancer diagnosis". eng. In: *Med Phys* 47.9 (Sept. 2020), pp. 4125–4136. ISSN: 2473-4209. DOI: 10.1002/mp.14308.

[115] Burak Koçak et al. "Radiomics with artificial intelligence: a practical guide for beginners". In: *Diagn Interv Radiol* 25.6 (Nov. 2019), pp. 485–495. ISSN: 1305-3825. DOI: 10.5152/dir.2019.19321. URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6837295/ (visited on 01/09/2023).

[116] Fabian Pedregosa et al. "Scikit-learn: Machine Learning in Python". In: *Journal of Machine Learning Research* 12.85 (2011), pp. 2825–2830. ISSN: 1533-7928. URL: http://jmlr.org/papers/v12/pedregosa11a.html (visited on 01/09/2023).

[117] Sebastian Pölsterl. "scikit-survival: A Library for Time-to-Event Analysis Built on Top of scikit-learn". In: *Journal of Machine Learning Research* 21.212 (2020), pp. 1–6. ISSN: 1533-7928. URL: http://jmlr.org/papers/v21/20-729.html (visited on 02/17/2022).

[118] R. Stuart Geiger et al. ""Garbage in, garbage out" revisited: What do machine learning application papers report about human-labeled training data?" In: *Quantitative Science Studies* 2.3 (Nov. 2021), pp. 795–827. ISSN: 2641-3337. DOI: 10.1162/qss_a_00144. URL: https://doi.org/10.1162/qss_a_00144 (visited on 01/09/2023).

[119] Jason Brownlee. *Data Preparation for Machine Learning: Data Cleaning, Feature Selection, and Data Transforms in Python*. en. Google-Books-ID: uAPuDwAAQBAJ. Machine Learning Mastery, June 2020.

[120] Payam Refaeilzadeh, Lei Tang, and Huan Liu. "Cross-Validation". en. In: *Encyclopedia of Database Systems*. Ed. by LING LIU and M. TAMER ÖZSU. Boston, MA: Springer US, 2009, pp. 532–538. ISBN: 978-0-387-39940-9. DOI: 10.1007/978-0-387-39940-9_565. URL: https://doi.org/10.1007/978-0-387-39940-9_565 (visited on 01/09/2023).

[121] Alexandre Carré. "Radiomics and Artificial Intelligence Applied to Conventional MRI : Towards Personalized Treatment of Brain Tumors". These de doctorat. université Paris-Saclay, Dec. 2021. URL: https://www.theses.fr/2021UPASL108 (visited on 01/09/2023).

[122] Max Kuhn and Kjell Johnson. *Applied Predictive Modeling*. en. New York, NY: Springer, 2013. ISBN: 978-1-4614-6848-6 978-1-4614-6849-3. DOI: 10.1007/978-1-4614-6849-3. URL: http://link.springer.com/10.1007/978-1-4614-6849-3 (visited on 01/09/2023).

[123] Gary Smith. “Step away from stepwise”. In: *Journal of Big Data* 5.1 (Sept. 2018), p. 32. ISSN: 2196-1115. DOI: 10.1186/s40537-018-0143-6. URL: https://doi.org/10.1186/s40537-018-0143-6 (visited on 01/09/2023).

[124] Ian H. Witten et al. *Data Mining: Practical Machine Learning Tools and Techniques*. Anglais. 4e édition. Amsterdam: Morgan Kaufmann, Nov. 2016. ISBN: 978-0-12-804291-5.

[125] Kevin P. Murphy. *Machine Learning: A Probabilistic Perspective*. Anglais. Cambridge, MA: The MIT Press, 2012. ISBN: 978-0-262-01802-9.

[126] Trevor Hastie, Jerome Friedman, and Robert Tibshirani. *The Elements of Statistical Learning*. Springer Series in Statistics. New York, NY: Springer, 2001. ISBN: 978-1-4899-0519-2 978-0-387-21606-5. DOI: 10.1007/978-0-387-21606-5. URL: http://link.springer.com/10.1007/978-0-387-21606-5 (visited on 01/10/2023).

[127] Ana Barragán-Montero et al. “Artificial intelligence and machine learning for medical imaging: A technology review”. en. In: *Physica Medica* 83 (Mar. 2021), pp. 242–256. ISSN: 1120-1797. DOI: 10.1016/j.ejmp.2021.04.016. URL: https://www.sciencedirect.com/science/article/pii/S1120179721001733 (visited on 10/15/2021).

[128] Christian S. Perone et al. *Unsupervised domain adaptation for medical imaging segmentation with self-ensembling*. arXiv:1811.06042 [cs]. Jan. 2019. DOI: 10.48550/arXiv.1811.06042. URL: http://arxiv.org/abs/1811.06042 (visited on 03/20/2023).

[129] Yingzi Liu et al. “MRI-based treatment planning for proton radiotherapy: dosimetric validation of a deep learning-based liver synthetic CT generation method”. eng. In: *Phys Med Biol* 64.14 (July 2019), p. 145015. ISSN: 1361-6560. DOI: 10.1088/1361-6560/ab25bc.

[130] Emilie Alvarez Andres et al. “Dosimetry-Driven Quality Measure of Brain Pseudo Computed Tomography Generated From Deep Learning for MRI-Only Radiation Therapy Treatment Planning”. en. In: *International Journal of Radiation Oncology*Biology*Physics* 108.3 (Nov. 2020), pp. 813–823. ISSN: 0360-3016. DOI: 10.1016/j.ijrobp.2020.05.006. URL: https://www.sciencedirect.com/science/article/pii/S0360301620311305 (visited on 01/09/2023).

[131] Iman Aganj et al. “Unsupervised Medical Image Segmentation Based on the Local Center of Mass”. en. In: *Sci Rep* 8.1 (Aug. 2018). Number: 1 Publisher: Nature Publishing Group, p. 13012. ISSN: 2045-2322. DOI: 10.1038/s41598-018-31333-5. URL: https://www.nature.com/articles/s41598-018-31333-5 (visited on 03/20/2023).

[132] Cheng-Bin Jin et al. “DC2Anet: Generating Lumbar Spine MR Images from CT Scan Data Based on Semi-Supervised Learning”. en. In: *Applied Sciences* 9.12 (Jan. 2019). Number: 12 Publisher: Multidisciplinary Digital Publishing Institute, p. 2521. ISSN: 2076-3417. DOI: 10.3390/app9122521. URL: https://www.mdpi.com/2076-3417/9/12/2521 (visited on 03/20/2023).

[133] William Burton, Casey Myers, and Paul Rullkoetter. "Semi-supervised learning for automatic segmentation of the knee from MRI with convolutional neural networks". eng. In: *Comput Methods Programs Biomed* 189 (June 2020), p. 105328. ISSN: 1872-7565. DOI: 10.1016/j.cmpb.2020.105328.

[134] Yang Lei et al. "MRI-Only Based Synthetic CT Generation Using Dense Cycle Consistent Generative Adversarial Networks". en. In: *Medical physics* 46.8 (Aug. 2019). Publisher: NIH Public Access, p. 3565. DOI: 10.1002/mp.13617. URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6692192/ (visited on 11/09/2022).

[135] Dong Nie et al. "Medical Image Synthesis with Context-Aware Generative Adversarial Networks". In: *Med Image Comput Comput Assist Interv* 10435 (Sept. 2017), pp. 417–425. DOI: 10.1007/978-3-319-66179-7_48. URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6044459/ (visited on 11/04/2022).

[136] Nesrine Bnouni et al. "Context-Aware Synergetic Multiplex Network for Multi-organ Segmentation of Cervical Cancer MRI". en. In: *Predictive Intelligence in Medicine*. Ed. by Islem Rekik et al. Lecture Notes in Computer Science. Cham: Springer International Publishing, 2020, pp. 1–11. ISBN: 978-3-030-59354-4. DOI: 10.1007/978-3-030-59354-4_1.

[137] Gourav Modanwal et al. "MRI image harmonization using cycle-consistent generative adversarial network". In: *Medical Imaging 2020: Computer-Aided Diagnosis*. Vol. 11314. SPIE, Mar. 2020, pp. 259–264. DOI: 10.1117/12.2551301. URL: https://www.spiedigitallibrary.org/conference-proceedings-of-spie/11314/1131413/MRI-image-harmonization-using-cycle-consistent-generative-adversarial-network/10.1117/12.2551301.full (visited on 01/18/2023).

[138] Erick Costa de Farias et al. "Impact of GAN-based lesion-focused medical image super-resolution on the robustness of radiomic features". en. In: *Sci Rep* 11.1 (Nov. 2021). Number: 1 Publisher: Nature Publishing Group, p. 21361. ISSN: 2045-2322. DOI: 10.1038/s41598-021-00898-z. URL: https://www.nature.com/articles/s41598-021-00898-z (visited on 01/27/2023).

[139] Junhua Chen et al. "Improving reproducibility and performance of radiomics in low-dose CT using cycle GANs". eng. In: *J Appl Clin Med Phys* 23.10 (Oct. 2022), e13739. ISSN: 1526-9914. DOI: 10.1002/acm2.13739.

[140] Chenyang Shen et al. "Intelligent inverse treatment planning via deep reinforcement learning, a proof-of-principle study in high dose-rate brachytherapy for cervical cancer". eng. In: *Phys Med Biol* 64.11 (May 2019), p. 115013. ISSN: 1361-6560. DOI: 10.1088/1361-6560/ab18bf.

[141] Zhe Li and Yong Xia. "Deep Reinforcement Learning for Weakly-Supervised Lymph Node Segmentation in CT Images". In: *IEEE Journal of Biomedical and Health Informatics*

25.3 (Mar. 2021). Conference Name: IEEE Journal of Biomedical and Health Informatics, pp. 774–783. ISSN: 2168-2208. DOI: 10.1109/JBHI.2020.3008759.

[142] Solveig Badillo et al. "An Introduction to Machine Learning". en. In: *Clinical Pharmacology & Therapeutics* 107.4 (2020). _eprint: https://ascpt.onlinelibrary.wiley.com/doi/pdf/10.1002/cpt.1796, pp. 871–885. ISSN: 1532-6535. DOI: 10.1002/cpt.1796. URL: https://onlinelibrary.wiley.com/doi/abs/10.1002/cpt.1796 (visited on 01/09/2023).

[143] James Bergstra and Yoshua Bengio. "Random Search for Hyper-Parameter Optimization". en. In: ().

[144] Randal S. Olson et al. "Evaluation of a Tree-based Pipeline Optimization Tool for Automating Data Science". In: *Proceedings of the Genetic and Evolutionary Computation Conference 2016*. GECCO '16. New York, NY, USA: Association for Computing Machinery, July 2016, pp. 485–492. ISBN: 978-1-4503-4206-3. DOI: 10.1145/2908812.2908918. URL: https://dl.acm.org/doi/10.1145/2908812.2908918 (visited on 03/20/2023).

[145] Warren S. McCulloch and Walter Pitts. "A logical calculus of the ideas immanent in nervous activity". en. In: *Bulletin of Mathematical Biophysics* 5.4 (Dec. 1943), pp. 115–133. ISSN: 1522-9602. DOI: 10.1007/BF02478259. URL: https://doi.org/10.1007/BF02478259 (visited on 03/22/2023).

[146] F. Rosenblatt. "The perceptron: A probabilistic model for information storage and organization in the brain." en. In: *Psychological Review* 65.6 (1958), pp. 386–408. ISSN: 1939-1471, 0033-295X. DOI: 10.1037/h0042519. URL: http://doi.apa.org/getdoi.cfm?doi=10.1037/h0042519 (visited on 03/20/2023).

[147] Achim Hekler et al. "Deep learning outperformed 11 pathologists in the classification of histopathological melanoma images". en. In: *European Journal of Cancer* 118 (Sept. 2019), pp. 91–96. ISSN: 0959-8049. DOI: 10.1016/j.ejca.2019.06.012. URL: https://www.sciencedirect.com/science/article/pii/S0959804919303806 (visited on 03/20/2023).

[148] Titus J. Brinker et al. "Deep learning outperformed 136 of 157 dermatologists in a head-to-head dermoscopic melanoma image classification task". English. In: *European Journal of Cancer* 113 (May 2019). Publisher: Elsevier, pp. 47–54. ISSN: 0959-8049, 1879-0852. DOI: 10.1016/j.ejca.2019.04.001. URL: https://www.ejcancer.com/article/S0959-8049%2819%2930221-7/fulltext (visited on 03/20/2023).

[149] David Killock. "AI outperforms radiologists in mammographic screening". en. In: *Nat Rev Clin Oncol* 17.3 (Mar. 2020). Number: 3 Publisher: Nature Publishing Group, pp. 134–134. ISSN: 1759-4782. DOI: 10.1038/s41571-020-0329-7. URL: https://www.nature.com/articles/s41571-020-0329-7 (visited on 03/20/2023).

[150] Anna Braghetto et al. “Radiomics and deep learning methods for the prediction of 2-year overall survival in LUNG1 dataset”. en. In: *Sci Rep* 12.1 (Aug. 2022). Number: 1 Publisher: Nature Publishing Group, p. 14132. ISSN: 2045-2322. DOI: 10.1038/s41598-022-18085-z. URL: https://www.nature.com/articles/s41598-022-18085-z (visited on 01/10/2023).

[151] Jean-Emmanuel Bibault et al. “Deep Learning and Radiomics predict complete response after neo-adjuvant chemoradiation for locally advanced rectal cancer”. en. In: *Sci Rep* 8.1 (Aug. 2018). Number: 1 Publisher: Nature Publishing Group, pp. 1–8. ISSN: 2045-2322. DOI: 10.1038/s41598-018-30657-6. URL: https://www.nature.com/articles/s41598-018-30657-6 (visited on 01/10/2023).

[152] Taotao Dong et al. “Development and Validation of a Deep Learning Radiomics Model Predicting Lymph Node Status in Operable Cervical Cancer”. In: *Frontiers in Oncology* 10 (2020). ISSN: 2234-943X. URL: https://www.frontiersin.org/articles/10.3389/fonc.2020.00464 (visited on 01/10/2023).

[153] Mingxu Huang et al. “Segmentation of Clinical Target Volume From CT Images for Cervical Cancer Using Deep Learning”. en. In: *Technol Cancer Res Treat* 22 (Jan. 2023). Publisher: SAGE Publications Inc, p. 15330338221139164. ISSN: 1533-0346. DOI: 10.1177/15330338221139164. URL: https://doi.org/10.1177/15330338221139164 (visited on 01/10/2023).

[154] Lise Wei et al. “A deep survival interpretable radiomics model of hepatocellular carcinoma patients”. eng. In: *Phys Med* 82 (Feb. 2021), pp. 295–305. ISSN: 1724-191X. DOI: 10.1016/j.ejmp.2021.02.013.

[155] Kavi Fatania et al. “Harmonisation of scanner-dependent contrast variations in magnetic resonance imaging for radiation oncology, using style-blind auto-encoders”. en. In: *Physics and Imaging in Radiation Oncology* 22 (Apr. 2022), pp. 115–122. ISSN: 2405-6316. DOI: 10.1016/j.phro.2022.05.005. URL: https://www.sciencedirect.com/science/article/pii/S2405631622000458 (visited on 01/10/2023).

[156] Li Wang et al. *Convolutional Restricted Boltzmann Machine Based-Radiomics for Prediction of Pathological Complete Response to Neoadjuvant Chemotherapy in Breast Cancer*. arXiv:1905.13312 [cs, eess]. May 2019. DOI: 10.48550/arXiv.1905.13312. URL: http://arxiv.org/abs/1905.13312 (visited on 01/10/2023).

[157] Y. LeCun et al. “Backpropagation Applied to Handwritten Zip Code Recognition”. In: *Neural Computation* 1.4 (Dec. 1989). Conference Name: Neural Computation, pp. 541–551. ISSN: 0899-7667. DOI: 10.1162/neco.1989.1.4.541.

[158] Zewen Li et al. “A Survey of Convolutional Neural Networks: Analysis, Applications, and Prospects”. en. In: *IEEE Trans. Neural Netw. Learning Syst.* 33.12 (Dec. 2022), pp. 6999–7019. ISSN: 2162-237X, 2162-2388. DOI: 10.1109/TNNLS.2021.3084827. URL: https://ieeexplore.ieee.org/document/9451544/ (visited on 03/22/2023).

[159] Ian J. Goodfellow et al. *Generative Adversarial Networks*. arXiv:1406.2661 [cs, stat]. June 2014. DOI: 10.48550/arXiv.1406.2661. URL: http://arxiv.org/abs/1406.2661 (visited on 03/22/2023).

[160] Phillip Isola et al. *Image-to-Image Translation with Conditional Adversarial Networks*. arXiv:1611.07004 [cs]. Nov. 2018. DOI: 10.48550/arXiv.1611.07004. URL: http://arxiv.org/abs/1611.07004 (visited on 11/04/2022).

[161] Maayan Frid-Adar et al. "Synthetic data augmentation using GAN for improved liver lesion classification". In: Apr. 2018, pp. 289–293. DOI: 10.1109/ISBI.2018.8363576.

[162] Jelmer M. Wolterink et al. "Deep MR to CT Synthesis Using Unpaired Data". en. In: *Simulation and Synthesis in Medical Imaging*. Ed. by Sotirios A. Tsaftaris et al. Lecture Notes in Computer Science. Cham: Springer International Publishing, 2017, pp. 14–23. ISBN: 978-3-319-68127-6. DOI: 10.1007/978-3-319-68127-6_2.

[163] Axel Sauer, Katja Schwarz, and Andreas Geiger. *StyleGAN-XL: Scaling StyleGAN to Large Diverse Datasets*. arXiv:2202.00273 [cs] version: 2. May 2022. DOI: 10.48550/arXiv.2202.00273. URL: http://arxiv.org/abs/2202.00273 (visited on 03/22/2023).

[164] Mehdi Mirza and Simon Osindero. *Conditional Generative Adversarial Nets*. arXiv:1411.1784 [cs, stat]. Nov. 2014. DOI: 10.48550/arXiv.1411.1784. URL: http://arxiv.org/abs/1411.1784 (visited on 04/25/2023).

[165] Jun-Yan Zhu et al. "Unpaired Image-to-Image Translation Using Cycle-Consistent Adversarial Networks". In: *2017 IEEE International Conference on Computer Vision (ICCV)*. ISSN: 2380-7504. Oct. 2017, pp. 2242–2251. DOI: 10.1109/ICCV.2017.244.

[166] Théo Estienne et al. "U-ReSNet: Ultimate Coupling of Registration and Segmentation with Deep Nets". en. In: *Medical Image Computing and Computer Assisted Intervention – MICCAI 2019*. Ed. by Dinggang Shen et al. Lecture Notes in Computer Science. Cham: Springer International Publishing, 2019, pp. 310–319. ISBN: 978-3-030-32248-9. DOI: 10.1007/978-3-030-32248-9_35.

[167] Jesse Davis and Mark Goadrich. "The relationship between Precision-Recall and ROC curves". In: *Proceedings of the 23rd international conference on Machine learning*. ICML '06. New York, NY, USA: Association for Computing Machinery, June 2006, pp. 233–240. ISBN: 978-1-59593-383-6. DOI: 10.1145/1143844.1143874. URL: https://doi.org/10.1145/1143844.1143874 (visited on 04/03/2023).

[168] Yunqian Ma and Haibo He, eds. *Imbalanced Learning: Foundations, Algorithms, and Applications*. English. 1st edition. Hoboken, NJ: Wiley-IEEE Press, July 2013. ISBN: 978-1-118-07462-6.

[169] Hassan Bagher-Ebadian et al. "Application of radiomics for the prediction of HPV status for patients with head and neck cancers". eng. In: *Med Phys* 47.2 (Feb. 2020), pp. 563–575. ISSN: 2473-4209. DOI: 10.1002/mp.13977.

[170] Bolin Song et al. “Radiomic Features Associated With HPV Status on Pretreatment Computed Tomography in Oropharyngeal Squamous Cell Carcinoma Inform Clinical Prognosis”. eng. In: *Front Oncol* 11 (2021), p. 744250. ISSN: 2234-943X. DOI: 10.3389/fonc.2021.744250.

[171] Xueming Xia et al. “Radiomics Based on Nomogram Predict Pelvic Lymphnode Metastasis in Early-Stage Cervical Cancer”. en. In: *Diagnostics* 12.10 (Oct. 2022). Number: 10 Publisher: Multidisciplinary Digital Publishing Institute, p. 2446. ISSN: 2075-4418. DOI: 10.3390/diagnostics12102446. URL: https://www.mdpi.com/2075-4418/12/10/2446 (visited on 01/05/2023).

[172] Yu-Chun Lin et al. “Deep learning for fully automated tumor segmentation and extraction of magnetic resonance radiomics features in cervical cancer”. eng. In: *Eur Radiol* 30.3 (Mar. 2020), pp. 1297–1305. ISSN: 1432-1084. DOI: 10.1007/s00330-019-06467-3.

[173] Hanna Sartor et al. “Auto-segmentations by convolutional neural network in cervical and anorectal cancer with clinical structure sets as the ground truth”. en. In: *Clinical and Translational Radiation Oncology* 25 (Nov. 2020), pp. 37–45. ISSN: 2405-6308. DOI: 10.1016/j.ctro.2020.09.004. URL: https://www.sciencedirect.com/science/article/pii/S2405630820300756 (visited on 01/03/2023).

[174] I. El Naqa et al. “Exploring feature-based approaches in PET images for predicting cancer treatment outcomes”. eng. In: *Pattern Recognit* 42.6 (June 2009), pp. 1162–1171. ISSN: 0031-3203. DOI: 10.1016/j.patcog.2008.08.011.

[175] Sylvain Reuzé et al. “Prediction of cervical cancer recurrence using textural features extracted from 18 F-FDG PET images acquired with different scanners”. In: *Oncotarget* 8.26 (May 2017). Publisher: Impact Journals, pp. 43169–43179. ISSN: 1949-2553. DOI: 10.18632/oncotarget.17856. URL: https://www.oncotarget.com/article/17856/text/ (visited on 01/26/2021).

[176] P. J. Brown et al. “Prediction of outcome in anal squamous cell carcinoma using radiomic feature analysis of pre-treatment FDG PET-CT”. eng. In: *Eur J Nucl Med Mol Imaging* 46.13 (Dec. 2019), pp. 2790–2799. ISSN: 1619-7089. DOI: 10.1007/s00259-019-04495-1.

[177] Hu Cong et al. “FDG-PET/CT Radiomics Models for The Early Prediction of Locoregional Recurrence in Head and Neck Cancer”. eng. In: *Curr Med Imaging* 17.3 (2021), pp. 374–383. ISSN: 1573-4056. DOI: 10.2174/1573405616666200712181135.

[178] Kathinka S. Slørdahl et al. “Treatment outcomes and prognostic factors after chemoradiotherapy for anal cancer”. eng. In: *Acta Oncol* 60.7 (July 2021), pp. 921–930. ISSN: 1651-226X. DOI: 10.1080/0284186X.2021.1918763.

[179] Haowei Zhang et al. “Locally advanced squamous cell carcinoma of the head and neck: CT texture and histogram analysis allow independent prediction of overall survival in patients treated with induction chemotherapy”. eng. In: *Radiology* 269.3 (Dec. 2013), pp. 801–809. ISSN: 1527-1315. DOI: 10.1148/radiol.13130110.

[180] James S. Lewis Jr et al. “Human Papillomavirus Testing in Head and Neck Carcinomas: Guideline From the College of American Pathologists”. In: *Archives of Pathology & Laboratory Medicine* 142.5 (Dec. 2017), pp. 559–597. ISSN: 0003-9985. DOI: 10.5858/arpa.2017-0286-CP. URL: https://doi.org/10.5858/arpa.2017-0286-CP (visited on 03/08/2023).

[181] Hassan Bagher-Ebadian et al. “Radiomics outperforms clinical factors in characterizing human papilloma virus (HPV) for patients with oropharyngeal squamous cell carcinomas”. eng. In: *Biomed Phys Eng Express* 8.4 (June 2022). ISSN: 2057-1976. DOI: 10.1088/2057-1976/ac39ab.

[182] S. Ramkumar et al. “MRI-Based Texture Analysis to Differentiate Sinonasal Squamous Cell Carcinoma from Inverted Papilloma”. eng. In: *AJNR Am J Neuroradiol* 38.5 (May 2017), pp. 1019–1025. ISSN: 1936-959X. DOI: 10.3174/ajnr.A5106.

[183] Hugo J. W. L. Aerts et al. “Decoding tumour phenotype by noninvasive imaging using a quantitative radiomics approach”. en. In: *Nat Commun* 5.1 (June 2014). Bandiera_abtest: a Cg_type: Nature Research Journals Number: 1 Primary_atype: Research Publisher: Nature Publishing Group Subject_term: Cancer imaging;Medical research Subject_term_id: cancer-imaging;medical-research, p. 4006. ISSN: 2041-1723. DOI: 10.1038/ncomms5006. URL: http://www.nature.com/articles/ncomms5006 (visited on 09/20/2021).

[184] Chao Huang et al. “Development and validation of radiomic signatures of head and neck squamous cell carcinoma molecular features and subtypes”. eng. In: *EBioMedicine* 45 (July 2019), pp. 70–80. ISSN: 2352-3964. DOI: 10.1016/j.ebiom.2019.06.034.

[185] Yitan Zhu et al. “Imaging-Genomic Study of Head and Neck Squamous Cell Carcinoma: Associations Between Radiomic Phenotypes and Genomic Mechanisms via Integration of The Cancer Genome Atlas and The Cancer Imaging Archive”. eng. In: *JCO Clin Cancer Inform* 3 (Feb. 2019), pp. 1–9. ISSN: 2473-4276. DOI: 10.1200/CCI.18.00073.

[186] Elizabeth A. Kidd et al. “Lymph node staging by positron emission tomography in cervical cancer: relationship to prognosis”. eng. In: *J Clin Oncol* 28.12 (Apr. 2010), pp. 2108–2113. ISSN: 1527-7755. DOI: 10.1200/JCO.2009.25.4151.

[187] Wei Mu et al. “Staging of cervical cancer based on tumor heterogeneity characterized by texture features on 18F-FDG PET images”. en. In: *Phys. Med. Biol.* 60.13 (June 2015). Publisher: IOP Publishing, p. 5123. ISSN: 0031-9155. DOI: 10.1088/0031-9155/60/13/5123. URL: https://dx.doi.org/10.1088/0031-9155/60/13/5123 (visited on 03/08/2023).

[188] Longchao Li et al. "A meta-analysis of MRI-based radiomic features for predicting lymph node metastasis in patients with cervical cancer". en. In: *European Journal of Radiology* 151 (June 2022), p. 110243. ISSN: 0720-048X. DOI: 10.1016/j.ejrad.2022.110243. URL: https://www.sciencedirect.com/science/article/pii/S0720048X22000936 (visited on 04/22/2022).

[189] Wei-Chih Shen et al. "[18]Fluorodeoxyglucose Positron Emission Tomography for the Textural Features of Cervical Cancer Associated with Lymph Node Metastasis and Histological Type". en. In: *Eur J Nucl Med Mol Imaging* 44.10 (Sept. 2017), pp. 1721–1731. ISSN: 1619-7089. DOI: 10.1007/s00259-017-3697-1. URL: https://doi.org/10.1007/s00259-017-3697-1 (visited on 03/08/2023).

[190] Tetsuya Tsujikawa et al. "18F-FDG PET radiomics approaches: comparing and clustering features in cervical cancer". en. In: *Ann Nucl Med* 31.9 (Nov. 2017), pp. 678–685. ISSN: 1864-6433. DOI: 10.1007/s12149-017-1199-7. URL: https://doi.org/10.1007/s12149-017-1199-7 (visited on 03/08/2023).

[191] Rui-Yun Chen et al. "Associations of Tumor PD-1 Ligands, Immunohistochemical Studies, and Textural Features in 18F-FDG PET in Squamous Cell Carcinoma of the Head and Neck". en. In: *Sci Rep* 8.1 (Jan. 2018). Number: 1 Publisher: Nature Publishing Group, p. 105. ISSN: 2045-2322. DOI: 10.1038/s41598-017-18489-2. URL: https://www.nature.com/articles/s41598-017-18489-2 (visited on 03/08/2023).

[192] Roger Sun et al. "A radiomics approach to assess tumour-infiltrating CD8 cells and response to anti-PD-1 or anti-PD-L1 immunotherapy: an imaging biomarker, retrospective multicohort study". eng. In: *Lancet Oncol* 19.9 (Sept. 2018), pp. 1180–1191. ISSN: 1474-5488. DOI: 10.1016/S1470-2045(18)30413-3.

[193] Roger Sun et al. "Radiomics to predict outcomes and abscopal response of patients with cancer treated with immunotherapy combined with radiotherapy using a validated signature of CD8 cells". In: *J Immunother Cancer* 8.2 (Nov. 2020), e001429. ISSN: 2051-1426. DOI: 10.1136/jitc-2020-001429. URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7668366/ (visited on 05/23/2022).

[194] Kristy K. Brock. *Image Processing in Radiation Therapy*. fr. URL: https://www.lavoisier.fr/livre/medecine/image-processing-in-radiation-therapy/descriptif_4381303 (visited on 01/03/2023).

[195] Michael Lempart et al. "Pelvic U-Net: multi-label semantic segmentation of pelvic organs at risk for radiation therapy anal cancer patients using a deeply supervised shuffle attention convolutional neural network". In: *Radiation Oncology* 17.1 (June 2022), p. 114. ISSN: 1748-717X. DOI: 10.1186/s13014-022-02088-1. URL: https://doi.org/10.1186/s13014-022-02088-1 (visited on 01/03/2023).

[196] Tim Lustberg et al. “Clinical evaluation of atlas and deep learning based automatic contouring for lung cancer”. eng. In: *Radiother Oncol* 126.2 (Feb. 2018), pp. 312–317. ISSN: 1879-0887. DOI: 10.1016/j.radonc.2017.11.012.

[197] W. Jeffrey Zabel et al. “Clinical Evaluation of Deep Learning and Atlas-Based Auto-Contouring of Bladder and Rectum for Prostate Radiation Therapy”. eng. In: *Pract Radiat Oncol* 11.1 (2021), e80–e89. ISSN: 1879-8519. DOI: 10.1016/j.prro.2020.05.013.

[198] Wei Huang, Kap Luk Chan, and Jiayin Zhou. “Region-Based Nasopharyngeal Carcinoma Lesion Segmentation from MRI Using Clustering- and Classification-Based Methods with Learning”. en. In: *J Digit Imaging* 26.3 (June 2013), pp. 472–482. ISSN: 1618-727X. DOI: 10.1007/s10278-012-9520-4. URL: https://doi.org/10.1007/s10278-012-9520-4 (visited on 01/03/2023).

[199] Turid Torheim et al. “Autodelineation of cervical cancers using multiparametric magnetic resonance imaging and machine learning”. eng. In: *Acta Oncol* 56.6 (June 2017), pp. 806–812. ISSN: 1651-226X. DOI: 10.1080/0284186X.2017.1285499.

[200] Francesco Bianconi et al. “Comparative evaluation of conventional and deep learning methods for semi-automated segmentation of pulmonary nodules on CT”. In: *Quant Imaging Med Surg* 11.7 (July 2021), pp. 3286–3305. ISSN: 2223-4292. DOI: 10.21037/qims-20-1356. URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC8250017/ (visited on 03/07/2023).

[201] Ashirbani Saha et al. “Interobserver variability in identification of breast tumors in MRI and its implications for prognostic biomarkers and radiogenomics”. eng. In: *Med Phys* 43.8 (Aug. 2016), p. 4558. ISSN: 2473-4209. DOI: 10.1118/1.4955435.

[202] Wei Deng et al. “Head and Neck Cancer Tumor Segmentation Using Support Vector Machine in Dynamic Contrast-Enhanced MRI”. eng. In: *Contrast Media Mol Imaging* 2017 (2017), p. 8612519. ISSN: 1555-4317. DOI: 10.1155/2017/8612519.

[203] Kuo Men, Jianrong Dai, and Yexiong Li. “Automatic segmentation of the clinical target volume and organs at risk in the planning CT for rectal cancer using deep dilated convolutional neural networks”. eng. In: *Med Phys* 44.12 (Dec. 2017), pp. 6377–6389. ISSN: 2473-4209. DOI: 10.1002/mp.12602.

[204] Jiazhou Wang et al. “Technical Note: A deep learning-based autosegmentation of rectal tumors in MR images”. eng. In: *Med Phys* 45.6 (June 2018), pp. 2560–2564. ISSN: 2473-4209. DOI: 10.1002/mp.12918.

[205] Kuo Men et al. “A study of positioning orientation effect on segmentation accuracy using convolutional neural networks for rectal cancer”. eng. In: *J Appl Clin Med Phys* 20.1 (Jan. 2019), pp. 110–117. ISSN: 1526-9914. DOI: 10.1002/acm2.12494.

[206] Andrew J. Wong et al. “Radiomics in head and neck cancer: from exploration to application”. en. In: *Translational Cancer Research* 5.4 (Aug. 2016). Publisher: AME Publishing Company. ISSN: 2219-6803, 2218-676X. DOI: 10.21037/tcr.2016.07.18. URL: https://tcr.amegroups.com/article/view/8805 (visited on 01/05/2023).

[207] Khadija Sheikh et al. “Predicting acute radiation induced xerostomia in head and neck Cancer using MR and CT Radiomics of parotid and submandibular glands”. en. In: *Radiat Oncol* 14.1 (July 2019), p. 131. ISSN: 1748-717X. DOI: 10.1186/s13014-019-1339-4. URL: https://doi.org/10.1186/s13014-019-1339-4 (visited on 03/08/2023).

[208] Thomas Berger et al. “Predicting radiotherapy-induced xerostomia in head and neck cancer patients using day-to-day kinetics of radiomics features”. eng. In: *Phys Imaging Radiat Oncol* 24 (Oct. 2022), pp. 95–101. ISSN: 2405-6316. DOI: 10.1016/j.phro.2022.10.004.

[209] Thomas Berger et al. “Sub-regional analysis of the parotid glands: model development for predicting late xerostomia with radiomics features in head and neck cancer patients”. eng. In: *Acta Oncol* (Feb. 2023), pp. 1–8. ISSN: 1651-226X. DOI: 10.1080/0284186X.2023.2179895.

[210] Jungsu S. Oh et al. “Intratumor Textural Heterogeneity on Pretreatment (18)F-FDG PET Images Predicts Response and Survival After Chemoradiotherapy for Hypopharyngeal Cancer”. eng. In: *Ann Surg Oncol* 22.8 (Aug. 2015), pp. 2746–2754. ISSN: 1534-4681. DOI: 10.1245/s10434-014-4284-3.

[211] Nai-Ming Cheng et al. “Textural Features of Pretreatment 18F-FDG PET/CT Images: Prognostic Significance in Patients with Advanced T-Stage Oropharyngeal Squamous Cell Carcinoma”. en. In: *Journal of Nuclear Medicine* 54.10 (Oct. 2013). Publisher: Society of Nuclear Medicine Section: Clinical Investigations, pp. 1703–1709. ISSN: 0161-5505, 2159-662X. DOI: 10.2967/jnumed.112.119289. URL: https://jnm.snmjournals.org/content/54/10/1703 (visited on 01/03/2023).

[212] Nai-Ming Cheng et al. “Zone-size nonuniformity of 18F-FDG PET regional textural features predicts survival in patients with oropharyngeal cancer”. eng. In: *Eur J Nucl Med Mol Imaging* 42.3 (Mar. 2015), pp. 419–428. ISSN: 1619-7089. DOI: 10.1007/s00259-014-2933-1.

[213] Martin Vallières et al. “Radiomics strategies for risk assessment of tumour failure in head-and-neck cancer”. eng. In: *Sci Rep* 7.1 (Aug. 2017), p. 10117. ISSN: 2045-2322. DOI: 10.1038/s41598-017-10371-5.

[214] En-Hong Zhuo et al. “Radiomics on multi-modalities MR sequences can subtype patients with non-metastatic nasopharyngeal carcinoma (NPC) into distinct survival subgroups”. en. In: *Eur Radiol* 29.10 (Oct. 2019), pp. 5590–5599. ISSN: 1432-1084. DOI: 10.1007/s00330-019-06075-1. URL: https://doi.org/10.1007/s00330-019-06075-1 (visited on 03/07/2023).

[215] Sebastian Starke et al. “Longitudinal and Multimodal Radiomics Models for Head and Neck Cancer Outcome Prediction”. en. In: *Cancers* 15.3 (Jan. 2023). Number: 3 Publisher: Multidisciplinary Digital Publishing Institute, p. 673. ISSN: 2072-6694. DOI: 10.3390/cancers15030673. URL: https://www.mdpi.com/2072-6694/15/3/673 (visited on 03/07/2023).

[216] François Lucia et al. “External validation of a combined PET and MRI radiomics model for prediction of recurrence in cervical cancer patients treated with chemoradiotherapy”. eng. In: *Eur J Nucl Med Mol Imaging* 46.4 (Apr. 2019), pp. 864–877. ISSN: 1619-7089. DOI: 10.1007/s00259-018-4231-9.

[217] Arnaud Hocquelet et al. “Pre-treatment magnetic resonance-based texture features as potential imaging biomarkers for predicting event free survival in anal cancer treated by chemoradiotherapy”. en. In: *Eur Radiol* 28.7 (July 2018), pp. 2801–2811. ISSN: 1432-1084. DOI: 10.1007/s00330-017-5284-z. URL: https://doi.org/10.1007/s00330-017-5284-z (visited on 04/22/2022).

[218] Kasia Owczarczyk et al. “MRI heterogeneity analysis for prediction of recurrence and disease free survival in anal cancer”. English. In: *Radiotherapy and Oncology* 134 (May 2019). Publisher: Elsevier, pp. 119–126. ISSN: 0167-8140, 1879-0887. DOI: 10.1016/j.radonc.2019.01.022. URL: https://www.thegreenjournal.com/article/S0167-8140(19)30027-1/fulltext#%20 (visited on 04/22/2022).

[219] Nicolas Giraud et al. “MRI-Based Radiomics Input for Prediction of 2-Year Disease Recurrence in Anal Squamous Cell Carcinoma”. In: *Cancers (Basel)* 13.2 (Jan. 2021), p. 193. ISSN: 2072-6694. DOI: 10.3390/cancers13020193. URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7827348/ (visited on 04/22/2022).

[220] Mario Tortora et al. “Radiomics Applications in Head and Neck Tumor Imaging: A Narrative Review”. en. In: *Cancers* 15.4 (Jan. 2023). Number: 4 Publisher: Multidisciplinary Digital Publishing Institute, p. 1174. ISSN: 2072-6694. DOI: 10.3390/cancers15041174. URL: https://www.mdpi.com/2072-6694/15/4/1174 (visited on 03/07/2023).

[221] Joonsang Lee et al. “Radiomics feature robustness as measured using an MRI phantom”. eng. In: *Sci Rep* 11.1 (Feb. 2021), p. 3973. ISSN: 2045-2322. DOI: 10.1038/s41598-021-83593-3.

[222] Jing Yuan et al. “Quantitative assessment of acquisition imaging parameters on MRI radiomics features: a prospective anthropomorphic phantom study using a 3D-T2W-TSE sequence for MR-guided-radiotherapy”. In: *Quant Imaging Med Surg* 11.5 (May 2021), pp. 1870–1887. ISSN: 2223-4292. DOI: 10.21037/qims-20-865. URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC8047358/ (visited on 01/05/2023).

[223] Linda Bianchini et al. “A multicenter study on radiomic features from T2 -weighted images of a customized MR pelvic phantom setting the basis for robust radiomic models in clinics”. eng. In: *Magn Reson Med* 85.3 (Mar. 2021), pp. 1713–1726. ISSN: 1522-2594. DOI: 10.1002/mrm.28521.

[224] Paulina E. Galavis et al. “Variability of textural features in FDG PET images due to different acquisition modes and reconstruction parameters”. eng. In: *Acta Oncol* 49.7 (Oct. 2010), pp. 1012–1016. ISSN: 1651-226X. DOI: 10.3109/0284186X.2010.498437.

[225] Jianhua Yan et al. “Impact of Image Reconstruction Settings on Texture Features in 18F-FDG PET”. eng. In: *J Nucl Med* 56.11 (Nov. 2015), pp. 1667–1673. ISSN: 1535-5667. DOI: 10.2967/jnumed.115.156927.

[226] Isaac Shiri et al. “The impact of image reconstruction settings on 18F-FDG PET radiomic features: multi-scanner phantom and patient studies”. eng. In: *Eur Radiol* 27.11 (Nov. 2017), pp. 4498–4509. ISSN: 1432-1084. DOI: 10.1007/s00330-017-4859-z.

[227] Spyridon Bakas et al. *Identifying the Best Machine Learning Algorithms for Brain Tumor Segmentation, Progression Assessment, and Overall Survival Prediction in the BRATS Challenge*. arXiv:1811.02629 [cs, stat]. Apr. 2019. DOI: 10.48550/arXiv.1811.02629. URL: http://arxiv.org/abs/1811.02629 (visited on 03/08/2023).

[228] Vincent Andrearczyk et al. *Overview of the HECKTOR Challenge at MICCAI 2021: Automatic Head and Neck Tumor Segmentation and Outcome Prediction in PET/CT Images*. arXiv:2201.04138 [cs, eess]. Feb. 2022. DOI: 10.48550/arXiv.2201.04138. URL: http://arxiv.org/abs/2201.04138 (visited on 03/08/2023).

[229] Bjoern H. Menze et al. “The Multimodal Brain Tumor Image Segmentation Benchmark (BRATS)”. In: *IEEE Transactions on Medical Imaging* 34.10 (Oct. 2015). Conference Name: IEEE Transactions on Medical Imaging, pp. 1993–2024. ISSN: 1558-254X. DOI: 10.1109/TMI.2014.2377694.

[230] Thibault Escobar et al. “An original voxel-wise supervised analysis of tumors with multimodal radiomics to highlight predictive biological patterns”. en. In: *Journal of Nuclear Medicine* 62.supplement 1 (May 2021). Publisher: Society of Nuclear Medicine Section: Poster - PhysicianPharm, pp. 1404–1404. ISSN: 0161-5505, 2159-662X. URL: https://jnm.snmjournals.org/content/62/supplement_1/1404 (visited on 04/25/2023).

[231] Thibault Escobar et al. “Voxel-wise supervised analysis of tumors with multimodal engineered features to highlight interpretable biological patterns”. en. In: *Medical Physics* 49.6 (2022). _eprint: https://aapm.onlinelibrary.wiley.com/doi/pdf/10.1002/mp.15603, pp. 3816–3829. ISSN: 2473-4209. DOI: 10.1002/mp.15603. URL: https://onlinelibrary.wiley.com/doi/abs/10.1002/mp.15603 (visited on 04/25/2023).

Chapter 4

A common [18F]-FDG PET radiomic signature to predict survival in patients with HPV-induced cancers

Abstract

Locally advanced cervical cancer (LACC), anal and oropharyngeal squamous cell carcinoma (ASCC and OPSCC) are mostly caused by oncogenic human papillomaviruses (HPV). In this paper, we developed machine learning (ML) models based on clinical, biological and radiomic features extracted from pre-treatment fluorine-18-fluorodeoxyglucose positron emission tomography ([18F]-FDG PET) images to predict survival of patients with HPV induced cancers. For this purpose, cohorts from five institutions were used: two cohorts of patients treated for LACC including 104 patients from Gustave Roussy Campus Cancer (Center 1) and 90 patients from Leeds Teaching Hospitals NHS Trust (Center 2), two datasets of patients treated for ASCC composed of 66 patients from Institut du Cancer de Montpellier (Center 3) and 67 patients from Oslo University Hospital (Center 4) and one dataset of OPSCC patients from University Hospital of Zurich (Center 5). Radiomic features were extracted from baseline [18F]-FDG PET images. The ComBat technique was applied to mitigate intra-scanner variability. A modified consensus nested cross-validation for feature selection and hyperparameter tuning was applied on four machine-learning (ML) models to predict progression free survival (PFS) and overall survival (OS) using harmonized imaging features or/and clinical and biological variables as inputs. Each model was trained and optimized on Center 1 & Center 3 cohorts, and tested on Center 2, Center 4 and Center 5 cohorts. The radiomic-based CoxNet model achieved C-index values of 0.75, 0.78 for PFS and 0.76, 0.74, 0.75 for OS on the test sets. Radiomic feature-based models had superior performance compared to the bioclinical ones and combining radiomic and bioclinical variables did not improve the performances. Metabolic tumor volume (MTV) based models obtained lower C-index values for a majority of the tested configurations, but the same or better performance for the identification of early relapses (td-AUC). The results demonstrate the possibility of identifying common PET-based image signatures for predicting the response of patients with induced HPV pathology, validated on multi-center multiconstructor data.

Keywords— Humanopapillomavirus, LACC, ASCC, OPSCC, [18F]-FDG PET, Radiomics, ComBat

4.1 Introduction

Oncogenic human papillomaviruses (HPV), one of the most sexually transmitted infections, are established to be the primary causes of locally advanced cervical cancer (LACC) and locally advanced anal squamous cell carcinomas (ASCC) [1], [2], accounting for up to 99% of cervical cancers[3] and being associated with 88-100% of anal cancers (AC)[4]. A clear pathogenic role for this virus in head and neck squamous cell carcinoma (HNSCC) is also emerging, as it plays a part, especially for oropharyngeal squamous cell carcinoma (OPSCC)[5]. Over the past two decades, several studies have shown a substantial increase in the number of HPV-positive OPSCC, currently accounting for over 60% [5, 6, 7, 8]. Based on epidemiological and biological data, nearly all squamous intraepithelial lesions of the cervix, anus and oropharynx are caused by alpha HPV genotype, mostly HPV 16 and HPV 18 [9]. Despite differences in epidemiology, management and treatment, these types of cancer share many similarities both cytologically and histologically [10]. Patients with LACC are treated with external beam radiotherapy (EBRT) with concomitant chemotherapy followed by brachytherapy (BT) whereas anal cancer treatment consists of chemoradiotherapy (CRT) [11, 12, 13, 14] combined with a radical abdominoperineal resection (APR) for non-responding cases [12]. OPSCC treatment is dependent on tumor stage and location. In early stage (I–II), either conservative surgery or radiotherapy is preferred. Combined concomitant chemoradiation is the standard treatment in non-resectable patients [15]. Sturdza et al. reported 3-year local control (LC), pelvic control (PC), and overall survival (OS), of 91%, 87%, and 74%, respectively, for LACC [16]. Das et al. showed that ASCC 3-year rates of locoregional control (LRC), distant control (DC), and OS were 81%, 88%, and 84%, respectively [17]. More than 75% of recurrences occur within 2-3 years in LACC [18] and ASCC [19]. Taylor et al. found that the rates of three-year OS were 94% in HPV-positive OPSCC [8], with approximately two-thirds of events occurring within the first 6 months of follow-up [20].

Technological improvement in diagnostic imaging has allowed better characterization of cancerous tissues, which has led to better cancer management. Imaging such as Positron Emission Tomography (PET) with 18F-fluorodeoxyglucose ([18F]-FDG PET) is particularly useful to distinguish pelvic lymph nodes and in initial staging of patients with LACC and ASCC [21, 22, 23]. Besides, it plays a crucial role in HNSCC as it is utilized before treatment in refinement of neck staging and evaluation of distant metastases or second primaries[24, 25]. In addition, features extracted from PET images have already shown promising prognostic results. Multiple classical semi-quantitative pre-treatment PET features such as Standardized Uptake Value (SUV), Metabolic Tumor Volume (MTV), and Total Lesion Glycolysis (TLG), have shown prognostic value in squamous cell cancers [26, 27, 28]. Extracting more complex quantitative features from diagnostic images, known as radiomics, characterizing not only the tumor's intensities and volume but also its shape and heterogeneity, has the potential to get deeper tumor characterization. It allows for stratification of patients into distinct phenotypic subgroups with the added advantage of being non-invasive[29]. Radiomic features have also been successfully assessed for recurrence and survival prediction not only in LACC but also in ASCC and OPSCC[30, 31, 32, 33, 34, 35, 36]. Reuzé et al. [37] developed PET-based four-feature radiomic signature that could predict local recurrence in LACC. Brown et al. [38] showed that radiomic features extracted from [18F]-FDG PET/CT could improve prediction of progression-free survival (PFS) in ASCC compared to conventional clinical variables. Bogowicz et al.[39] showed that PET radiomics are potential prognostic biomarkers in HNSCC for local tumor control. Moreover, several studies have shown that combining radiomic and clinicopathological features might improve predictions [27, 38, 40, 41, 42, 43]. In the era of personalized medicine, the early identification of patients most likely to experience relapse may aid better treatment selection to improve outcomes. Therefore, the purpose

of this study is to develop a multi-parametric signature including clinical data, baseline biological data, and imaging features extracted from pre-treatment [18F]-FDG PET exams to predict PFS and OS of HPV-induced cancers. To the best of our knowledge, this work is the first multicentric study to evaluate the generalizability of a single signature to multiple HPV-induced cancers.

4.2 Materials and methods

4.2.1 Patients

In this retrospective study, cohorts from five institutions were gathered. Patients treated for LACC were enrolled: 118 patients from Gustave Roussy Cancer Campus in France (Center 1) and 93 patients from Leeds Teaching Hospitals NHS Trust in the United Kingdom (Center 2). Furthermore, two datasets of patients treated for ASCC were also collected: 88 patients from Institut du Cancer de Montpellier in France (Center 3) and 93 patients from Oslo University Hospital in Norway (Center 4). Additionally, a cohort from University Hospital of Zurich in Switzerland (Center 5) with 45 patients treated for OPSCC was retrieved.

For each patient, specific inclusion criteria were defined in our study: i) availability of baseline PET scan ii) histologically-confirmed LACC, ASCC, or OPSCC, iii) HPV p16 positive (or p18 positive subtypes for ASCC) iv) no cervical conization performed before baseline PET acquisition for LACC patients vi) minimum follow-up period of 36 months after treatment for patients without recurrence. Patients who did not experience any event with follow-up < 36 months were also excluded: 14 patients from Center 1, 3 patients from Center 2 and 2 patients from Center 3. At the end, 104, 90, 66, 67 and 45 patients from Center 1, Center 2, Center 3, Center 4 and Center 5 respectively, were considered in this study.

This study was performed in accordance with the General Data Protection Regulation (GDPR) and approved by the local ethical committees for Centers 1 and 3; all alive patients at the time of the study were informed of the use of their data. In Center 2, this retrospective study was considered with no impact on patients: the review board considered it as a service evaluation and issued a waiver for informed consent/formal ethics approval. For Center 4, the study was approved by the Regional Ethical Committee and all patients had given written informed consent. For center 5, data analysis was approved by the Swissethics and was carried out in accordance with Swissethics guidelines and regulations.

LACC patients from Center 1 and Center 2 were treated with chemoradiotherapy (weekly intravenous cisplatin 40 mg/m^2, 5–6 cycles, 1 day per cycle, plus 45–50 Gy external-beam radiotherapy delivered in 1.8–2 Gy fractions) +/- simultaneously integrated boost to metastatic lymph nodes, followed by brachytherapy (BT) based on recommendations published by the Groupe Européen de Curiethérapie - European Society for Radiation Oncology (GEC-ESTRO) [44]. For 30 patients from Center 2 with issues preventing intra-cavity treatment or with poor initial response, an external beam boost of 12-18 Gy was given instead of BT. LACC patients' characteristics are displayed in Table 1.

While there is no definitive consensus on treatment, patients with ASCC are usually treated by CRT with at least 45 Gy to the clinical target volume (CTV) and concomitant fluoroucil(5-FU)/mitomycin(MMC) during the first and the fifth week of RT, followed by a boost limited to the macroscopic tumor given by either EBRT or brachytherapy. Guidelines on management of ASCC can be found in [14]. Patients from Center 3 were treated with 45(n=85) or 50 (n=2) Gy to the CTV and 55.8-67 Gy (n=18) or 61-67 Gy (n=67) or without additional boost (n=2) to the GTV and lymph nodes if present. In Center 3, 5-FU and MMC were administered

	Center 1	Center 2
Variables	mean[min-max]	mean[min-max]
Age	48.0 [27.0-82.0]	48.0 [25.0-75.0]
Hemoglobin (g/dL)	12.0 [6.5-15.3]	12.8 [7.3-15.7]
Platelets (G/L)	302.0 [89-662]	338.4 [118-699]
Leukocytes (G/L)	8.2 [1.8-18.0]	8.8 [4.7-20.2]
Neutrophils (G/L)	5.8 [1.2-14.0]	5.9 [2.0-15.3]
Lymphocytes (G/L)	1.7 [0.2-4.1]	1.9 [0.8-4.1]
Monocytes (G/L)	0.6 [0.1-1.5]	0.5 [0.2-1.1]
NLR	4.2 [0.9-24.0]	3.4 [0.2-10.2]
	Value(%)	Value(%)
HPV Status		
Positive	104 (100%)	90 (100%)
Stage		
I	33 (31.7%)	9 (10%)
II	59 (56.7%)	63 (70.0%)
III	8 (7.7%)	8 (8.9%)
IV	4 (3.9%)	10 (11.1%)
Histology		
SCC	87 (83.7%)	71 (78.8%)
Adenocarcinoma	17 (16.3%)	19 (21.2%)
PET device		
Siemens Biograph	67 (64.4%)	-
GE Discovery 690	37 (35.6%)	-
Philips Gemini TF TOF 64	-	90 (100%)

Table 4.1: LACC patients' characteristics.

NLR: neutrophil-to-lymphocytes ratio. HPV: human papillomavirus. SCC: squamous cell carcinoma. *According to the cancer stage classification by the TNM 8th edition

intravenously during the first and the fifth week of RT at a dose of 800 mg/m^2 daily during 4 days and 80 mg/m^2 daily during 2 days, respectively. Delivered dose during RT in Center 4 was similar to that in Center 3 with 88 patients treated with 46 Gy sequential radiotherapy to the CTV and 54 or 58 Gy to the tumor and pathologic lymph nodes if present. Five patients were treated with simultaneous integrated boost giving 48.6 Gy to the CTV and 57.5 Gy to the GTV. Most patients received concomitant chemotherapy with MMC 10 mg/m^2 Day 1 (and Day 29 if T3-4 or N+) and 5-FU 1000 mg/m^2 days 1–4 (and days 29–32 if T3-4 or N+). More details are given in [19, 27]. Clinical characteristics of ASCC patients are shown in Table 2.

The HPV status of 35 patients in center 3 was missing. These patients were not excluded since 88-100% of ASCCs are HPV-related[4]. The results of the survival analysis excluding these patients are presented in the discussion.

OPSCC from Center 5 were all treated with RT and concomitant chemotherapy, receiving a total dose of 66-72 Gy to the planning target volume (PTV) and a minimum dose of 55 Gy to the elective lymph nodes delivered in 32–35 fractions of 2-2.2 Gy, following guidelines from Gregoire et al[45]. Concomitant chemotherapy for 25 of the patients consisted of cycles of cisplatin and cetuximab. Twelve patients received cisplatin only, and 8 patients received cetuximab only. Clinical characteristics of OPSCC patients are shown in Table 3. Blood count information was not retrieved for OPSCC patients.

4.2.2 Model inputs

Age, lymph node status, cancer stage and blood count, i.e neutrophils, leukocytes, monocytes, lymphocytes, platelets, hemoglobin, neutrophil-to-lymphocyte ratio (NLR), before treatment were available and used as bio-clinical variables. Cancer stage was assessed using the concordance between TNM classification of malignant tumours[46] and FIGO stage for LACC[47].

Images were acquired on different scanners. Patients from Center 1 were scanned either on a Siemens Biograph mCT (Siemens AG, Erlangen, Germany) (n=67) or on a GE Discovery 690 (GE Healthcare, Waukesha, WI) (n=37). In Center 2, all images (n=90) were scanned with a Philips Gemini (Philips Healthcare, Best, The Netherlands). GE Discovery 690 and Siemens Biograph mCT were utilized in Center 3 (n=66) and Center 4 (n=67) respectively. In addition, two devices, GE Discovery STE (n=30) and GE Discovery RX (n=15) were used to scan patients in Center 5. The difference in acquisition and reconstruction parameters did not allow us to consider each machine as a separate "batch". Thus, three scanner image series were considered for LACC, two for ASCC, and two for OPSCC.

Image intensities were converted from activity concentration to parametric SUV units. A threshold-based segmentation was applied, considering the threshold as 40% of the maximum radiopharmaceutical uptake in the lesion (SUVmax), to delineate the volume of interest (VOI). This delineation was performed by a certified nuclear medicine physician (R.D.S.with 5 years of expertise), and an experienced Clinical Scientist (G.M. with 19 years of expertise) under the supervision of a dual certified Radiologist/Nuclear Medicine Physician (AS, >15 yrs experience). They both used using the LifeX software, V6.30, and V4, respectively.

To ensure comparability and rotational invariance of extracted radiomic features, spatial resampling was applied, setting voxel size on all images to 3x3x3 mm^3 using b-spline interpolation implemented in the Advanced Normalization Tools (ANTs)[48]. A fixed bin width of 0.3 SUV was selected for gray-level discretization [49]. With PyRadiomics v3.0.1, an open-source python package [50], radiomic features (n=107) were extracted from baseline [18F]-FDG PET images: first-order (n=18), gray-level co-occurrence matrix (GLCM)(n=24), gray-level

	Center 3	Center 4
Variables	mean[min-max]	mean[min-max]
Age	63 [41-92]	62 [40-90]
Hemoglobin (g/dL)	12.7 [8.6-15.6]	13.5 [8.6-16.9]
Platelets (G/L)	282.3 [60.0-628.0]	289.8 [83.0-849]
Leukocytes (G/L)	7.6 [2.3-17.4]	8.0 [3.3-16.9]
Neutrophils (G/L)	5.1 [1.3-13.4]	5.3 [1.8-13.7]
Lymphocytes (G/L)	1.7 [0-3.8]	2.1 [0.3-4.1]
Monocytes (G/L)	0.6 [0.1-1.5]	0.6 [0.1-1.4]
NLR	3.7 [0.6-17.3]	3.3 [0.7-27.3]
	Value(%)	Value(%)
HPV Status		
Positive	31 (47.0%)	67 (100%)
Missing	35 (53.0%)	-
TNM*		
T0	-	-
T1	10 (15.1%)	1 (1.5%)
T2	24 (36.4%)	39 (58.2%)
T3	19 (28.8%)	11 (16.4%)
T4	13 (19.7%)	16 (22.4%)
N0	26 (39.4%)	36 (53.7%)
N1	11 (16.7%)	7 (10.4%)
N2	16 (24.2%)	13 (19.4%)
N3	13 (19.%)	11 (16.4%)
M0	44 (66.7%)	66 (98.5)
M1	22 (33.3%)	1 (1.5%)
Stage		
I	8 (12.1%)	1 (1.5%)
II	17 (25.8%)	32 (47.8%)
III	19 (28.8%)	33 (49.2%)
IV	22 (33.3%)	1 (1.5%)
Histology		
SCC	66 (100%)	67 (100%)
PET device		
GE Discovery 690	66 (100%)	-
Siemens Biograph mCT	-	67 (100%)

Table 4.2: ASCC patients' characteristics.

NLR: neutrophil-to-lymphocyte ratio. HPV: human papillomavirus. SCC: squamous cells carcinoma. *According to the cancer stage classification by the TNM 8th edition

Variables	Center 5 mean[min-max]
Age	59 [43-75]
	Value (%)
HPV Status	
Positive	45 (100%)
TNM*	
T0	-
T1	3 (6.7%)
T2	14 (31.1%)
T3	14 (31.1%)
T4	14 (31.1%)
N0	5 (11.1%)
N1	1 (2.2%)
N2	37 (82.2%)
N3	2 (4.4%)
M0	40 (88.9%)
M1	5 (11.1%)
Stage	
I	-
II	24 (53.3%)
III	16 (35.5%)
IV	5 (11.1%)
Histology	
SCC	45 (100%)
PET device	
GE Discovery STE	30 (67%)
GE Discovery RX	15 (33%)

Table 4.3: OPSCC patients' characteristics.
HPV: human papillomavirus. SCC: squamous cells carcinoma. *According to the cancer stage classification by the TNM 8th edition

dependence matrix (GLDM)(n=14), gray-level run length matrix (GLRLM)(n=16), gray-level size zone matrix (GLSZM)(n=16), neighboring gray-tone difference matrix (NGTDM)(n=5), and shape (n=14) features.

To tackle differences in scanner acquisition and image reconstruction, the ComBat method was applied. Originally developed for combating the batch effect in microarray genomic analysis [51], it was later demonstrated that the ComBat harmonization technique successfully removes inter-site variability while preserving inter-patient biological variability on MR images[52, 53]. When applied to radiomics, the raw value of each feature is considered the sum of a mean value, an additive scanner effect, and a multiplicative scanner effect [49]. The purpose of this technique is to eliminate the differences between batches, by a location (mean) and scale (variance) (L/S) transform. Radiomic features values in each batch are adjusted to have the same mean and/or variance. To improve fit across batches, ComBat can be run with an empirical Bayes (EB) procedure that pools the values of radiomic features from each batch to "reduce" the error on the batch effect parameter [51]. ComBat applied with EB estimates increases robustness especially if the number of samples in each batch is small. In this study, ComBat with EB estimates and non-parametric prior method was applied. In addition to eliminating the so-called scanner effect, ComBat can also take into account covariates that should not be changed when removing the scanner effect. In this study, two covariates were considered. A covariate indicating the tumor location, whether LACC, ASCC, or OPSCC, was set since it was seen that the intensities distribution was statistically different among batches depending on tumor location. The other covariate was cancer stage, as it has been stressed out by Orlhac et al. [54] that feature value distributions might be different in patients with different tumor stages. Center 3 GE Discovery images were selected as the reference cohort.

The workflow used in this paper is described in Figure4.1.

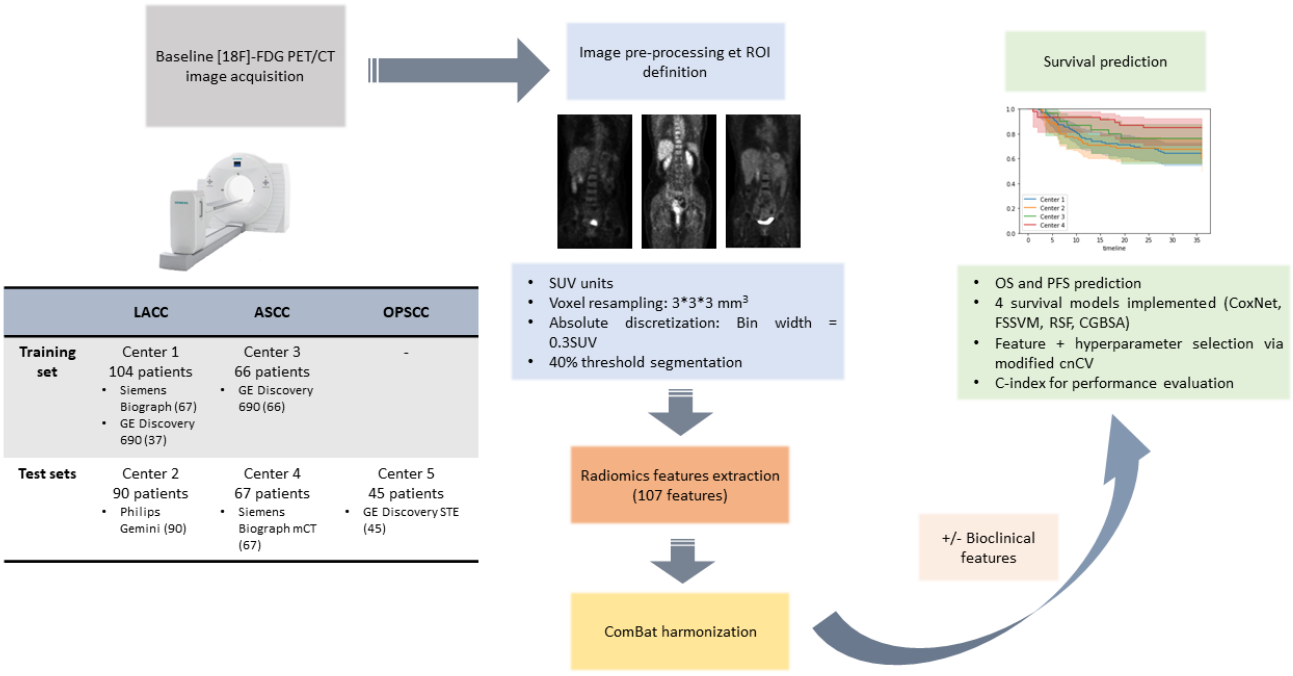


Figure 4.1: Overview of the workflow, including image acquisition, data splitting, data pre-processing, extraction of radiomic features and final survival analyses

4.2.3 Survival prediction

Four machine-learning (ML) models for survival analysis were built to predict OS and PFS based on clinical and biological variables only, radiomic features only and a combination of clinical variables and radiomics: Cox's proportional hazard's model with elastic net penalty (CoxNet), Fast Survival Support Vector Machine (FSSVM) and two ensemble methods: Random Survival Forest (RSF) and Component-wise Gradient Boosting Survival Analysis (CGBSA). The different models were implemented with scikit-survival [55], a Python module for survival analysis built on top of scikit-learn library. Ensemble method is a machine learning technique built on several learners in order to improve generalizability and robustness over a single estimator [56, 57]. RSF and CGBSA are built on an ensemble of tree-based learners, considering Gradient-boosted Cox proportional hazard loss with regression trees and Gradient boosting with component-wise least squares as base learner, respectively. Patients from Center 1 and Center 3 were used as training sets. The trained models were applied on independent test sets,i.e., Center 2, Center 4, and Center 5 to test models generalizability. Since the information on PFS in OPSCC patients from Center 5 was not available, we evaluated the trained models for PFS on patients from centers 2 and 4 only. The performance of the trained models was evaluated using the concordance-index (C-index), a metric assessing the model's ability to correctly provide a reliable ranking of the survival times based on the individual risk scores.

Prediction with bioclinical variables

The first model aimed to assess the prognostic value of clinical and biological variables for survival and thus build a baseline model. Three clinical variables (age, lymph node status, cancer stage), and seven biological variables, i.e blood count (neutrophils, leucocytes, monocytes, lymphocytes, platelets, hemoglobin and NLR) were available. Performance of clinical features and blood count features were evaluated independently before the combined model. It has been shown in several studies that MTV, and TLG are prognostic factors for survival in SCC [27, 58]. A grid search with 5-fold cross-validation (CV) was used to train the models and tune hyper-parameters.

Prediction with radiomic features

Regarding the radiomic features-based model, a modified consensus nested cross-validation (cnCV)[59] was implemented for feature selection and hyper-parameter tuning (Figure 2). Although MTV and TLG are commonly used in clinical routine, they were considered as radiomic features in this study, being directly extracted from the images.

To this end, a grid search strategy was implemented. The parameter space used for tuning the hyper-parameters is presented in the supplementary materials (Table S1). The C-index was used as performance metric. In this study, instead of choosing the best outer model like in a classical nCV, all models yielding a C-index $\geq$ 0.7 on the outer folds were retained. The modified cnCV was repeated ten times to increase robustness of the selected features and hyper-parameters. The most common features (appearing at least in 60 models) and the most selected hyper-parameters within these models were then selected to train the final model on the full training cohort (Center 1 and Center 3) and tested on the test sets (Center 2, Center 4 and Center 5). Using the predicted risk scores, we plotted the area under the receiver operating characteristic (ROC) curve (AUC) extended to survival by defining sensitivity (true positive rate) and specificity (true negative rate)

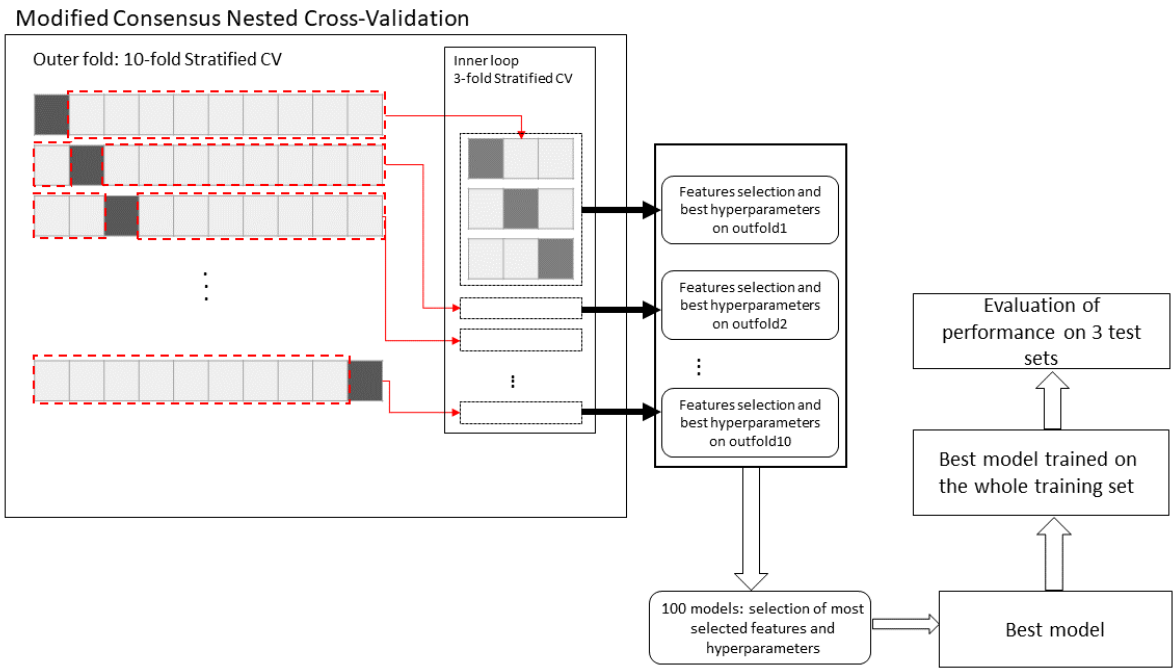


Figure 4.2: Modified consensus Nested Cross-Validation (cnCV) implemented for feature selection and hyperparameter tuning. Features are selected in the inner loops, and robustness is evaluated on the outer loops. The final radiomic features are a consensus on the most selected features.

as time-dependent measures. This was done to evaluate how well estimated risk scores can separate patients experiencing death or progression from healthy patients n months after the end of treatment on Center 2 and 4. As MTV has been already identified as a prognostic feature for survival[58, 60, 61], time-dependent AUC of a CoxNet model based on MTV was also evaluated.

Prediction with clinical variables and radiomic features

For each ML model, a model combining the clinical variables and new radiomic features, again selected by the modified cnCV, was trained and compared with the respective models built on bioclinical variables only or radiomic features only.

4.3 Results

4.3.1 Prediction with bioclinical variables

Baseline models considering age, cancer stage and lymph node status as inputs, were found to give the best performance on OS and PFS (Table 3) compared to blood count-based models or a combination of both. Blood counts alone were found to be weak prognostic features (C-index$\leq$0.55). The C-index did not improve when these variables were introduced into the OS and PFS prognosis models for all ML models developed (results not shown). The baseline models gave C-index values between 0.60 and 0.67 on Centers 2 and 4. On Center 5, only CoxNet model achieved a C-index of 0.60, when other models showed poor prognostic values (C-index$<$0.50).

Table 4.4: Performance of baseline models based on clinical variables alone, including age, lymph node status and cancer stage.

CoxNet	PFS	OS
Train	0.68 ± 0.01	0.66 ± 0.02
Validation	0.66 ± 0.09	0.62 ± 0.09
Test (Center 2: LACC)	0.64	0.65
Test (Center 4: ASCC)	0.66	0.67
Test (Center 5: OPSCC)	-	0.60
FSSVM	**PFS**	**OS**
Train	0.71 ± 0.06	0.70 ± 0.07
Validation	0.66 ± 0.08	0.63 ± 0.08
Test (Center 2: LACC)	0.62	0.64
Test (Center 4: ASCC)	0.64	0.66
Test (Center 5: OPSCC)	-	0.46
RSF	**PFS**	**OS**
Train	0.75 ± 0.07	0.74 ± 0.08
Validation	0.65 ± 0.07	0.63 ± 0.09
Test (Center 2: LACC)	0.61	0.60
Test (Center 4: ASCC)	0.63	0.64
Test (Center 5: OPSCC)	-	0.50
CGBSA	**PFS**	**OS**
Train	0.72 ± 0.06	0.71 ± 0.08
Validation	0.66 ± 0.08	0.63 ± 0.08
Test (Center 2: LACC)	0.61	0.61
Test (Center 4: ASCC)	0.62	0.63
Test (Center 5: OPSCC)	-	0.48

LACC: locally advanced cervical cancer. ASCC: anal squamous cells carcinoma. OPSCC: oropharyngeal squamous cell carcinoma. Coxnet: Cox's proportional hazard's model with elastic net penalty, FSSVM: Fast Survival Support Vector Machine, RSF: Random Survival Forest, CGBSA: Component-wise Gradient Boosting Survival Analysis.

4.3.2 Prediction with radiomic features

C-index values obtained for the models incorporating the radiomic features are presented in Table 4. For PFS, CoxNet yielded C-index values of 0.75 and 0.78, CGBSA achieved C-indexes of 0.72 and 0.74, and FSSVM 0.75 and 0.76 on Centers 2 and 4 respectively. Using MTV only, these models achieved C-index of 0.75 on Center 2 and values ranging from 0.65 to 0.70 on Center 4. Likewise, these models had similar C-indexes on OS prediction on center 2 and 4. However, on OPSCC (Center 5), we obtained lower results for OS prediction on all trained models except for CoxNet model with a C-index value equal to 0.75. FSSVM had 0.69, and CGBSA had 0.65 as C-index values. The MTV based model had lower results on Center 5, C-index values were between 0.56 and 0.59. On all the test sets in general, results were consistent on LACC and ASCC patients for both OS and PFS across CoxNet, FSSVM and CGBSA ($0.72 \geq$ C-index ≤ 0.78). RSF was prone to overfitting as it performed remarkably well on the training cohort but decreased in performance on the validation and test sets.

The time-dependent AUC is shown in Figure 3. Regarding PFS prediction for Center 2, the MTV based model and FSSVM achieved time dependent AUC values of 0.78 and 0.77 respectively. CoxNet and CGBSA had similar results with 0.75 and 0.74, respectively. Similarly, all models achieved similar results on Center 4: MTV based model, CoxNet and FSSVM had mean AUC values of 0.75 while CGBSA reached 0.73. On OS prediction, the time dependent AUC values were comparable across all models on Centers 2 and 4. FSSVM model and MTV based model presented mean AUC values of 0.72 and 0.71 on Center 2 respectively, when CoxNet and CGBSA had 0.70 and 0.69 respectively. On Center 4, CGBSA, with the highest score, achieved 0.78 as mean AUC value, CoxNet and MTV based model achieved 0.76 and 0.77 respectively, while FSSVM had an AUC of 0.74. On Center 5, the MTV based model had an AUC of 0.65 whereas CoxNet, FSSVM and CGBSA yielded 0.61, 0.60 and 0.53 respectively.

4.3.3 Prediction with clinical variables and radiomic features

The modified cnCV built on radiomic and bioclinical variables selected radiomic features only as relevant features for OS and PFS prediction. However, to evaluate the added value of bioclinical variables, we added age, stage cancer and lymph node status to the previously selected radiomic features. Combining clinical and radiomic features improved slightly the results on all the outcomes on the training set, but the performances remained comparable to radiomic based models on the test sets.

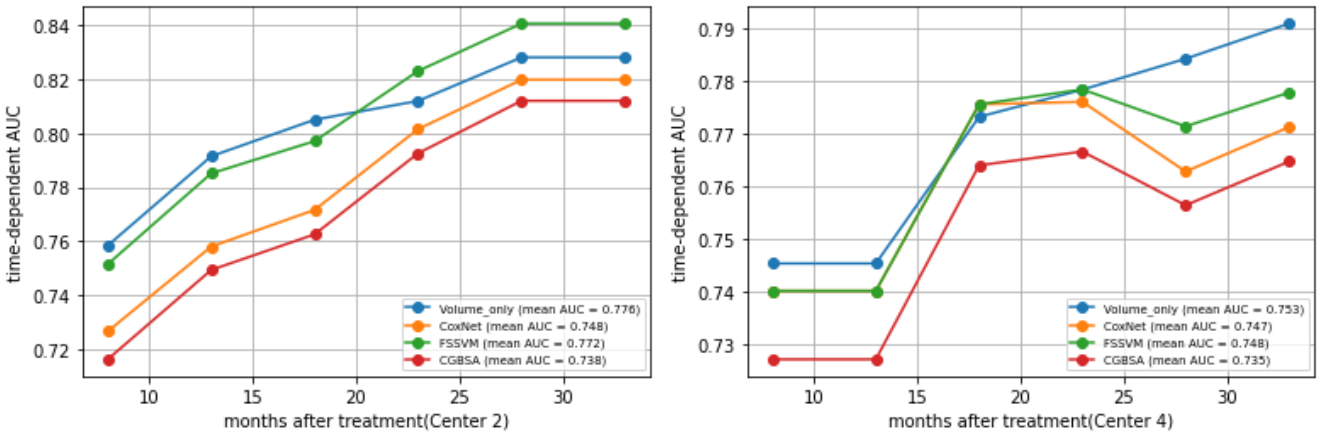


Figure 4.3: Time-dependent AUC performance metrics assessed over 36 months in terms of progression-free survival for Center 2 (left, LACC) and Center 4 (right, ASCC)
LACC: locally advanced cervical cancer, ASCC: anal squamous cells carcinoma, OPSCC: oropharyngeal squamous cell carcinoma, Coxnet: Cox's proportional hazard's model with elastic net penalty, FSSVM: Fast Survival Support Vector Machine, RSF: Random Survival Forest, CGBSA: Component-wise Gradient Boosting Survival Analysis

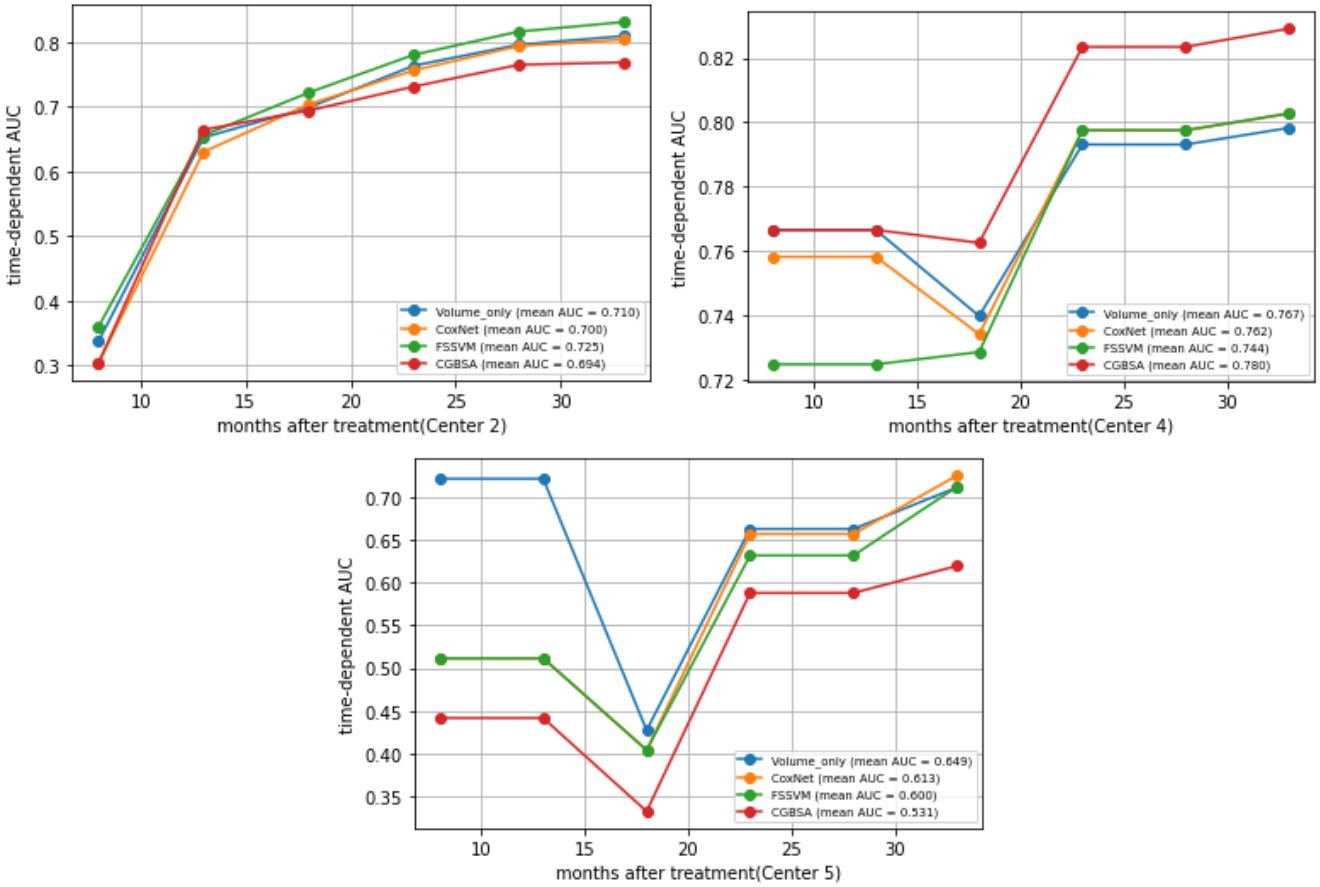


Figure 4.4: Time-dependent AUC performance metrics assessed over 36 months in terms of overall survival for Center 2 (top left, LACC), Center 4 (top right, ASCC), and OPSCC (bottom)
LACC: locally advanced cervical cancer, ASCC: anal squamous cells carcinoma, OPSCC: oropharyngeal squamous cell carcinoma, Coxnet: Cox's proportional hazard's model with elastic net penalty, FSSVM: Fast Survival Support Vector Machine, RSF: Random Survival Forest, CGBSA: Component-wise Gradient Boosting Survival Analysis

Table 4.5: C-index of models based on radiomic features alone and combined features (radiomic features combined to clinical variables).

	Radiomic features		Combined features	
Model	PFS	OS	PFS	OS
CoxNet				
Train	0.74 ± 0.02	0.77 ± 0.01	0.76 ± 0.02	0.76 ± 0.01
Validation	0.73 ± 0.07	0.74 ± 0.08	0.71 ± 0.07	0.70 ± 0.06
Test(Center 2: LACC)	0.75	0.76	0.72	0.75
Test(Center 4: ASCC)	0.78	0.74	0.76	0.75
Test(Center 5: OPSCC)	-	0.75	-	0.67
FSSVM				
Train	0.74 ± 0.01	0.77 ± 0.02	0.76 ± 0.02	0.81 ± 0.01
Validation	0.72 ± 0.06	0.72 ± 0.07	0.72 ± 0.08	0.75 ± 0.08
Test(Center 2: LACC)	0.75	0.75	0.72	0.72
Test(Center 4: ASCC)	0.76	0.78	0.73	0.72
Test(Center 5: OPSCC)	-	0.69	-	0.60
RSF				
Train	0.93 ± 0.01	0.94 ± 0.01	0.90 ± 0.01	0.91 ± 0.01
Validation	0.72 ± 0.06	0.72 ± 0.10	0.72 ± 0.06	0.73 ± 0.05
Test(Center 2: LACC)	0.65	0.69	0.65	0.71
Test(Center 4: ASCC)	0.69	0.70	0.63	0.69
Test(Center 5: OPSCC)	-	0.51	-	0.51
CGBSA				
Train	0.74 ± 0.01	0.76 ± 0.01	0.75 ± 0.02	0.79 ± 0.02
Validation	0.73 ± 0.05	0.74 ± 0.07	0.72 ± 0.07	0.72 ± 0.06
Test(Center 2: LACC)	0.72	0.72	0.69	0.64
Test(Center 4: ASCC)	0.74	0.75	0.69	0.72
Test(Center 5: OPSCC)	-	0.65	-	0.64

LACC: locally advanced cervical cancer, ASCC: anal squamous cells carcinoma, OPSCC: oropharyngeal squamous cell carcinoma, Coxnet: Cox's proportional hazard's model with elastic net penalty, FSSVM: Fast Survival Support Vector Machine, RSF: Random Survival Forest, CGBSA: Component-wise Gradient Boosting Survival Analysis.

4.4 Discussion

In this study, a common [18F]-FDG PET radiomic signature was developed to predict OS and PFS for patients with HPV induced cervical cancer, anal and oropharyngeal squamous carcinomas. For this purpose, five independent cohorts were used: two cohorts of patients with LACC, two cohorts of patients treated for ASCC and a cohort of patients with OPSCC. Four ML models were trained on LACC patients from Center 1 and ASCC patients from Center 3. Center 2 (LACC), Center 4 (ASCC) and Center 5 (OPSCC) cohorts were used as independent test sets to evaluate the generalizability of the trained models. Models trained on the clinical variables including age, lymph nodes status, and cancer stage (baseline model) yielded results between 0.61 and 0.66 for PFS prediction, and between 0.60 and 0.67 for OS prediction in centers 2 and 4, respectively. OS prediction for OPSCC patients in center 5 showed C-index values < 0.5, except for Coxnet model, which achieved a C-index of 0.60. Using radiomics features only, results were consistent on LACC and ASCC cohorts for both OS and PFS with CoxNet, FSSVM and CGBSA achieving C-indexes comprised between 0.72 and 0.78, whereas RSF was prone to overfitting with C-index values≤ 0.70 on all the test sets. However, on OPSCC (Center 5), we obtained lower results for OS prediction on all trained models (C-index$\leq$0.69) except for CoxNet model (C-index=0.75). The radiomic based CoxNet model showed better ability to predict survival on the three independent cohorts with C-index scores between 0.74 and 0.78 regardless of the predicted task. Combining radiomic and bioclinical variables gave better results than the baseline model but remained inferior or equal to models based on radiomics alone.

C-index is a measure of the model's ability to correctly provide a trustworthy ranking of survival times based on individual risk scores. It is the fraction of pairs in a cohort where the observation with the higher survival time has the higher probability of survival predicted by the model. Therefore, C-index should allow for discrimination between early events, associated with higher-risk subjects, and later occurrences. To evaluate the ability of the trained models to classify patients who will experience death or relapse n months after the treatment, we computed the time dependent AUC. The MTV based model was slightly superior in distinguishing patients prone to experience progression in the first months after treatment, but over time all models had similar results. For OS, all models achieved similar performances except on Center 5 where the MTV based model proved to be superior in the first months after treatment and achieved a mean AUC of 0.65. In all cases, the mean AUC remained stable and comparable regardless of the ML model, except for the prediction of OS on Center 5 during the first months after treatment. These results can be explained by the fact that the selected radiomic features for FSSVM and CGBSA were found to be correlated to MTV ($r \geq 0.6$) except for two features of texture, i.e, dependence entropy calculated from GLDM and zone entropy calculated from GLSZM. The predicted risk scores of the FSSVM model showed a linear correlation with MTV (Figure S2). We note, furthermore, that combining second-order features (gldm DependenceEntropy and glszm ZoneEntropy), not correlated to MTV, provided the best survival ranking (CoxNet model). The C-index values of all radiomic features-based ML models were found to be higher than the prediction based on MTV alone for PFS and OS (Table S3). These results highlight the importance of the performance metric used for model construction according to the clinical purpose of the developed tool. To ensure that forcing the signature to be generalizable to 2 types of HPV-induced cancers did not worsen the performance of radiomics and only favored the selection of volume-correlated features, a CoxNet model based on radiomic features was trained on cervical cancer patients only (Center 1) and its performance evaluated on Center 2. The same process was applied to Center 3 and 4. For PFS, the model achieved C-index values of 0.72 and 0.73 on Center 2 and 4 respectively. For OS, the model yielded 0.73 and 0.75 on Center 2

and 4 respectively. All these results are in favor of a common phenotype, highlighted by the generalizability of the results obtained.

Several studies have highlighted the interest of combining radiomic and clinical features for performance improvements[27, 38, 40, 41, 42, 43]. In this project, combining radiomic and clinical features did not improve the C-index values on the test cohorts. In addition, some radiomic features were found to be better predictors of survival than conventional SUV measures, such as maximum SUV, minimum SUV, mean SUV and peak SUV. In line with our findings, other studies have pointed out the superiority of radiomic features over SUV measurements [30, 62, 63]. Since the first publications on radiomics [29, 64, 65], the main limitation remains the non-standardization of imaging protocols, acquisition and reconstruction parameters. Several studies have assessed the impact of these parameters and multiple solutions have been proposed [66, 67, 68, 69]. We applied the ComBat method to mitigate the variability in extracted features and to improve the generalizability of response prediction signatures in a multi-center context. The main limitation of using ComBat is the way batch allocation is determined. As acquisition and reconstruction parameters have an impact on image quality, batch allocation was done in such a way that one batch had to share the same device, the same acquisition parameters, and the same reconstruction algorithm and parameters. The results show that the approach seems to have been correctly applied with generalizable results from one cohort to another.

The amount of variables used as inputs to train ML algorithms plays a key role on its performance. Suboptimal number of variables can lead to over-fitting or under-fitting [70]. Combining a high number of variables can limit the performance by diluting the predictive power of certain features by the noise present in others. In most radiomic studies, the number of patients is lower than the number of features, which generates a high-dimensional data problem, leading to overly optimistic results and preventing the generalizability of the trained models. Therefore, feature selection is an indispensable step to avoid irrelevant features and redundancy. It allows for better comprehension and explainability of the trained models[70]. Several methods, each with its limitations, for feature selection have been assessed but no consensus has been reached. In this study, a modified cnCV was implemented to ensure for quality and robustness of the trained models performance on the independent test sets. Given that, feature selection is data dependent, feature selection was carried out in the inner folds of the modified cnCV, while assessing the stability of model performance in the outer folds[59]. The selected features are presented in Table S2

In this study, the HPV status of 35 patients with ASCC treated in Center 3 was missing. However, these patients were included in the study under the assumption that 88-100% [4]of anal canal cancers are HPV-induced. To ensure that the addition of these patients did not constitute a bias, the performance of the trained models was evaluated by removing these patients. The C-index values decreased slightly but remained above 0.70 for all models (Table S4).

This study had some limitations. First, patients with OPSCC were only 45 and the number of events, here death, occurring in the following 36 months after treatment, was only six. The size of this cohort is clearly a limitation for survival prediction using the C-index metric. Nevertheless, the results for OS prediction are promising and further investigation should be conducted by adding OPSCC patients in the training and/or test cohorts. Second, the tumors on the PET images were segmented by two board-certified nuclear medicine physicians. This was shown to add variability in radiomic features but for the purposes of this study, we considered it negligible. Third, impact of voxel resampling was not assessed. As there is no consensus on the best method for voxel resampling, further analysis should be conducted to ensure that the signature developed is independent of resampling settings. Adding other information besides imaging such as histology, genomics or

dose-based variables could lead to better prediction.

In conclusion, survival prediction models using radiomic features extracted from [18F]-FDG PET images could predict survival outcomes using a common HPV signature in LACC, ASCC and HNSCC. Results have been validated on external independent data sets, but the added value remains to be demonstrated compared to conventional quantities such as MTV extracted from PET in particular for early event detection. Radiomics may therefore pave the way to optimization of HPV induced cancers treatment, identifying patients who might benefit from intensified or deintensified dose delivery.

4.5 Supplementary materials

Table S1: Grid search space used for hyper-parameters tuning of the different ML models

ML models	**Parameters**	
	Name	***Space***
CoxNet	alphas	[1e-04 – 1e+4]
FSSVM	alpha	[2e-4 - 4e+3]
	optimizer	[avltree, rbtree, direct-count]
RSF	n_estimators	[100, 150, 200]
	max_features	[auto, sqrt, log2, None]
	min_samples_split	[4, 6, 10]
CGBSA	loss	[coxph, squared]
	learning_rate	[0.01, 0.1, 0.5, 1]
	n_estimators	[100, 150, 200, 250]

Coxnet: Cox's proportional hazard's model with elastic net penalty
FSSVM: Fast Survival Support Vector Machine
RSF: Random Survival Forest
CGBSA: Component wise Gradient Boosting Survival Analysis

Table S2: Selected features after consensus Nested Cross-Validation

ML models	***PFS***	***OS***
CoxNet		
	gldm DependenceEntropy	gldm DependenceEntropy
	glszm ZoneEntropy	glszm ZoneEntropy
	TLG	TLG
	shape MinorAxisLength	
FSSVM		
	shape MinorAxisLength	shape MinorAxisLength
	shape SurfaceArea	shape SurfaceArea
	shape MeshVolume	shape MeshVolume
	shape LeastAxisLength	shape LeastAxisLength
	shape Maximum2DDiameterColumn	shape Maximum2DDiameterColumn
	shape Maximum2DDiameterSlice	glcm Correlation
	shape VoxelVolume	gldm DependenceEntropy
	glcm Correlation	glszm ZoneEntropy
	gldm DependenceEntropy	TLG
	glszm ZoneEntropy	
	TLG	
	shape SurfaceVolumeRatio	
RSF		
	glszm ZoneEntropy	glszm ZoneEntropy
	glcm JointENtropy	glcm JointEntropy
	gldm DependenceEntropy	gldm DependenceEntropy
	glrlm RunEntropy	glrlm RunEntropy
	firstorder Variance	firstorder Variance
	TLG	TLG
	glszm SmallAreaEmphasis	
	shape Elongation	
CGBSA		
	gldm DependenceNonUniformity	gldm DependenceNonUniformity
	gldm DependenceEntropy	gldm DependenceEntropy
	glszm ZoneEntropy	glszm ZoneEntropy
	firstorder TotalEnergy	firstorder TotalEnergy
	TLG	TLG
	shape Maximum2DDiameterSlice	glrlm RunLengthNonUniformity
	shape MeshVolume	ngtdm Coarseness

Coxnet: Cox's proportional hazard's model with elastic net penalty
FSSVM: Fast Survival Support Vector Machine
RSF: Random Survival Forest
CGBSA: Component wise Gradient Boosting Survival Analysis
GLDM: Gray Level Dependence Matrix
GLSZM: Gray Level Size Zone Matrix
TLG: Total Lesion Glycolysis
ngtdm: Neighbouring Gray Tone Difference Matrix
glrlm: Gray Level Run Length Matrix

Table S3: C-index of models based on MTV only

	PFS	***OS***
CoxNet		
Train	0.69 ± 0.02	0.72 ± 0.02
Validation	0.69 ± 0.07	0.73 ± 0.08
Test (Center 2: LACC)	0.75	0.74
Test (Center 4: ASCC)	0.70	0.71
Test (Center 5: OPSCC)		0.56
FSSVM		
Train	0.69 ± 0.02	0.72 ± 0.07
Validation	0.69 ± 0.08	0.72 ± 0.08
Test (Center 2: LACC)	0.75	0.74
Test (Center 4: ASCC)	0.65	0.71
Test (Center 5: OPSCC)		0.59
RSF		
Train	0.77 ± 0.03	0.73 ± 0.06
Validation	0.66 ± 0.06	0.65 ± 0.06
Test Center 2	0.60	0.60
Test Center 4	0.61	0.59
Test Center 5		0.52
CGBSA		
Train	0.69 ± 0.01	0.74 ± 0.02
Validation	0.70 ± 0.05	0.74 ± 0.07
Test (Center 2: LACC)	0.75	0.75
Test (Center 4: ASCC)	0.65	0.71
Test (Center 5: OPSCC)		0.56

Coxnet: Cox's proportional hazard's model with elastic net penalty
FSSVM: Fast Survival Support Vector Machine
RSF: Random Survival Forest
CGBSA: Component wise Gradient Boosting Survival Analysis

Table S4: C-index of models based on radiomic features: exclusion of patients with missing information on HPV status in Center 3

	PFS	***OS***
CoxNet		
Train	0.71 ± 0.03	0.73 ± 0.02
Validation	0.72 ± 0.08	0.70 ± 0.07
Test (Center 2: LACC)	0.71	0.72
Test (Center 4: ASCC)	0.72	0.73
Test (Center 5: OPSCC)		0.64
FSSVM		
Train	0.73 ± 0.04	0.72 ± 0.07
Validation	0.70 ± 0.08	0.71 ± 0.08
Test (Center 2: LACC)	0.70	0.72
Test (Center 4: ASCC)	0.71	0.71
Test (Center 5: OPSCC)		0.68
RSF		
Train	0.75 ± 0.04	0.73 ± 0.05
Validation	0.72 ± 0.09	0.70 ± 0.08
Test (Center 2: LACC)	0.64	0.66
Test (Center 4: ASCC)	0.65	0.66
Test (Center 5: OPSCC)		0.62
CGBSA		
Train	0.74 ± 0.02	0.72 ± 0.06
Validation	0.72 ± 0.08	0.73 ± 0.09
Test (Center 2: LACC)	0.71	0.71
Test (Center 4: ASCC)	0.71	0.70
Test (Center 5: OPSCC)		0.61

Coxnet: Cox's proportional hazard's model with elastic net penalty
FSSVM: Fast Survival Support Vector Machine
RSF: Random Survival Forest
CGBSA: Component wise Gradient Boosting Survival Analysis

Figure S1: CoxNet risk score predictions and tumor volume

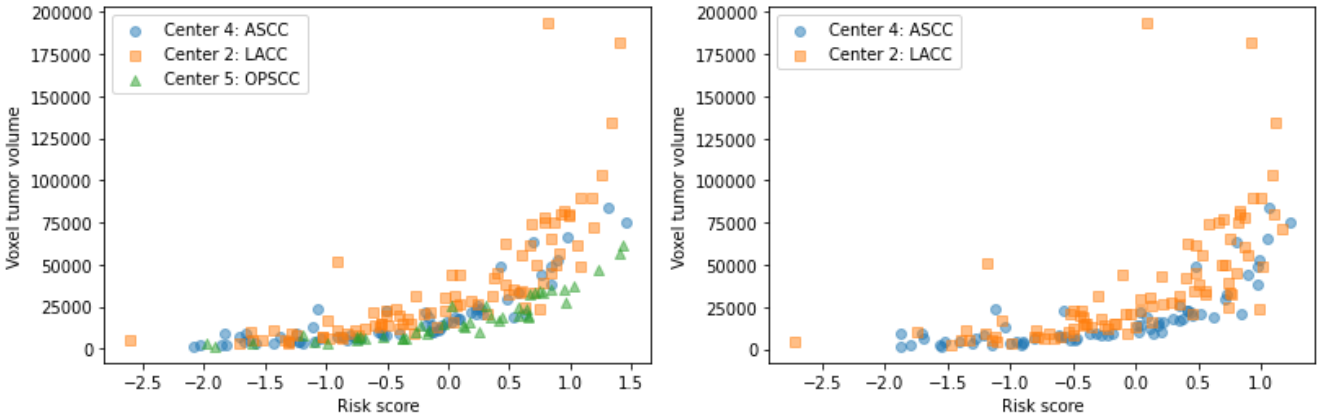


Figure S2: FSSVM risk score predictions and tumor volume

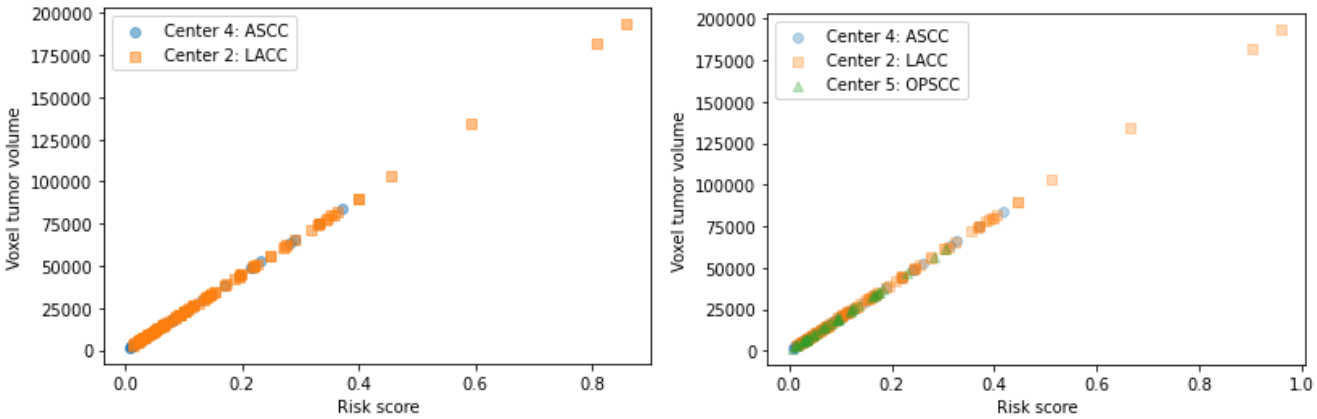
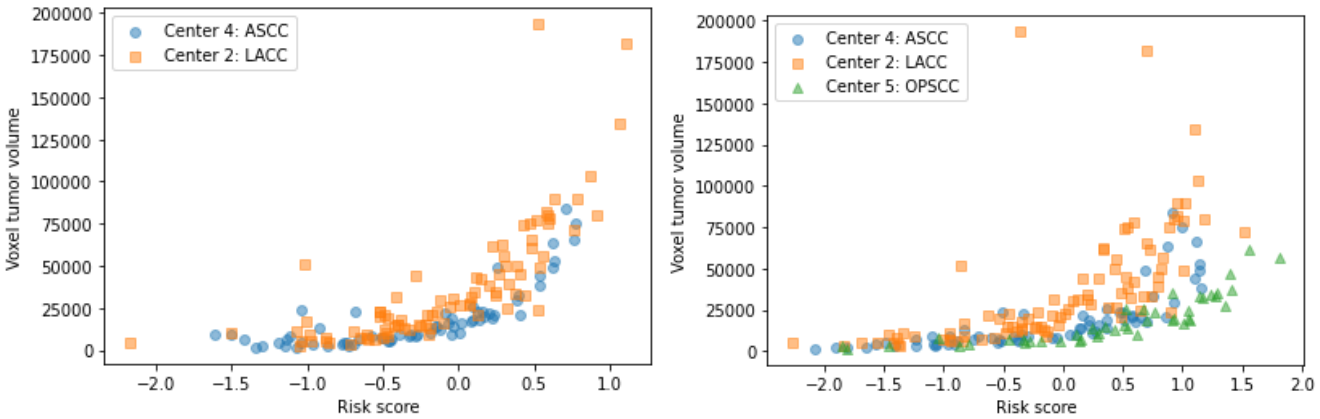


Figure S3: CGBSA risk score predictions and tumor volume



LACC: Locally advanced cervical cancer
ASCC: Anal squamous cell carcinoma
OPSCC: Oropharyngeal squamous cell carcinoma
Coxnet: Cox's proportional hazard's model with elastic net penalty
FSSVM: Fast Survival Support Vector Machine
RSF: Random Survival Forest
CGBSA: Component wise Gradient Boosting Survival Analysis

Bibliography

[1] Pina Brianti, Eduardo De Flammineis, and Santo Raffaele Mercuri. "Review of HPV-related diseases and cancers". eng. In: *New Microbiol* 40.2 (Apr. 2017), pp. 80–85. ISSN: 1121-7138.

[2] Catherine de Martel et al. "Worldwide burden of cancer attributable to HPV by site, country and HPV type". eng. In: *Int J Cancer* 141.4 (Aug. 2017), pp. 664–670. ISSN: 1097-0215. DOI: 10.1002/ijc.30716.

[3] Rodrigo Pinheiro Araldi et al. "The human papillomavirus (HPV)-related cancer biology: An overview". eng. In: *Biomed Pharmacother* 106 (Oct. 2018), pp. 1537–1556. ISSN: 1950-6007. DOI: 10.1016/j.biopha.2018.06.149.

[4] Ulrike Wieland and Alexander Kreuter. "Anal cancer risk: HPV-based cervical screening programmes". English. In: *The Lancet Infectious Diseases* 19.8 (Aug. 2019). Publisher: Elsevier, pp. 799–800. ISSN: 1473-3099, 1474-4457. DOI: 10.1016/S1473-3099(19)30296-8. URL: https://www.thelancet.com/journals/laninf/article/PIIS1473-3099(19)30296-8/fulltext (visited on 11/26/2021).

[5] Maria Elisa Sabatini and Susanna Chiocca. "Human papillomavirus as a driver of head and neck cancers". en. In: *Br J Cancer* 122.3 (Feb. 2020). Number: 3 Publisher: Nature Publishing Group, pp. 306–314. ISSN: 1532-1827. DOI: 10.1038/s41416-019-0602-7. URL: http://www.nature.com/articles/s41416-019-0602-7 (visited on 07/27/2022).

[6] Xavier Castellsagué et al. "HPV Involvement in Head and Neck Cancers: Comprehensive Assessment of Biomarkers in 3680 Patients". In: *JNCI: Journal of the National Cancer Institute* 108.6 (June 2016), djv403. ISSN: 0027-8874. DOI: 10.1093/jnci/djv403. URL: https://doi.org/10.1093/jnci/djv403 (visited on 07/27/2022).

[7] Angela Hong et al. "Rising prevalence of human papillomavirus-related oropharyngeal cancer in Australia over the last 2 decades". eng. In: *Head Neck* 38.5 (May 2016), pp. 743–750. ISSN: 1097-0347. DOI: 10.1002/hed.23942.

[8] Alon Taylor et al. "Human papillomavirus and oropharyngeal squamous cell carcinoma: a 12-year retrospective review in a New South Wales tertiary referral centre". en. In: *Australian Journal of Otolaryngology* 2.0 (Jan. 2019). Number: 0 Publisher: AME Publishing Company. ISSN: 2616-2792. URL: https://www.theajo.com/article/view/4143 (visited on 07/27/2022).

[9] Tarik Gheit. “Mucosal and Cutaneous Human Papillomavirus Infections and Cancer Biology”. eng. In: *Front Oncol* 9 (2019), p. 355. ISSN: 2234-943X. DOI: 10.3389/fonc.2019.00355.

[10] Teresa M. Darragh and Barbara Winkler. “Anal cancer and cervical cancer screening: key differences”. eng. In: *Cancer Cytopathol* 119.1 (Feb. 2011), pp. 5–19. ISSN: 1934-662X. DOI: 10.1002/cncy.20126.

[11] N. D. Nigro et al. “Combined therapy for cancer of the anal canal”. eng. In: *Dis Colon Rectum* 24.2 (Apr. 1981), pp. 73–75. ISSN: 0012-3706. DOI: 10.1007/BF02604287.

[12] H Bartelink et al. “Concomitant radiotherapy and chemotherapy is superior to radiotherapy alone in the treatment of locally advanced anal cancer: results of a phase III randomized trial of the European Organization for Research and Treatment of Cancer Radiotherapy and Gastrointestinal Cooperative Groups.” In: *JCO* 15.5 (May 1997). Publisher: Wolters Kluwer, pp. 2040–2049. ISSN: 0732-183X. DOI: 10.1200/JCO.1997.15.5.2040. URL: https://ascopubs.org/doi/10.1200/JCO.1997.15.5.2040 (visited on 02/08/2021).

[13] Eleonor Rivin Del Campo et al. “Pooled Analysis of external-beam RADiotherapy parameters in phase II and phase III trials in radiochemotherapy in Anal Cancer (PARADAC)”. eng. In: *Eur J Cancer* 121 (Nov. 2019), pp. 130–143. ISSN: 1879-0852. DOI: 10.1016/j.ejca.2019.08.022.

[14] S. Rao et al. “Anal cancer: ESMO Clinical Practice Guidelines for diagnosis, treatment and follow-up”. English. In: *Annals of Oncology* 32.9 (Sept. 2021). Publisher: Elsevier, pp. 1087–1100. ISSN: 0923-7534, 1569-8041. DOI: 10.1016/j.annonc.2021.06.015. URL: https://www.annalsofoncology.org/article/S0923-7534(21)02064-0/fulltext#secsectitle0095 (visited on 11/25/2022).

[15] V. Grégoire et al. “Squamous cell carcinoma of the head and neck: EHNS-ESMO-ESTRO Clinical Practice Guidelines for diagnosis, treatment and follow-up”. eng. In: *Ann Oncol* 21 Suppl 5 (May 2010), pp. v184–186. ISSN: 1569-8041. DOI: 10.1093/annonc/mdq185.

[16] Alina Sturdza et al. “Image guided brachytherapy in locally advanced cervical cancer: Improved pelvic control and survival in RetroEMBRACE, a multicenter cohort study”. eng. In: *Radiother Oncol* 120.3 (Sept. 2016), pp. 428–433. ISSN: 1879-0887. DOI: 10.1016/j.radonc.2016.03.011.

[17] Prajnan Das et al. “Predictors and patterns of recurrence after definitive chemoradiation for anal cancer”. eng. In: *Int J Radiat Oncol Biol Phys* 68.3 (July 2007), pp. 794–800. ISSN: 0360-3016. DOI: 10.1016/j.ijrobp.2006.12.052.

[18] Ritu Salani et al. “An update on post-treatment surveillance and diagnosis of recurrence in women with gynecologic malignancies: Society of Gynecologic Oncology (SGO) recommendations”. eng. In: *Gynecol Oncol* 146.1 (July 2017), pp. 3–10. ISSN: 1095-6859. DOI: 10.1016/j.ygyno.2017.03.022.

[19] Kathinka S. Slørdahl et al. “Treatment outcomes and prognostic factors after chemoradiotherapy for anal cancer”. eng. In: *Acta Oncol* 60.7 (July 2021), pp. 921–930. ISSN: 1651-226X. DOI: 10.1080/0284186X.2021.1918763.

[20] Jessica M. Frakes et al. “Determining optimal follow-up in the management of human papillomavirus-positive oropharyngeal cancer”. en. In: *Cancer* 122.4 (2016). _eprint: https://onlinelibra pp. 634–641. ISSN: 1097-0142. DOI: 10.1002/cncr.29782. URL: https://onlinelibrary.wiley.com/doi/abs/10.1002/cncr.29782 (visited on 08/29/2022).

[21] N. J. Bhuva et al. “To PET or not to PET? That is the question. Staging in anal cancer”. English. In: *Annals of Oncology* 23.8 (Aug. 2012). Publisher: Elsevier, pp. 2078–2082. ISSN: 0923-7534, 1569-8041. DOI: 10.1093/annonc/mdr599. URL: https://www.annalsofoncology.org/article/S0923-7534(19)38107-4/fulltext#%20 (visited on 02/16/2022).

[22] Carmelo Caldarella et al. “Diagnostic Performance of Positron Emission Tomography/Computed Tomography Using Fluorine-18 Fluorodeoxyglucose in Detecting Locoregional Nodal Involvement in Patients with Anal Canal Cancer: A Systematic Review and Meta-Analysis”. In: *ScientificWorldJournal* 2014 (Feb. 2014), p. 196068. ISSN: 2356-6140. DOI: 10.1155/2014/196068. URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3932262/ (visited on 02/16/2022).

[23] Thomas G. Trautmann and James H. Zuger. “Positron Emission Tomography for pretreatment staging and posttreatment evaluation in cancer of the anal canal”. eng. In: *Mol Imaging Biol* 7.4 (Aug. 2005), pp. 309–313. ISSN: 1536-1632. DOI: 10.1007/s11307-005-0003-6.

[24] P. CASTALDI et al. “Role of 18F-FDG PET-CT in head and neck squamous cell carcinoma”. In: *Acta Otorhinolaryngol Ital* 33.1 (Feb. 2013), pp. 1–8. ISSN: 0392-100X. URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3631810/ (visited on 08/04/2022).

[25] Sven van den Bosch et al. “18F-FDG-PET/CT-based treatment planning for definitive (chemo)radiotherapy in patients with head and neck squamous cell carcinoma improves regional control and survival”. English. In: *Radiotherapy and Oncology* 142 (Jan. 2020). Publisher: Elsevier, pp. 107–114. ISSN: 0167-8140, 1879-0887. DOI: 10.1016/j.radonc.2019.07.025. URL: https://www.thegreenjournal.com/article/S0167-8140(19)33018-X/fulltext (visited on 08/04/2022).

[26] Fernanda G. Herrera et al. “[(18)F]FDG-PET/CT metabolic parameters as useful prognostic factors in cervical cancer patients treated with chemo-radiotherapy”. eng. In: *Radiat Oncol* 11 (Mar. 2016), p. 43. ISSN: 1748-717X. DOI: 10.1186/s13014-016-0614-x.

[27] Espen Rusten et al. “Anal cancer chemoradiotherapy outcome prediction using 18F-fluorodeoxyglucose positron emission tomography and clinicopathological factors”. In: *BJR* 92.1097 (May 2019). Publisher: The British Institute of Radiology, p. 20181006. ISSN: 0007-1285. DOI: 10.1259/

bjr.20181006. URL: https://www.birpublications.org/doi/full/10.1259/bjr.20181006 (visited on 02/16/2022).

[28] J. Castelli et al. "Overview of the predictive value of quantitative 18 FDG PET in head and neck cancer treated with chemoradiotherapy". eng. In: *Crit Rev Oncol Hematol* 108 (Dec. 2016), pp. 40–51. ISSN: 1879-0461. DOI: 10.1016/j.critrevonc.2016.10.009.

[29] Philippe Lambin et al. "Radiomics: extracting more information from medical images using advanced feature analysis". eng. In: *Eur J Cancer* 48.4 (Mar. 2012), pp. 441–446. ISSN: 1879-0852. DOI: 10.1016/j.ejca.2011.11.036.

[30] Baderaldeen A. Altazi et al. "Investigating multi-radiomic models for enhancing prediction power of cervical cancer treatment outcomes". en. In: *Physica Medica* 46 (Feb. 2018), pp. 180–188. ISSN: 1120-1797. DOI: 10.1016/j.ejmp.2017.10.009. URL: https://www.sciencedirect.com/science/article/pii/S1120179717304787 (visited on 04/22/2022).

[31] François Lucia et al. "External validation of a combined PET and MRI radiomics model for prediction of recurrence in cervical cancer patients treated with chemoradiotherapy". eng. In: *Eur J Nucl Med Mol Imaging* 46.4 (Apr. 2019), pp. 864–877. ISSN: 1619-7089. DOI: 10.1007/s00259-018-4231-9.

[32] Arnaud Hocquelet et al. "Pre-treatment magnetic resonance-based texture features as potential imaging biomarkers for predicting event free survival in anal cancer treated by chemoradiotherapy". en. In: *Eur Radiol* 28.7 (July 2018), pp. 2801–2811. ISSN: 1432-1084. DOI: 10.1007/s00330-017-5284-z. URL: https://doi.org/10.1007/s00330-017-5284-z (visited on 04/22/2022).

[33] Kasia Owczarczyk et al. "MRI heterogeneity analysis for prediction of recurrence and disease free survival in anal cancer". English. In: *Radiotherapy and Oncology* 134 (May 2019). Publisher: Elsevier, pp. 119–126. ISSN: 0167-8140, 1879-0887. DOI: 10.1016/j.radonc.2019.01.022. URL: https://www.thegreenjournal.com/article/S0167-8140(19)30027-1/fulltext#%20 (visited on 04/22/2022).

[34] Nicolas Giraud et al. "MRI-Based Radiomics Input for Prediction of 2-Year Disease Recurrence in Anal Squamous Cell Carcinoma". In: *Cancers (Basel)* 13.2 (Jan. 2021), p. 193. ISSN: 2072-6694. DOI: 10.3390/cancers13020193. URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7827348/ (visited on 04/22/2022).

[35] J. Beaumont et al. "Voxel-based identification of local recurrence sub-regions from pretreatment PET/CT for locally advanced head and neck cancers". eng. In: *EJNMMI Res* 9.1 (Sept. 2019), p. 90. ISSN: 2191-219X. DOI: 10.1186/s13550-019-0556-z.

[36] Reinhard R. Beichel et al. “FDG PET based prediction of response in head and neck cancer treatment: Assessment of new quantitative imaging features”. In: *PLoS One* 14.4 (Apr. 2019), e0215465. ISSN: 1932-6203. DOI: 10.1371/journal.pone.0215465. URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6474600/ (visited on 08/04/2022).

[37] Sylvain Reuzé et al. “Prediction of cervical cancer recurrence using textural features extracted from 18 F-FDG PET images acquired with different scanners”. In: *Oncotarget* 8.26 (May 2017). Publisher: Impact Journals, pp. 43169–43179. ISSN: 1949-2553. DOI: 10.18632/oncotarget.17856. URL: https://www.oncotarget.com/article/17856/text/ (visited on 01/26/2021).

[38] P. J. Brown et al. “Prediction of outcome in anal squamous cell carcinoma using radiomic feature analysis of pre-treatment FDG PET-CT”. eng. In: *Eur J Nucl Med Mol Imaging* 46.13 (Dec. 2019), pp. 2790–2799. ISSN: 1619-7089. DOI: 10.1007/s00259-019-04495-1.

[39] Marta Bogowicz et al. “Comparison of PET and CT radiomics for prediction of local tumor control in head and neck squamous cell carcinoma”. eng. In: *Acta Oncol* 56.11 (Nov. 2017), pp. 1531–1536. ISSN: 1651-226X. DOI: 10.1080/0284186X.2017.1346382.

[40] Wei Mu et al. “18F-FDG PET/CT Habitat Radiomics Predicts Outcome of Patients with Cervical Cancer Treated with Chemoradiotherapy”. eng. In: *Radiol Artif Intell* 2.6 (Nov. 2020), e190218. ISSN: 2638-6100. DOI: 10.1148/ryai.2020190218.

[41] Marta Ferreira et al. “[18F]FDG PET radiomics to predict disease-free survival in cervical cancer: a multi-scanner/center study with external validation”. en. In: *Eur J Nucl Med Mol Imaging* 48.11 (Oct. 2021), pp. 3432–3443. ISSN: 1619-7089. DOI: 10.1007/s00259-021-05303-5. URL: https://doi.org/10.1007/s00259-021-05303-5 (visited on 04/22/2022).

[42] Hu Cong et al. “FDG-PET/CT Radiomics Models for The Early Prediction of Locoregional Recurrence in Head and Neck Cancer”. eng. In: *Curr Med Imaging* 17.3 (2021), pp. 374–383. ISSN: 1573-4056. DOI: 10.2174/1573405616666200712181135.

[43] Martin Vallières et al. “Radiomics strategies for risk assessment of tumour failure in head-and-neck cancer”. eng. In: *Sci Rep* 7.1 (Aug. 2017), p. 10117. ISSN: 2045-2322. DOI: 10.1038/s41598-017-10371-5.

[44] Richard Pötter et al. “Recommendations from gynaecological (GYN) GEC ESTRO working group (II): Concepts and terms in 3D image-based treatment planning in cervix cancer brachytherapy—3D dose volume parameters and aspects of 3D image-based anatomy, radiation physics, radiobiology”. en. In: *Radiotherapy and Oncology* 78.1 (Jan. 2006), pp. 67–77. ISSN: 0167-8140. DOI: 10.1016/j.radonc.2005.11.014. URL: https://www.sciencedirect.com/science/article/pii/S0167814005005463 (visited on 03/16/2022).

[45] Vincent Grégoire et al. “Delineation of the primary tumour Clinical Target Volumes (CTV-P) in laryngeal, hypopharyngeal, oropharyngeal and oral cavity squamous cell carcinoma: AIRO, CACA, DAHANCA, EORTC, GEORCC, GORTEC, HKNPCSG, HNCIG, IAG-KHT, LPRHHT, NCIC CTG, NCRI, NRG Oncology, PHNS, SBRT, SOMERA, SRO, SSHNO, TROG consensus guidelines”. eng. In: *Radiother Oncol* 126.1 (Jan. 2018), pp. 3–24. ISSN: 1879-0887. DOI: 10.1016/j.radonc.2017.10.016.

[46] James D. Brierley, Mary K. Gospodarowicz, and Christian Wittekind. *TNM Classification of Malignant Tumours*. en. Google-Books-ID: 642GDQAAQBAJ. John Wiley & Sons, Jan. 2017. ISBN: 978-1-119-26357-9.

[47] Sergio Pecorelli. “Revised FIGO staging for carcinoma of the vulva, cervix, and endometrium”. en. In: *International Journal of Gynecology & Obstetrics* 105.2 (2009). _eprint: https://onlinelibrary.wil pp. 103–104. ISSN: 1879-3479. DOI: 10.1016/j.ijgo.2009.02.012. URL: https://onlinelibrary.wiley.com/doi/abs/10.1016/j.ijgo.2009.02.012 (visited on 04/22/2022).

[48] *ANTs by stnava*. URL: http://stnava.github.io/ANTs/ (visited on 08/29/2022).

[49] Fanny Orlhac et al. “Radiomics in PET Imaging: A Practical Guide for Newcomers”. English. In: *PET Clinics* 16.4 (Oct. 2021). Publisher: Elsevier, pp. 597–612. ISSN: 1556-8598, 1879-9809. DOI: 10.1016/j.cpet.2021.06.007. URL: https://www.pet.theclinics.com/article/S1556-8598(21)00046-8/fulltext (visited on 09/13/2022).

[50] Joost J. M. van Griethuysen et al. “Computational Radiomics System to Decode the Radiographic Phenotype”. eng. In: *Cancer Res* 77.21 (Nov. 2017), e104–e107. ISSN: 1538-7445. DOI: 10.1158/0008-5472.CAN-17-0339.

[51] W. Evan Johnson, Cheng Li, and Ariel Rabinovic. “Adjusting batch effects in microarray expression data using empirical Bayes methods”. In: *Biostatistics* 8.1 (Jan. 2007), pp. 118–127. ISSN: 1465-4644. DOI: 10.1093/biostatistics/kxj037. URL: https://doi.org/10.1093/biostatistics/kxj037 (visited on 09/20/2021).

[52] Jean-Philippe Fortin et al. “Harmonization of multi-site diffusion tensor imaging data”. eng. In: *Neuroimage* 161 (Nov. 2017), pp. 149–170. ISSN: 1095-9572. DOI: 10.1016/j.neuroimage.2017.08.047.

[53] Jean-Philippe Fortin et al. “Harmonization of cortical thickness measurements across scanners and sites”. en. In: *NeuroImage* 167 (Feb. 2018), pp. 104–120. ISSN: 1053-8119. DOI: 10.1016/j.neuroimage.2017.11.024. URL: https://www.sciencedirect.com/science/article/pii/S105381191730931X (visited on 03/22/2022).

[54] Fanny Orlhac et al. “A guide to ComBat harmonization of imaging biomarkers in multicenter studies”. en. In: *Journal of Nuclear Medicine* (Sept. 2021). Publisher: Society of Nuclear Medicine Section: State-of-the-Art (Invitation Only). ISSN: 0161-5505, 2159-662X.

DOI: 10.2967/jnumed.121.262464. URL: https://jnm.snmjournals.org/content/early/2021/09/15/jnumed.121.262464 (visited on 04/28/2022).

[55] Sebastian Pölsterl. "scikit-survival: A Library for Time-to-Event Analysis Built on Top of scikit-learn". In: *Journal of Machine Learning Research* 21.212 (2020), pp. 1–6. ISSN: 1533-7928. URL: http://jmlr.org/papers/v21/20-729.html (visited on 02/17/2022).

[56] *Ensemble Methods: Foundations and Algorithms - 1st Edition - Zhi-Hua*. URL: https://www.routledge.com/Ensemble-Methods-Foundations-and-Algorithms/Zhou/p/book/9781439830031 (visited on 02/17/2022).

[57] Ludmila I Kuncheva. "Ensemble Methods". en. In: *Combining Pattern Classifiers*. Section: 6 _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/9781118914564.ch6. John Wiley & Sons, Ltd, 2014, pp. 186–229. ISBN: 978-1-118-91456-4. DOI: 10.1002/9781118914564.ch6. URL: http://onlinelibrary.wiley.com/doi/abs/10.1002/9781118914564.ch6 (visited on 08/08/2022).

[58] Giacomo Maria Lima et al. "Prognostic value of posttreatment 18F-FDG PET/CT and predictors of metabolic response to therapy in patients with locally advanced cervical cancer treated with concomitant chemoradiation therapy: an analysis of intensity- and volume-based PET parameters". en. In: *Eur J Nucl Med Mol Imaging* 45.12 (Nov. 2018), pp. 2139–2146. ISSN: 1619-7089. DOI: 10.1007/s00259-018-4077-1. URL: https://doi.org/10.1007/s00259-018-4077-1 (visited on 04/04/2022).

[59] Saeid Parvandeh et al. "Consensus features nested cross-validation". In: *Bioinformatics* 36.10 (May 2020), pp. 3093–3098. ISSN: 1367-4803. DOI: 10.1093/bioinformatics/btaa046. URL: https://doi.org/10.1093/bioinformatics/btaa046 (visited on 05/19/2022).

[60] Mathieu Gauthé et al. "Prognostic impact of tumour burden assessed by metabolic tumour volume on FDG PET/CT in anal canal cancer". eng. In: *Eur J Nucl Med Mol Imaging* 44.1 (Jan. 2017), pp. 63–70. ISSN: 1619-7089. DOI: 10.1007/s00259-016-3475-5.

[61] Jefferson Rijo-Cedeño et al. "Metabolic tumor volume and total lesion glycolysis as prognostic factors in head and neck cancer: Systematic review and meta-analysis". eng. In: *Head Neck* 42.12 (Dec. 2020), pp. 3744–3754. ISSN: 1097-0347. DOI: 10.1002/hed.26461.

[62] I. El Naqa et al. "Exploring feature-based approaches in PET images for predicting cancer treatment outcomes". eng. In: *Pattern Recognit* 42.6 (June 2009), pp. 1162–1171. ISSN: 0031-3203. DOI: 10.1016/j.patcog.2008.08.011.

[63] Letizia Deantonio et al. "Anal cancer FDG-PET standard uptake value: correlation with tumor characteristics, treatment response and survival". eng. In: *Radiol Med* 121.1 (Jan. 2016), pp. 54–59. ISSN: 1826-6983. DOI: 10.1007/s11547-015-0562-9.

[64] R. J. Gillies et al. “The biology underlying molecular imaging in oncology: from genome to anatome and back again”. eng. In: *Clin Radiol* 65.7 (July 2010), pp. 517–521. ISSN: 1365-229X. DOI: 10.1016/j.crad.2010.04.005.

[65] Hugo J. W. L. Aerts et al. “Decoding tumour phenotype by noninvasive imaging using a quantitative radiomics approach”. en. In: *Nat Commun* 5.1 (June 2014). Bandiera_abtest: a Cg_type: Nature Research Journals Number: 1 Primary_atype: Research Publisher: Nature Publishing Group Subject_term: Cancer imaging;Medical research Subject_term_id: cancer-imaging;medical-research, p. 4006. ISSN: 2041-1723. DOI: 10.1038/ncomms5006. URL: http://www.nature.com/articles/ncomms5006 (visited on 09/20/2021).

[66] Alexandre Carré et al. “Standardization of brain MR images across machines and protocols: bridging the gap for MRI-based radiomics”. en. In: *Sci Rep* 10.1 (July 2020). Bandiera_abtest: a Cc_license_type: cc_by Cg_type: Nature Research Journals Number: 1 Primary_atype: Research Publisher: Nature Publishing Group Subject_term: Cancer imaging;CNS cancer;Computational science;Diagnostic markers Subject_term_id: cancer-imaging;cns-cancer;computational-science;diagnostic-markers, p. 12340. ISSN: 2045-2322. DOI: 10.1038/s41598-020-69298-z. URL: https://www.nature.com/articles/s41598-020-69298-z (visited on 09/20/2021).

[67] Alex Zwanenburg et al. “The Image Biomarker Standardization Initiative: Standardized Quantitative Radiomics for High-Throughput Image-based Phenotyping”. eng. In: *Radiology* 295.2 (May 2020), pp. 328–338. ISSN: 1527-1315. DOI: 10.1148/radiol.2020191145.

[68] R. Da-ano et al. “Performance comparison of modified ComBat for harmonization of radiomic features for multicenter studies”. en. In: *Sci Rep* 10.1 (June 2020). Number: 1 Publisher: Nature Publishing Group, p. 10248. ISSN: 2045-2322. DOI: 10.1038/s41598-020-66110-w. URL: https://www.nature.com/articles/s41598-020-66110-w (visited on 02/17/2022).

[69] Maxime Lacroix et al. “Correction for Magnetic Field Inhomogeneities and Normalization of Voxel Values Are Needed to Better Reveal the Potential of MR Radiomic Features in Lung Cancer”. eng. In: *Front Oncol* 10 (2020), p. 43. ISSN: 2234-943X. DOI: 10.3389/fonc.2020.00043.

[70] H. Liu et al. “Evolving feature selection”. In: *IEEE Intelligent Systems* 20.6 (Nov. 2005). Conference Name: IEEE Intelligent Systems, pp. 64–76. ISSN: 1941-1294. DOI: 10.1109/MIS.2005.105.

Chapter 5

MR image normalization in locally advanced cervical cancer: what is the optimal method for developing multicenter radiomic models?

Abstract

MRI-based radiomic studies have shown promising results in locally advanced cervical cancer (LACC) but have been negatively impacted by the lack of robustness of radiomic features depending on the device used and parameters selected for image acquisition. In this paper, we proposed a conditional generative adversarial network (cGAN) and a CycleGAN to address this problem and compared it to conventional preprocessing and a posteriori methods proposed in the literature. For this purpose, T2w MR images from 30 patients treated for LACC were acquired prospectively, as Cohort 1. For each patient, three images were taken sequentially on the same scanner with fixed values of repetition time (TR) and voxel size (VS). A retrospective cohort of 216 LACC patients (Cohort 2) was also gathered, including 86 and 160 T2w MR images taken before radiotherapy (RT) and brachytherapy (BT), respectively. A cGAN was trained on paired images from Cohort 1 to generate images robust to the impact of TR and VS modulation, and its performance was evaluated on an unseen test set considering 10 patients. A CycleGAN strategy was applied on unpaired images from Cohort 2. Histogram-matching standardization (Nyul), z-score normalization and ComBat harmonization methods were also implemented. Different quality metrics (e.g. mean, range, variance, Peak signal to noise ratio, etc) were extracted from Cohort 1 images and the impact of standardization methods assessed by means of principal component analysis (PCA). Using Intra-Class Correlation (ICC) and Concordance Correlation Coefficient (CCC), robust features were also characterized (CCC and ICC $\geq$ 0.75). Similarly, pre-RT and pre-BT MR images of Cohort 2 were standardized using Nyul, z-score, and CycleGAN, and different machine learning (ML) models were trained to investigate the impact of these harmonization methods on stage classification and relapse prediction, respectively. PCA on quality metrics showed that TR and VS changes were mitigated the most with cGAN. The maximum variance was effectively reduced by Nyul and z-score but only cGAN managed to reduce

also the maximum variance direction. Regarding TR/VS modulation on first-order features, cGAN achieved the best results with 100% (18/18) robust features after standardization follwed by Nyul combined with ComBat. Regarding TR/VS modulation on second order features, cGAN yielded 77%(58/75) of robust features. On pre-RT images, stage classification showed improved performance whatever the harmonization strategy implemented. Average accuracy was improved on the test set from 0.68 ± 0.16 to 0.83 ± 0.07, 0.78 ± 0.04, and 0.88 ± 0.09 with Nyul standardization, z-score standardization, and CycleGAN standardizations, respectively. Likewise, relapse prediction based on pre-BT images achieved higher AUC on the test set after image standardization with the best model achieving 0.44 before standardization, 0.52, 0.56, and 0.60 after z-score, Nyul, and CycleGAN standardization, respectively. Our results suggest that neural network-based normalization could better limit the impact of acquisition parameters on MRI radiomic features compared to conventional methods in LACC. This result reinforces the growing popularity of DL in medical imaging field, providing here an answer to the not-so-common problem of image normalization. This study provides an additional building block for the deployment of radiomic signatures at a multi-center scale.

Keywords— LACC, MRI, Radiomics, image standardization, cGAN, CycleGAN

5.1 Introduction

Magnetic resonance imaging (MRI) thanks to its high soft tissue contrast is a capital imaging modality in diagnosis, treatment planning, and follow-up of patients with locally advanced cervical cancer (LACC)[1, 2, 3, 4, 5, 6]. MRI T2 Weighted (T2w) sequence is particularly essential for RT and BT target volume and organs at risk (OARs) definition, as well as for the assessment of nodal involvement and distant metastasis[7]. Furthermore, it has been shown by several studies that T2w MR images-based processing methods, such as radiomic analysis, could lead to a better prognosis of tumor response in LACC [8, 9, 10, 11].

MR images manipulated on a daily basis by radiologists are generally contrast images, expressed in arbitrary units, meaning that the intensity value of one tissue can vary from scanner to scanner and from one acquisition to another[12]. Quantitative MRI sequences are emerging as potential imaging biomarkers for better diagnosis and characterization of tumors, in a similar fashion to standard uptake value (SUV) in positron emission tomography (PET[13, 14]). Diffusion-weighted imaging in MRI has the particularity to measure the mobility of water molecules for complementary information to convetional MRI. MR T2 mapping (or T1 mapping) measures transverse magnetization T2 relaxation time (or longitudinal magnetization T1 relaxation time) in milliseconds. There are still relatively few studies on quantitative MR images and these modalities are still scarcely used in clinical routine. Therefore, most radiomic studies are still based on conventional anatomical sequences such as T2w MR images. In contrast to the T2 MR mapping values, which are absolute and therefore more or less comparable, signal intensities in T2w MRI are highly dependent on imaging parameters, which will in turn impact radiomic features[15, 16, 17]. The first publications in the field of radiomics[18, 19] were met with strong enthusiasm, but the interest is now experiencing a relative decline. The main reason is the lack of reproducibility and robustness of radiomic features. If the first papers were content to develop signatures on very homogeneous datasets, these must now demonstrate their generalisability in multicentre conditions and on a large number of patients in order to be considered solid, and robust enough to be applicable to clinical practice. Various methods to tackle the lack of standardization of imaging protocols and inter-device variability have been proposed but, to this time, no consensus was reached. These methods might be applied directly on

MR images such as histogram matching [20, 21] or image intensity normalization; or a posteriori on radiomic features to eliminate the so-called 'batch-effect' [**carre_autoComBat_2022**, 22, 17].

Deep learning (DL) is becoming very popular in computer vision and medical imaging, achieving state-of-the-art results in both fields without the need of defining hand-crafted features. In radiotherapy, deep-learning is now widely used in clinical routine for a variety of tasks, ranging from automatic segmentation to the generation of synthetic images[23]. Generative adversarial networks (GANs) have been successfully applied for almost 10 years, and have demonstrated their ability to increase the realism of the generated images, i.e generation of synthetic CT from MRI, their utility in data augmentation in machine-learning (ML), noise removal, etc[24, 25, 26, 27, 28] Since its first publication[29], GANs have been broadly applied, and several improvements have been proposed. The conditional GAN (cGAN) is based on the original GAN with additional inputs to guide the generation of image. GAN and cGAN architectures are based on unsupervised learning and require a large amount of paired images. This means that to train successfully a GAN/cGAN to generate MR images from domain X (machine X with specific parameters) to domain Y (Machine X or machine Y, with different parameters), one has to gather paired data from domain X and domain Y. This would be an insurmountable limitation in radiomic analysis, since, in a retrospective context, images are not paired. Therefore, to overcome this limitation, unsupervised cycle generative adversarial networks (CycleGAN) to translate an image from a source domain X to a target domain Y in the absence of paired data were proposed[30]. CycleGANs have been applied for many purposes in MRI including intensity standardization, or generation of synthetic CT from MRI [31, 32, 33, 34, 35]. CycleGANs have shown superior visual similarities between image domains both quantitatively and qualitatively compared with other normalization techniques and thus increase the overlap of feature distribution not only among unharmonized MR images[32, 33, 36], but also in digital pathology, methods[37, 38]. Few studies have applied GAN or CycleGan for MR image standardization so far, despite the high resolution of generated images and preservation of the semantic information within the data[34, 39].

This study had two main objectives. On the one hand, on a prospective cohort with fixed acquisition parameters and paired images, a conditional GAN (cGAN) was trained to generate standardized MR images robust to the impact of repetition time (TR) and voxel size (VS). Its performance was compared to conventional standardization techniques: two standardization methods applied on MR images (a histogram matching method called Nyul and z-score normalization) and the ComBat method applied to extracted features. On the other hand, on a retrospective dataset, to evaluate the performance of these standardization methods, ML algorithms were trained on T2w MR images on two clinical tasks: tumor grade classification and relapse prediction. Here, MR images were unpaired with very large acquisition parameters. Therefore, instead of a cGAN, a CycleGAN method was implemented for MR image standardization.

5.2 Material and Methods

5.2.1 Input Data

T2w MR images from 30 patients treated for LACC at different steps of their cancer management (5 pre-RT, 7 pre-BT, 18 follow-up) were acquired prospectively, as Cohort 1. For each patient, T2w MR images were taken sequentially using a 1.5T GE Optima MR450w (GE Healthcare, Waukesha, WI) with specific values of TR and VS: 1) $TR \geq 5700ms$ and VS=0.58mm×0.58mm×4.0mm, 2) $TR = 3300ms$ and VS=0.78mmx0.78mm×4.0mm, 3) $TR = 3300ms$ and VS=0.58mm×0.58mm×4.0mm. All other parameters were kept constant (magnetic field, echo time, slice thickness, acquisition matrix). Image parameters of Cohort 1 are shown in Table 5.1. On each image, the whole cervix, excluding the disease, was manually segmented by a board-certified radiation oncologist (P-A.L) with 3 years of expertise. This was done to ensure that the region from which the radiomic features were extracted was the same across patients.

Acquisition	Manufacturer - Model name	Matrix dimension	TE (ms)	TR (ms)	Slice thickness (mm)	Pixel size (mmxmm)	FOV (mm)
1^{st}	GE Medical Systems - Optima MR450w	512x512	108	≥ 5700	4.0	0.58x0.58	300
2^{nd}				3300	4.0	0.78x0.78	400
3^{rd}				3300	4.0	0.58x0.58	300

Table 5.1: Prospective cohort MR images (n=30) acquisition parameters. TE: Echo time, TR: Repetition time, FOV: Field of view.

A retrospective cohort, Cohort 2, of 216 LACC patients was gathered. Patients were treated by a combination of external beam RT (EBRT) with concomitant chemotherapy, followed by a BT boost. T2w MR images for 86 patients and 216 patients, taken before EBRT and BT, respectively, were available as shown in Table 4.1. The Gross Tumor Volume (GTV) and the high-risk clinical target volume (CTVHR) were manually segmented on pre-RT and pre-BT MR images respectively. After quality checks, 56 patients were excluded from the pre-BT cohort: 22 patients presented artifacts in the segmented CTVHR and clinical follow-up was missing for 34 patients.

Patient characteristics and MRI acquisition parameters of Cohort 2 are detailed in Tables 2-4. Tables 6.2 and 5.3 refer to pre-RT and pre-BT MRI acquisition parameters, respectively. Table5.4 details patients' clinical characteristics. Patients experiencing local, regional (pelvic), or distant (metastasis) recurrence were considered relapsing patients. Patients were followed up at least until the occurrence of the relapse or at least for 30 months for non-relapsing patients.

5.2.2 Standardization methods

Two standardization techniques applied to images prior to radiomic features extraction and one a posteriori harmonization method applied to extracted features were implemented. GAN-based standardization was also assessed on Cohort 1 and Cohort 2 with cGAN and CycleGAN respectively. Prior to image standardization, MR images were corrected from magnetic field inhomogeneity using the N4ITK correction algorithm [40]. The images

Manufacturer	Model (n)	Magnetic Field (T)	Matrix dimension	TE (ms)	TR (ms)	Slice thickness (mm)	Pixel Spacing Range (mmxmm)
			min*min-max*max	min-max	min-max	min-max	min*min-max*max
GE Medical Systems	SIGNA EXCITE (34)	1.5	512x512	88-149	2720-6750	3-8	0.39x0.39-0.78x0.78
	SIGNA HDxt (14)	1.5	512x512	103-124	4330-5460	3.0-7.0	0.47x0.47-0.58x0.58
	SIGNA HDx (3)	1.5	512x512	100-108	4330-5460	4.0-5.0	0.47x0.47
	SIGNA HDe (1)	1.5	512x512	100-108	4330-5460	5.0	0.47x0.47
	Genesis Sigma (1)	1.5	512x512	95-123	2400-3000	5.0	0.47x0.47-0.62x0.62
	Optima MR450w (1)	1.5	512x512	124	6036	3.5	0.49x0.49
	Optima MR360 (1)	1.5	512x512	117	4177	5.0	0.64x0.64
	Discovery MR750 (1)	3.0	512x512	113	5055	5.0	0.43x0.43
HITACHI Medical Corporation	Airis Elite (1) ($*$)	3.0	256x256	100	4720	6.0	1.17x1.17
Philips Medical System	Panorama HFOx (2)	1.5	512x512	117	3200-3600	4.0	0.47x0.47
	Intera (1)	1.5	512x512	80	3000	3.0	0.66x0.66
	Panorama HFO (2)	1.5	512x512	117	4177	5.0	0.64x0.64
	Achieva (2)	1.5	320x320-512x512	90-125	3774-8260	2.0-5.0	0.48x0.48-0.75x0.75
Siemens	Avanto (7)	1.5	320x320-384x384	108-149	3550-7570	3.5-6.0	0.75x0.75-1.0x1.0
	Symphony (6)	1.5	372x372-512x512	88-110	5960-7510	4.0-5.0	0.74x0.74-0.83x0.83
Toshiba	MRT200PP3 (7)	1.5	288x288-512x512	75-120	4800-7320	3.5-4.0	0.58x0.58-0.76x0.76
	MRT200SP6 (3)	1.5	512x512-768x768	108-144	4200-8134	3.0-3.5	0.32x0.32-0.47x0.47

Table 5.2: Retrospective pre-RT MR images (n=86) acquisition parameters. RT: Radiotherapy, T: Tesla, TE: Echo time, TR: Repetition time, ms: milliseconds. n is the number of patients for whom images were acquired on the specified model. $*$: MR image selected as as target for CycleGAN training.

and associated masks were then resampled to isotropic voxels of $1 \times 1 \times 1\ mm^3$ using b-spline interpolation and nearest-neighbor interpolation, respectively, implemented in the Advanced Normalization Tools (ANTs)[41].

Methods based on standardization of MR image intensities

In this paper, two methods proposed in the literature for MR images intensity standardization were assessed. The first method was a histogram-matching method detailed for the first time by Nyul et al.[20] and modified

Manufacturer	Model	Magnetic Field T)	Matrix dimension	TE (ms)	Slice thickness (mm)	PixelSpacing Range (mmxmm)	TR (ms) (min-max)	n patients
GE Medical Systems	SIGNA EXCITE	1.5	512x512	128	3	0.58x0.58	2340	1
							3300-3320 (∗)	16
							3440-3900	60
							4460	1
							5200-5800	4
							6000-6640	41
						0.60x0.60-0.68x0.68	3300-3440	22
							4260	1
							6600	5
						0.70x0.70-0.78x0.78	3400	7
							6600	2

Table 5.3: Retrospective pre-BT MR images (n=160) acquisition parameters (Cohort 2). BT: Brachytherapy, T: Tesla, TE: Echo time, TR: Repetition time, ms: milliseconds. ∗: MR images with TR=3300ms (16 patients) were selected as target for CycleGAN training. n is the number of patients for whom images were acquired with the specified parameters.

later by Shah et al.[21]. The main idea underlying the method is to deform the histogram of a source image so that it matches a target histogram determined over the training set in an ML context. The matching is based on intensity-landmark configuration C_L as:

$$C_L = [p_{low}, m_{10}, m_{20}, m_{30}, m_{40}, m_{50}, m_{60}, m_{70}, m_{80}, m_{90}, p_{high}] \tag{5.1}$$

where each m_i, $i \in$ 10, 20, ..., 90 denotes the mode m at each ith percentile of the histogram associated with the foreground of the image. The tails of the image histogram are generally pruned so as to make the algorithm robust against artifacts and outliers that may result in inter-patient and scanner variations. This pruning results in an intensity range called intensity of interest (IOI).p_{low} and p_{high} denote the minimum and maximum percentile values of the overall intensity range of the IOI. The method consists of a first stage to find the parameters of the standard scale and a transformation stage that maps the histograms of candidate volumes to the standard histogram scale in a piece-wise linear manner[21].

The second technique was z-score normalization. Here, the average intensity and standard deviation of each MR image foreground was set to zero and one, respectively, by subtracting the average intensity of the entire image foreground from each voxel value and dividing it by the corresponding standard deviation.

Conditional Generative Adversarial Network - Cohort 1

Generative Adversarial networks family consists of two neural networks trained simultaneously: a generative model G that captures the data distribution and a discriminative model D that estimates the probability that a sample came from the real data rather than generated by G. Through a minimax two-player game, in the space of arbitrary functions G and D, a unique solution exists, with G recovering the training data distribution and D

Variables	Pre-RT mean[min-max]	Pre-BT mean[min-max]
Age (years)	48.0[26.5-79.9]	48.6[25.5-82.7]
	Value(%)	Value(%)
FIGO stage		
Ib1	1 (2.2%)	5 (3.1%)
Ib2	29 (32.9%)	43 (26.9%)
IIa	5 (5.7%)	13 (8.1%)
IIb	34 (45.5%)	78 (48.7%)
IIIa	1 (2.2%)	3 (1.9%)
IIIb	8 (9.1%)	14 (8.7%)
IVa	2 (2.2%)	4 (2.5%)
IVb	6 (1%)	-
Nodal status		
Yes	33 (38%)	72 (45%)
No	53 (62%)	88 (55%)
Histology		
SCC	72 (81.8%)	132 (82.5%)
ADK	12 (13.6%)	23 (14.4%)
ASC	2 (3.4%)	5 (3.1%)
Relapse		
Total	18 (20.9%)	51 (31.9%)
Local	7 (8.1%)	26 (16.2%)
Pelvic	7 (8.1%)	18 (11.2%)
Metastasis	9 (10.5%)	20 (12.5%)

Table 5.4: Patients' clinical characteristics. SCC: squamous cell carcinoma, ADK: adenocarcinoma, ASC: adenosquamous carcinoma

equal to 0.5 everywhere[29].

On Cohort 1, the proposed standardization method is an image-to-image translation inspired by the Pix2Pix cGAN developed by Isola et al.[42]. Their method for image-to-image translation is not task-specific and is relatively easy to implement. While GANs learn a mapping from random noise vector z to output an image y (G: z $\rightarrow$ y), cGANs learn a mapping from observed image x and random noise vector z, to output an image y, (G: (x,z) $\rightarrow$ y). This makes cGANs suitable for image-to-image translation tasks, where we condition on an input image and generate a corresponding output image[42]. The cGAN was used to translate domain information from the 1^{st} and 2^{nd} acquisitions to the 3^{rd} acquisition, reducing variabilities due to different TR and VS in Cohort 1. The cGAN consisted here of a U-Net-based[43] G and a convolutional neural network D. The generator was trained via adversarial loss, namely binary cross-entropy (BCE), to generate plausible images in the target space. The discriminator consisted of four convolutional layers with different filter sizes but the same kernel sizes and strides, followed by three fully connected layers. ReLU was used as the activation function.

The training process followed a two-player minimax game with loss function $\mathbb{L}_{cGAN}(G,D)$:

$$\min_{G} \max_{D} \mathbb{L}_{cGAN}(G,D) = \mathbb{E}_{x,y}[logD(x,y)] + \mathbb{E}_{x,z}[log(1-D(x,G(x,z)))] \tag{5.2}$$

where G tries to minimize this function against an adversarial D that tries to maximize it.

It has been shown that combining the GAN objective with another loss function, such as a distance loss function can improve its performance. Doing this implies that G has to generate an image similar to the ground truth in regard to distance. Here similarly to Isola et al., [42], we used L1 distance rather than L2 as L1 encourages higher spatial resolutions. The loss function applied to D was BCE. The final objective becomes:

$$\min_G \max_D \mathbb{L}_{cGAN}(G,D) = \mathbb{E}_{x,y}[logD(x,y)] + \mathbb{E}_{x,z}[log(1-D(x,G(x,z)))] + \lambda \mathbb{L}_{L1}(G) \tag{5.3}$$

The cGAN was trained using 40 paired images from the 1^{st} or 2^{nd} acquisitions to generate MR images standardized to the 3^{rd} acquisition (20 patients). The network performance was evaluated on the remaining 10 patients. More details on cGAN can be found in [29] and [42]. To accelerate convergence, a slice-paired training strategy was applied, making the hypothesis that there was no patient motion between acquisitions. Here images were resized to (256,256,32). Due to constraints, we considered a batch size of 1, corresponding to each slice along the z-axis. The model was trained for 700 epochs until D achieved a random probability when estimating if the generated image was from 3^{rd} acquisition. The Adam solver, with a learning rate of 0.0002, and momentum parameters 1 = 0.5, 2 = 0.999 were applied, corresponding to hyperparameters selected in the original work of Isola et al.[42] The workflow is described in Figure 5.1.

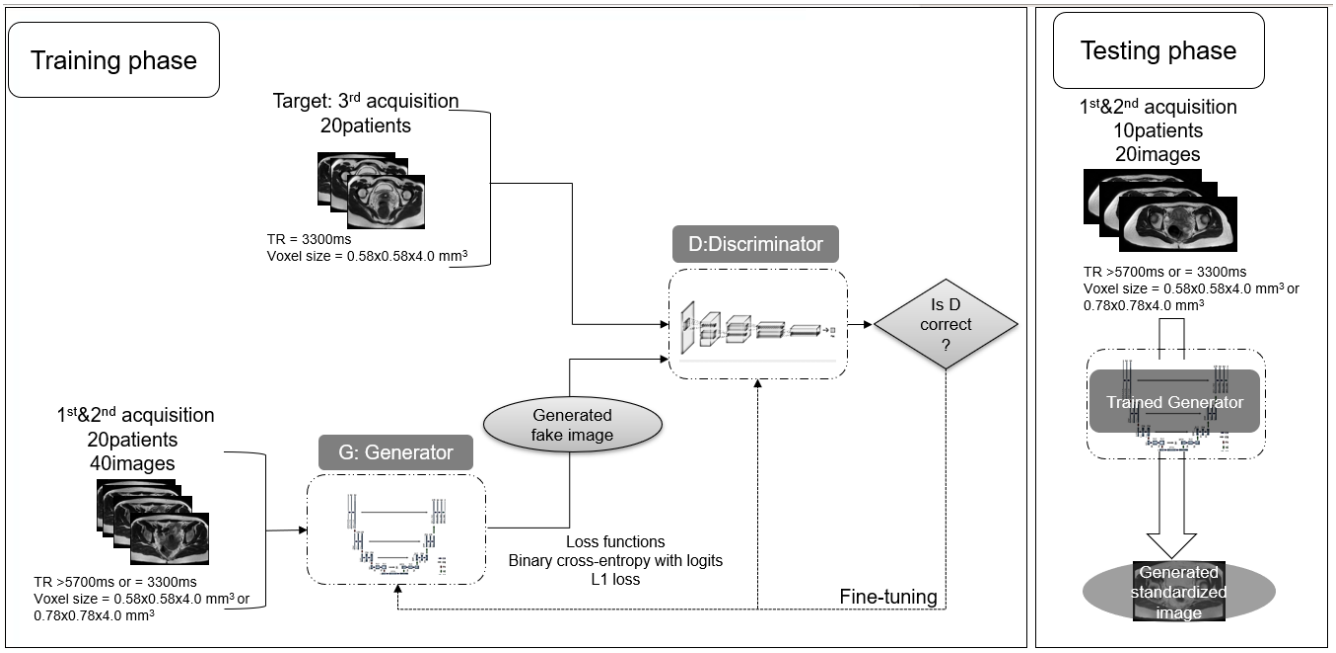


Figure 5.1: cGAN architecture for MR image standardization applied to Cohort 1. The input MR images have a TR or VS different from 3300ms or $0.58x0.58x4.0mm^3$, respectively, and are transformed into a new domain, in terms of image quality ($TR = 3300ms, VS = 0.58x0.58x4.0mm^3$). Matched data from 20 patients were used for training and the network performance was tested on 10 new patients.

CycleGAN - Cohort 2

Given that in Cohort 2, the use of the GAN was impossible because of unpaired data, a CycleGAN was trained. Mathematically, CycleGAN consists of training a translator G: X→Y and another translator F: Y→X, with

G and F being inverses of each, both mappings should be bijections. X and Y are the domain source and domain target respectively. We applied this structural assumption by training both the mapping G and F simultaneously and adding a cycle consistency loss that encourages $F(G(x)) \approx x$ and $G(F(y)) \approx y$. Combining this loss with adversarial losses as BCE on domains X and Y yields our full objective for unpaired image-to-image translation[30]. The GAN training process described in 3.11 becomes in CycleGAN:

$$\min_{G,F} \max_{D_X,D_Y} \mathbb{L}_{CycleGAN}(G,F,D_X,D_Y) = \min_{G} \max_{D_Y} \mathbb{L}_{GAN}(G,D_Y,X,Y) + \min_{F} \max_{D_X} \mathbb{L}_{GAN}(F,D_X,Y,X) + \lambda \min_{G,F} \mathbb{L}_{cyc}(G,F)$$

, where λ controls the relative importance of the two objectives and

$$\min_{G,F} \mathbb{L}_{cyc}(G,F) = \mathbb{E}_{x\sim p_{data}}(x)[||F(G(x)) - x||_1] + \mathbb{E}_{z\sim p_z}(z)[||G(F(y)) - y||_1] \tag{5.4}$$

The generators and discriminators used in the CycleGAN share the same specificities as those of the cGAN trained on the prospective cohort. To accelerate convergence during the CycleGAN training, a slice training strategy was implemented. CycleGAN was trained to generate MR images considering MRI with specific acquisition parameters ($*$ in Table **??**) as the target domain for the pre-RT, and pre-BT images respectively. In Table 6.2, the 3 tesla MRI was chosen as target domain while in Table5.3 MR images with parameters similar to the target domain of the prospective cohort (with TR 3300ms and pixel size:0.58x0.58) were selected. Similarly to cGAN, images were resized to (256,256,32) and a batch size of 1 was considered. The model was trained for 700 epochs until convergence using Adam optimizer, with a learning rate of 0.0002, and momentum parameters $\beta 1 = 0.5, \beta 2 = 0.999$.

ComBat harmonization

Originally developed for ComBating the batch effect in microarray genomic analysis [22], it was later demonstrated that the ComBat harmonization technique successfully removes inter-site technical variability, while preserving inter-patient biological variability on MR images[44, 45]. ComBat harmonization implements the location (mean) and scale (variance) (L/S) method. The main goal of the L/S method is to set the same mean and variance on all batches. In the context of radiomics analysis, correcting for the batch effect consists of estimating additive and multiplicative parameters accounting for batch effects for each radiomic feature while taking into account present covariates. To improve the estimations of these parameters, ComBat harmonization technique can be applied with an empirical Bayes (EB) method, which assumes that the batch effect affects the radiomic features in a similar way, thus allowing for more robustness in the evaluation of the corrective parameters when the batch size is small. The EB method can be used with a parametric or a non-parametric approach. Here, we applied ComBat harmonization with EB estimates and non-parametric prior method.

5.2.3 Effect of standardization on robustness of radiomic features

Image Clustering with Principal component analysis

On the one hand, to evaluate the impact of each step (pre-processing and standardization) in the radiomic workflow (see Figure**??**) on Cohort 1, different quality metrics (Table 5.5) were extracted from MR images before and after each step, i.e bias field correction, bias field correction and resampling, bias field correction, resampling, and a standardization technique (Nyul, z-score or cGAN). These metrics have been proposed to quantify the batch scanner effect in MRI as well as to detect artifacts[46]. In total, 15 quality metrics were extracted including first-order statistical measures (e.g., range, variance, coefficient of variation) as well as second-order statistics and filter-based measures (e.g., contrast per pixel (CPP), entropic focus criterion (EFC), signal-to-noise ratio). Detailed formulation on the calculation of these metrics can be found in [46]. We then used Principal Component Analysis (PCA) for image quality metrics clustering for an easier interpretation of the information contained across acquisitions.

Table 5.5: Summary table of quality measures extracted before and after each image processing step prior to extraction of radiometric features[46].

Name	Description
mean	Mean of the foreground
range	Range of the foreground
variance	Variance of the foreground
PCV	Percent coefficient of variation: coefficient of variation of the foreground for shadowing and inhomogeneity artifacts
CPP	Contrast per pixel: mean of the foreground filtered by a 3×3 2D Laplacian kernel for shadowing artifacts
PSNR	Peak signal-to-noise ratio of the foreground
SNR1	Foreground standard deviation (SD) divided by background SD
SNR2	Mean of the foreground patch divided by background SD
SNR3	Foreground patch SD divided by the centered foreground patch SD
SNR4	Mean of the foreground patch divided by mean of the background patch
CNR	Contrast-to-noise ratio for shadowing and noise artifacts: mean of the foreground and background patches difference divided by background patch SD
CVP	Coefficient of variation of the foreground patch for shading artifacts: foreground patch SD divided by foreground patch mean
CJV	Coefficient of joint variation between the foreground and background for aliasing and inhomogeneity artifacts
EFC	Entropy focus criterion for motion artifacts
FBER	Foreground-background energy ratio for ringing artifacts

Radiomic features robustness

With the goal to evaluate radiomic features robustness, Nyul method and z-score normalisation were applied to the spatially resampled bias field corrected MR images before radiomic features extraction. The a posteriori ComBat method was also implemented on the radiomic features extracted from these MRI images, as well as to radiomic features extracted after Nyul or z-score image standardization. Lastly, the trained cGAN was applied on the remaining 10 patients and radiomic features were extracted from the generated images. Radiomic features were extracted after image standardization using a relative discretization with 32 binswith PyRadiomics (v3.0.1) [47]. First-order (n=18), gray-level co-occurrence matrix (GLCM)(n=24), gray-level dependence matrix (GLDM)(n=14), gray-level run length matrix (GLRLM)(n=16), gray-level size zone matrix (GLSZM)(n=16), neighboring gray-tone difference matrix (NGTDM)(n=5), and shape (n=14) features for a total of 107 radiomic features were extracted. By means of Intra-Class Correlation (ICC) and Concordance Correlation Coefficient (CCC), robust features were assessed (CCC and ICC ≥ 0.75) for each standardization method.

5.2.4 Stage classification and relapse prediction

Pre-RT and pre-BT MR images (Cohort 2) were standardized using Nyul, z-score, and CycleGAN methods. The ComBat method was not applied here due to the diversity of acquisition parameters. Patients for whom pre-RT MR images were available were separated into two classes: low (FIGO stage <IIb) and high (FIGO stage $\geq$ IIb) stages. Low stages cancer are characterized by small tumor size and no invasion of the vagina while high stages refer to tumors that have spread to the upper and/or lower vagina[48].

Four ML algorithms were trained with the goal to define a radiomic signature able to correctly categorize the patients depending on their stage: logistic regression, random forest, k-nearest neighbors classifier, and support vector machine (SVM), using a stratified five-fold cross-validation repeated 10 times. To address the class imbalance in the pre-RT dataset, the Synthetic Minority Over-sampling Technique (SMOTE)[49] was used to generate synthetic patient samples from the minority class on the training set of each fold. Classification performance was evaluated using the area under the receiver operating characteristic curve (ROC-AUC)[50]. The default settings of the different classifiers were used.

The other task was the prediction of relapse based on pre-BT images, considered here as a binary variable. In this aim, pre-BT images were stratified split into 80% for training (n=128) and 20% for testing (n=32). As shown in Table 4, only 51 patients were prone to relapse. Therefore, we also applied SMOTE to tackle class imbalance in the pre-BT dataset before training 4 ML algorithms: logistic regression, random forest classifier, decision tree classifier, and SVM with a stratified 5-fold cross-validation. The trained models were then evaluated on the test set using a ROC-AUC.

5.3 Results

5.3.1 Clustering with PCA

Results from the PCA are shown in Figure 2. Visually, using quality metrics as inputs, PCA showed that 2^{nd} and 3^{rd} acquisitions were much closer unlike 1^{st} acquisition after bias field correction and image resampling. This suggests that VS modulation induces less variability than TR in quality metrics. Nyul standardization added similarity among acquisitions, especially between 1^{st} and 2^{nd}. The variance of the principal component 2 (PC2) was similar among the three acquisitions after Nyul standardization but several elements from 3^{rd} acquisition could still be regarded as outliers in regard to the variance of principal component 1 (PC1). Z-score normalization applied after pre-processing reduced divergence among groups regarding PC1 compared to Nyul standardization. However, variations were greater for PC2 and the different acquisitions could be better differentiated comparatively to Nyul method. cGAN standardization compacted all acquisitions making it almost impossible to separate them both with respect to PC1 and PC2.

5.3.2 Radiomic features robustness

According to ICC and CCC values, we found first-order features to be very sensitive to TR modulation since none was robust (ICC and CCC < 0.75) between acquisitions 1 and 3. Only 22% (4/18) were robust to VS changes before any standardization. For first-order features, application of Nyul method followed by ComBat increased the features robust to TR and VS changes to 100% and 67% respectively. cGAN standardization resulted in 100% robust first-order features for both TR and VS modulations. For second-order features, 21% (16/75) and 15% (11/75) features were robust to TR and VS shifts before standardization. Regarding TR modulation, ComBat, z-score, z-score combined with ComBat, Nyul and Nyul combined with ComBat achieved 40% (30/75), 13% (10/75), 27% (20/75), 40% (30/75), 72% (54/75) of robust features respectively. Regarding VS modulation, ComBat harmonization achieved 15/75 (20%), z-score only 9/75 (12%), z-score combined to ComBat and Nyul 12/75 (16%), while Nyul combined to ComBat had only 10/75 (13%) of robust features. cGAN normalization resulted in the highest scores with 77% (58/75) features robust to TR and VS modulation (Figure 4).

5.3.3 Stage classification and relapse prediction

Classification results based on pre-RT and pre-BT MR images are shown in Tables 5.6 and 5.7. On pre-RT images (Cohort 1), grade classification showed better results after standardization for all ML algorithms. After standardization, the ROC-AUC was improved from 0.68±0.16 to 0.83±0.07 with Nyul standardization, 0.78±0.04 with z-score standardization and 0.88±0.09 with CycleGAN standardization for the logistic regression classifier, which in all cases achieved the best performance. Despite poor performance, relapse prediction based on pre-BT images achieved higher AUC after image standardization whatever the ML model considered. In the cross-validation, logistic regression was found to be the best model, with ROC-AUC of 0.54±0.11, 0.58±0.10, 0.63±0.11 and 0.68±0.14 before standardization, after z-score, Nyul, and CycleGAN standardization, respectively. On the test set, the logistic regression model's performance was equal to 0.44 before standardization, to be compared to 0.52, 0.56, and 0.60 after z-score, Nyul, and CycleGAN standardization, respectively.

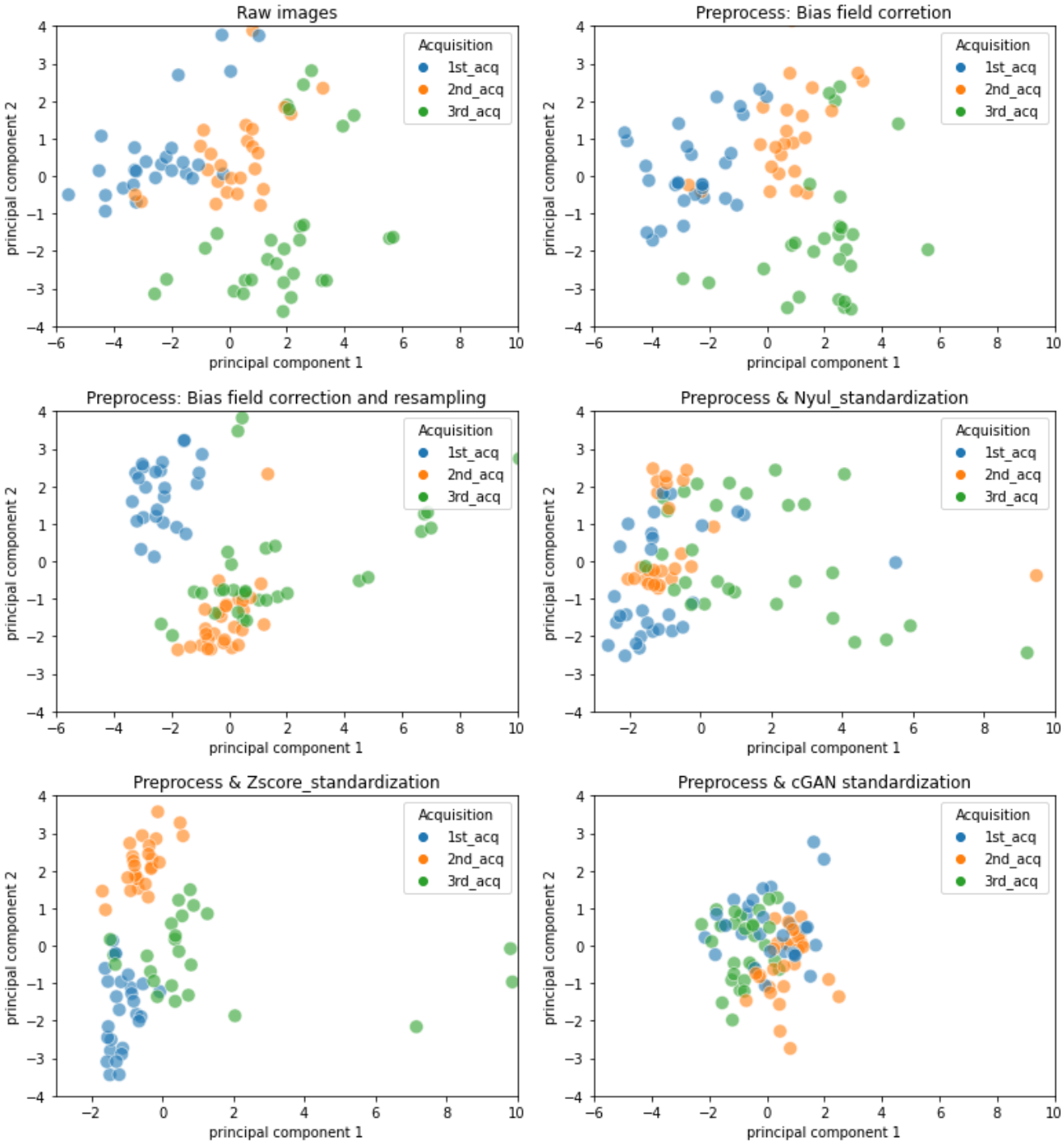


Figure 5.2: Principal component analysis based on the quality metrics extracted from raw images(wo_preprocess), after bias field correction, after preprocess (bias field correction and spatial resampling), after preprocess and Nyul standardization, after preprocess and z-score normalization, and preprocess and cGAN standardization

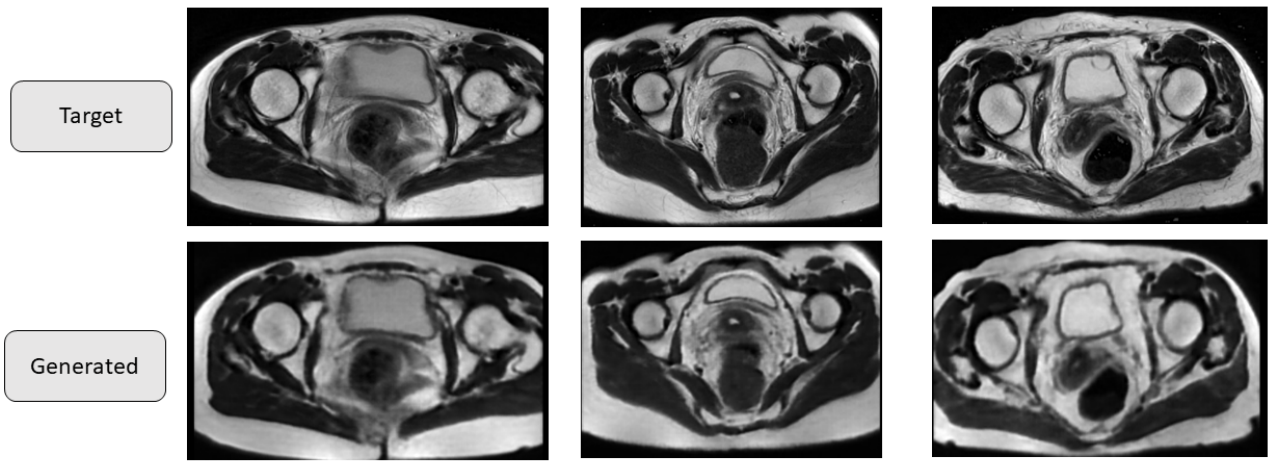


Figure 5.3: Samples of slices generated by the cGAN

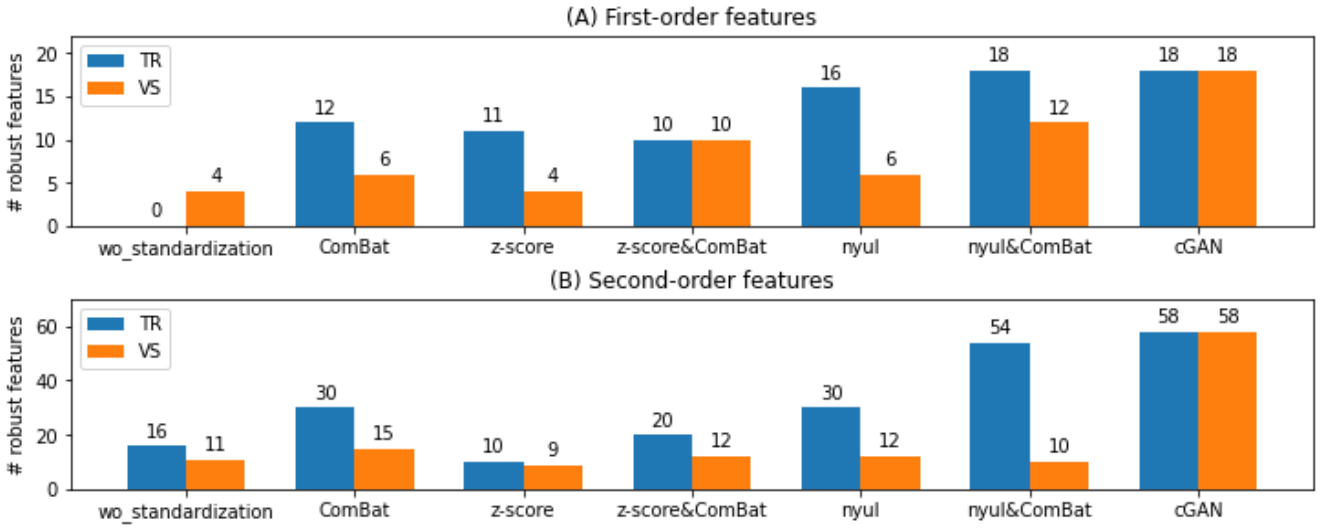


Figure 5.4: Number of robust features before and after standardization: A) first-order radiomic features (n=18), B) Second-order radiomic features (n=75)

				Logistic regression
	wo_standardization	z-score	Nyul	CycleGAN
Train	0.81 ± 0.03	0.88 ± 0.07	0.94 ± 0.05	0.93 ± 0.06
Validation	0.68 ± 0.16	0.78 ± 0.09	0.83 ± 0.07	0.88 ± 0.09
				Random forest
	wo_standardization	z-score	Nyul	CycleGAN
Train	0.90 ± 0.09	0.84 ± 0.08	0.96 ± 0.06	0.97 ± 0.08
Validation	0.62 ± 0.11	0.72 ± 0.13	0.76 ± 0.11	0.77 ± 0.09
				K-nearest neighbors
	wo_standardization	z-score	Nyul	CycleGAN
Train	0.91 ± 0.08	0.92 ± 0.09	0.99 ± 0.09	0.95 ± 0.08
Validation	0.64 ± 0.16	0.66 ± 0.10	0.64 ± 0.09	0.71 ± 0.13
				SVM
	wo_standardization	z-score	Nyul	CycleGAN
Train	0.93 ± 0.08	0.85 ± 0.08	0.94 ± 0.07	0.92 ± 0.06
Validation	0.61 ± 0.18	0.70 ± 0.13	0.72 ± 0.14	0.71 ± 0.09

Table 5.6: Tumor stage classification into low ($< IIa$) and high grade before and after standardization($\geq IIa$). wo: without, SVM: Support-vector machine.

				Logistic regression
	wo_standardization	z-score	Nyul	CycleGAN
Train	0.81 ± 0.13	0.78 ± 0.09	0.79 ± 0.09	0.76 ± 0.06
Validation	0.54 ± 0.11	0.58 ± 0.10	0.63 ± 0.11	0.68 ± 0.14
Test	0.44	0.52	0.56	0.60
				Random forest
	wo_standardization	z-score	Nyul	CycleGAN
Train	0.96 ± 0.07	0.92 ± 0.04	0.95 ± 0.09	0.96 ± 0.07
Validation	0.47 ± 0.09	0.55 ± 0.11	0.53 ± 0.13	0.55 ± 0.10
Test	0.49	0.53	0.55	0.55
				Decision Tree
	wo_standardization	z-score	Nyul	CycleGAN
Train	0.84 ± 0.04	0.87 ± 0.06	0.91 ± 0.09	0.86 ± 0.04
Validation	0.51 ± 0.11	0.57 ± 0.09	0.58 ± 0.10	0.59 ± 0.08
Test	0.47	0.53	0.58	0.57
				SVM
	wo_standardization	z-score	Nyul	CycleGAN
Train	0.94 ± 0.08	0.89 ± 0.02	0.93 ± 0.08	0.89 ± 0.04
Validation	0.49 ± 0.14	0.55 ± 0.11	0.59 ± 0.11	0.57 ± 0.13
Test	0.46	0.52	0.55	0.54

Table 5.7: Relapse prediction on pre-BT MR images before and after standardization. wo: without, SVM: Support-vector machine.

5.4 Discussion

In this study, different standardisation methods applied to MR images obtained from LACC patients were evaluated with different complementary objectives. Firstly, on a prospective cohort, image preprocessing techniques applied before extraction of radiomic features (Nyul and z-score) and the a posteriori ComBat harmonisation method were implemented. In addition, a cGAN-based standardization method was proposed to deal with the variability of TR and VS. The PCA analysis using image quality variables as input demonstrated the very positive impact of applying a GAN strategy for the transition from one domain to another. This positive effect was confirmed quantitatively by the identification of robust radiomic features, both first and second order. In a second step, on two retrospective cohorts of patients treated for LACC, the impact of the image normalization step was evaluated on two clinical tasks. Since in a retrospective scenario, normalization by cGAN is often unrealistic, due to the lack of matched data for training, a normalization method based on CycleGAN was proposed. For this purpose, 86 and 160 images taken before RT and BT, retrospectively, were collected and ML algorithms to classify patients with high and low stage tumours, on the one hand, and to predict relapsing patients, on the other hand, were developed. In both cases, the application of a CycleGAN strategy significantly improves the results compared to conventional approaches used in the literature.

In the considered scenarii, TR had the greatest impact on the raw MR images compared to VS as reflected by the PCA analysis. This effect is, however, more easily overcome regardless of the method of standardization applied. Pre-processing combined with Nyul standardization made cluster distinction less clear than the consecutive application of pre-processing and z-score as seen in Figure, illustrating a beneficial effect of this strategy5.2. This conclusion is further supported by the subsequent analysis of the robustness of the first-order radiomics features based on the ICC and CCC indices (Figure 5.4). For TR modulation, Nyul standardization gave better results than z-score normalization with 16/18 robust features and 18/18 when combined to ComBat harmonization. Similarly, for VS modulation, Nyul combined with ComBat achieved 12/18 of robust features on first-order features while z-score combined with ComBat achieved 10/18. Logically, the change in parameters impacts even more drastically the robustness of 2^{nd}-order features, again with a greater ease of features harmonization in the case of a change in TR than in VS. cGAN-based method increased robust features from 15% to 77% (11/75 to 58/75) in the case of VS modulation while other methods did not exceed 20% (15/75). Considering TR changes, conventional normalization techniques improved the robustness of the features, but their performance was inferior to that of the cGAN-based method, except for the combination of Nyul and ComBat which performed about equally here.

This 2-step analysis allowed us to confirm the interest of the image quality metrics proposed by Sadri et al.[46] in the context of a radiomic evaluation, which have the advantage of being applicable to the image without the need to delineate a volume of interest. If a visual analysis on the first 2 components has only been considered here, it is possible to envisage inter-cluster and intra-cluster distance evaluations to better quantify the harmonization capacities of the different strategies. Different metrics could be considered in this aim, including Euclidean distance or cosine distances. It should be noted that these metrics had already demonstrated their added value for automated batch definition for the implementation of ComBat in highly heterogeneous multicentre brain cancer studies[**carre_autoComBat_2022**].

Given the large number of MRI machines and acquisition parameters present in Cohort 2 (Table 6.2 and 5.3), ComBat harmonization was not suitable since the determination of batches with a minimum of five patients was impossible. Furthermore, the variability in magnetic fields, TR, TE, VS and slice thickness values is such

that ComBat would have been fastidious and time-consuming to implement. As mentioned earlier, Carré et al[**carre_autoComBat_2022**] have developed AutoComBat, which aims to automatically determine the batch labels, using either MRI metadata or MR image quality metrics. While AutoComBat could have been useful in this configuration as it does not require a statistically representative sample of each batch, the method deployed here has the advantage of being almost instantaneous to apply at the inference stage. In addition to providing very promising results for robust radiomic studies, adversarial networks could be used for multiple and varied tasks in combination with augmentation and normalization techniques with complementary benefits[51]. As GANs have been used to generate complementary sequences from a single MRI sequence [52, 53], thereby reducing acquisition times and increasing patient comfort, CycleGAN could be used to move from one image quality space to another, by changing target domain selection. This generation of MR images with different properties in terms of noise, contrast and spatial resolution, would give physicians access to additional information that could help to improve the accuracy of manual and automatic segmentation in the context of radiation therapy. For example, Conte et al. [54] implemented two GANs to generate missing T1-weighted images from post-contrast T1w images and FLAIR T2 images from T2w images. They demonstrated that the generated images could be used as inputs for automatic segmentation in case of missing MRI sequences. This strategy could also be explored for multi-modality radiomic studies, as it has been found that radiomic images from different modalities can improve model prediction [17]. Access to images with different properties could also be an alternative to the application of filters, i.e. extraction of higher-order radiomic features, used in a large number of studies. In all these virtual image generation strategies however, care must be taken with the generated images, which may be polluted by unrealistic artifacts. Advances in the DL field such as style GANs[55] and their derivates should also be investigated for multi-centric radiomic studies.

On a similar topic, Andrearczyk et al.[56] showed that DL could be applied to normalize hand-crafted and deep features extracted from images and could be generalized to features coming from unknown scanners. Their approach was a posteriori standardization comparable to ComBat harmonization. Wei et al.[57] used a 3D GAN to normalize CT images obtained from different slice thickness and dosage scenarios. Their method used a specific spectral-norm layer to perform global regularization in the feature matrices. The GAN-based approach improved CT scans' perceptual appearance and reduced the variability across radiomic features. Despite being denoising and super-resolution simultaneously, this approach was not evaluated on a clinical task. Similarly, Chen et al.[36] have applied GAN and CycleGAN on simulated and real CT images to improve radiomic features reproducibility in simulated low-dose CT images but also their performance in survival prediction. For both simulated and real CT images, the trained CycleGANs increased the CCC of radiomic features and the AUC of survival prediction. Marcadent et al.[58] developed a CycleGAN to translate texture information from chest radiographs acquired using one manufacturer to another. This CycleGAN approach showed decreased inter-manufacturer radiomic features variability and experienced radiologists were not able to identify generated radiographs. Despite these promising results, vanishing gradient renders GAN training time-consuming and data-hungry. Similarly to our approach, Du et al[39] developed a CNN-based harmonization method for CT images, and found their method to outperform ComBat, centering-scaling method, matrix factorization harmonization methods, and independent component analysis for lymph node metastasis (LNM) prediction and classification between lung cancer and pulmonary tuberculosis. They also showed that CNN harmonization methods improve feature reproducibility not only between specific kernels from the same scanner but also between unobserved kernels from different scanners of different vendors. Similarly to our study, Chang et al.[59] developed a cycleGAN coupled to a histogram matching module for female pelvic T2w MR image standardization. The

proposed method overperformed ComBat and contrast limited adaptive histogram equalization (CLAHE) in reducing variabilities in radiomic features and in predicting LNM in LACC (AUC of proposed method: 0.89, ComBat: 0.685, and CLAHE: 0.770).

On 86 pre-RT T2w MR images of LACC patients, radiomic features were extracted before and after image standardization in order to classify tumor patients into low and high stages. For all ML algorithms considered, the ROC-AUC increased after image standardization. Across standardization methods, the CycleGAN-based method gave the best results with a ROC-AUC of 0.88 ± 0.09 for the logistic regression (Table 5.6). Staging of cervical cancer is performed on MRI by radiologists in a first attempt. However, histopathological examination is usually needed as definitive diagnosis. Early-stage cervical cancer are usually treated with hysterectomy or cervical resection, while high stages are treated by chemoradiotherapy (CRT). Staging based on MR images alone would eliminate the need for invasive procedures to confirm the stage of the tumor and, above all, would drastically speed up patient management. For such a strategy to be considered in clinical routine, however, it is important that the performance of the models is perfect or at least similar to the current classification reliability. On 160 pre-BT T2w MR images, we also extracted radiomic features to predict relapse. For this task, a binary classification of the relapse event was considered. This assumption is meaningful in LACC as any recurrence (local, regional or distant) was considered as relapse. Without any standardization, the classification performances were below 0.50 for all ML algorithms on the test set but were increased after standardization, although the final results remain rather poor (Table5.7).

This study has several limitations. The analysis on the prospective cohort was done on the cervix and not the tumor, due to the heterogeneity of the patients recruited, as the images were acquired at the diagnostic, radiotherapy treatment preparation, or follow-up stage. Further work is also needed to identify the limits of application of this generative neural network-based methodology to bridge the differences in acquisition properties, i.e. whether there are too extreme configurations for which the model is not applicable [60]. While the issue of differences in scanner model, slice thickness, magnetic field and TE has been indirectly assessed, a larger prospective study should also be implemented to collect larger data sets for proper validation of these models The number of patients in the pre-RT cohort did not allow us to split the data into training and testing cohorts. An external test set is needed to validate our findings. Furthermore, pre-BT images were taken on the same machine and despite the large variability of acquisition parameters, the test set used to evaluate CycleGAN standardization shared acquisition parameters with data from the training set. Some of the limitations are also inherent to generative networks i.e. evaluation metrics of the generated images. Here, radiomic features were extracted and robustness was assessed using ICC and CCC. The cGAN/CycleGAN-based normalization methods presented in this study only allows translation of an image from domain X to domain Y in order to get rid of the limitations due to the differences in acquisition parameters but are in no way a method to improve the quality of native images. Thus, artifacts present in the input images will also be present in the generated images. In this study, only one loss function for the discriminator (BCE) and two for the generator (BCE and L1 loss) have been implemented. In particular, a loss function to compare the similarity of structures, such as structural similarity loss, could increase the structural correspondence between the generated and the source images. Finally, different neural network architectures should have been compared. However, the objective of this paper was to demonstrate the feasibility and superiority of a neural network-based approach for MRI image standardization and this was outside the scope of the study.

In conclusion, we have shown that GAN-based normalization method could better limit the impact of acquisition parameters on MRI radiomics features compared to conventional methods, and increase the performances

of predictive signatures in LACC. Considering that DL is becoming popular in computer vision and medical imaging, one would argue that adversarial networks could present an "all in one" solution to several problems, from image standardization to treatment outcome prediction (survival, relapse, complication after treatment, etc).

Bibliography

[1] Corinne Balleyguier et al. "Staging of uterine cervical cancer with MRI: guidelines of the European Society of Urogenital Radiology". eng. In: *Eur Radiol* 21.5 (May 2011), pp. 1102–1110. ISSN: 1432-1084. DOI: 10.1007/s00330-010-1998-x.

[2] Kari Tanderup et al. "From point A to the sculpted pear: MR image guidance significantly improves tumour dose and sparing of organs at risk in brachytherapy of cervical cancer". eng. In: *Radiother Oncol* 94.2 (Feb. 2010), pp. 173–180. ISSN: 1879-0887. DOI: 10.1016/j.radonc.2010.01.001.

[3] Emma C. Fields and Elisabeth Weiss. "A practical review of magnetic resonance imaging for the evaluation and management of cervical cancer". In: *Radiat Oncol* 11 (Feb. 2016), p. 15. ISSN: 1748-717X. DOI: 10.1186/s13014-016-0591-0. URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4736634/ (visited on 10/10/2022).

[4] Evis Sala et al. "The Added Role of MR Imaging in Treatment Stratification of Patients with Gynecologic Malignancies: What the Radiologist Needs to Know". In: *Radiology* 266.3 (Mar. 2013). Publisher: Radiological Society of North America, pp. 717–740. ISSN: 0033-8419. DOI: 10.1148/radiol.12120315. URL: https://pubs.rsna.org/doi/10.1148/radiol.12120315 (visited on 10/10/2022).

[5] Alina Sturdza et al. "Image guided brachytherapy in locally advanced cervical cancer: Improved pelvic control and survival in RetroEMBRACE, a multicenter cohort study". eng. In: *Radiother Oncol* 120.3 (Sept. 2016), pp. 428–433. ISSN: 1879-0887. DOI: 10.1016/j.radonc.2016.03.011.

[6] Richard Pötter et al. "MRI-guided adaptive brachytherapy in locally advanced cervical cancer (EMBRACE-I): a multicentre prospective cohort study". English. In: *The Lancet Oncology* 22.4 (Apr. 2021). Publisher: Elsevier, pp. 538–547. ISSN: 1470-2045, 1474-5488. DOI: 10.1016/S1470-2045(20)30753-1. URL: https://www.thelancet.com/journals/lanonc/article/PIIS1470-2045(20)30753-1/fulltext (visited on 03/16/2022).

[7] Richard Pötter et al. "Recommendations from gynaecological (GYN) GEC ESTRO working group (II): Concepts and terms in 3D image-based treatment planning in cervix cancer brachytherapy—3D dose volume parameters and aspects of 3D image-based anatomy, radiation physics, radiobiology". en. In: *Radiotherapy and Oncology* 78.1 (Jan. 2006), pp. 67–

77. ISSN: 0167-8140. DOI: 10.1016/j.radonc.2005.11.014. URL: https://www.sciencedirect.com/science/article/pii/S0167814005005463 (visited on 03/16/2022).

[8] Kasia Owczarczyk et al. "MRI heterogeneity analysis for prediction of recurrence and disease free survival in anal cancer". English. In: *Radiotherapy and Oncology* 134 (May 2019). Publisher: Elsevier, pp. 119–126. ISSN: 0167-8140, 1879-0887. DOI: 10.1016/j.radonc.2019.01.022. URL: https://www.thegreenjournal.com/article/S0167-8140(19)30027-1/fulltext#%20 (visited on 04/22/2022).

[9] François Lucia et al. "External validation of a combined PET and MRI radiomics model for prediction of recurrence in cervical cancer patients treated with chemoradiotherapy". eng. In: *Eur J Nucl Med Mol Imaging* 46.4 (Apr. 2019), pp. 864–877. ISSN: 1619-7089. DOI: 10.1007/s00259-018-4231-9.

[10] Meiling Xiao et al. "Multiparametric MRI-Based Radiomics Nomogram for Predicting Lymph Node Metastasis in Early-Stage Cervical Cancer". en. In: *J Magn Reson Imaging* (Feb. 2020), jmri.27101. ISSN: 1053-1807, 1522-2586. DOI: 10.1002/jmri.27101. URL: https://onlinelibrary.wiley.com/doi/abs/10.1002/jmri.27101 (visited on 03/19/2020).

[11] Hitoshi Ikushima et al. "Prediction of out-of-field recurrence after chemoradiotherapy for cervical cancer using a combination model of clinical parameters and magnetic resonance imaging radiomics: a multi-institutional study of the Japanese Radiation Oncology Study Group". In: *Journal of Radiation Research* 63.1 (Jan. 2022), pp. 98–106. ISSN: 1349-9157. DOI: 10.1093/jrr/rrab104. URL: https://doi.org/10.1093/jrr/rrab104 (visited on 04/22/2022).

[12] Russell T. Shinohara et al. "Statistical normalization techniques for magnetic resonance imaging". en. In: *NeuroImage: Clinical* 6 (Jan. 2014), pp. 9–19. ISSN: 2213-1582. DOI: guo. URL: https://www.sciencedirect.com/science/article/pii/S221315821400117X (visited on 10/13/2022).

[13] P. Tsialios et al. "Evaluation of MRI sequences for quantitative T1 brain mapping". en. In: *J. Phys.: Conf. Ser.* 931.1 (Nov. 2017). Publisher: IOP Publishing, p. 012038. ISSN: 1742-6596. DOI: 10.1088/1742-6596/931/1/012038. URL: https://dx.doi.org/10.1088/1742-6596/931/1/012038 (visited on 03/27/2023).

[14] Alexander Seiler et al. "Multiparametric Quantitative MRI in Neurological Diseases". In: *Frontiers in Neurology* 12 (2021). ISSN: 1664-2295. URL: https://www.frontiersin.org/articles/10.3389/fneur.2021.640239 (visited on 03/27/2023).

[15] Alexandre Carré et al. "Standardization of brain MR images across machines and protocols: bridging the gap for MRI-based radiomics". en. In: *Sci Rep* 10.1 (July 2020). Bandiera_abtest: a Cc_license_type: cc_by Cg_type: Nature Research Journals Number: 1 Primary_atype: Research Publisher: Nature Publishing Group Subject_term: Cancer

imaging;CNS cancer;Computational science;Diagnostic markers Subject_term_id: cancer-imaging;cns-cancer;computational-science;diagnostic-markers, p. 12340. ISSN: 2045-2322. DOI: 10.1038/s41598-020-69298-z. URL: https://www.nature.com/articles/s41598-020-69298-z (visited on 09/20/2021).

[16] Alex Zwanenburg et al. "The Image Biomarker Standardization Initiative: Standardized Quantitative Radiomics for High-Throughput Image-based Phenotyping". eng. In: *Radiology* 295.2 (May 2020), pp. 328–338. ISSN: 1527-1315. DOI: 10.1148/radiol.2020191145.

[17] R. Da-Ano et al. "Performance comparison of modified ComBat for harmonization of radiomic features for multicenter studies". eng. In: *Sci Rep* 10.1 (June 2020), p. 10248. ISSN: 2045-2322. DOI: 10.1038/s41598-020-66110-w.

[18] Philippe Lambin et al. "Radiomics: extracting more information from medical images using advanced feature analysis". eng. In: *Eur J Cancer* 48.4 (Mar. 2012), pp. 441–446. ISSN: 1879-0852. DOI: 10.1016/j.ejca.2011.11.036.

[19] Hugo J. W. L. Aerts et al. "Decoding tumour phenotype by noninvasive imaging using a quantitative radiomics approach". en. In: *Nat Commun* 5.1 (June 2014). Bandiera_abtest: a Cg_type: Nature Research Journals Number: 1 Primary_atype: Research Publisher: Nature Publishing Group Subject_term: Cancer imaging;Medical research Subject_term_id: cancer-imaging;medical-research, p. 4006. ISSN: 2041-1723. DOI: 10.1038/ncomms5006. URL: http://www.nature.com/articles/ncomms5006 (visited on 09/20/2021).

[20] L. G. Nyúl, J. K. Udupa, and X. Zhang. "New variants of a method of MRI scale standardization". eng. In: *IEEE Trans Med Imaging* 19.2 (Feb. 2000), pp. 143–150. ISSN: 0278-0062. DOI: 10.1109/42.836373.

[21] Mohak Shah et al. "Evaluating intensity normalization on MRIs of human brain with multiple sclerosis". eng. In: *Med Image Anal* 15.2 (Apr. 2011), pp. 267–282. ISSN: 1361-8423. DOI: 10.1016/j.media.2010.12.003.

[22] W. Evan Johnson, Cheng Li, and Ariel Rabinovic. "Adjusting batch effects in microarray expression data using empirical Bayes methods". In: *Biostatistics* 8.1 (Jan. 2007), pp. 118–127. ISSN: 1465-4644. DOI: 10.1093/biostatistics/kxj037. URL: https://doi.org/10.1093/biostatistics/kxj037 (visited on 09/20/2021).

[23] M. Boulanger et al. "Deep learning methods to generate synthetic CT from MRI in radiotherapy: A literature review". en. In: *Physica Medica* 89 (Sept. 2021), pp. 265–281. ISSN: 1120-1797. DOI: 10.1016/j.ejmp.2021.07.027. URL: https://www.sciencedirect.com/science/article/pii/S1120179721002714 (visited on 04/14/2023).

[24] Cam Bermudez et al. "Learning Implicit Brain MRI Manifolds with Deep Learning". In: (Jan. 2018).

[25] Dong Nie et al. “Medical Image Synthesis with Context-Aware Generative Adversarial Networks”. In: *Med Image Comput Comput Assist Interv* 10435 (Sept. 2017), pp. 417–425. DOI: 10.1007/978-3-319-66179-7_48. URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6044459/ (visited on 11/04/2022).

[26] Qianqian Zhang et al. “Medical Image Synthesis with Generative Adversarial Networks for Tissue Recognition”. In: June 2018. DOI: 10.1109/ICHI.2018.00030.

[27] Maayan Frid-Adar et al. “Synthetic data augmentation using GAN for improved liver lesion classification”. In: Apr. 2018, pp. 289–293. DOI: 10.1109/ISBI.2018.8363576.

[28] Jelmer M. Wolterink et al. “Deep MR to CT Synthesis Using Unpaired Data”. en. In: *Simulation and Synthesis in Medical Imaging*. Ed. by Sotirios A. Tsaftaris et al. Lecture Notes in Computer Science. Cham: Springer International Publishing, 2017, pp. 14–23. ISBN: 978-3-319-68127-6. DOI: 10.1007/978-3-319-68127-6_2.

[29] Ian Goodfellow et al. “Generative adversarial networks”. In: *Commun. ACM* 63.11 (Oct. 2020), pp. 139–144. ISSN: 0001-0782. DOI: 10.1145/3422622. URL: https://doi.org/10.1145/3422622 (visited on 11/04/2022).

[30] Jun-Yan Zhu et al. “Unpaired Image-to-Image Translation Using Cycle-Consistent Adversarial Networks”. In: *2017 IEEE International Conference on Computer Vision (ICCV)*. ISSN: 2380-7504. Oct. 2017, pp. 2242–2251. DOI: 10.1109/ICCV.2017.244.

[31] Hidetoshi Matsuo et al. “Unsupervised-learning-based method for chest MRI–CT transformation using structure constrained unsupervised generative attention networks”. en. In: *Sci Rep* 12.1 (June 2022). Number: 1 Publisher: Nature Publishing Group, p. 11090. ISSN: 2045-2322. DOI: 10.1038/s41598-022-14677-x. URL: https://www.nature.com/articles/s41598-022-14677-x (visited on 11/09/2022).

[32] Yuan Gao et al. “A Universal Intensity Standardization Method Based on a Many-to-One Weak-Paired Cycle Generative Adversarial Network for Magnetic Resonance Images”. In: *IEEE Transactions on Medical Imaging* 38.9 (Sept. 2019). Conference Name: IEEE Transactions on Medical Imaging, pp. 2059–2069. ISSN: 1558-254X. DOI: 10.1109/TMI.2019.2894692.

[33] Yae Won Park et al. “Cycle-consistent adversarial networks improves generalizability of radiomics model in grading meningiomas on external validation”. en. In: *Sci Rep* 12.1 (Apr. 2022). Number: 1 Publisher: Nature Publishing Group, p. 7042. ISSN: 2045-2322. DOI: 10.1038/s41598-022-10956-9. URL: https://www.nature.com/articles/s41598-022-10956-9 (visited on 01/18/2023).

[34] Gourav Modanwal et al. “MRI image harmonization using cycle-consistent generative adversarial network”. In: *Medical Imaging 2020: Computer-Aided Diagnosis*. Vol. 11314. SPIE, Mar. 2020, pp. 259–264. DOI: 10.1117/12.2551301. URL: https://www.spiedigitallibrary.org/conference-proceedings-of-spie/11314/1131413/MRI-image-harmonization-

using-cycle-consistent-generative-adversarial-network/10.1117/12.2551301.full (visited on 01/18/2023).

[35] Yang Lei et al. "MRI-Only Based Synthetic CT Generation Using Dense Cycle Consistent Generative Adversarial Networks". en. In: *Medical physics* 46.8 (Aug. 2019). Publisher: NIH Public Access, p. 3565. DOI: 10.1002/mp.13617. URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6692192/ (visited on 11/09/2022).

[36] Junhua Chen et al. "Improving reproducibility and performance of radiomics in low-dose CT using cycle GANs". eng. In: *J Appl Clin Med Phys* 23.10 (Oct. 2022), e13739. ISSN: 1526-9914. DOI: 10.1002/acm2.13739.

[37] M. Tarek Shaban et al. *StainGAN: Stain Style Transfer for Digital Histological Images*. arXiv:1804.01601 [cs]. Apr. 2018. DOI: 10.48550/arXiv.1804.01601. URL: http://arxiv.org/abs/1804.01601 (visited on 01/18/2023).

[38] Seo Jeong Shin et al. "Style transfer strategy for developing a generalizable deep learning application in digital pathology". eng. In: *Comput Methods Programs Biomed* 198 (Jan. 2021), p. 105815. ISSN: 1872-7565. DOI: 10.1016/j.cmpb.2020.105815.

[39] Dongyang Du et al. "Deep learning-based harmonization of CT reconstruction kernels towards improved clinical task performance". eng. In: *Eur Radiol* 33.4 (Apr. 2023), pp. 2426–2438. ISSN: 1432-1084. DOI: 10.1007/s00330-022-09229-w.

[40] Nicholas J. Tustison et al. "N4ITK: Improved N3 Bias Correction". In: *IEEE Trans Med Imaging* 29.6 (June 2010), pp. 1310–1320. ISSN: 0278-0062. DOI: 10.1109/TMI.2010.2046908. URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3071855/ (visited on 10/13/2022).

[41] *ANTs by stnava*. URL: http://stnava.github.io/ANTs/ (visited on 08/29/2022).

[42] Phillip Isola et al. *Image-to-Image Translation with Conditional Adversarial Networks*. arXiv:1611.07004 [cs]. Nov. 2018. DOI: 10.48550/arXiv.1611.07004. URL: http://arxiv.org/abs/1611.07004 (visited on 11/04/2022).

[43] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. *U-Net: Convolutional Networks for Biomedical Image Segmentation*. arXiv:1505.04597 [cs]. May 2015. DOI: 10.48550/arXiv.1505.04597. URL: http://arxiv.org/abs/1505.04597 (visited on 01/18/2023).

[44] Jean-Philippe Fortin et al. "Harmonization of multi-site diffusion tensor imaging data". eng. In: *Neuroimage* 161 (Nov. 2017), pp. 149–170. ISSN: 1095-9572. DOI: 10.1016/j.neuroimage.2017.08.047.

[45] Jean-Philippe Fortin et al. "Harmonization of cortical thickness measurements across scanners and sites". en. In: *NeuroImage* 167 (Feb. 2018), pp. 104–120. ISSN: 1053-8119. DOI: 10.1016/j.neuroimage.2017.11.024. URL: https://www.sciencedirect.com/science/article/pii/S105381191730931X (visited on 03/22/2022).

[46] Amir Reza Sadri et al. “Technical Note: MRQy - An open-source tool for quality control of MR imaging data”. eng. In: *Med Phys* 47.12 (Dec. 2020), pp. 6029–6038. ISSN: 2473-4209. DOI: 10.1002/mp.14593.

[47] Joost J. M. van Griethuysen et al. “Computational Radiomics System to Decode the Radiographic Phenotype”. eng. In: *Cancer Res* 77.21 (Nov. 2017), e104–e107. ISSN: 1538-7445. DOI: 10.1158/0008-5472.CAN-17-0339.

[48] James D. Brierley, Mary K. Gospodarowicz, and Christian Wittekind. *TNM Classification of Malignant Tumours*. en. Google-Books-ID: 642GDQAAQBAJ. John Wiley & Sons, Jan. 2017. ISBN: 978-1-119-26357-9.

[49] N. V. Chawla et al. “SMOTE: Synthetic Minority Over-sampling Technique”. In: *jair* 16 (June 2002). arXiv:1106.1813 [cs], pp. 321–357. ISSN: 1076-9757. DOI: 10.1613/jair.953. URL: http://arxiv.org/abs/1106.1813 (visited on 01/18/2023).

[50] David J. Hand and Robert J. Till. “A Simple Generalisation of the Area Under the ROC Curve for Multiple Class Classification Problems”. en. In: *Machine Learning* 45.2 (Nov. 2001), pp. 171–186. ISSN: 1573-0565. DOI: 10.1023/A:1010920819831. URL: https://doi.org/10.1023/A:1010920819831 (visited on 01/18/2023).

[51] M. Jorge Cardoso et al., eds. *Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support: Third International Workshop, DLMIA 2017, and 7th International Workshop, ML-CDS 2017, Held in Conjunction with MICCAI 2017, Québec City, QC, Canada, September 14, Proceedings*. en. Vol. 10553. Lecture Notes in Computer Science. Cham: Springer International Publishing, 2017. ISBN: 978-3-319-67557-2 978-3-319-67558-9. DOI: 10.1007/978-3-319-67558-9. URL: http://link.springer.com/10.1007/978-3-319-67558-9 (visited on 04/14/2023).

[52] Daisuke Kawahara and Yasushi Nagata. “T1-weighted and T2-weighted MRI image synthesis with convolutional generative adversarial networks”. In: *Rep Pract Oncol Radiother* 26.1 (Feb. 2021), pp. 35–42. ISSN: 1507-1367. DOI: 10.5603/RPOR.a2021.0005. URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC8086713/ (visited on 04/14/2023).

[53] Liliana Valencia et al. “Evaluating the use of synthetic T1-w images in new T2 lesion detection in multiple sclerosis”. In: *Frontiers in Neuroscience* 16 (2022). ISSN: 1662-453X. URL: https://www.frontiersin.org/articles/10.3389/fnins.2022.954662 (visited on 04/14/2023).

[54] Gian Marco Conte et al. “Generative Adversarial Networks to Synthesize Missing T1 and FLAIR MRI Sequences for Use in a Multisequence Brain Tumor Segmentation Model”. In: *Radiology* 299.2 (May 2021). Publisher: Radiological Society of North America, pp. 313–323. ISSN: 0033-8419. DOI: 10.1148/radiol.2021203786. URL: https://pubs.rsna.org/doi/full/10.1148/radiol.2021203786 (visited on 04/14/2023).

[55] Tero Karras, Samuli Laine, and Timo Aila. *A Style-Based Generator Architecture for Generative Adversarial Networks*. arXiv:1812.04948 [cs, stat]. Mar. 2019. DOI: 10.48550/arXiv.1812.04948. URL: http://arxiv.org/abs/1812.04948 (visited on 01/27/2023).

[56] Vincent Andrearczyk, Adrien Depeursinge, and Henning Müller. "Neural network training for cross-protocol radiomic feature standardization in computed tomography". In: *JMI* 6.2 (June 2019). Publisher: SPIE, p. 024008. ISSN: 2329-4302, 2329-4310. DOI: 10.1117/1.JMI.6.2.024008. URL: https://www.spiedigitallibrary.org/journals/journal-of-medical-imaging/volume-6/issue-2/024008/Neural-network-training-for-cross-protocol-radiomic-feature-standardization-in/10.1117/1.JMI.6.2.024008.full (visited on 01/18/2023).

[57] Leihao Wei, Yannan Lin, and William Hsu. "Using a Generative Adversarial Network for CT Normalization and Its Impact on Radiomic Features". In: *2020 IEEE 17th International Symposium on Biomedical Imaging (ISBI)*. ISSN: 1945-8452. Apr. 2020, pp. 844–848. DOI: 10.1109/ISBI45749.2020.9098724.

[58] Sandra Marcadent et al. "Generative Adversarial Networks Improve the Reproducibility and Discriminative Power of Radiomic Features". In: *Radiol Artif Intell* 2.3 (May 2020), e190035. ISSN: 2638-6100. DOI: 10.1148/ryai.2020190035. URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC8082326/ (visited on 01/27/2023).

[59] Xiao Chang et al. "Self-supervised learning for multi-center magnetic resonance imaging harmonization without traveling phantoms". eng. In: *Phys Med Biol* 67.14 (July 2022). ISSN: 1361-6560. DOI: 10.1088/1361-6560/ac7b66.

[60] Shruti Atul Mali et al. "Making Radiomics More Reproducible across Scanner and Imaging Protocol Variations: A Review of Harmonization Methods". en. In: *Journal of Personalized Medicine* 11.9 (Sept. 2021). Number: 9 Publisher: Multidisciplinary Digital Publishing Institute, p. 842. ISSN: 2075-4426. DOI: 10.3390/jpm11090842. URL: https://www.mdpi.com/2075-4426/11/9/842 (visited on 01/18/2023).

Chapter 6

Radiomic signature of lymphocyte infiltrate in LACC treated with Radiotherapy/Immunotherapy

6.1 Stimulating the immune system for cancer treatment

"Drugs can only repress symptoms: they cannot eradicate disease. The true remedy for all diseases is Nature's remedy. There is at bottom only one genuinely scientific treatment for all diseases, and that is to stimulate the phagocytes. Stimulate the phagocytes. Drugs are a delusion" as it can be read in George Bernard Shaw's 1906 play The Doctor's Dilemma. Immunotherapy as the basis for a new paradigm in cancer treatment was born by the end of the 19th century when Dr. William Coley decided to inject preparations of streptococci directly into the tumors of patients with sarcomas after having observed a correlation between the regression of this type of tumor and acute infection with Streptococcus pyogenes[1]. Following the work of Ehrlich in 1909, later taken up in 1957 by Lewis Thomas and Frank McFarlane Burnet, the hypothesis that the immune response after infection could prevent the development of cancers was put forward and the concept of immunosurveillance was developed according to which the immune system could control, detect and eliminate tumor cells. Later on, after a century of controversy, an overwhelming amount of data from animal models, together with compelling data from human patients, indicate that a functional cancer immunosurveillance process indeed exists that acts as an extrinsic tumor suppressor. However, it has also become clear that the immune system can facilitate tumor progression, at least in part, by sculpting the immunogenic phenotype of tumors as they develop. The recognition that immunity plays a dual role in the complex interactions between tumors and the host prompted a refinement of the cancer immunosurveillance hypothesis into one termed "cancer immunoediting." [2] The concept of cancer immunoediting was defined using the 3Es, i.e, elimination, equilibrium, and escape. During the elimination phase, the immune system detects and destroys tumor cells by the immune system. The equilibrium phase corresponds to the control of the proliferation of tumor cells without eradicating them. During this phase, the immune system exerts a selective pressure leading to clones that are not very immunogenic: this is the immunoediting process. Finally, the tumor cells hijack the immune system to promote their development: this is the escape phase.

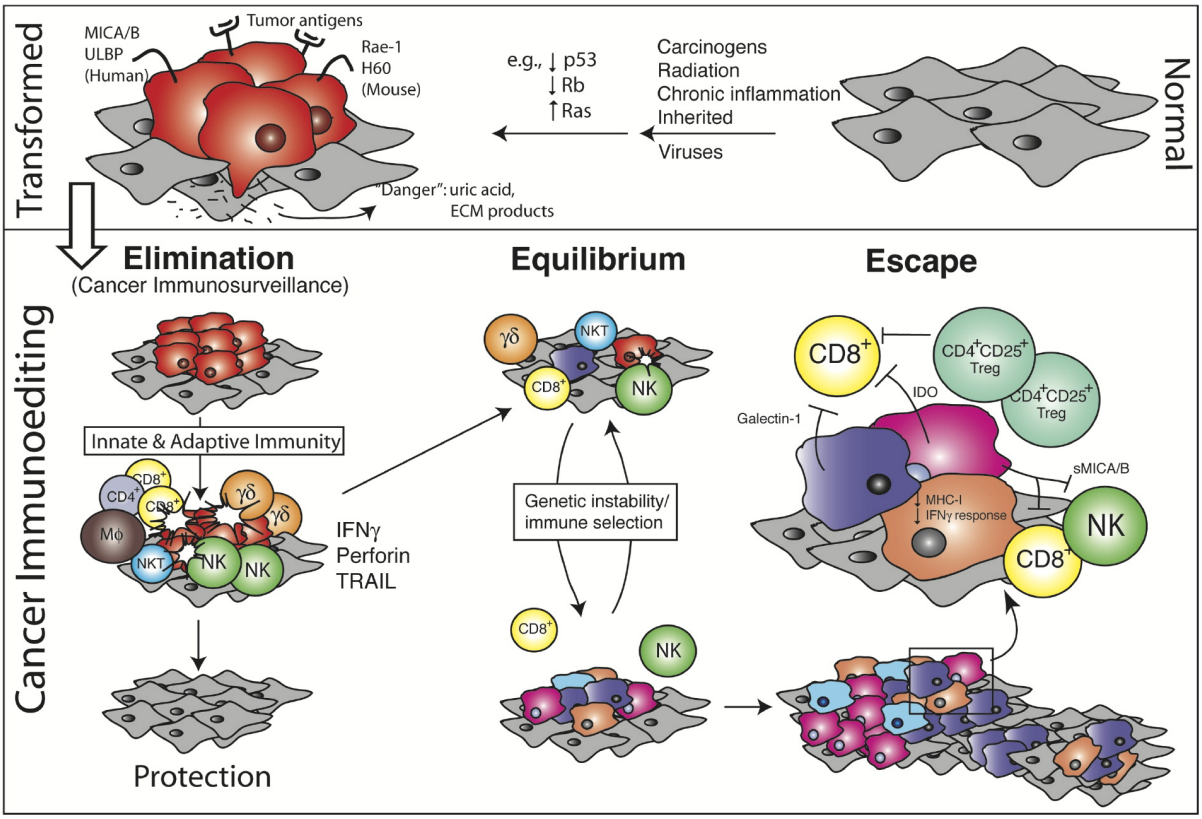


Figure 6.1: Cancer Immunoediting Process. In the elimination process, normal cells (here in gray) subject to common oncogenic stimuli transform into tumor cells (here in red). These cells may express distinct tumor-specific markers and generate proinflammatory signals that trigger the cancer immunoediting process (bottom). In this phase, cells and molecules of innate and adaptive immunity, which comprise the cancer immunosurveillance network, may eradicate the developing tumor and protect the host from tumor formation. However, if this process is not successful, the equilibrium phase can be initiated. Tumor cells may be either maintained chronically or immunologically sculpted by immune "editors" to produce new populations of tumor variants. These variants may eventually evade the immune system by a variety of mechanisms and become clinically detectable in the escape phase. *Reprinted from [3] under the Creative Commons Attribution-NonCommercial 4.0 International.*

In the presence of cancer cells, the immune system might act as sentinel in order to identify and possibly eliminate somatic cells transformed by mutations, yet, cancer cells can escape identification, therefore destruction by the immune system. This provided a strong rationale for immunotherapy. Efforts to do so and to stimulate the immune response were made for treatment of cancer. T-cells are among the main cellular actors involved in the anti-tumor response through the specific recognition of their receptor, T-cell receptor (TCR) present in the major histocompatibility complex (MHC) of the tumor peptide. For a complete activation of the T-cell, the latter must also integrate activating or inhibiting signals from co-stimulatory molecules. Called immune checkpoints (ICs), these co-stimulatory molecules are receptors and ligands located respectively on the surface of lymphocytes and on the surface of antigen-presenting cells and tumor cells. ICs play a key role in limiting the immune response and inducing autoimmunity. Several ICs have been identified such as: cytotoxic T-lymphocyte-associated protein 4 (CTLA-4), programed death 1 (PD-1), programed death-ligand 1 et 2 (PD-L1 et PD-L2) [4, 5, 6]. The CTLA-4 protein, expressed by effector T-cells (Teffs) and regulatory T-cells (Tregs), is an inhibitory

co-receptor that has the same ligands, B7-1 and B7-2 also called CD80 and CD86, as the activating co-receptor CD28. It is involved early in the immune response, on naive T-cells in secondary lymphoid organs. PD-1 is expressed by T-cells and Tregs. It is involved later in the immune response by inhibiting the activation of T-cells during the effector phase in tissues and in the tumor microenvironment. PD-1, by binding to PD-L1, a protein found on some normal (and cancer) cells, prevents T-cells from attacking other cells in the body. Some cancer cells have large amounts of PD-L1, which allows them to escape from the immune response. The development of monoclonal antibodies targeting these ICs, the immune checkpoint inhibitors (ICIs), in the treatment of certain cancers, has shown impressive results [7, 8, 9, 10]. The efficacy of ICIs relies on the release of T-cells that are no longer subject to the inhibitory effects of the ICs. Thus, in order for T-cells to be effective, it is essential that the therapeutic targets, CTLA-4, PD-1 and/or PD-L1, be present to allow the immune system to regain sustained anti-tumor activity.

Immunotherapy has shown success in several types of cancers including lung cancer, head and neck cancer, bladder cancer, kidney cancer, and Hodgkin lymphoma[8, 7, 9, 10, 11, 12]. Immunotherapies based on anti-PD-1 and anti-PD-L1 have shown promising results in long-term cancer remissions. However, these drugs work better in some cancers than others. Moreover, only a fraction of patients with cancer responds to immunotherapy, and currently, available immunotherapeutic agents are expensive and generally associated with considerable toxicities, calling for the identification of robust predictive biomarkers. This calls for the development of regimens that combine various interventions. RT stands out as a particularly promising candidate in this setting, not only because of its established safety profile but also because radiotherapy has the potential ability to mediate robust immunostimulatory effects that could synergize with immunotherapy in systemic tumor control[13]. Ionizing radiation can cause direct damage to cancer cells, leading to cell death, but also trigger CD8+ T-cell activation. Numerous preclinical studies have suggested that combining immunotherapy with radiotherapy could be a promising strategy for synergistic enhancement of treatment efficacy[8, 14, 15, 13]. Preclinical studies demonstrated that the combination of RT and targeted PD-1/PD-L1 therapy activated cytotoxic T-cells, reduced myeloid-derived suppressor cells, and induced an abscopal response[16, 17]. Numerous clinical trials are in progress exploring the combination of standard, full-dose (chemo)radiotherapy plus immunotherapy[18]. Several ongoing clinical trials have assessed and confirmed the benefit of the combination of PD-1/PD-L1 inhibition and RT in different localizations[19, 20]. Recent results of the phase III randomized trial (PACIFIC, NCT02125461) testing the role of the PD-L1 antibody durvalumab versus placebo as consolidation therapy after chemoradiation for stage III non-small cell lung cancer (NSCLC) demonstrated substantial improvement in progression-free survival (PFS) with durvalumab (16.8 months versus 5.6 months with placebo), with similar types and severity of side effects [21, 22]. However, it is important to note that clinical trials have also reported negative results of immunotherapy. combined to other standard-of-care or that some immunotherapy regimens seem to work better than others. In a phase II trial, Sun et al. found that tolerance of pembrolizumab concomitant to RT was good compared to cetuximab concomitant to RT in locally advanced squamous cell carcinoma HNC[23]. In a double-blind, placebo-controlled, phase III trial (NCT02952586), Lee et al. reported that prolonging PFS with avelumab plus CRT followed by avelumab maintenance in patients with locally advanced squamous cell carcinoma of the head and neck was not met[24].

Tumoral and peritumoral immune infiltration were shown to be correlated with patient response to anti-

PD-1/PD-L1 immunotherapy. Depending on the immune phenotypes, i.e. immune-inflamed, immune-excluded, and immune-desert, patient will answer differently to treatment. Immune-inflamed tumors are characterized by dense, functional CD8 T-cell infiltration[25] and showed better response to immunotherapy. Both immune-excluded and immune-deserted phenotypes, present very low CD8 T-cells infiltration. CD8 T-cells infiltration could provide a powerful biomarker for predicting clinical outcomes of patients treated with immunotherapy alone or combined with conventional treatment. In addition to CD8 T-cells, radiomics, clinical data, hematology data, patient demographics and driver mutations have also been identified as potential biomarkers to predict response to immunotherapy[26, 27, 28, 29].

Several ways have been explored to tackle these limitations and in the era of big data, radiomics could allow the identification of responders and non-responders prior to immunotherapy as a standalone treatment or combined with other strategies. This remains a major focus in cancer studies. The tools developed in this thesis will be applied to the prediction of cancer treatment responses based on MR images from the ATEZOLACC clinical trial, to develop a radiomic signature assessing the amount of tumor-infiltrating CD8 T-cells in LACC. As a result, the differences between responding and non-responding patients will be characterized.

6.2 ATEZOLACC clinical trial

6.2.1 Overview

A multi-centric randomized phase II clinical trial evaluating chemo-radiotherapy±ICI against PD-L1 [1] was set up starting from August 2018 at Gustave Roussy Cancer Campus; inclusion of patients is now closed. The primary objective of this clinical trial is to evaluate the clinical benefits of the addition of atezolizumab to standard CRT (arm 2) compared with CRT alone (arm1), on investigator-assessed progression-free survival (PFS), as per Response Evaluation Criteria in Solid Tumors version 1.1. Additionally to the exams already planned in the clinical trial, MR exams (T1, T1-Gd, T2, DWI) have been realized for 40 patients at week 4 of the radiotherapy treatment in arm 2 (see Figure6.2). In addition to MR images, 3 tumor biopsies has been performed as well, before treatment, at weeks 3 and 7 (pre-brachytherapy) to assess tumor immune infiltration and tumor response. In both arms, CRT consisted of weekly intravenous cisplatin 40 mg/m^2 (5–6 cycles, 1 day per cycle) plus 45 Gy external-beam radiotherapy (EBRT) delivered in 1.8 Gy fractions ± simultaneously integrated boost to macroscopically involved lymph nodes, if any, followed by brachytherapy (BT) on week 7 with the objective to deliver a minimal total dose of 85 Gy (equivalent dose in 2-Gy fractions with $\alpha/\beta = 10Gy$) to 80% of the High Risk-Clinical Target Volume (HR-CTV). Atezolizumab was administered (IV 1200 mg Q3W), starting on the same week as EBRT (Week 1), and continued as an adjuvant for a total maximum of 20 cycles (approximately 14 months total of treatment). Tumor biopsies were analyzed and characterized for immunological parameters such as Tumor Infiltrating Lymphocytes (TIL). Briefly, samples were processed, formalin-fixed, and paraffin-embedded according to conventional procedures. Serial 4 µm-thick tissue sections were prepared for histological and immunohistochemical examinations. The first section was stained for histological evaluation in order to assess the representativity of the sample and the presence of possible confounding factors: the presence of tumor

[1] (ATEZOLACC, https://clinicaltrials.gov/ct2/show/NCT03612791)

cells, type of tumor, tumor cellularity, amount of stroma, ulceration, and necrosis. In the same section, the density of tumor intra-epithelial lymphocytes was evaluated according to current recommendations[30]. The next sections were used for immunohistochemical evaluation, including the immunodetection of CD8+ lymphocytes (clone SP16, automated immunoperoxidase technique, Ventana Benchmark Ultra stainer). The density in CD8+ lymphocytes was evaluated by two approaches: (a) visual counting by an experienced pathologist, and (b) automated counting by an algorithm available on the platform (Tribvn Healthcare, Chatillon, France). All sections were digitalized (Hamamatsu scanner) and stored for further analysis.

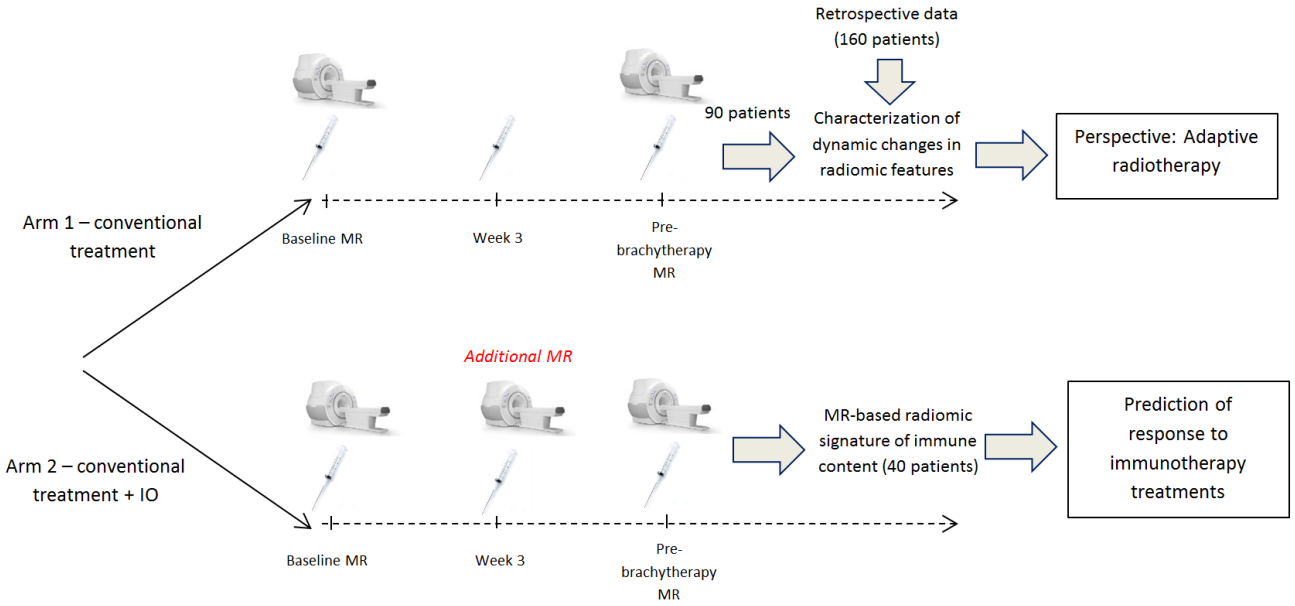


Figure 6.2: AtezoLACC clinical trial. Patients in Arm 1 received the conventional treatment, i.e. radiotherapy with concomitant chemotherapy followed by brachytherapy. Patients in Arm 2 (experimental arm) received in addition to the conventional treatment an immunotherapy treatment with Atezolizumab. In addition to patients in Arm 1, retrospective data from 160 patients were also gathered.

6.2.2 Data acquisition

Patient inclusion in the ATEZOLACC clinical trial started in the fourth quarter of 2018, involved 20 centers, and was closed in 2022. 110, 64, and 84 biopsies were performed in both arms before treatment, at weeks 3 and 7 (pre-brachytherapy) respectively. 102, 56, and 78 biopsies taken before treatment, at weeks 3 and 7, respectively, were retrieved by Gustave Roussy Cancer Campus. 137, 12, and 127 MR images (T1, T1-Gd, T2, DWI) were taken before treatment, at weeks 4 and 7 respectively. 106, 7, and 102 MR images taken before treatment, at weeks 4 and 7 respectively, were repatriated at Gustave Roussy Cancer Campus. Details about Arm 1 and Arm 2 are given in Table 6.1.

6.2.3 Tumor segmentation

In AtezoLACC clinical trial, several MR images were acquired at different time points. To study the evolution of tumor heterogeneity throughout the treatment or to characterize the profile of responding patients to RT $\pm$

			Biopsies
	Baseline	Week 3	Week 7
Arm 1	54 (52)	33 (29)	44 (42)
Arm 2	56 (50)	31 (27)	40 (36)
			MR images
	Baseline	Week 3	Week 7
Arm 1	70 (52)	5	65 (55)
Arm 2	67 (54)	7 (7)	62 (47)

Table 6.1: Biopsies and MR images in Arm 1(standard arm) and Arm 2(experimental arm) collected from the AtezoLACC clinical trial. In (), the number of samples already at Gustave Roussy Cancer Campus.

immunotherapy, using radiomic features, it is essential to delineate the tumor on all the acquired images. As stated in the previous chapters, one of the difficulties in radiomic analysis in general and in clinical routine in particular, lies in the segmentation of the tumor on medical images. Indeed, in addition to generating inter-observer variability, segmentation remains an extremely time-consuming task. To overcome this limitation, efforts have been undertaken to develop an automatic method for GTV detection and segmentation on T2 MRI images. This work was done in collaboration with Rahimeh Rouhi, a post-doctoral researcher in our team. This study was designed by Charlotte Robert, Rahimeh Rouhi, and I. I performed image pre-processing and Rahimeh Rouhi developed the automatic segmentation tool. Rahimeh Rouhi, Charlotte Robert, and I conducted the analysis of the results.

For this purpose, 160 pre-RT T2w MR images from patients treated for LACC were gathered retrospectively, as the training set for the automatic segmentation model. 45 images were excluded for the presence of artifacts in the image, or poor image quality. These images were acquired on 30 different scanners and had multiple acquisition parameters. We also gathered prospectively, 51 T2w MR images from the ATEZOLACC clinical trial from 8 centers, as the testing set. MRI device characteristics and acquisition parameters for test set are detailed in tables and 6.2.

Manual segmentation of the training cohort was done by two board-certified radiation oncologists P-A.L and S.A, respectively with 3 and 7 years of expertise. Manual segmentation of the test cohort was done by MB.B, a radiation oncologist with 5 years of expertise, and the delineations were checked by R.S., with 5 years of experience.

Deep learning model training, validation, and testing

Several deep neural network architectures were trained for 2D and 3D segmentation (UNet [31, 32], VNet [33], SegResNet and SegResNetVAE [34]). In total, 8 different models were evaluated. A 5-fold cross-validation was applied to the training set and two strategies were assessed. In the first strategy, the best model on the five folds was selected and then applied to the test set. In the second strategy, an ensembling model was put in place, where all five trained models were applied to each MR image in the test cohort, and the ensemble prediction was obtained by the majority of voting among the results of all the models. All models were trained for 300 epochs, rectified linear unit (ReLU) as the activation function, and Adam optimizer were selected for all

Manufacturer	Manufacturer model name	Magnetic Field Strength (T)	Repetition time (ms)	Echo time (ms)	Slice thickness (mm)	Pixel spacing (mm)	Matrix dimension	Number of patients
GE Medical Systems	Optima MR450W	1.5	5235, 7348, 7756, 7799, 8289, 8319, 8552, 8581, 8981, 9012, 10858, 10903	81, 86, 87, 93, 102, 111, 122, 125	4	0.58x0.58 0.74x0.74, 0.82x0.82	512x512	14
GE Medical Systems	Optima MR360	1.5	5665, 10838, 12789	101, 102	3, 4.5	0.47x0.74, 0.55x0.55, 0.58x0.58	512*512	7
GE Medical Systems	DISCOVERY MR750w	1.5	10598, 10800, 11815, 13019	79, 128	4	0.78x0.78, 0.82x0.82	512*512	4
GE Medical Systems	Signa Explorer	1.5	6421	102	5	0.78x0.78	512*512	4
GE Medical Systems	Signa Premier	1.5	7267, 7483, 14149	119, 121, 122	5	0.78x0.78	512*512	4
Siemens	Avanto_fit	1.5	4820, 5540, 7230	118	5	1.0x1.0	320x320	13
Siemens	Aera	1.5	1200, 3400, 4040, 4080, 7530	78, 89, 158	3, 3.5, 5	0.62x0.62, 0.75x0.75, 1.17x1.17, 1.18x1.18	260x260, 384x384	5

Table 6.2: Prospective pre-RT MR images (n=51) acquisition parameter, for testing cohort. RT: Radiotherapy, ms: milliseconds.

the applied networks. To tackle the overfitting issue and increase the generalizability of the segmentation, data augmentation was performed on the fly, prior to each optimization iteration during the training, to decrease the excessive storage requirements. More specifically, random rotation, flip, zoom with 0.5 probability, contrast adjustment, Gaussian noise, and Gaussian smoothing with 0.2 probability were applied for data augmentation. During the validation and testing, to include the GTVs completely in the segmentation inference for both single and ensemble models, the model already trained with the image size of 256×256×64 was applied to images with the size of 256×256×128, with a sliding window and a batch size equal to 2. We did also post-processing on the predicted segmentation, during the validation and testing, by applying morphological operations and keeping the largest connected components.

Evaluation metrics

The dice similarity coefficient (DSC) was used to evaluate the concordance between manual and automatic segmentation. DSC has become the most broadly used metric for automatic segmentation evaluation in AI. Given two sets X and Y, DSC measures the ratio between twice the overlap of sets X and Y divided by the sum of of both sets:

$$\text{DSC} = \frac{2|X \cap Y|}{|X| + |Y|} \tag{6.1}$$

where X and Y represent the ground truth and the automatically segmented target volume, respectively.

The surface DSC (SDSC)[35], which measures the overlap between two surfaces instead of volumes, was

also applied to evaluate manual and automatic segmentation:

$$\text{SDSC} = \frac{|S_g \cap B_o^\tau| + |S_o \cap B_g^\tau|}{|S_g| + |S_o|} \tag{6.2}$$

where the output and the ground truth binary volumes are converted to polygon meshes, representing their surfaces depicted by S_o and S_g, respectively, and B_g and B_o are the boundaries of ground truth and output surfaces at a tolerance τ ($\tau = 3$mm in our evaluation).

The Hausdorff distance (HD) is another commonly used metric to evaluate how far two subsets are from each other. The HD is the maximum distance of a set to the nearest point in the other set. For our purpose, the 95^{th} percentile of the distances between boundary points in X and Y, or 95 Hausdorff Distance(95HD) was calculated:

$$\text{95HD}(X,Y) = \text{percentile}(h(X,Y) \cup h(Y,X), \text{95th}) \tag{6.3}$$

$$h(X,Y) = \max_{y \in Y} \left(\min_{x \in X} ||x - y|| \right) \tag{6.4}$$

where $h(X,Y)$ is the largest distance from a point in X to the nearest point in Y and $||.||$ is the Euclidean norm of the points of X and Y.

For 2D segmentation, 3D volume final segmentation was reconstructed before calculating DSC, SDSC, and 95HD. In our experiments, we computed the aforementioned metrics based on deepmind Python library[2]. Since some models may fail in segmentation, resulting in no output and an infinitive value of 95HD, we also reported F as the number of segmentation failures.

Results

Table 6.3 presents the average validation results obtained from the 5 trained models in the 5-fold cross-validation for each network in 2D and 3D segmentation separately on the training set. The 2D-SegResNetVAE resulted in the best values of DSC $= 0.61 \pm 0.03$, SDSC $= 0.57 \pm 0.04$, $95HD = 16.17 \pm 2.65mm$, and F=0, respectively compared with the other models. Besides, Table 6.4 shows the results of segmentation on the test cohort based on single models, i.e. the best model obtained from 5-fold cross-validation. The results of SegResNet and SegResNetVAE in the 2D case were the best when considering the output of the single models with/without failure i.e., F=1/0. Furthermore, Table 6.5 presents the results of the final segmentation by applying the ensemble models, obtained from 5-fold cross-validation, to the test samples in cohort 2. Accordingly, 2D-SegResNet achieved the best performance compared to all the other models.

Conclusion on tumor segmentation

The total time required for automatic segmentation was 209 min (with an average of 4.26 ± 3.62 min), which was approximately less than one-third of the time required for manual segmentation of images in the test set (cohort 2). A radiation oncologist performed qualitative results of the automatic segmentation. 62% (32/51), 27% (14/51), and 5% (3/51) of the images in the test set were scored as A (good with no/with minor corrections),

[2] https://github.com/deepmind/surface-distance

Table 6.3: Average results of DSC(%), SDSC(%), and 95HD(mm) and total number of segmentation failures F of the 5 models obtained from 5-fold cross-validation on cohort 1 for different network architectures in 2D and 3D segmentation.

	2D				3D			
Network	DSC ± SD(%)	SDSC ± SD	95HD ± SD	F	DSC ± SD	SDSC ± SD	95HD ± SD	F
UNet	0.59 ± 0.05	0.57 ± 0.06	19.42 ± 1.27	1	0.40 ± 0.08	0.34 ± 0.06	37.20 ± 6.49	8
VNet	0.57 ± 0.07	0.54 ± 0.06	19.53 ± 3.85	5	0.38 ± 0.10	0.28 ± 0.10	38.58 ± 16.58	13
SegResNet	0.60 ± 0.03	0.56 ± 0.03	17.45 ± 1.81	**0**	0.37 ± 0.09	0.31 ± 0.07	52.44 ± 16.94	**0**
SegResNetVAE	**0.61 ± 0.03**	**0.57 ± 0.04**	**16.17 ± 2.65**	**0**	0.41 ± 0.11	0.35 ± 0.08	45.87 ± 14.71	**0**

Table 6.4: Average results of DSC(%), SDSC(%), and 95HD(mm) and total number of segmentation failures F for single models on cohort 2 for different network architectures in 2D and 3D segmentation.

	2D				3D			
Network	DSC ± SD(%)	SDSC ± SD	95HD ± SD	F	DSC ± SD	SDSC ± SD	95HD ± SD	F
UNet	0.66 ± 0.25	0.63 ± 0.25	20.09 ± 29.83	**0**	0.51 ± 0.25	0.37 ± 0.21	27.61 ± 18.01	2
VNet	0.67 ± 0.23	0.63 ± 0.23	22.60 ± 34.08	1	0.43 ± 0.23	0.25 ± 0.18	37.23 ± 21.72	2
SegResNet	**0.70 ± 0.16**	**0.66 ± 0.16**	**15.68 ± 9.50**	1	0.49 ± 0.27	0.37 ± 0.22	34.27 ± 27.59	**0**
SegResNetVAE	0.68 ± 0.21	0.63 ± 0.22	16.26 ± 12.36	**0**	0.56 ± 0.26	0.45 ± 0.22	27.16 ± 32.68	**0**

Table 6.5: Average results of DSC(%), SDSC(%), and 95HD(mm) and the total number of segmentation failures F for ensemble models on cohort 2 for different network architectures in 2D and 3D segmentation.

	2D				3D			
Network	DSC ± SD(%)	SDSC ± SD	95HD ± SD	F	DSC ± SD	SDSC ± SD	95HD ± SD	F
UNet	0.69 ± 0.23	0.65 ± 0.23	18.18 ± 24.19	**0**	0.54 ± 0.28	0.44 ± 0.25	32.04 ± 36.10	**0**
VNet	0.71 ± 0.20	0.66 ± 0.19	14.65± 11.71	1	0.50 ± 0.22	0.32 ± 0.19	30.99 ± 24.01	1
SegResNet	**0.72 ± 0.16**	**0.66 ± 0.17**	**14.62 ± 9.04**	**0**	0.57 ± 0.26	0.45 ± 0.23	37.41 ± 53.33	**0**
SegResNetVAE	0.70 ± 0.18	0.63 ± 0.18	15.09 ± 10.89	**0**	0.63 ± 0.21	0.52 ± 0.19	20.59 ± 16.47	**0**

B (satisfactory with major corrections), and C (Bad with a complete correction or delineation from scratch), respectively. The oncologist refused to score 3.9% (2/51) images due to their low quality. Similarly to our findings, Breto et al. developed a region-based convolutional neural network (R-CNN) for the segmentation of GTV+Cervix and OARs in 646 onboard MR images taken before each treatment fraction from five patients [36] and showed a good agreement between the auto-contoured and manual segmentation with (DSC) $= 0.84$ for GTV+Cervix and DSC > 0.7 for OAR. In another similar study, Breto et al.[37] used planning and daily treatment fraction MR images, from 15 LACC patients for OARs and GTV+Cervix segmentation. The best results for GTV+Cervix segmentation achieved DSC $= 0.67 \pm 0.30$ and $95HD = 2.77 \pm 1.73mm$ and the DSC for rectum, femur, and bladder was greater than 0.8. These results suggest that the developed tools for GTV automatic segmentation could be used for image segmentation in the AtezoLACC clinical trial for harmonization of practices. Nevertheless, despite promising results, the models' performances should be improved by combining different planes and MR sequences.

6.2.4 Are biopsy samples representative of the whole tumor?

In the AtezoLACC trial, several biopsy samples have been collected. These samples were acquired in different configurations, which might not be representative of the tumor cells and would create biases in our study.

With the help of two board-certified pathologists (M. C. and C. G.), we conducted a preliminary study that consisted in validating the hypothesis that the tumor infiltrate on the surface biopsy would be representative of the overall tumor infiltrate. Indeed, the tumor biopsy is firstly performed to confirm the presence of tumor cells and secondly to evaluate the type and stage of the tumor. From a sample of 10 cone biopsies (conization), peritumoral, stroma and global TILs were analyzed. Polymorphonuclear neutrophils (PMN) were also assessed on the superior (at the surface) and the inferior slice of the volume. The visual assessment showed that the infiltrate was quite homogeneous between the surface and the depth of the tissues from the conizations. PMN scores between the superior and the inferior part of the conization and TILs quantification (Figure 6.3) were also found not statistically different.

Afterward, pre-BT biopsies of 17 patients with a majority of squamous cell carcinomas except for three cases (adenocarcinomas), treated at Gustave Roussy (retrospective biological material) for cervical cancer were retrieved for a first multiplex CD3 (T-cells), CD20 (B-cells), CD68 (macrophages), CD57(natural killer (NK)), with the goal to exhaustively analyze the tumor microenvironment in the frame of the ATEZOLACC clinical trial. B-cells, T-cells, and NK cells are all involved in the immune system response but while NK cells respond to signals of immune stress such as inflammation, B and T-cells specifically recognize and eliminate foreign antigens. Observations showed that CD3 populations were superior in number to macrophage populations, which were superior to CD20. NK populations were not observed. These results led to the development of a new multiplex: CD8 (T-cells), PD-L1 (tumor cells/response to anti-PD-1 or anti-PD-L1 antibodies), FoxP3 (T-reg), and CK (tumor cells). The first labeling was launched in March 2022 following the repatriation of the biological material.

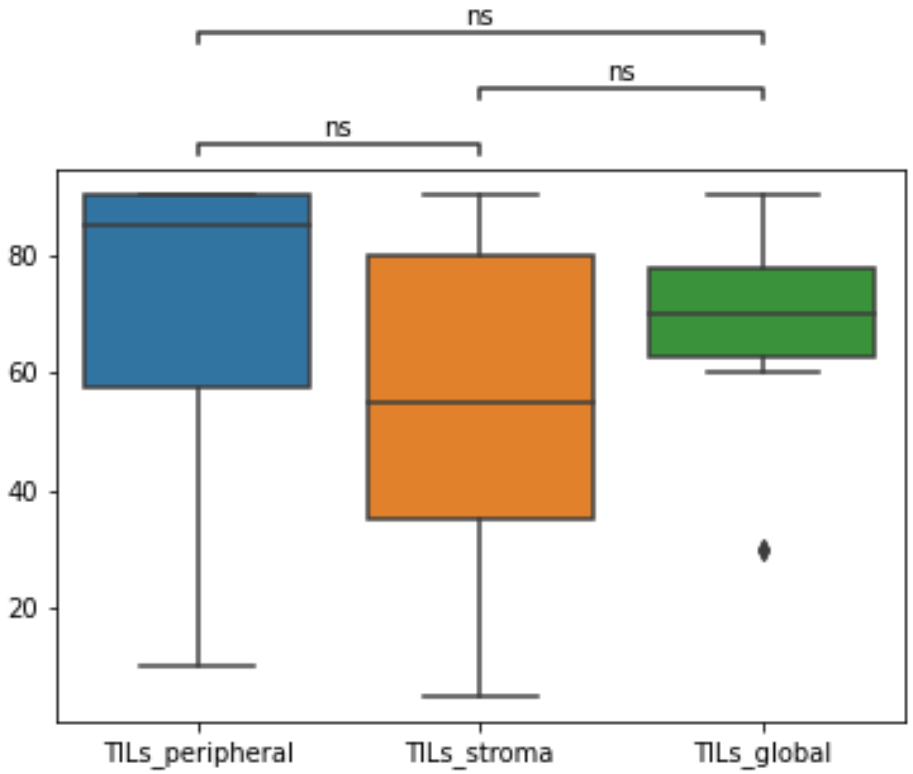


Figure 6.3: TILs quantification in conizations from pre-BT LACC patients. The biopsy is well representative of the whole tumor lymphocyte infiltration in LACC.

6.2.5 Evolution of CD8 T-cells prevalence troughout the treatment

The approach explored in order to distinguish the phenotype of responders and non-responders to immunotherapy combined or not with CRT, was the quantification of the lymphocyte cell infiltrate in the tumor, in particular CD8 T-cells. After having validated the hypothesis that the infiltrate analyzed on a sample was well representative of the tumor infiltrate, this quantification was done on biopsies acquired at different times to characterize the evolution of CD8 T-cells over time in the two arms (experimental vs standard).

In the standard arm (arm 1), 56 biopsies were analyzed: 28, 10, and 18 taken respectively before treatment, at weeks 3 and 7. In arm 2, 56 biopsies were gathered: 25, 11, and 20 were taken respectively before treatment, at weeks 4 and 7. The estimation of CD8 cell infiltrate was done on each biopsy using the validated multiplex described above using the HALO algorithm and was validated by a board-certified pathologist.

As shown in Figure 6.4, there was no statistical difference (Wilcoxon rank-sum test) between arm 1 (standard arm) and arm 2 (experimental arm) during the course of the treatment up to week 7.

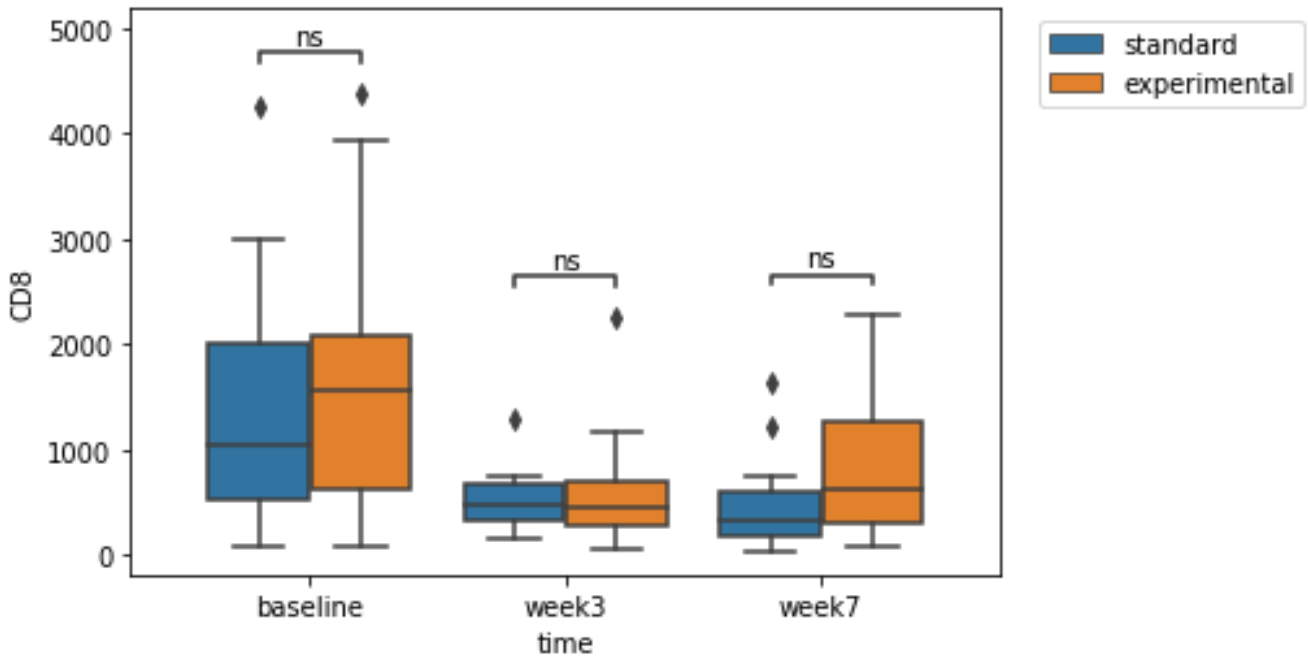


Figure 6.4: Evolution of CD8 T-cells prevalence throughout the treatment between arm 1 and arm 2.

6.3 Future work and final results

Although no differentiated CD8 modulation was observed in the 2 arms at 7 weeks, it remains to be seen whether a link between response to treatment and CD8 modulation can be established and, if so, whether a radiomic signature of the infiltrate could be developed, probably common to both arms. A radiomic score will be then determined to assess CD8 T-cell content based on MR images. To do so, the tumor core as well as a peripheral ring will be delineated on the MR images from which radiomic features will be extracted. For that purpose, the automatic segmentation tool described above will be applied before verification by a board-certified radiation oncologist. This study could be done in >24 months after the last inclusion, given the 2-year follow-up required in the clinical trial. The pre-trained model on images acquired before RT will be transferred to the pre-BT images via transfer learning. For this purpose, 160 T2w pre-BT MR images were retrospectively collected, have been contoured, and will serve as the training cohort. The AtezoLACC pre-BT images will serve as the testing cohort. The standardization method based on CycleGAN developed in this thesis will also be used on the different MR images for robust radiomic feature extraction, as well as the developed modified consensus nested-cross validation (cnCV) for an appropriate feature selection. Several ML algorithms will be tested (classifiers and predictors) such as random forest, penalized regression (lasso, elastic-net), support vector machine (SVC), etc., and the performance of each will be evaluated. Analysis of the association of the CD8 T-cell signature with other immune cell populations (neutrophils, CD4) and PD-L1 assessed by immunohistochemistry will be also performed. The last step would be to analyze the correlation between dynamic changes in radiomic features and immunological content for conventional RT treatments with the goal to develop predictive signatures of local, loco-regional, distant recurrence-free, and overall survivals. Optimizing the information generated by routine imaging performed during treatment will provide insights for potential biomarkers, which will constitute valuable tools to be used in a personalized therapeutic approach. In parallel, the analysis of intra-patients, per treatment dynamic changes, is a first step for a better understanding of tumor heterogeneity and eventually will serve for adaptive radiotherapy purposes as allowed by innovative MR-Linacs.

Bibliography

[1] William B. Coley. "The Treatment of Inoperable Sarcoma by Bacterial Toxins (the Mixed Toxins of the Streptococcus erysipelas and the Bacillus prodigiosus)". In: *Proc R Soc Med* 3.Surg Sect (1910), pp. 1–48. ISSN: 0035-9157. URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1961042/ (visited on 12/13/2022).

[2] Gavin P. Dunn, Lloyd J. Old, and Robert D. Schreiber. "The Three Es of Cancer Immunoediting". en. In: *Annu. Rev. Immunol.* 22.1 (Apr. 2004), pp. 329–360. ISSN: 0732-0582, 1545-3278. DOI: 10.1146/annurev.immunol.22.012703.104803. URL: https://www.annualreviews.org/doi/10.1146/annurev.immunol.22.012703.104803 (visited on 12/13/2022).

[3] Gavin P. Dunn, Lloyd J. Old, and Robert D. Schreiber. "The Immunobiology of Cancer Immunosurveillance and Immunoediting". en. In: *Immunity* 21.2 (Aug. 2004), pp. 137–148. ISSN: 10747613. DOI: 10.1016/j.immuni.2004.07.017. URL: https://linkinghub.elsevier.com/retrieve/pii/S1074761304002092 (visited on 12/13/2022).

[4] J. F. Brunet et al. "A new member of the immunoglobulin superfamily–CTLA-4". eng. In: *Nature* 328.6127 (July 1987), pp. 267–270. ISSN: 0028-0836. DOI: 10.1038/328267a0.

[5] Y. Ishida et al. "Induced expression of PD-1, a novel member of the immunoglobulin gene superfamily, upon programmed cell death". eng. In: *EMBO J* 11.11 (Nov. 1992), pp. 3887–3895. ISSN: 0261-4189. DOI: 10.1002/j.1460-2075.1992.tb05481.x.

[6] G. J. Freeman et al. "Engagement of the PD-1 immunoinhibitory receptor by a novel B7 family member leads to negative regulation of lymphocyte activation". eng. In: *J Exp Med* 192.7 (Oct. 2000), pp. 1027–1034. ISSN: 0022-1007. DOI: 10.1084/jem.192.7.1027.

[7] F. Stephen Hodi et al. "Improved survival with ipilimumab in patients with metastatic melanoma". eng. In: *N Engl J Med* 363.8 (Aug. 2010), pp. 711–723. ISSN: 1533-4406. DOI: 10.1056/NEJMoa1003466.

[8] Caroline Robert et al. "Anti-programmed-death-receptor-1 treatment with pembrolizumab in ipilimumab-refractory advanced melanoma: a randomised dose-comparison cohort of a phase 1 trial". eng. In: *Lancet* 384.9948 (Sept. 2014), pp. 1109–1117. ISSN: 1474-547X. DOI: 10.1016/S0140-6736(14)60958-2.

[9] Jeffrey S. Weber et al. “Nivolumab versus chemotherapy in patients with advanced melanoma who progressed after anti-CTLA-4 treatment (CheckMate 037): a randomised, controlled, open-label, phase 3 trial”. eng. In: *Lancet Oncol* 16.4 (Apr. 2015), pp. 375–384. ISSN: 1474-5488. DOI: 10.1016/S1470-2045(15)70076-8.

[10] Jonathan E. Rosenberg et al. “Atezolizumab in patients with locally advanced and metastatic urothelial carcinoma who have progressed following treatment with platinum-based chemotherapy: a single-arm, multicentre, phase 2 trial”. eng. In: *Lancet* 387.10031 (May 2016), pp. 1909–1920. ISSN: 1474-547X. DOI: 10.1016/S0140-6736(16)00561-4.

[11] Stephen M. Ansell. “Hodgkin lymphoma: MOPP chemotherapy to PD-1 blockade and beyond”. en. In: *American Journal of Hematology* 91.1 (2016). _eprint: https://onlinelibrary.wiley.com/do pp. 109–112. ISSN: 1096-8652. DOI: 10.1002/ajh.24226. URL: http://onlinelibrary.wiley.com/doi/abs/10.1002/ajh.24226 (visited on 12/13/2022).

[12] Martin Reck et al. “Pembrolizumab versus Chemotherapy for PD-L1–Positive Non–Small-Cell Lung Cancer”. In: *New England Journal of Medicine* 375.19 (Nov. 2016). Publisher: Massachusetts Medical Society _eprint: https://doi.org/10.1056/NEJMoa1606774, pp. 1823–1833. ISSN: 0028-4793. DOI: 10.1056/NEJMoa1606774. URL: https://doi.org/10.1056/NEJMoa1606774 (visited on 12/13/2022).

[13] Eric Deutsch et al. “Optimising efficacy and reducing toxicity of anticancer radioimmunotherapy”. eng. In: *Lancet Oncol* 20.8 (Aug. 2019), e452–e463. ISSN: 1474-5488. DOI: 10.1016/S1470-2045(19)30171-8.

[14] Yifan Wang et al. “Combining Immunotherapy and Radiotherapy for Cancer Treatment: Current Challenges and Future Directions”. In: *Front Pharmacol* 9 (Mar. 2018), p. 185. ISSN: 1663-9812. DOI: 10.3389/fphar.2018.00185. URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5844965/ (visited on 12/13/2022).

[15] Guillaume Louvel et al. “Immunotherapy and pulmonary toxicities: can concomitant immune-checkpoint inhibitors with radiotherapy increase the risk of radiation pneumonitis?” en. In: *European Respiratory Journal* 51.1 (Jan. 2018). Publisher: European Respiratory Society Section: Agora. ISSN: 0903-1936, 1399-3003. DOI: 10.1183/13993003.01737-2017. URL: http://erj.ersjournals.com/content/51/1/1701737 (visited on 12/13/2022).

[16] Simon J. Dovedi et al. “Acquired Resistance to Fractionated Radiotherapy Can Be Overcome by Concurrent PD-L1 Blockade”. In: *Cancer Research* 74.19 (Sept. 2014), pp. 5458–5468. ISSN: 0008-5472. DOI: 10.1158/0008-5472.CAN-14-1258. URL: https://doi.org/10.1158/0008-5472.CAN-14-1258 (visited on 04/19/2023).

[17] Liufu Deng et al. “Irradiation and anti-PD-L1 treatment synergistically promote antitumor immunity in mice”. eng. In: *J Clin Invest* 124.2 (Feb. 2014), pp. 687–695. ISSN: 1558-8238. DOI: 10.1172/JCI67313.

[18] Lorenzo Galluzzi et al. “The hallmarks of successful anticancer immunotherapy”. In: *Science Translational Medicine* 10.459 (Sept. 2018). Publisher: American Association for the Advancement of Science, eaat7807. DOI: 10.1126/scitranslmed.aat7807. URL: https://www.science.org/doi/abs/10.1126/scitranslmed.aat7807 (visited on 03/01/2023).

[19] Josephine Kang, Sandra Demaria, and Silvia Formenti. “Current clinical trials testing the combination of immunotherapy with radiotherapy”. en. In: *J Immunother Cancer* 4.1 (Dec. 2016). Publisher: BMJ Specialist Journals Section: Clinical Trials Monitor, p. 51. ISSN: 2051-1426. DOI: 10.1186/s40425-016-0156-7. URL: https://jitc.bmj.com/content/4/1/51 (visited on 04/19/2023).

[20] Salma K. Jabbour et al. “Pembrolizumab Plus Concurrent Chemoradiation Therapy in Patients With Unresectable, Locally Advanced, Stage III Non–Small Cell Lung Cancer: The Phase 2 KEYNOTE-799 Nonrandomized Trial”. In: *JAMA Oncology* 7.9 (Sept. 2021), pp. 1351–1359. ISSN: 2374-2437. DOI: 10.1001/jamaoncol.2021.2301. URL: https://doi.org/10.1001/jamaoncol.2021.2301 (visited on 04/19/2023).

[21] Scott J. Antonia et al. “Overall Survival with Durvalumab after Chemoradiotherapy in Stage III NSCLC”. In: *New England Journal of Medicine* 379.24 (Dec. 2018). Publisher: Massachusetts Medical Society _eprint: https://doi.org/10.1056/NEJMoa1809697, pp. 2342–2350. ISSN: 0028-4793. DOI: 10.1056/NEJMoa1809697. URL: https://doi.org/10.1056/NEJMoa1809697 (visited on 12/13/2022).

[22] Corinne Faivre-Finn et al. “Four-Year Survival With Durvalumab After Chemoradiotherapy in Stage III NSCLC-an Update From the PACIFIC Trial”. eng. In: *J Thorac Oncol* 16.5 (May 2021), pp. 860–867. ISSN: 1556-1380. DOI: 10.1016/j.jtho.2020.12.015.

[23] Xu Shan Sun et al. “A phase II randomized trial of pembrolizumab versus cetuximab, concomitant with radiotherapy (RT) in locally advanced (LA) squamous cell carcinoma of the head and neck (SCCHN): First results of the GORTEC 2015-01 “PembroRad” trial.” In: *JCO* 36.15_suppl (May 2018). Publisher: Wolters Kluwer, pp. 6018–6018. ISSN: 0732-183X. DOI: 10.1200/JCO.2018.36.15_suppl.6018. URL: https://ascopubs.org/doi/10.1200/JCO.2018.36.15_suppl.6018 (visited on 04/25/2023).

[24] Nancy Y. Lee et al. “Avelumab plus standard-of-care chemoradiotherapy versus chemoradiotherapy alone in patients with locally advanced squamous cell carcinoma of the head and neck: a randomised, double-blind, placebo-controlled, multicentre, phase 3 trial”. eng. In: *Lancet Oncol* 22.4 (Apr. 2021), pp. 450–462. ISSN: 1474-5488. DOI: 10.1016/S1470-2045(20)30737-3.

[25] Roger Sun et al. “A radiomics approach to assess tumour-infiltrating CD8 cells and response to anti-PD-1 or anti-PD-L1 immunotherapy: an imaging biomarker, retrospective multicohort study”. eng. In: *Lancet Oncol* 19.9 (Sept. 2018), pp. 1180–1191. ISSN: 1474-5488. DOI: 10.1016/S1470-2045(18)30413-3.

[26] Roger Sun et al. “Radiomics to predict outcomes and abscopal response of patients with cancer treated with immunotherapy combined with radiotherapy using a validated signature of CD8 cells”. In: *J Immunother Cancer* 8.2 (Nov. 2020), e001429. ISSN: 2051-1426. DOI: 10.1136/jitc-2020-001429. URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7668366/ (visited on 05/23/2022).

[27] S. Trebeschi et al. “Predicting response to cancer immunotherapy using noninvasive radiomic biomarkers”. eng. In: *Ann Oncol* 30.6 (June 2019), pp. 998–1004. ISSN: 1569-8041. DOI: 10.1093/annonc/mdz108.

[28] Mohammadhadi Khorrami et al. “Changes in CT Radiomic Features Associated with Lymphocyte Distribution Predict Overall Survival and Response to Immunotherapy in Non–Small Cell Lung Cancer”. In: *Cancer Immunology Research* 8.1 (Jan. 2020), pp. 108–119. ISSN: 2326-6066. DOI: 10.1158/2326-6066.CIR-19-0476. URL: https://doi.org/10.1158/2326-6066.CIR-19-0476 (visited on 04/25/2023).

[29] Ilke Tunali et al. “Novel clinical and radiomic predictors of rapid disease progression phenotypes among lung cancer patients treated with immunotherapy: An early report”. en. In: *Lung Cancer* 129 (Mar. 2019), pp. 75–79. ISSN: 0169-5002. DOI: 10.1016/j.lungcan.2019.01.010. URL: https://www.sciencedirect.com/science/article/pii/S016950021930296X (visited on 04/25/2023).

[30] Shona Hendry et al. “Assessing Tumor-Infiltrating Lymphocytes in Solid Tumors: A Practical Review for Pathologists and Proposal for a Standardized Method from the International Immuno-Oncology Biomarkers Working Group: Part 2: TILs in Melanoma, Gastrointestinal Tract Carcinomas, Non-Small Cell Lung Carcinoma and Mesothelioma, Endometrial and Ovarian Carcinomas, Squamous Cell Carcinoma of the Head and Neck, Genitourinary Carcinomas, and Primary Brain Tumors”. eng. In: *Adv Anat Pathol* 24.6 (Nov. 2017), pp. 311–335. ISSN: 1533-4031. DOI: 10.1097/PAP.0000000000000161.

[31] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. *U-Net: Convolutional Networks for Biomedical Image Segmentation*. arXiv:1505.04597 [cs]. May 2015. DOI: 10.48550/arXiv.1505.04597. URL: http://arxiv.org/abs/1505.04597 (visited on 01/18/2023).

[32] Özgün Çiçek et al. *3D U-Net: Learning Dense Volumetric Segmentation from Sparse Annotation*. arXiv:1606.06650 [cs]. June 2016. DOI: 10.48550/arXiv.1606.06650. URL: http://arxiv.org/abs/1606.06650 (visited on 04/19/2023).

[33] Fausto Milletari, Nassir Navab, and Seyed-Ahmad Ahmadi. *V-Net: Fully Convolutional Neural Networks for Volumetric Medical Image Segmentation*. arXiv:1606.04797 [cs]. June 2016. DOI: 10.48550/arXiv.1606.04797. URL: http://arxiv.org/abs/1606.04797 (visited on 04/19/2023).

[34] Andriy Myronenko. *3D MRI brain tumor segmentation using autoencoder regularization*. arXiv:1810.11654 [cs, q-bio]. Nov. 2018. DOI: 10.48550/arXiv.1810.11654. URL: http://arxiv.org/abs/1810.11654 (visited on 04/19/2023).

[35] Stanislav Nikolov et al. *Deep learning to achieve clinically applicable segmentation of head and neck anatomy for radiotherapy*. arXiv:1809.04430 [physics, stat]. Jan. 2021. DOI: 10.48550/arXiv.1809.04430. URL: http://arxiv.org/abs/1809.04430 (visited on 02/01/2023).

[36] A.L. Breto et al. "A Deep Learning Pipeline for per-Fraction Automatic Segmentation of GTV and OAR in cervical cancer". en. In: *International Journal of Radiation Oncology*Biology*Physics* 105.1 (Sept. 2019), S202. ISSN: 03603016. DOI: 10.1016/j.ijrobp.2019.06.267. URL: https://linkinghub.elsevier.com/retrieve/pii/S0360301619311022 (visited on 04/26/2023).

[37] Adrian L. Breto et al. "Deep Learning for Per-Fraction Automatic Segmentation of Gross Tumor Volume (GTV) and Organs at Risk (OARs) in Adaptive Radiotherapy of Cervical Cancer". eng. In: *Front Oncol* 12 (2022), p. 854349. ISSN: 2234-943X. DOI: 10.3389/fonc.2022.854349.

Chapter 7

Conclusion and perspectives

7.1 Conclusion

The last past few years have seen the development of new lines of treatment towards precision medicine through immunotherapy, targeted therapies, and efforts have been made for personalized chemotherapy (CT), and RT regimens. Despite these advancements, in most cases, patients with the same type of cancer are still treated similarly, especially in RT and CT where they are given the same doses in a standard manner. Developments of effective personalized treatment involve understanding and characterization of tumor phenotypes using a large panel of high dimensional data such as genomics, proteomics, radiomics, dosiomics in addition to clinical and biological information.

Radiomics, specifically AI applied to radiomics, is a highly promising and evolving field for personalized cancer treatment as shown by the high number of studies on the subject. Multiple lines of evidence demonstrated how radiomics is revealing real information regarding tumor cells that is above and beyond visual analysis, which could be taken advantage of for cancer early detection, treatment optimization, and prediction of treatment outcome. However, some limitations still hinder the field of radiomics. The main limitation to clinical implementation is by large the lack of large, structured, high-quality datasets that are well annotated with respect to the relevant outcome. This strongly and negatively impacts the generalizability of the developed models to real life. The other limitation in radiomic studies is the lack of standardization tools for multi-institutional images affected by the 'center-effect'. In this thesis, several approaches were made to tackle these limitations.

In the first project, efforts have been made to collect a large dataset of [18F]-FDG PET images from 5 independent institutions with the objective of building a robust model for survival prediction in patients with HPV-induced cancers, namely LACC, ASCC, and OPSCC. PET images of patients treated for LACC from Gustave Roussy Campus Cancer (104 patients) and Leeds Teaching Hospitals NHS Trust (90 patients) were gathered. We also collected ASCC PET images from Institut du Cancer de Montpellier (66 patients) and from Oslo University Hospital (67 patients). An additional cohort of patients with OPSCC was also gathered from University Hospital of Zurich (45 patients). The ComBat harmonization technique was applied to radiomic features. We implemented a robust feature selection method coupled with hyperparameters tuning, the modified consensus nested cross-validation, within four ML models to predict PFS and OS. Each model was trained and optimized on LACC and ASCC cohorts and tested on independent LACC, ASCC, and OPSCC patients. The

radiomic-based CoxNet model achieved the best results with C-index values of 0.75, 0.78 for PFS, and 0.76, 0.74, 0.75 for OS on the test sets. Radiomic feature-based models had superior performance compared to the bioclinical ones and combining radiomic and bioclinical variables did not improve the performances. Metabolic tumor volume (MTV)-based models obtained lower C-index values for a majority of the tested configurations, but the same or better performance for the identification of early relapses. The results demonstrate the possibility of identifying common PET-based image signatures for predicting the response of patients with induced HPV pathology, validated on multi-center multiconstructor data. However, the added value of radiomics remains to be demonstrated compared to conventional quantitative metrics. To our knowledge, this is the first time that a common radiomic signature has been developed for different HPV-positive cancers. This study may therefore pave the way for the optimization of HPV-induced cancers treatment, by guiding the clinician on the dose coverage trade-offs to be implemented in the planning stages of RT and/or BT.

In the second project, we developed a standardization method based on GAN for T2w MR images in patients with LACC. A cGAN and a CycleGAN method to address MR standardization were proposed and compared to conventional preprocessing and a posteriori methods proposed in the literature (Nyul, z-score normalization, and ComBat harmonization). In a first step, T2w MR images from 30 patients treated for LACC were acquired prospectively. For each patient, three images were taken sequentially on the same scanner with fixed values of repetition time (TR) and voxel size (VS). A cGAN was trained on these paired images to generate images robust to the impact of TR and VS modulation. The impact of standardization methods was assessed by means of PCA on image quality metrics extracted from the images before and after standardization. Using ICC and CCC, robust features were also characterized (CCC and ICC 0.75). PCA on image quality metrics showed that TR and VS changes were mitigated the most with cGAN-based standardization. Regarding TR/VS modulation, cGAN achieved the best results with 100% (18/18) and 77%(58/75) of robust features after standardization, on first-order and second-order features, respectively. In a second step, a retrospective cohort of 216 LACC patients was also gathered, including 86 and 160 T2w MR images taken before radiotherapy (RT) and brachytherapy (BT), respectively. A cycleGAN standardization method for unpaired images was trained but also Nyul and z-score normalization on the retrospective images. Different ML models were trained to investigate the impact of these standardization methods on stage classification and relapse prediction, respectively. CycleGAN achieved the best results on both tasks compared to other standardization methods. On pre-RT images, stage classification average accuracy was improved on the test set from 0.68 ± 0.16 to 0.83 ± 0.07, 0.78 ± 0.04, and 0.88 ± 0.09 with Nyul standardization, z-score standardization, and CycleGAN standardization, respectively. Likewise, relapse prediction based on pre-BT images achieved higher AUC on the test set after image standardization with the best model achieving 0.44 before standardization, 0.52, 0.56, and 0.60 after z-score, Nyul, and CycleGAN standardization, respectively. Our results suggest that neural network-based normalization could better limit the impact of acquisition parameters on MRI radiomic features compared to conventional methods in LACC. Such approach could be used in a single harmonization strategy encompassing image pre-processing steps without the need for a priori information on imaging acquisition and reconstruction parameters. In addition to providing very promising results for robust radiomic studies, adversarial networks could be used for multiple and varied tasks in combination with augmentation and normalization techniques with complementary benefits[1]. CycleGAN could be used to move from one image quality space to another, by changing target domain selection. which

would generate MR images with different properties in terms of noise, contrast and spatial resolution. This would help improve the accuracy of manual and automatic segmentation in the context of radiation therapy. It may also reduce the image acquisition time, thus increasing patient comfort.

7.2 Perspectives

The developed model and tools will be further applied to unseen data. An additional cohort of PET images of patients with HNC has been already collected at Gustave Roussy Campus Cancer. The developed radiomics signature will be evaluated on it to validate furthermore our findings, given the small size of OPSCC patients used in this thesis. In addition to OPSCC, it will also be tested on other types of HPV-related HNCs. The CycleGAN-based standardization tool will be applied to different sequences of MR images from the AtezoLACC clinical trial. It will allow for the development of a multi-sequence radiomic signature which will hopefully ensure the generalizability of the radiomic signature of CD8 T-cells infiltrate. Additional loss, such as structural similarity loss, could increase the structural correspondence between the generated and the source images and should be assessed. Advances in the field of GAN such as style GANs and their derivates should also be investigated for multi-centric radiomic studies. Several studies have proven the power of radiomic features for personalized treatment. However, a question remains: "How accurate should a radiomic or DL-based model be to be considered relevant in clinical routine?" To answer this question, prospective studies must be set up. In addition, a close collaboration among physicists, physicians, researchers but also patients, is essential to clearly define the relevant metrics for an efficient evaluation of a radiomic model, for example, before moving to a dose escalation strategy in RT.

Appendix A

List of publications

- Alexandre Carré, Guillaume Klausner, Myriam Edjlali, Marvin Lerousseau, Jade Briend-Diop, Roger Sun, Samy Ammari, Sylvain Reuzé, Emilie Alvarez Andres, Théo Estienne, **Stéphane Niyoteka**, Enzo Battistella, Maria Vakalopoulou, Frédéric Dhermain, Nikos Paragios, Eric Deutsch, Catherine Oppenheim, Johan Pallud, and Charlotte Robert. Standardization of brain MR images across machines and protocols: Bridging the gap for MRI-based radiomics. Scientific Reports, 10(1):12340, July 2020.

- Emilie Alvarez Andres, Lucas Fidon, Maria Vakalopoulou, Marvin Lerousseau, Alexandre Carré, Roger Sun, Guillaume Klausner, Samy Ammari, Nathan Benzazon, Sylvain Reuzé, Théo Estienne, **Stéphane Niyoteka**, Enzo Battistella, Angéla Rouyar, Georges Noël, Anne Beaudre, Frédéric Dhermain, Eric Deutsch, Nikos Paragios, and Charlotte Robert. 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, 108(3):813–823, November 2020.

- Alexandre Carré, Guillaume Klausner, Myriam Edjlali, Marvin Lerousseau, Jade Briend-Diop, Roger Sun, Samy Ammari, Sylvain Reuzé, Emilie Alvarez Andres, Théo Estienne, **Stéphane Niyoteka**, Enzo Battistella, Maria Vakalopoulou, Frédéric Dhermain, Nikos Paragios, Eric Deutsch, Catherine Oppenheim, Johan Pallud, and Charlotte Robert. Standardization of brain MR images across machines and protocols: Bridging the gap for MRI-based radiomics. Scientific Reports, 10(1):12340, July 2020.

- Roger Sun, Nora Sundahl, Markus Hecht, Florian Putz, Andrea Lancia, Angela Rouyar, Marina Milic, Alexandre Carré, Enzo Battistella, Emilie Alvarez Andres, **Stéphane Niyoteka**, Edouard Romano, Guillaume Louvel, Jérôme Durand-Labrunie, Sophie Bockel, Rastilav Bahleda, Charlotte Robert, Celine Boutros, Maria Vakalopoulou, Nikos Paragios, Benjamin Frey, Jean-Charles Soria, Christophe Massard, Charles Ferté, Rainer Fietkau, Piet Ost, Udo Gaipl, and Eric Deutsch. Radiomics to predict outcomes and abscopal response of patients with cancer treatedwith immunotherapy combined with radiotherapy using a validated signature of CD8 cells. Journal for ImmunoTherapy of Cancer, 8(2):e001429, November 2020.

- Roger Sun, Marvin Lerousseau, Theophraste Henry, Alexandre Carré, Amaury Leroy, Théo Estienne, **Stéphane Niyoteka**, Sophie Bockel, Angela Rouyar, Émilie Alvarez Andres, Nathan Benzazon, Enzo

Battistella, Marion Classe, Charlotte Robert, Jean Yves Scoazec, and Éric Deutsch. Intelligence artificielle en radiothérapie : radiomique, pathomique, et prédiction de la survie et de la réponse aux traitements. Cancer/Radiothérapie, 25(6):630–637, October 2021.

- Alexandre Carré, Enzo Battistella, **Stéphane Niyoteka**, Eric Deutsch, and Charlotte Robert. AutoComBat: a generic method for harmonizing MRI-based radiomic features. Scientific Reports, 10.1038/s41598-022-16609-1, July 2022.

Appendix B

Oral communication and Poster

- **Stéphane Niyoteka**, S. Achkar, I. Coric, R. Bourdais, E. Manea, I. Dumas, R. Marini-Silva, E. Ullmann, A. Carré, N. Paragios, E. Deutsch, C. Chargari, and C. Robert. PO-1667: Are current margins in locally advanced cervical cancers treated by tomotherapy appropriate? Radiotherapy and Oncology, 152:S915-916, November 2020.
- **Stéphane Niyoteka**, Romain-David Seban, Rahimeh Rouhi, Andrew Scarsbrook, Catherine Genestie, Marion Classe, Alexandre Carré, Roger Sun, Cyrus Chargari, Jack McKenna, Garry McDermott, Claire Lemanski, Eric Deutsch, and Charlotte Robert [18F]-FDG PET radiomics to predict survival in locally advanced cervical cancers (LACC) and anal squamous cancer cells (ASCC). INanoTheRad International Conference, November 2021.
- Rahimeh Rouhi, **Stéphane Niyoteka**, Alexandre Carre, Samir Achkar, Pierre-Antoine Laurent, Mouhamadou Bachir Ba, Cristina Veres, Théophraste Henry, Maria Vakalopoulou, Roger Sun, Sophie Espenel, Adrien Laville, Cyrus Chargari, Eric Deutsch, Charlotte Robert. Automatic Detection and Segmentation of GTV for Locally Advanced Cervical Cancer in T2-Weighted MRI Images. ESTRO Congress, May 2022.
- Nathan Benzazon, Alexandre Carré, François de Kermenguy, **Stéphane Niyoteka**, Cristina Veres, Rodrigue Allodji, Florent de Vathaire, Eric Deutsch, Ibrahima Diallo, Charlotte Robert.Deep-learning for instantaneous and universal estimation of the out-of-field dose in external radiotherapy. ESTRO Congress, May 2023/ ISORED Conference, May2023.
- **Stéphane Niyoteka**, Rahimeh Rouhi, Pierre-Antoine Laurent, Samir Achkar, Alexandre Carré, Roger Sun, Sebastien Diffetocq, Corinne Balleyguier, Cyrus Chargari, Eric Deutsch, Charlotte Robert. MR image normalization in locally advanced cervical cancer: what is the optimal method for developing multicenter radiomic models? ESTRO Congress, May 2023

Appendix C

Submitted papers

- **Stéphane Niyoteka**, Romain-David Seban, Rahimeh Rouhi, Andrew Scarsbrook, Catherine Genestie, Marion Classe, Alexandre Carré, Roger Sun, Agustina La Greca Saint-Esteven, Cyrus Chargari, Jack McKenna, Garry McDermott, Eirik Malinen, Stephanie Tanadini-Lang, Matthias Guckenberger, Marianne G. Guren, Claire Lemanski, Eric Deutsch, and Charlotte Robert. A common [18F]-FDG PET radiomic signature to predict survival in patients with HPV-induced cancers. Submitted to European Journal of Nuclear Medicine and Molecular Imaging, December 2022.
- **Stéphane Niyoteka**, Rahimeh Rouhi, Pierre-Antoine Laurent, Samir Achkar, Alexandre Carré, Roger Sun, Sebastien Diffetocq, Corinne Balleyguier, Cyrus Chargari, Eric Deutsch, Charlotte Robert. MR image normalization in locally advanced cervical cancer: what is the optimal method for developing multicenter radiomic models? Submitted to Radiology AI, April 2023.
- Rahimeh Rouhi, **Stéphane Niyoteka**, Alexandre Carre, Samir Achkar, Pierre-Antoine Laurent, Mouhamadou Bachir Ba, Cristina Veres, Théophraste Henry, Maria Vakalopoulou, Roger Sun, Sophie Espenel, Adrien Laville, Cyrus Chargari, Eric Deutsch, Charlotte Robert. Automatic GTV segmentation with failure detection for safe implementation in locally advanced cervical cancer. Submitted to International Journal of Radiation Oncology, Biology, Physics, March 2023.

Appendix D

Substantial French Summary

Il est établi que les papillomavirus humains (HPV) oncogènes, l'une des infections les plus sexuellement transmissibles, sont responsables d'une forte proportion des cancers du col de l'utérus localement avancés (LACC), des cancers de l'anus (ASCC) et de l'oropharynx (OPSCC). Le dépistage pour une détection précoce et la campagne de vaccination contre les virus HPV font partie des solutions pour réduire le nombre de cancers HPV-induits, mais des efforts supplémentaires doivent être faits pour améliorer la prise en charge et le traitement de ces cancers. Le développement de schémas thérapeutiques combinant des modalités de traitement complémentaires telles que la chimioradiothérapie (CRT) et l'immunothérapie semble être une option thérapeutique particulièrement prometteuse. Dans ce contexte, la recherche de biomarqueurs spécifiques à la réponse au traitement reste un enjeu majeur, notamment pour identifier les patients non-répondeurs aux thérapies conventionnelles. Les récentes innovations technologiques en informatique, à savoir le Machine Learning (ML) et surtout le Deep-Learning (DL) basé sur l'analyse d'images médicales, i.e la radiomique, pourraient fournir des outils complémentaires et puissants pour une meilleure compréhension du cancer. La radiomique consiste à extraire des caractéristiques quantitatives d'images médicales, caractérisant en particulier la forme de la tumeur, ses intensités sur l'image et son hétérogénéité spatiale. Les algorithmes de ML basés sur les caractéristiques radiomiques ont déjà démontré leur pouvoir d'identifier des motifs complexes permettant l'élaboration de modèles phénotypiques, pronostiques et prédictifs robustes. Ces dernières années ont vu le développement de nouvelles techniques de traitement allant de plus en plus vers une médecine de précision notamment parl'immunothérapie, les thérapies ciblées, et des efforts ont été faits pour personnaliser les schémas standard de chimiothérapie (CT) et de radiothérapie (RT). Malgré ces avancées, dans la plupart des cas, les patients atteints du même type de cancer sont toujours traités de la même manière, en particulier en ce qui concerne la RT et la CT, où ils reçoivent les mêmes doses de manière standard.

La radiomique, et plus précisément l'IA appliquée à la radiomique, est un domaine très prometteur et en pleine évolution pour le traitement personnalisé du cancer, comme le montre le grand nombre d'études sur le sujet. De nombreuses sources de données ont démontré que la radiomique révèle des informations réelles sur les cellules tumorales qui vont au-delà de l'analyse visuelle et qui pourraient être exploitées pour la détection précoce du cancer, l'optimisation du traitement et la prédiction de l'issue du traitement. Cependant, certaines limitations entravent encore le domaine de la radiomique. La principale limite à la mise en œuvre clinique est, de manière générale, le manque de grands ensembles de données structurés et de haute qualité qui sont

correctement annotés. Cela a un impact fort et négatif sur la généralisation des modèles développés dans la vie réelle. L'autre limite des études radiomiques est le manque d'outils de standardisation pour les images multi-institutionnelles affectées par 'l'effet de centre'. Dans cette thèse, plusieurs approches ont été mises en œuvre pour remédier à ces limitations.

Dans le premier projet, des efforts ont été faits pour collecter un large ensemble de données d'images [18F]-FDG TEP provenant de 5 institutions indépendantes avec l'objectif de construire un modèle robuste pour la prédiction de la survie chez les patients atteints de cancers induits par le HPV, à savoir LACC, ASCC, et OPSCC. Nous avons recueilli des images TEP de patients traités pour LACC au Gustave Roussy Campus Cancer (104 patients) et au Leeds Teaching Hospitals NHS Trust (90 patients). Nous avons également recueilli des images TEP d'ASCC de l'Institut du cancer de Montpellier (66 patients) et de Oslo University Hospital (67 patients). Une cohorte supplémentaire de patients atteints d'OPSCC a également été recueillie auprès de University Hospital of Zurich (45 patients). La technique d'harmonisation ComBat a été appliquée aux caractéristiques radiomiques. Nous avons mis en œuvre une méthode robuste de sélection des caractéristiques couplée à l'ajustement des hyperparamètres, le "modified consensus nested cross-validation", au sein de quatre modèles ML pour prédire la survie sans progression (PFS) et la survie globale (OS). Chaque modèle a été entraîné et optimisé sur les cohortes LACC et ASCC et testé sur des patients LACC, ASCC et OPSCC indépendants. Le modèle CoxNet basé sur la radiomique a obtenu les meilleurs résultats avec des valeurs de C-index de 0,75, 0,78 pour la PFS, et 0,76, 0,74, 0,75 pour l'OS sur les ensembles de test. Les modèles basés sur les caractéristiques radiomiques ont eu des performances supérieures à celles des modèles biocliniques et la combinaison des variables radiomiques et biocliniques n'a pas amélioré les performances. Les modèles basés sur le volume tumoral métabolique (MTV) ont obtenu des valeurs de C-index inférieures pour la majorité des configurations testées, mais des performances identiques ou supérieures pour l'identification des rechutes précoces. Les résultats démontrent la possibilité d'identifier des signatures d'image communes basées sur la TEP pour prédire la réponse des patients présentant une pathologie HPV induite, validée sur des données multiconstructeurs multicentriques. Cependant, la valeur ajoutée de la radiomique reste à démontrer par rapport aux mesures quantitatives conventionnelles. À notre connaissance, c'est la première fois qu'une signature radiomique commune a été développée pour différents cancers HPV-positifs. Cette étude pourrait donc ouvrir la voie à l'optimisation du traitement des cancers induits par le HPV, en guidant le clinicien sur les compromis de couverture de dose à mettre en œuvre dans les étapes de planification de la RT et/ou de la curiethérapie.

Dans le second projet, nous avons développé une méthode de standardisation basée sur le GAN pour les images IRM T2 chez des patients atteints de LACC. Nous avons proposé une méthode basée sur un cGAN et une méthode basée sur un CycleGAN pour la standardisation des images MR et les avons comparées aux méthodes conventionnelles de prétraitement et a posteriori proposées dans la littérature (Nyul, normalisation z-score et ComBat). Dans un premier temps, les images RM T2w de 30 patients traités pour un LACC ont été acquises de manière prospective. Pour chaque patient, trois images ont été prises séquentiellement sur le même scanner avec des valeurs fixes de temps de répétition (TR) et de taille de voxel (VS). Un cGAN a été entraîné sur ces images appariées pour générer des images robustes à l'impact de la modulation du TR et de la VS. L'impact des méthodes de standardisation a été évalué au moyen d'une analyse en composantes principales (PCA) sur les mesures de qualité d'image extraites des images avant et après la standardisation. En utilisant le

coefficient de corrélation intraclasse (ICC) et le coefficient de corrélation de concordance (CCC), des variables radiomiques robustes ont également été analysées (CCC et ICC 0,75). La PCA sur les métriques de qualité d'image a montré que les changements de TR et de VS étaient les plus atténués avec la standardisation basée sur le cGAN. En ce qui concerne la modulation TR/VS, le cGAN a obtenu les meilleurs résultats avec 100% (18/18) et 77% (58/75) de caractéristiques radiomiques robustes après standardisation, sur les caractéristiques de premier ordre et de second ordre, respectivement. Dans un deuxième temps, une cohorte rétrospective de 216 patients atteints de LACC a également été collectée, comprenant 86 et 160 images IRM T2 prises avant la radiothérapie (RT) et la curiethérapie (BT), respectivement. Une méthode de standardisation basé sur le CycleGAN pour les images non appariées a été entraînée, ainsi qu'une standardisation Nyul et z-score sur les images rétrospectives. Différents modèles ML ont été entraînés pour étudier l'impact de ces méthodes de standardisation sur la classification des stades et la prédiction des rechutes, respectivement. CycleGAN a obtenu les meilleurs résultats sur les deux tâches par rapport aux autres méthodes de standardisation. Sur les images pré-RT, la précision moyenne de la classification des stades a été améliorée sur le jeu de test de 0.68 ± 0.16 à 0.83 ± 0.07, 0.78 ± 0.04, et 0.88 ± 0.09 avec la standardisation Nyul, la normalisation z-score, et la standardisation basé sur le CycleGAN, respectivement. De même, la prédiction des rechutes basée sur les images pré-BT a obtenu une meilleure AUC sur l'ensemble des tests après la standardisation des images, le meilleur modèle atteignant 0,44 avant la standardisation, 0,52, 0,56, et 0,60 après la standardisation par z-score, Nyul, et CycleGAN, respectivement. Nos résultats suggèrent que la standardisation basée sur un réseau neuronal pourrait mieux limiter l'impact des paramètres d'acquisition sur les caractéristiques radiomiques de l'IRM que les méthodes conventionnelles dans le cadre du LACC. Cette approche pourrait être utilisée dans une stratégie d'harmonisation unique englobant les étapes de prétraitement des images sans qu'il soit nécessaire de disposer d'informations a priori sur les paramètres d'acquisition et de reconstruction de l'imagerie.