POLITECNICO
MILANO 1863

SCUOLA DI INGEGNERIA INDUSTRIALE
E DELL'INFORMAZIONE

Dosiomics analysis to predict local recurrence in patients affected by skull-base chordoma treated with proton therapy

TESI DI LAUREA MAGISTRALE IN
BIOMEDICAL ENGINEERING
INGEGNERIA BIOMEDICA

Author: **Gloria Rizzato**

Student ID: 967036
Advisor: Chiara Paganelli
Co-advisors: Letizia Morelli, Giovanni Parrella
Academic Year: 2021-22

Summary

Skull-base chordoma (SBC) is a rare and radioresistant tumour and its complete surgical resection is often unfeasible. Particle therapy with proton beams is a promising approach to control SBC tumour progression thanks to its improved physical properties (i.e. linear energy transfer, LET) and radiobiological effectiveness (RBE) with respect to conventional radiotherapy. At the same time, the extraction of quantitative features from patient-specific medical imaging has become a promising approach for treatment outcome prediction. Quantitative features can in principle be extracted from any volumetric map, such as also dose maps defined during the radiotherapy treatment planning, with the so-called dosiomics framework. The purpose of this thesis was to investigate dosiomics in predicting local recurrence for SBCs patients treated with proton therapy, by exploring how different features extraction settings can influence model's performance. Features were extracted from dose-averaged LET (LET$_d$), physical and RBE-weighted dose maps by varying the setting of parameters related to dose map intensity discretization (binwidth/bincount), and then fed into a prediction model of local recurrences. Two features selection approaches have been implemented to evaluate models' performance by varying the binwidth/bincount values across each dose map type (approach 1) and by fixing the binwidth/bincount value across different dose maps (approach 2). In both cases, stratified 5-fold cross-validation routine repeated 10 times was used and Harrell Concordance index (CI) employed to evaluate models' performance. In approach 1, a greater stability was obtained in models trained by varying the bincount parameter with respect to the binwidth parameter for physical dose (CI 0.62) and LET$_d$ (CI 0.60); for RBE-weighted dose no marked difference in models' performance stability was present between the two intensity discretization methods. In approach 2, a higher number of significant first-order features were found in LET$_d$ for both binwidth/bincount methods with respect to physical and RBE-weighted dose maps. Concerning models' performance, by fixing the binwidth value, better models' performance was obtained for physical and RBE-weighted dose maps; whereas, by fixing the bincount value, better performance was obtained for LET$_d$ maps rather than physical dose maps in most cases. Nevertheless, no relevant findings have been found on the optimal value of binwidth/bincount to be used in the dosiomics pipeline, but further investigations are needed, also considering a larger population.

Key-words: dosiomics, local recurrence, proton therapy, skull-base chordoma, dose maps, intensity discretization.

Sommario

Il cordoma della base cranica (SBC) è un tumore raro e radioresistente e la resezione totale è spesso infattibile. La terapia con raggi protonici è un approccio promettente per controllare la progressione del tumore grazie alle sue migliori proprietà fisiche (i.e. trasferimento lineare di energia, LET) e all'efficacia radiobiologica (RBE) rispetto alla radioterapia convenzionale. Allo stesso tempo, l'estrazione di caratteristiche quantitative da immagini biomediche paziente-specifiche si è rivelata un approccio promettente per la predizione di outcome del trattamento. In linea di principio, caratteristiche quantitative possono essere estratte da qualsiasi mappa volumetrica, come anche da mappe di dose definite durante la pianificazione del trattamento di radioterapia, attraverso la cosiddetta dosiomica. Lo scopo di questa tesi è quello di investigare la dosiomica nella predizione della recidiva locale in pazienti affetti da SBC trattati con protonterapia, esplorando come diversi settings di estrazione delle caratteristiche possano influenzare le performance dei modelli. Le caratteristiche sono state estratte da mappe di LET mediate sulla dose (LET_d), mappe di dose fisica e di dose ponderate su RBE variando il setting dei parametri relativi alla discretizzazione dell'intensità della mappa della dose (binwidth/bincount), e poi inserite in un modello di previsione delle recidive locali. Due approcci di selezione delle caratteristiche sono stati implementati per valutare la prestazione dei modelli variando i valori binwidth/bincount per ogni tipo di mappa di dose (approccio 1) e fissando il valore binwidth/bincount per diverse mappe di dose (approccio 2). In entrambi i casi è stata utilizzata una routine di cross-validazione con un k-fold di 5 ripetuta 10 volte e l'indice di concordanza di Harrell (CI) è stato impiegato per valutare le prestazioni dei modelli. Nell'approccio 1, è stata ottenuta una maggiore stabilità nei modelli addestrati variando il parametro bincount rispetto al parametro binwidth per la dose fisica (CI 0.62) e per LET_d (CI 0.60); per la mappa di dose ponderata su RBE nessuna differenza marcata nella stabilità delle prestazioni dei modelli era presente tra i due metodi di discretizzazione dell'intensità. Per quanto riguarda la prestazione dei modelli, fissando il valore binwidth, sono state ottenute prestazioni migliori per le mappe di dose fisiche e di dose pesate su RBE; mentre, fissando il valore bincount, prestazioni migliori sono state ottenute per le mappe LET_d rispetto alle mappe di dose fisiche nella maggior parte dei casi. Tuttavia, non sono stati trovati risultati rilevanti sul valore ottimale di binwidth/bincount da utilizzare nella pipeline dosiomica, ma sono necessarie ulteriori indagini, anche considerando una popolazione più ampia.

Parole chiave: dosiomica, recidiva locale, protonterapia, cordoma della base cranica, mappe di dose, discretizzazione dell'intensità.

Contents

1 Introduction

1.1. Chordoma

Definition and origin

Chordomas are neoplasms which develop at the level of the spine; mostly they arise at the ends of it, i.e. in skull-base or at the level of sacrum, but they can grow from any anatomical location along the length of spine. Chordomas are radioresistant malignant cancers, locally aggressive, invasive and characterized by slow-growing [1].

Their origin may derive from undifferentiated notochordal remnants which remain in vertebral bodies and axial skeleton. The notochord is a flexible rod-shaped embryologic structure which develops in the third week of gestation and serves as primitive axial skeleton. As the vertebral column develops, the notochord degrades in what later becomes the nucleus pulposus: after birth, notochordal cells are replaced by small cartilage-like nucleus pulposus cells. In early adulthood this process is usually concluded. However, there are cases where foetal notochord cells still persist in the axial skeleton, and especially at the caudal and cranial ends, through life: it is thought that those remnants may proliferate into chordoma [2] [3]. During years, different benign tumours with notochordal cell origin have been discovered: it is still a discussion whether those benign tumours may undergo to malignant transformation leading to chordoma.

Incidence

The incidence of chordoma is 0.08/100,000 people/year and it most arise in patients between 40 and 60 years of age, but it can occur also in childhood and teenagerhood (less than 5% of cases occurs before the age of 20 years). In Europe, as for primary site distribution, 36% of chordoma cases are sacral, 30% cranial, 23% spinal and 11% have other primary sites distribution [4] [5].

Clinical presentation

Chordomas show a slow growth rate, hence they are clinically silent in early stage: the lack of symptoms in the first phase makes the tumour to be quite large when manifestations appear. Depending on the anatomical location of the lesion, patients affected by chordoma have different signs: head-ache, neck pain and cranial nerve

palsies arise in case of skull-base chordomas (SBCs), chronic back pain or urinary/bowel dysfunction due to nerve root compression are present in case of spinal and sacral chordomas [6].

In this thesis, a study on patients affected by skull-base chordoma was conducted. SBCs are deep-seated tumours, closely located to vital structures such as optic nerves and brainstem (Figure 1. 1 [7]). More in details, SBCs commonly arise in the spheno-occipital synchondrosis of the clivus, basiocciput clivus, or at the petrous apex, while intradural areas, sella turcica, sphenoid sinus, nasopharynx, maxilla, and paranasal sinuses are less common sites of origin. Patients with this kind of lesion show different symptoms based on the origin site: among the main ones there are dysphagia, chocking on water, decreased pharyngeal reflex due to cranial nerves compression [5] [7].

Figure 1. 1 - Contrast-enhanced sagittal T1 (a) and axial T2 (b) MRI exams of a 56-year-old male with a skull-base chordoma (white arrows) with brainstem compression (white stars).

The access to the lesion is often difficult due to its deep anatomical position. For this reason, a detailed histological analysis is not always possible leading to a poor characterization, which is fundamental for treatment decision making.

Diagnosis - Imaging

Computed tomography (CT) and magnetic resonance imaging (MRI) are the two imaging techniques which allow the characterization of SBC and surrounding tissues. In particular, the delineation of the lesion and surrounding soft tissues are crucial for the optimization of the treatment that will be delivered.

MRI allows the characterization of tumours' morphologic characteristics and extension thanks to the higher intrinsic soft tissue contrast than CT. This technique is able to identify the intradural extension and allows the delineation of margins of both tumour and relevant soft tissue structures which will be useful during the treatment phase. On contrary, CT is involved to discover osseous invasions and bone abnormalities: this technique is used to delineate the extent of bony involvement [7].

Skull-base chordoma: therapeutical approaches

Different kinds of therapeutical approaches can be considered for the treatment of SBC patients and their application depend on several factors such as the nature of the neoplasm, its localization, its diffusion and on patient's overall health condition. Typically, surgical resection is the first therapeutic strategy that is commonly applied: endoscopic endonasal approach (EEA) is commonly performed, but also minimally invasive open techniques such as microsurgical craniotomy are considered. SBCs are surrounded by critical anatomical structures such as optic system, carotid artery and brainstem that hamper the complete surgical eradication of the tumour, with the frequent identification of local residuals. It is known that SBC will recur from residuals of cells tumour or remnants (even if microscopic) located where surgical resection was applied. Therefore, radiotherapy is the recommended technique to be applied in combination with surgical resection [9] [10]. In literature different studies show the efficacy of combination of surgery and radiotherapy with respect to surgery alone in obtaining local control [11].

The ultimate objective in radiation oncology is the achievement of local control: local control of the primary site consists in the achievement of a condition of no-progression or complete eradication of the tumoral mass (favourable event), in contrast to tumour progression/local recurrence (adverse event). SBC is a radioresistant tumour if managed with conventional fractionated radiotherapy: this is challenging for the application of radiotherapy treatment, hence particle therapy has revealed to be a promising approach since it allows a dose escalation to the tumour while sparing healthy structures surrounding the lesion [6] [10]. Detailed explanations of

conventional fractionated radiotherapy and particle therapy are reported in paragraph 1.2 and 1.4, respectively.

Local control and survival rate

Although the combination of surgery with radiotherapy has been demonstrated to achieve greater local control and greater overall survival (i.e. the global survival, so if the patient will survive) and some improvements have been reached with particle therapy (such as proton and carbon ion radiotherapy), tumour response remains not satisfactory [4]. Considering the 5-year survival rate it is estimated to be 45% for conventional radiotherapy, 87% for carbon ion radiotherapy [12], while as for 5-year local control, it is 48–60% [13] and 72% [14], respectively. The high tendency to recur after the first treatment represents one of the major clinical challenges.

Regarding metastases, they may occur in bone, lungs, skin and liver with a percentage between 6 and 30, but their impact is still to be clarified [7].

1.2. Conventional radiotherapy

Radiation therapy is the use of directed X-rays or subatomic particles primarily for cancer management in both curative and palliative settings [15]. Ionizing radiations are targeted to malignant cells inducing apoptosis, one of cell death that can occur after irradiation [16]. The interaction between cell DNA and radiation provokes an irreparable DNA damage leading to cell death, thus avoiding proliferation and progression of tumour [17]. However, also healthy cells are involved in radiation damages, but their tolerance to radiations differs from the one of tumoral cells. The aim of radiotherapy is to reach maximal therapeutic benefit in terms of probability of tumour local control while minimizing side effects to surrounding healthy structures.

Conventional radiotherapy is performed by the usage of high-energy photons generated by linear accelerator mounted on a rotation gantry. Photons are referred to as indirectly ionizing radiation: the interactions between photons and tissue generate fast moving electrons which will directly provoke damages to DNA cells by propagating through tissue [18]. Radiation dose refers to the amount of energy absorbed per unit mass at a specific point, in other words, it describes how much energy from ionizing radiation has been absorbed in a small volume centred at a point and it is measured in Grays (1Gy=1J/kg) [19]. Considering the dose-depth curve, photons deposit a significant amount of dose at body surface entrance since the entrance dose decreases exponentially while increasing tissue depth. To better conform the dose to tumour volume while sparing healthy tissues, intensity modulated radiotherapy (IMRT) is one of the techniques which can be implemented: it allows irradiating the tumour from different directions while modulating beam intensities to vary delivery in depth. IMRT is realised with a moving multi-leaf collimator able to conform the dose to tumour volume and the dose is delivered by a medical linear accelerator mounted on a gantry able to rotate around the patient [20].

1.2.1. Imaging in radiotherapy

Imaging in radiotherapy plays a fundamental role in every phase of radiotherapy process. Firstly, images can be considered as a diagnostic tool to identify the presence of a tumour, then during treatment planning phase, they allow delineating the tumour and healthy structures. Concerning treatment delivery phase, images are involved to monitor the accuracy of treatment and lastly, during follow-up, imaging can be considered to verify if the desired treatment outcome has been achieved monitoring the treatment response.

To reach tumour local control while minimizing side effects to healthy organs, an accurate localization and extension of both tumour and normal structures must be performed. To accomplish this, imaging modality for delineation of structures must have high spatial resolution (for an accurate contouring), high sensitivity (to individuate all structures involved in lesion) and high specificity (to identify all normal tissues).

Concerning computed tomography, it plays a fundamental role in treatment planning. In a standard diagnostic CT scanner, an X-ray source and a detector are mounted on a rotational gantry which moves around the patient. Due to high spatial resolution, a three-dimensional and detailed image of the internal anatomy of the patient is produced by CT scan allowing the contouring of the lesion and healthy structures: still nowadays, CT is considered the mainstay in segmentation [21]. CT images are also given as input in treatment planning systems to calculate the dose which will be deposited in both tumour and healthy structures since they provide density tissue information (in Hounsfield units). However, the main disadvantages of this technique are the lack of contrast in soft tissue and artifacts (for example due to the presence of metal implants such as coils, stents and dental fillings).

Magnetic resonance imaging provides a great soft tissue contrast thanks to the magnetic properties of hydrogen protons within a tissue, thus MRI can complement CT images during treatment planning improving target and organs at risk delineation. Contrary to CT technique, MRI has the advantage to acquire images without exposing the patient to further ionizing radiation [21]. As for disadvantages, geometric distortions appear deforming images (including structures of interest). MRIs, nevertheless, can not be used alone in radiotherapy treatment planning due to the absence of information about tissue density needed to calculate dose deposition.

Positron emission tomography (PET) involves radioactive tracer (i.e. a positron emitter attached to targeted biologically active molecules) to construct three-dimensional images of tracer distribution within the body. PET images provide information about biological characteristics of tumour such as metabolism, oxygenation and proliferation allowing the identification of radio-resistance areas. Since the poor spatial resolution, PET images are never used alone for segmentation purpose, instead the integration of PET and CT scanners may lead to a better contouring of tumour with respect to CT alone [22].

As cited before, images are also involved in treatment delivery to improve tumour targeting while sparing healthy tissues and in follow-up to monitor treatment's results.

Before each treatment delivery, cone beam CT images are acquired to correct errors (such as patient positioning, gain/loss weight, tumour swelling/reduction) which may occur just before treatment delivery, thus influencing the success of treatment (see next paragraph for major details). Cone-beam CT technique is the most used technique to compensate for day to day variations, but recently novel systems based on MRI combined with a linear accelerator are emerging [23].

1.3. Radiotherapy workflow

The radiotherapy workflow can be summarized into two main phases: treatment planning and treatment delivery. Before these two stages, a phase of patient's diagnosis is conducted by collecting data such as patient's medical history, his symptoms and comorbidities, information related to previous diagnostic studies and images where cancer disease can be identified [24].

Treatment planning

Different stages characterize the treatment planning process. First of all, the acquisition of CT images (and MRIs if characterization of soft tissue is required when contouring) allows the construction of a three-dimensional patient specific model where tumour and relevant healthy organs are defined. Different are the techniques which can be employed in segmenting structures, but manual segmentation is still considered the ground truth. This technique requires a radiation oncologist which manually contour relevant structures. Focusing on tumour segmentation, four different volumes are generated according to International Commission on Radiation Units (ICRU, report 50 [25] and report 62):

- Gross Tumour Volume (GTV) representing the tumour volume which can be detected by CT and MRI and defined as the palpable or visible extent of the malignant tumour;

- Clinical Tumour Volume (CTV) made by GTV with additional margins including probable microscopic diseases and other areas (i.e. lymph nodes);

- Internal Target Volume (ITV) which includes CTV and internal margin taking into account possible uncertainties related to variations in size and position of CTV due to organ motions;

- Planning Target Volume (PTV) where are considered not only all possible geometrical uncertainties of the same nature of the ones of CTV and internal margin, but also all possible systematic (tumour reduction/swelling, loss/gain in patient's weight) and random set-up errors (patient positioning).

In Figure 1. 2 different volumes defined for radiation therapy planning are reported.
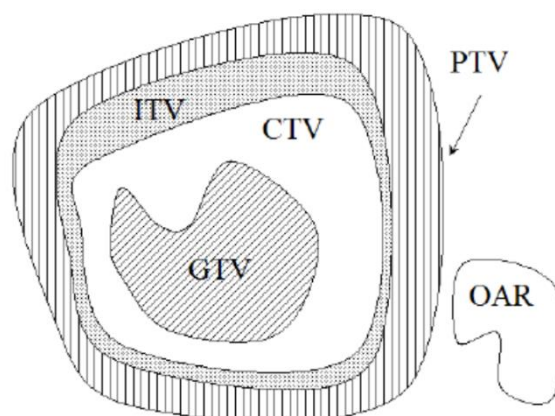
Figure 1. 2 – Defined volumes in radiation therapy planning. OAR = organ at risk

The segmentation of CTV plays an important role since this will be the irradiated target determining the achievement of local control, but also the reduction of set-up errors and organ motion understanding are essential to optimise the dose delivered to tumour while preserving organs at risk. During segmentation a critical aspect regards inter-operator variability: as reported in different studies (such as [26] and [27]), even if contouring is done by expert radiation oncologists, significant differences in target delineation are still present. Those uncertainties can lead to underdosing or overdosing: the result will be a decrease in the probability of obtaining local control or an increase in risk of toxicities. Semi-automatic and fully automatic segmentations methods are being developed to overcome inter-operator variability [28].

After having segmented all the relevant structures, the optimal patient treatment plan is generated involving treatment planning system (TPS) with the aim to achieve an optimal dose distribution according to prescribed dose objectives for target volume and doses constraints for surrounding healthy tissues. TPS makes use of CTs to generate treatment plan and it includes optimization of several parameters such as beam positioning, directions, shapes, but also dose calculation based on patient anatomy and type of particle involved to treat the tumour [29]. While designing a treatment plan, fractionation is also taken into account, indeed it can influence the effect of treatment. Fractionation is related to the way the dose is given to the patient over time. By fractionating the dose delivered to malignant cells in daily fractions, the number of tumoral cells that will survive decreases due to less ability to repair damages over short periods of time with respect to normal tissue cells (Figure 1. 3) [30].
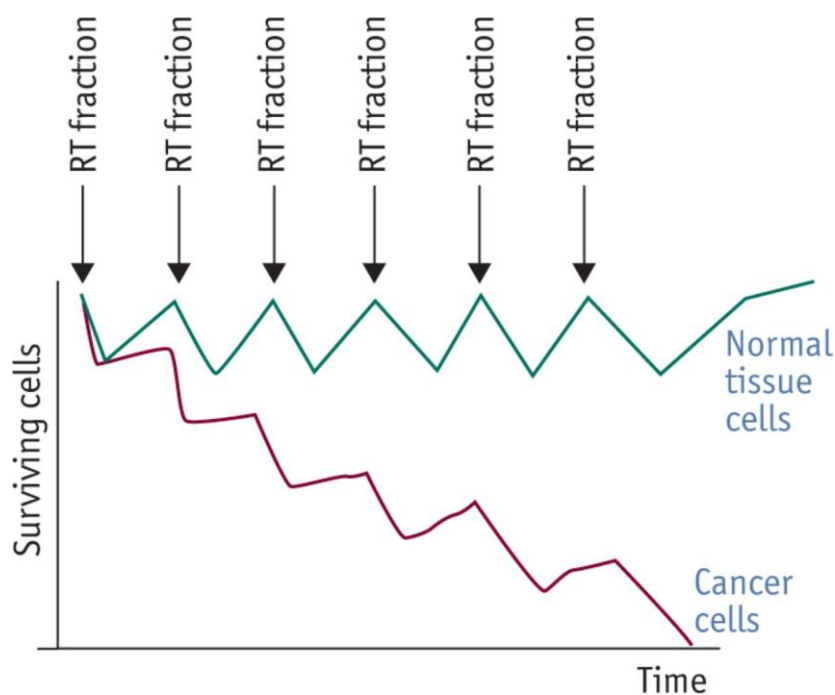
Figure 1. 3 [30] – Surviving cells curves of cancer cells and normal tissue cells over time. Fractionation technique for dose delivery enhances the difference between malignant and normal cells to repair damages provoked by irradiation.

Isodose maps representing lines of constant absorbed dose and dose volume histograms (i.e. a plot of cumulative dose-volume frequency distribution) are graphical tools used in treatment planning summarizing the simulated radiation dose distribution through target volume and anatomical structures of interest [31]. Examples of axial CT and three-dimensional view with isodose lines for a brain treatment plan are shown in Figure 1. 4 (left) and Figure 1. 4 (right) [32], instead in Figure 1. 5 [32] the cumulative dose volume histogram is reported.
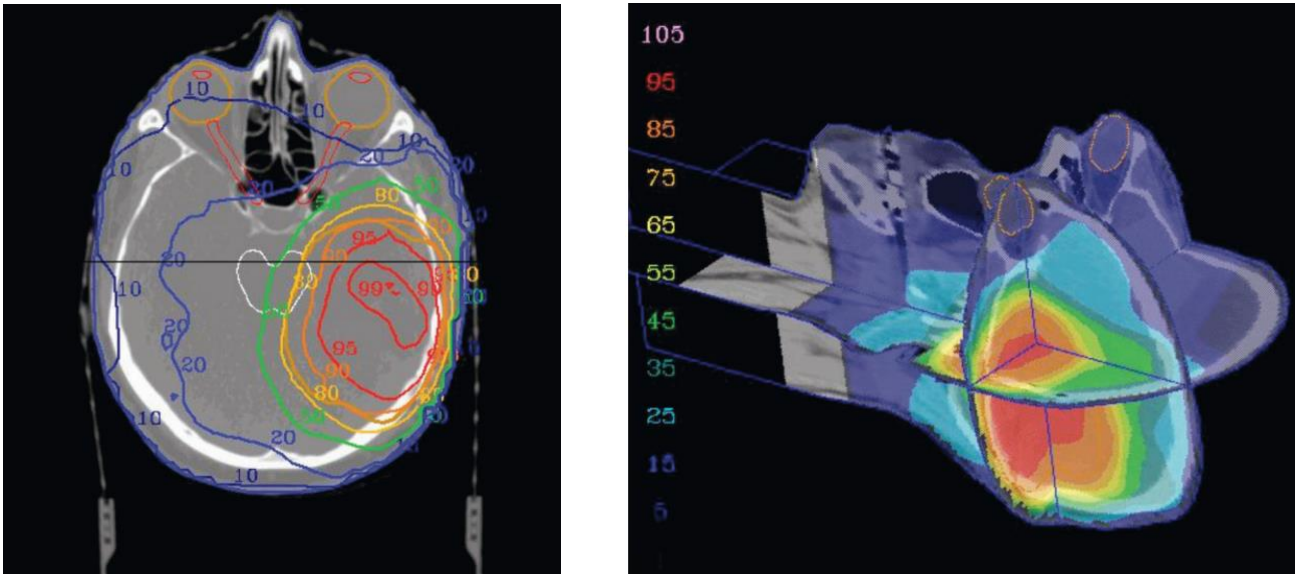
Figure 1. 4 [32] – left: Axial CT with isodose lines, right: Three-dimensional axonometric view with isodose map.
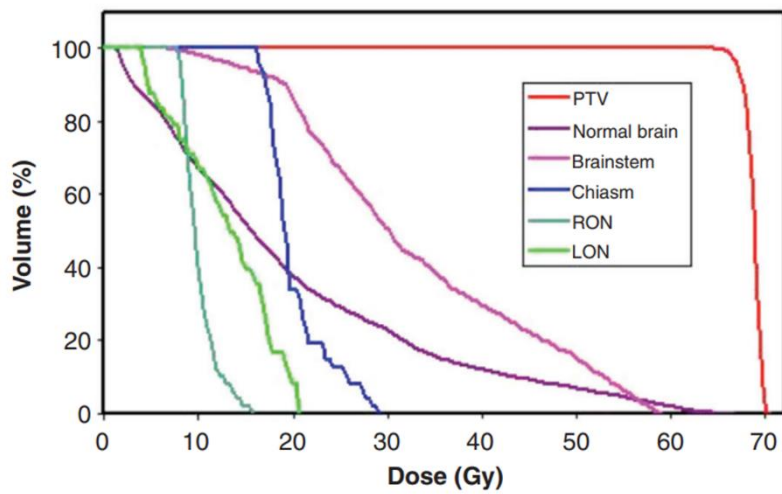

Figure 1. 5 [32] – Cumulative dose volume histogram for a brain treatment plan. PTV = planning target volume, LON = left optic nerve, RON = right optic nerve.

Treatment setup and delivery

After some checks performed by medical physicist and radiation oncologist (such as verification of plan made by radiation oncologist and quality assurance checks related to technical components involved in treatment delivery), the phase of treatment delivery starts. Treatment plan has to be precisely and accurately replicated during delivery phase and in every irradiation session, so imaging techniques are required for treatment verification. One of the crucial steps in delivery phase regards the patient's positioning: patient has to be placed in the same position at which treatment plan was built, indeed any difference in patient positioning with respect to the original plan leads to errors in dose delivery (both to the target and to healthy structures). The introduction of image guided radiotherapy (IGRT) helps in detecting and correcting deviations associated to patient positioning and motion by the acquisition of in-room images. In addition to patient set-up error, anatomical changes such as patient weight gain/loss and alterations in tumour geometries (shrinkage/growth) occur between different fractions. As previously mentioned, Cone-beam CT (CBCT) is the most used technique to quantify errors before treatment delivery: it allows the acquisition of in-room CT images which will be compared with the ones acquired in treatment planning to correct errors due to patient positioning and anatomo-pathological variations. It is useful to point out the fact that with CBCT the dose delivered to patient increases because of the usage of X-rays to collect CT images. If variations between treatment plan and delivery can not be ignored, treatment plan has to be adapted to those variations [29].

1.4. Particle Therapy

Particle therapy refers to the clinical use of ion beams (i.e. protons or carbon ions) to treat patients with tumours. Cyclotrons or synchrotrons (i.e. particle accelerators) are involved to produce monoenergetic beams which will be spread to conform the dose to target volume. With respect to conventional X-rays radiotherapy, particle therapy presents both physical and radiobiological advantages, which make this advanced treatment particularly suited for deep-seated and radioresistant tumours [33].

Physical characteristics

A substantial clinical advantage in using particle therapy with respect to conventional radiotherapy relies in the dose-depth characteristic of ion beam. As shown in Figure 1. 6a [34], while for photon the dose decreases exponentially as function of depth, charged particle beam can penetrate through tissues depositing low energy at the surface of the body and most of its energy near the end of the track, known as the Bragg peak. By changing the velocity of the ion beam it is possible to control the depth at which the maximum dose deposition occurs, indeed an ion beam with higher kinetic energy travels longer ranges within the patient body.

The physical characteristic of low entrance dose and no dose beyond a certain range of ion beam results in a highly-localized dose distribution, thus reduction of integral dose outside the target volume. Due to the fact that for a given energy of particles a narrow Bragg peak is created (the dose deposition regards a narrow depth in tissue), the necessity to spread particle energy to cover the entire volume of the lesion is needed: beams of different energies are superimposed thus creating a spread-out Bragg peak (SOBP). This effect can be seen in Figure 1. 6b [15]. SOBP can be obtained by active beam scanning, also called pencil beam scanning: this technique uses magnets (one for vertical, the other for horizontal steering) to steer charged particles of the beam towards the target so that he tumour is irradiated voxel by voxel and layer by layer. Due to dose conformality of particle therapy (i.e. a higher geometrical selectivity), a higher dose deposition in tumour region can be obtained as to optimize tumour local control probability, minimizing the dose deposited to surrounding healthy tissues [35]. This is the reason why charged particles are usually looked at when dealing with cancers at significative depth and near to sensitive healthy structures (e.g. brainstem). In addition to the physical advantages of particles due to the depth-dose distribution compared to the one of X-rays, particles show radiobiological advantages related to the density of dose deposition around Bragg-peak, indeed due to interactions between particles and cell's DNA, more severe

damages can be produced [36]. To account for biological aspects, relative biological effectiveness and linear energy transfer have to be introduced.
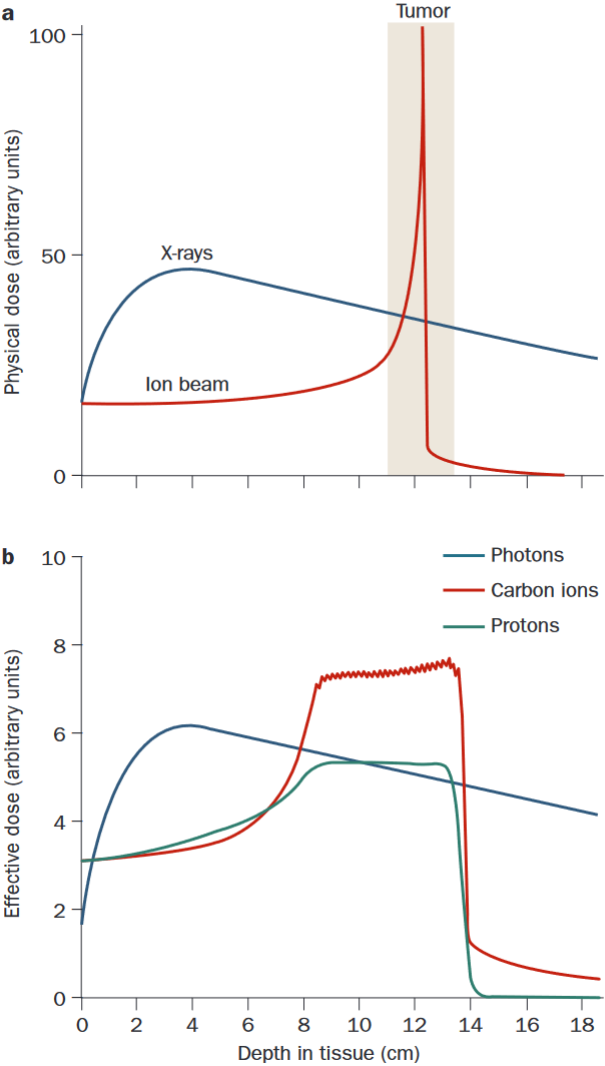

Figure 1. 6 [15] – a: Depth–dose relationships for X-rays and ion beams. Bragg peak is highlighted with rectangular box. b: Spread-out Bragg peak for particle beams.

Relative biological effectiveness

Different ionizing radiations leads to different biological effect. To understand the radiobiological response of different particles on tissues, relative biological effectiveness (RBE) has to be characterized. RBE is defined as the ratio between the absorbed dose of a reference beam (commonly the photon beam, e.g. 250 kVp X-rays or [60]Co γ-rays) and the one of a test radiation (protons, carbon ions) to produce the same biological effect [15]. In Equation 1. 1 RBE formula is reported.

$$RBE|_{biological\ effect} = \frac{Absorbed\ dose\ of\ reference\ beam}{Absorbed\ dose\ of\ test\ radiation} \qquad (1.1)$$

Equation 1. 1- RBE definition

For protons RBE is typically set at 1.1, whereas for carbon ions, an approximated value of 3.0 is taken into account. However, it must be pointed out that RBE is not a constant value, but it can vary depending on several different factors such as linear energy transfer (LET), particle species, tissue radiosensitivity and microenvironment [37]. Biologically, protons and carbon ions cause cellular damage more effectively than photons, this is the reason why protons and carbon ions are involved when dealing with radioresistant tumours [36].

Considering the different biological effect of different ionizing radiations, in clinical practise the absorbed radiation dose is weighted by RBE factor which depends on the type of radiation used. This quantity is called biologically effective dose and it is measured in Gy(RBE) (for example for X-rays the weighing factor is 1.0, so the equivalent dose is 1Gy(RBE)) [38].

Linear energy transfer

Linear energy transfer (LET) describes the amount of energy that a particle transfers to the material per unit length of track. In Equation 1. 2 LET formula is reported: dE represents the amount of energy that is lost by particle while travelling the distance dx. The LET unit of measure is keV/μm [37].

$$LET = \frac{dE}{dx} \qquad\qquad (1.2)$$

Equation 1. 2 – LET definition

LET value depends on the charge and on the velocity of the particle: if charged particle slows down, LET value increase.

Due the interactions occurring between particles and matter, energy loss and secondary ion emissions occur [39]. For this reason, different particle species has to be accounted for LET calculation: primary and secondary particles are taken into account in case of protons, while for carbon ions, LET is an averaged value resulting from different fragments until oxygen.

Particle energy spectrum has to be known to characterise beam at a specific location. Typically, track-averaging LET (LET_t) and dose-averaging LET (LET_d) are the considered averaged LET metrics to represent particle spectrum.

In Equation 1. 3 the definitions of LET_t and LET_d are reported [39].

$$LET_t = \frac{\sum_i \phi_i \cdot LET_i}{\sum_i \phi_i} \qquad\qquad (1.3a)$$

$$LET_d = \frac{\sum_i d_i \cdot LET_i}{\sum_i d_i} \qquad\qquad (1.3b)$$

Equation 1. 3—a: LET_t definition. Φ_i represents the fluence of charged particle *i* having a certain energy and charge and a certain LET_i for the medium of interest. b: LET_d definition. d_i represents the microscopic dose given in an infinitesimally small volume by the track of a single particle *i*.

In addition to higher RBE of protons and carbon ions with respect to the one of photons, another advantage of particle therapy is the fact that around Bragg peak

charged particles show a high LET [35]. The dense ionization gives rise to clustered DNA damages difficult to be repaired, thus a greater tumoral cell inactivation [36].

RBE vs LET relationship

In Figure 1. 7 the relationship between RBE and LET is reported: RBE increases as LET increases until a maximum RBE is reached, then decreases due to overkilling effect (i.e. a greater energy dose deposition in cell and DNA with respect to the one required to kill the cell leading to a less effective process per unit dose).

For clinical energy range (1.8-2.0 Gy once a day for 5 days/week over 4-6 week period), RBE and LET present a monotonic correlation: it increases toward the distal edge of the Bragg peak, reaching a maximum at the falloff region.
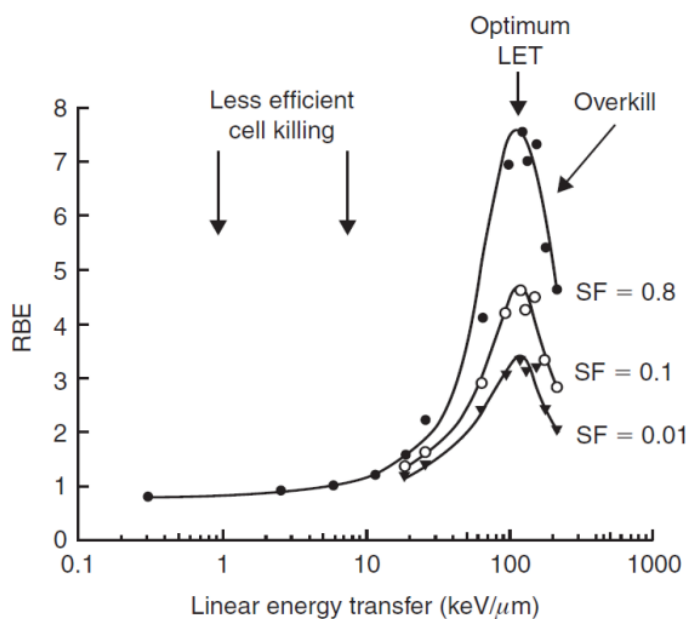

Figure 1. 7 [40] – Relationship between RBE vs LET.

Treatment planning and treatment delivery for particle therapy

As in conventional radiotherapy, also in particle therapy, a treatment planning system is employed to optimize the dose objectives and constraints for tumour and organs at risk.

Focusing on the physical properties of ion beam, by changing the velocity of particle beam it is possible to control the depth at which Bragg peak occurs within the body to deliver the maximum dose to tumour while preserving healthy tissues. However, electron density and tissue composition influence particle energy loss. In order to define particle range, proton stopping powers (i.e. how effective a given material is at slowing down protons) have to be determined for anatomy of the patient. Since CTs provides measurements in Hounsfield unit (HU) of anatomical structures crossed by ion beam during its travel towards the target, a conversion algorithm is used to determine the stopping powers, considering the composition of human tissues. The major sources of uncertainties during treatment planning phase are related to both the determination of patient's anatomy properties and to the algorithm used to calculate the range of particle beams [40]. Uncertainties in calculating particle range, even if of the order of millimetre, can give rise to overdosage in healthy tissues and underdosage in tumour volume, thus compromising the success of treatment.

Another critical aspect of particle treatment planning is related to RBE: regarding protons, even if RBE value is variable, in TPS it is considered as a constant value for biologically effective dose calculation. This leads to errors in dose deposition calculation, which could result in an increased risk of radiation-induced toxicity especially in high-LET locations. Studies related to models to predict RBE as function of tissue type, dose and LET have been published, however, RBE is still considered constant while planning the treatment [41]. Concerning carbon ions, radiobiological models accounting for RBE variations have been developed (e.g. local effect model, which can be described as a track-structure model characterizing the radiobiological features of ions as an effect of the localization of deposited dose around the ion path [42]), however the model's choice during treatment planning is not unique, but depends on the treating institution [36].

As in conventional radiotherapy, TPS provides optimization of parameters related to ion beams and dose calculations based on the type of particle involved. RBE-weighted dose maps and averaged LET_d dose maps are evaluated to select the best plan which will be delivered to the patient.

1.5. National Center of Oncological Hadrontherapy: treatment strategy for skull-based chordoma

The National Center of Oncological Hadrontherapy (CNAO, Pavia, Italy) is a cancer treatment structure where proton therapy and carbon ion therapy are offered to treat patients affected by radio-resistant tumours or tumours positioned in particularly difficult locations (such as skull-base chordomas, ocular melanomas, prostate adenocarcinoma,...) since physical and radiobiological advantages over conventional radiotherapy. Concerning carbon ion therapy, CNAO is the only center in Italy which provides this kind of treatment.

To treat tumours with hadrontherapy (i.e. the usage of hadrons such as protons and carbon ions to treat tumours), a particle accelerator is needed: at CNAO a ring-shaped synchrotron 80 meters long is employed for this purpose. Synchrotron is connected to different rooms where patients are treated. Particularly, before targeting the tumour, a particle beam travels 30.000 kilometres in half a second in the synchrotron to reach the energy required for the therapy. The number of treatment sessions depends on tumour characteristics (such as type, volume and location), but also on the type of particle involved: a common protocol involves one session per day, five days per week for a period ranging from two to seven weeks.

Once a patient is elected as treatable at CNAO, CT for treatment planning and MRIs for segmentation of target and organs at risk are acquired. Successively, a patient-specific treatment plan is built. Once the treatment plan has been optimized, the delivery phase starts.

Concerning SBC, characteristics such as tumour location and tumour radiosensitivity lead to focus on particle therapy to manage this lesion. Indeed, different studies report the benefits of particle therapy treatments over conventional radiotherapy approach both in terms of overall survival [12] and tumour local control [43], thus particle therapy turned out to be the most effective radiation modality for the treatment of SBC [44]. Patient affected by this kind of tumour are treated with particle therapy at CNAO after surgical resection (when it is possible). The choice of particle type depends on tumour characteristics: in a study published by Iannalfi et al. [44], indication about particle type to be used are provided based on the GTV after surgery. In this thesis, patients with SBCs treated with protons at CNAO were analysed. Typical at CNAO, a 74 Gy(RBE) delivered in 37 fractions is administered for proton therapy [44]. Once the radiant treatment has been concluded, a follow-up period starts involving imaging acquisition (typically MRIs) and clinical investigation to monitor treatment result.

Nowadays one of the main objectives in clinical field is the building of personalized treatment based on patient-specific characteristics to improve clinical outcomes. In this context, prognostic factor analysis is needed. Considering SBC, in literature different studies aim to this scope [44] [45] [6]: some clinical prognostic factors have been widely evaluated to be of predictive value such as GTV after surgical resection, dose target coverage and proximity of tumour to organs at risk.

1.6. Radiomics

In medical field, imaging plays an important role in clinical practice, indeed this technology can aid clinicians in decision making. During last years, the role of imaging has rapidly evolved from being primarily a diagnostic tool to being used as a source for the extraction of patient-specific quantitative information [46]. Due to the amount of information contained in imaging data, a timely evaluation in clinical routine is costly and hardly feasible without the support of computer tools. As a matter of fact, different studies have been focused on the discovery of quantitative features inside images to build mathematical models which can be used in clinical decision-making: all those steps can be summarized under the name of Radiomics.

Quoting Robert J. Gilles and colleagues, "images are more than pictures, they are data" [47]: this citation helps in understanding what Radiomics is. Radiomics consists both in automatic extraction of quantitative imaging characteristics from medical images (for example CT, MRI, positron emission tomography), in their analysis and in building mathematical models by combining those features with patient's characteristics (i.e. clinical information) in relation to prediction targets (for example clinical end points) to assist physicians in clinical decision making. Radiomics can be applied in different fields, especially in oncology where medical images are acquired since the beginning of radiotherapy workflow [46]. It has been demonstrated that radiomic features reveal not only information on tumour genomics [48], whose heterogeneity is theorized to be the major cause of failure and resistance to treatment, but also they can be useful in predicting treatment response and differentiating benign and malignant tumours [49].

The main hypothesis behind Radiomics is the fact that biomedical imaging characteristics contain physio-pathological tissue information which can not be detected by human eye, thus they are made accessible through quantitative imaging biomarkers [50] [51].

As cited in Paragraph 1.1, since the deep anatomical location of SBC and its proximity to vital structures, the access to the lesion is often hardly leading to a poor histopathological analysis. However, to make a decision about treatment that will be applied, it is necessary a characterization of the tumour. To deal with this, Radiomics can be applied to imaging data which can be considered as source of prognostic factors.

1.6.1. Radiomic features

As reported in image biomarker standardization initiative (IBSI) [52], which is an independent international collaboration working towards standardising the extraction of image biomarkers from imaging data to high-throughput quantitative image analysis, a biomarker is "a characteristic that is objectively measured and evaluated as an indicator of normal biological processes, pathogenic processes, or pharmacologic responses to a therapeutic intervention". Focusing on image biomarkers, they can be qualitative (requiring expert evaluation) and quantitative (based on mathematical definitions); particularly, the retrieval of the latter enables high-throughput analyses.

Radiomic features can be classified in four categories [53]: shape, first-order statistics, textural (or second-order statistics) and higher-order statistics features, particularly they are extracted from a volume of interest VOI (typically from gross tumour volume or clinical target volume) that is previously segmented [46].

Focusing on the description of different types of features, shape characteristics are descriptors of geometric relations and properties of the segmented VOI; first-order features describe the distribution of voxel intensity within the image VOI; textural features represent statistical relationships between intensities levels of neighbouring pixels or voxels and lastly higher-order statistics features are retrieved by statistical methods usually after the application of mathematical transformations. It is useful to point out the fact that mathematical definition of features is well known, so it is plausible to make a biological analogy interpretation. However, since most of image biomarkers are quite complex, a direct relationship between physio-pathological meaning and textural parameters results hard.

1.6.2. Radiomics workflow

A typical study of Radiomics is structured different steps [53]:

1. image pre-processing and tumour segmentation;
2. features extraction and selection;
3. model generation and evaluation.

Image pre-processing and tumour segmentation

In the first phase, as Radiomics is a retrospective study, heterogeneities in imaging protocols, such as acquisition and reconstruction of the image, appear leading to a loss of repeatability and reproducibility of radiomic features. Several studies highlight this problem aiming to analyse the impact of these settings on image biomarkers or trying

to reduce as much as possible their influence by removing the most influenced radiomic features. In addition, patient movements influence radiomic features leading to artifacts in images [50]. To overcome this issue, different pre-processing steps are applied such as intensity normalization and noise reduction. Then the definition of volume of interest VOI is performed: as previously cited, different methods to segment image exist, but manual segmentation of tumour structures is still considered the ground truth even if this method is time consuming and despite inter-observer variability[54]. Since the latter influences Radiomics analysis results (because features extraction is based on segmented structures), semi-automatically and fully automatically techniques are being developed.

Features extraction and selection

Regarding features extraction and feature selection steps, imaging characteristics can be retrieved by using different open-source packages like PyRadiomics [55]. Also, they depend upon several factors such as the pre-processing of the image and its reconstruction. Lack of reproducibility and validation of radiomic studies is the major challenge: to solve this problem IBSI works towards the standardization of imaging biomarkers extraction. Fixed bin number and fixed bin width discretization are the two common methods applied to calculate features from images and depending on those parameters, values of radiomics features vary: variability and influence on features reliability due to different discretization methods are still under research.

As reported in IBSI, in the fixed bin number method, intensities X_{gl} are discretised to a fixed number of N_g bins. Specifically, the intensity $X_{gl,k}$ of voxel k is corrected by the lowest occurring intensity $X_{gl,min}$ in the ROI, divided by the bin width $(X_{gl,max} - X_{gl,min}) / N_g$, and subsequently rounded down to the nearest integer (floor function). In Equation 1. 4 it is reported the correction applied to the intensity $X_{gl,k}$ of voxel k.

$$X_{d,k} = \begin{cases} \left\lfloor N_g \dfrac{X_{gl,k} - X_{gl,min}}{X_{gl,max} - X_{gl,min}} \right\rfloor + 1 & X_{gl,k} < X_{gl,max} \\ N_g & X_{gl,k} = X_{gl,max} \end{cases} \tag{1.4}$$

Equation 1. 4 – Definition of correction applied to the intensity $X_{gl,k}$ of voxel k

By the usage of this method, a normalising effect is implemented and if a link between image intensity and physiological meaning is present, it is broken by this method.

Fixing the bin number is considered helpful in different circumstances, such as in case of arbitrary intensity units, (for example in MRIs) since the introduction of a normalising effect. In addition, by discretizing images with fixed number of bins it is possible to directly compare features values between different regions of interest (values of some features, i.e. texture ones, depend on the number of grey levels present in the region of interest).

In the fixed bin size method, a new bin is assigned for every intensity interval with width w_b starting at a minimum $X_{gl,min}$, where w_b represents the bin width. In Equation 1. 5 the discretised intensities are reported: the minimum intensity $X_{gl,min}$ is subtracted from intensity $X_{gl,k}$ in voxel k and then divided by the bin width w_b. The resulting value is subsequently rounded down to the nearest integer (floor function), and 1 is added to arrive at the discretised intensity.

$$X_{d,k} = \left\lfloor \frac{X_{gl,k} - X_{gl,min}}{w_b} \right\rfloor + 1 \qquad (1.5)$$

Equation 1. 5 – Definition of the discretized intensities

Contrary to fixed number of bin method, the fixed bin width approach maintains direct relationship with the original intensity scale and this could be helpful in cases where functional imaging modalities are the analysed data (i.e. positron emission tomography and single positron emission tomography).

Due to the fact that the number of available samples (patients) for medical studies is usually lower than the number of extracted characteristics and since the number of radiomic features is often vast, feature selection is required before training the model to prevent overfitting. Most of the characteristics are indeed redundant and highly correlated, hence non robust. This is a crucial aspect, especially in the medical field. Feature selection has the purpose to eliminate a subset of variables which do not held relevant information for machine learning activities.

Feature selection is used to avoid overfitting (because model complexity is directly linked to the number of features: the higher the number of features, the higher the model complexity) and different approaches can be considered to skim features before moving on the following learning phase: a common one is called filter method. In this method, a subset of features is selected by considering a metric. An example of filter

method is Spearman's correlation coefficient which measures the degree of relationship between two variables: in case of high correlation one of the two characteristics is excluded.

The phenomenon of overfitting is related to the model complexity and the error made on training and test data. As shown in Figure 1. 8, as the model complexity increases the error committed on training data decreases, while the one on test data increases therefore the overfitting phenomenon occurs. In contrast, if the model complexity decreases too much, the error committed in both cases will be high: this phenomenon is called underfitting. The ideal case is the one where a balance between model complexity and committed error is found: this reduces both the computational time required in building the model and improves the model performances.
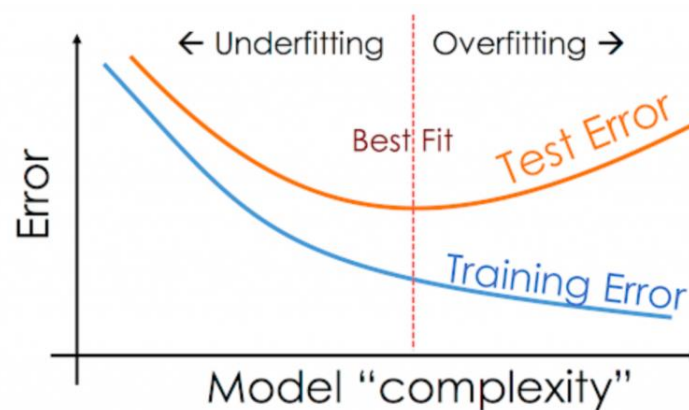

Figure 1. 8 – Error vs model complexity. Overfitting and underfitting problems.

Model generation and evaluation

The ultimate goal of Radiomics is the building of an accurate model able to predict clinical scores such as tumour diagnosis and treatment response: after having selected the relevant characteristics, statistical models or algorithms from machine learning can be trained depending on the goal of the study (see paragraph 1.6.3 for further details). After models' building, a step of validation is required: in this phase a prediction is made on data which were not used in models creation, then considering a metric such as accuracy, models performances are calculated [56]. The aim of this phase is to select the best model for actual future predictions.

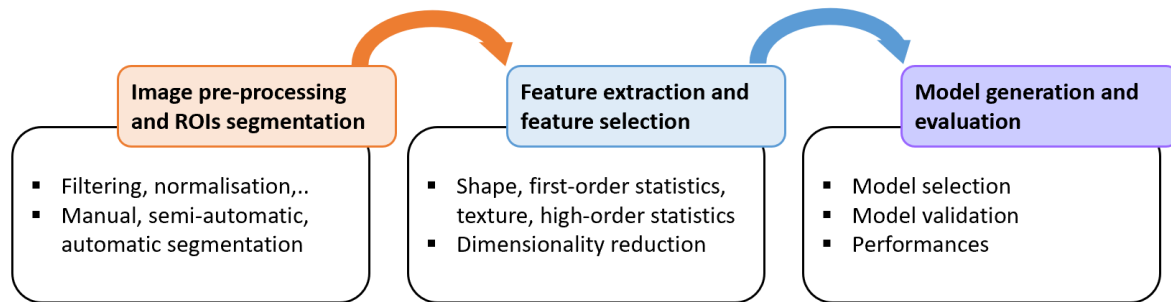A typical Radiomics workflow is summarized in Figure 1. 9 illustrating the main steps.

Figure 1. 9 – Workflow of a typical Radiomics study showing the main steps.

1.6.3. Machine learning

Radiomics can be seen as a machine learning based approach. Machine learning is a subfield of artificial intelligence whose aims are to extract relevant knowledge and to obtain meaningful rules by exploring and analysing a dataset. Starting from a dataset, machine learning explores the available data to build models trained on them which will be able to make predictions on future data, such as the prediction of tumour response to a particular treatment.

There are different ways to subdivide machine learning [57]:

- considering the presence or not of the target variable, it is possible to refer to supervised learning in case of the target attribute represents either a class to which a sample belongs or a measurable quantity. The aim of supervised learning is to find a general rule to associate the input with the output. Unsupervised learning refers to those cases where the target variable is not present, so the learning analysis is not guided by the target attribute. The goal, indeed, is to find a pattern or recurrencies in the dataset. Lastly semi-supervised learning refers to a learning problem involving a small portion of labelled samples with which a model is trained and a large number of unlabelled samples to which the just trained model is applied;

- regarding the aim of the machine learning problem, in classification models the goal is to predict a categorical target attribute (in those cases samples are divided into two or more classes), while if the target variable can assume continuous numerical values, regression problem is the case. Time series

analysis refers to cases where target attribute evolves over time, so the aim is the prediction of one or more values assumed by the target attribute in future periods; whereas clustering techniques are intended to identify homogeneous groups (clusters) of records made by samples which share similar characteristics. Lastly, if the goal is the investigation about the incidence of a certain event in a particular time window (e.g. tumour local recurrence during research time) survival analysis is applied. With the aim to build a model to predict SBC's local recurrence, survival analyses will be adopted for this thesis (specifically, Cox proportional hazards model will be built) and will be discussed in section 2.4.

The training process is made by three different steps: training, validation and testing. Starting from the original dataset, it is split into three subsets (i.e. training, validation and testing set) since in each phase of the training process different set of data are considered. This leads to the building of a robust model.

By looking deeper, training set is used to fit the model, so the algorithm sees and learns from this subset; after having fit the model, it is applied to the validation set to provide an unbiased evaluation of the model. During this phase model's hyperparameters tuning is performed.

Training and validation sets are usually exploited in cross-validation (CV). CV is commonly applied in case of limited available dataset and procedure widely used to evaluate the model performance. In a classical K-fold CV the original training dataset is split into K subsets having the same number of samples: in turn, K-1 subsets are considered as training set, while the remaining subset is used as validation set to evaluate prediction accuracy of the just trained model. This kind of training-validation operation is repeated K times so that each subset is used exactly once as validation set [58]. The final evaluation is given by the mean of a chosen metric which is estimated for each validation fold.

Lastly, concerning the test data, it is taken into account to provide an unbiased evaluation of final model which was fit on the training dataset.

It is important to highlight the fact that data belonging to test dataset have never been seen during model construction phases: in fact, a model applied to the same data used as training set will result in a perfect prediction, but will fail in predicting anything useful on unseen data. This leads also to overfitting problem.

In Figure 1. 10 a schema representing the partitioning of original data used during a machine learning study is reported. Particular attention is given to cross-validation technique.
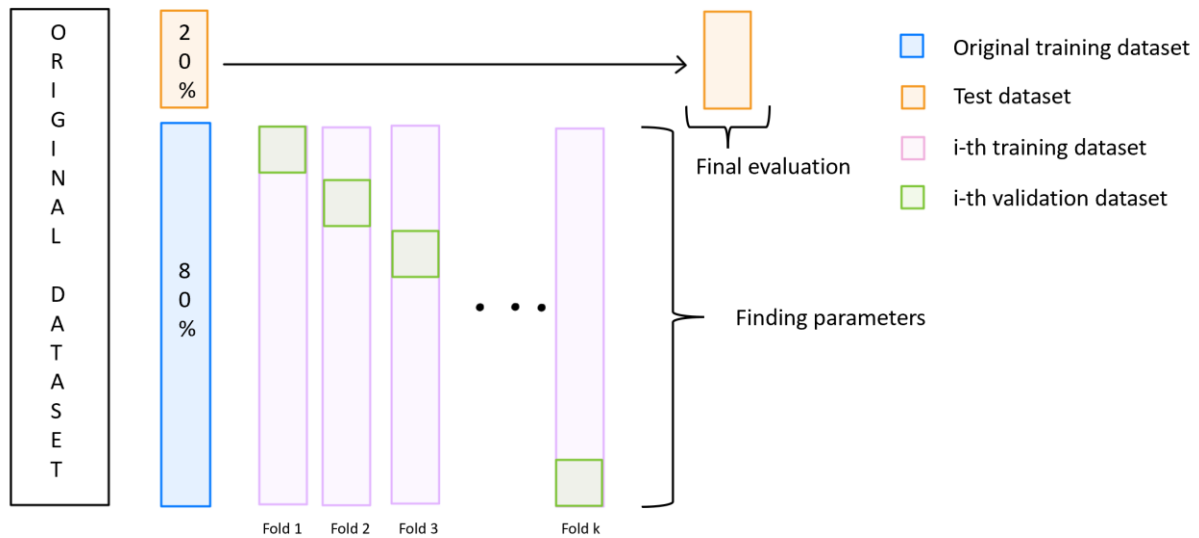

Figure 1. 10 - In this figure a focus on cross-validation and partitioning of original dataset is shown. Starting from the available dataset, it is divided into two subsets: usually 80% of the entire dataset is used as training set, while the remaining 20% as test set. During cross-validation procedure, the original training set is partitioned into k folds. In particular, in turn k-1 folds are used as training data, while the remaining fold as validation set. At the end of cross-validation phase, k performances are retrieved considering the validation set: the aim is to find the best set of parameters for the model. Then these parameters will be used in the model during the test phase to calculate model performance on test dataset.

1.6.4. Radiomics in oncology

In a review made by Song et al., it is reported that the annual grow rate of published articles related to Radiomics field from 2013 to 2018 is 177.82%. Particurarly, CT is the most frequent imaging modality considered followed by MRI, while concerning tuour site, lung, breast and prostate tumours are the ones much more investigated. Among the goals most often explored, characterization and monitoring of tumours are the most common [59]. To make some examples, Aerts et al. focused one of their works on radiomics features as prognostic factors in patients affected by lung and head-and-neck cancers using CTs founding that features related to intratumour heterogeneity can be used for that purpose [60], while in another research Urraro and collegues investigated reliability of MRI features in identifying patient affected by prostate cancers [61].

For chordomas, due to the rarity of tumour, the number of studies published in literature decreases drastically. By reducing the field of interest to articles where SBC is treated with particle therapy, even a more restricted number of works are present. To cite some studies, Xiyin and collegues investigated radiomics features combined with diagnostic characteristics to evaluate local control and toxicity in skull-base and cervical spine chordoma treated with proton and carbon ion radiations founding that tumour volume was a significant prognostic factor for overall survival and for progression-free survival (i.e. the length of time during and after the cancer treatment where patient lives with the disease but it does not get worse) [62]. Instead, Dominietto et al. made use of deep learning to investigate the role of radiomic features in predicting recurrence of SBC in patient treated with pencil beam scanning proton therapy concluding that shape and texture features play a potential role in predicting treatment outcome [63]. Still nowadays the potential role of radiomics features as prognostic and predictive factors in SBC is under investigation.

1.7. Dosiomics

During last years an extension of Radiomics has emerged: this new approach, known as Dosiomics, considers three-dimensional radiotherapy dose distributions to extract features which can be useful to predict treatment-related toxicity and prognosis, achievement of local control and overall survival in radiation oncology [64] [65]. Dosiomics workflow perfectly replicates the one of Radiomics, the only difference lies in considering dose distributions as input images.

Dosiomics studies are more rare with respect to Radiomics, since Dosiomics has emerged only in recent years. Wu and collegues investigated on the role of dosiomic features in predicting locoregional recurrence in cases of head and neck cancer treated with intensity modulated radiotherapy with respect to radiomic ones discovering that the integration of dosiomic features in modeling leads to a successful patient classification in high-risk and low-risk of recurrence [66]. Dai et al. analyzed patients with nasopharyngeal carcinoma treated with intensity modulated radiotherapy in order to predict recurrence-free survival (which is the period of time after primary treatment for the cancer ends that the patient survives without any signs or symptoms of the cancer) founding that the integration of dosiomic features with radiomic ones leads to better model's performance [67]. Concerning chordomas, only two literature works are present to the best of our knowledge: Buizza et al. applied Radiomics and Dosiomics in order to build survival models to stratify patients affected by SBC according to the risk of adverse local control treated with carbon ions discovering that dosiomic features reveal the most promising for the purpose [68], while Morelli et al. focused the analysis on Dosiomics considering patients affected by sacral chordomas and treated with carbon ion therapy showing that dosiomic features extracted from dose averaged LET_d maps result promising as prognostic factors [69].

To summarize, in literature a poor number of dosiomic studies are present, expecially if the population under investigation relates to patients affected by chordomas; in addition, none of them is related to treatment with proton therapy.

1.8. Aim of the thesis

In literature different radiomics studies exist, particularly they are focused on CTs and MRIs, which are the images that are mostly considered to investigate what kind of biomarkers are relevant in predicting local control and overall survival in radiotherapy. Dosiomics, instead, is a more recent research field and a limited number of studies are reported in the literature, especially on patient affected by chordomas and treated with particle therapy. Up to now, just two studies investigated Dosiomics on chordomas treated with carbon ion therapy, but no applications are present for proton therapy.

Due to the lack of researches focused on SBCs treated with proton therapy, the purpose of this thesis was to explore how different features extraction settings can influence model's performance in predicting local recurrence for SBCs patients treated with proton therapy.

The available dataset consists of physical dose, LET_d and RBE-weighted dose maps acquired from patients affected by SBC at CNAO. These patients were treated with protons after a partial surgical resection and CTV was the considered lesion segmentation from which image biomarkers were extracted.

To achieve the goal, different values of parameters were set to guide features extraction from dose maps, then, after a step of feature selection, a time-to-event model was fed with the chosen characteristics.

2 Materials and methods

2.1. Patient Data

Fifty patients affected by skull-base chordoma and treated with a macroscopic surgical resection followed by proton therapy at CNAO between 2012 and 2018 were considered for this study. Detailed clinical information of patient population are reported in Table 2. 1 and Table 2. 2.

Continuous Variables	Median (Range)
Age (years)	51.5 (14 - 81)
GTV (cm³)	4.06 (0.0 – 99.32)
Follow-up time (months)	63.22 (23.83 – 98.80)

Table 2. 1 – Continuous clinical variables describing the available population for the study reported as median (range).

Categorical Variables	Details	Frequency	Percentage %
Gender	Male	24	48
	Female	26	52
Local Control	Favorable (censored)	44	88
	Adverse (adverse event)	6	12

Table 2. 2 – Categorical clinical variables describing the available population for the study reported as recurrence and percentage.

Concerning proton therapy, the dose was delivered in 37 fractions for a prescribed biologically effective dose of 74 Gy(RBE); only in one case it was 72 Gy(RBE) delivered

in 36 fractions. Then, patients underwent to follow-up (median and range are reported in Table 2. 1), where treatment outcome was retrieved in terms of local control LC (no-progression or complete eradication of the tumoral mass) and tumour progression/local recurrence TR (accessed by imaging acquisition). LC and TR were classified as favourable and adverse events, respectively. After a median follow-up time of 63.22 months, TR was found in 6 patients (12% of population) with a median time-to-recurrence event of 26.82 months, while 44 patients (88% of population) were included in favourable event group with 65.20 months as median follow-up time. In Table 2. 2 are reported the number of patients which experienced LC or TR both in terms of percentage and frequency.

For each patient, CT images were acquired by the usage of Sensation Open CT scan (Siemens, Erlangen, Germany) to manually contour both the tumour and the organs at risk. In Figure 2. 1 examples of CTs overlapped with the correspondent CTVs (in red) are shown.
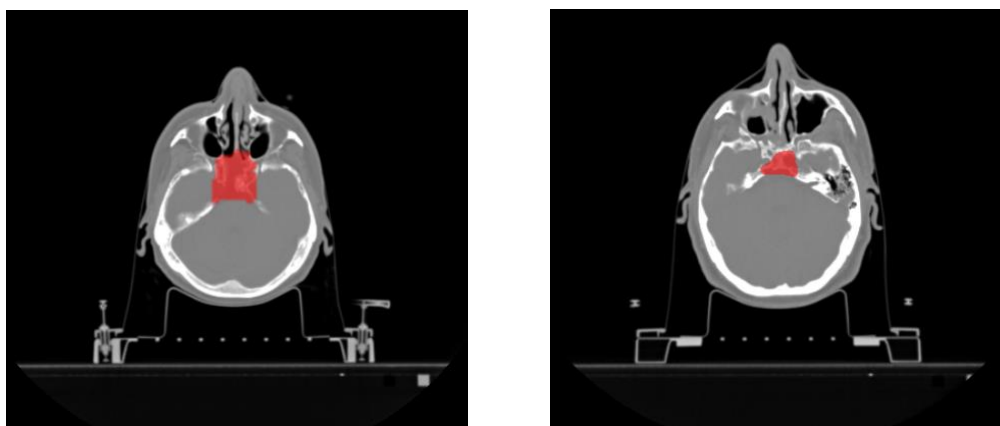

Figure 2. 1 – Examples of axial CT images overlapped with correspondent CTVs (red structures). CTs and CTVs belong to different patients.

Subsequently, CT images and relevant segmented structures were exploited to calculate the dose distribution which will be administered during the delivery phase: to optimize the dose which will be delivered to the target while sparing healthy structures, RayStation (10B version) was adopted.

With the aim of conducting a Dosiomics study, for each patient three different kinds of dose maps were analysed: physical dose, LET_d and RBE-weighted dose maps. Medians and ranges of the three dose maps are reported in

Table 2. 3. Since the treatment was delivered by the usage of two proton beams, for both physical dose and RBE-weighted dose maps the total physical dose map and the total RBE-weighted dose map were retrieved by summing maps relative to the two beams for each patient, while LET maps were elaborated into a dose-averaged LET map (LET$_d$) following Matsumoto's formula reported in Equation 2. 1 [70].

$$\overline{L_d}(\mathbf{r}) = \frac{\sum_k [n_k \cdot D_k(\mathbf{r}) \cdot L_k(\mathbf{r})]}{\sum_k [n_k \cdot D_k(\mathbf{r})]} \tag{2.1}$$

Equation 2. 1 – Matsumoto's formula to calculate the dose-averaged LET$_d$ map starting from LET distribution by beam k. $\overline{L_d}(\mathbf{r})$ is the dose-averaged LET distribution at location r, $D_k(\mathbf{r})$ represents the physical dose distribution by beam k, n_k is the number of fractionations and $L_k(\mathbf{r})$ is the LET distribution by beam k.

Type of map	Median (range)
Physical dose (Gy)	70.99 (58.96 – 78.32)
LET$_d$ (keV/μm)	2.9 (2.23 – 4.32)
RBE-weighted dose (Gy(RBE))	78.09 (64.86 – 86.16)

Table 2. 3 – Details about available dose maps shown as median (range) values.

In Figure 2. 2, Figure 2. 3 and Figure 2. 4 examples of physical dose, RBE-weighted dose and dose-averaged LET$_d$ maps belonging to the same patient are shown.
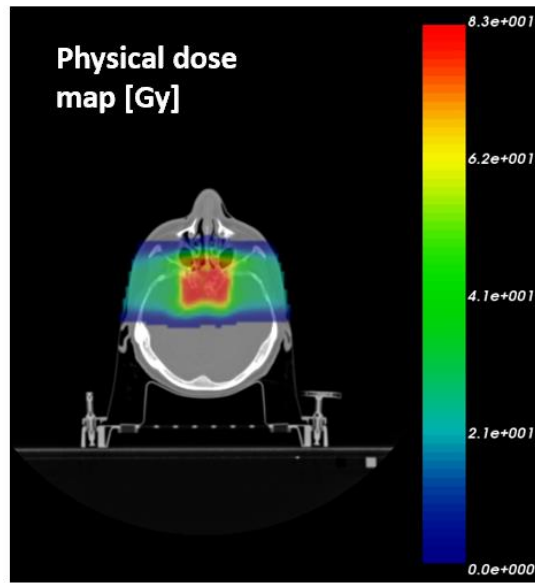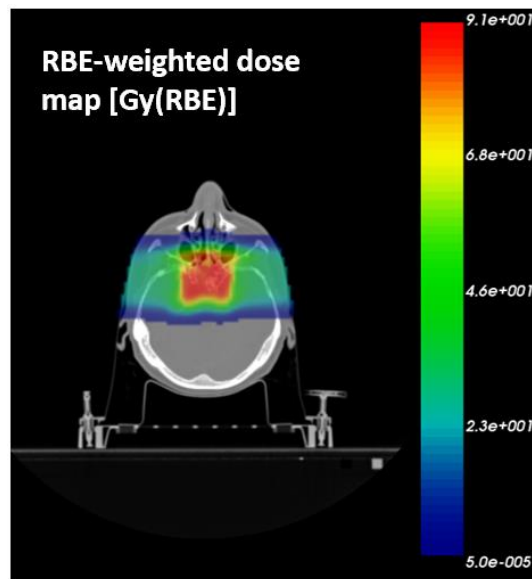
Figure 2. 2 - Physical dose map.
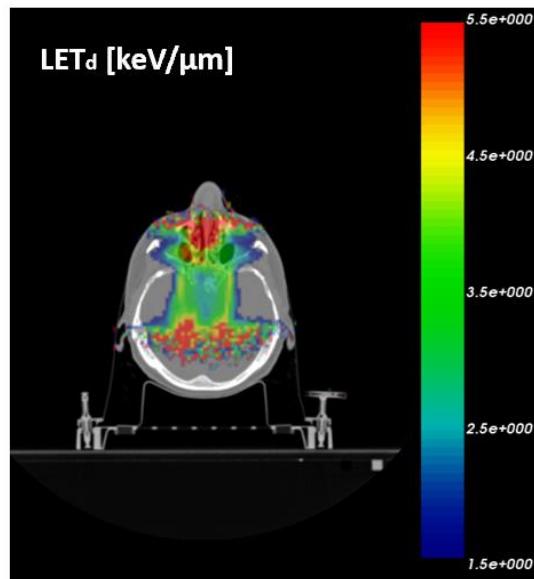

Figure 2. 3 – RBE-weighted dose map.

Figure 2. 4 – LET$_d$ map.

In Figure 2. 5 it is summarized the followed dosiomics workflow which will be discussed in the next paragraphs. Briefly, features were calculated from the available dose maps by the customization of some parameters related to discretization of image intensity. Then after a step of features selection (carried out by ICC, significance analysis and correlation), modelling was performed, particularly, survival analysis by the usage of Cox proportional hazards model was conducted. Lastly, performances of the built survival models were retrieved.
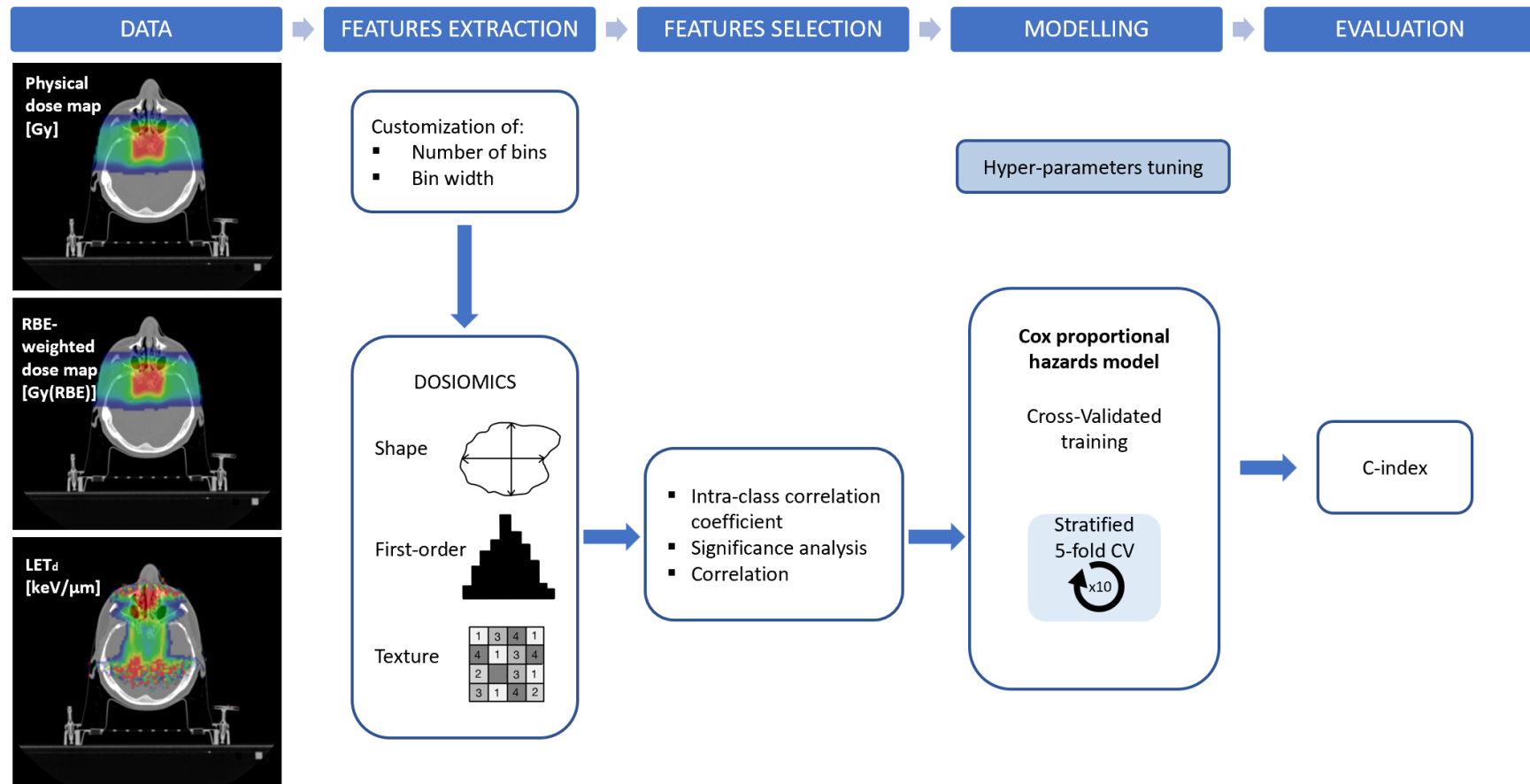
Figure 2. 5 – Followed dosiomics workflow.

2.2. Features extraction

Following the guidelines reported by IBSI [52], quantitative imaging biomarkers were extracted from CTV for each patient and for each dose map using PyRadiomics (v2.2.0). PyRadiomics is an open-source python package employed for the extraction of radiomic features from medical imaging. The goal of this package is to lend a standard for radiomic analysis proving a simple and reproducible way to extract imaging biomarkers. Specifically, image data are loaded and pre-processed using SimpleITK, which is an open-source python package developed for multi-dimensional image analysis. PyRadiomics allows both the image pre-processing and the application of different built-in filters to original or pre-processed image. In addition, by specifying some settings (such as discretization type of image intensities, features classes, image normalization and image resampling,…), it is also possible to customize features extraction.

Particularly, in this thesis:

- no image pre-processing was applied to any dose maps (neither filters, nor normalization or resampling);
- features were extracted in 3D from each original map.

In the next paragraph the customization of grey levels intensities discretization will be illustrated.

2.2.1. Customization of features extraction: fixed number of bins and fixed bin width

 Before the effective features extraction, discretization of image intensities inside the region of interest is required. Two are the common approaches which can be followed: discretisation with a fixed number of bins or discretization with a fixed bin width (definitions are reported in Paragraph 1.6.2, specifically in Equation 1. 4 and Equation 1. 5). By varying the number of bins and the bin width, different characteristics of the image can be revealed [52], indeed, considering wider bins, noise is reduced but at the same time too large bin leads to not enough differentiation. Whereas, the usage of narrow bins gives a greater precision, but it can bring to a poor data grouping [71]. For example, considering [0; 100] as the range of gray levels intensities, fixing to 20 the bin with, 5 discretization levels will be the obtained; instead considering 4 as bin width, 20 grey levels will be retrieved. In Figure 2. 6 a graphical example is shown.
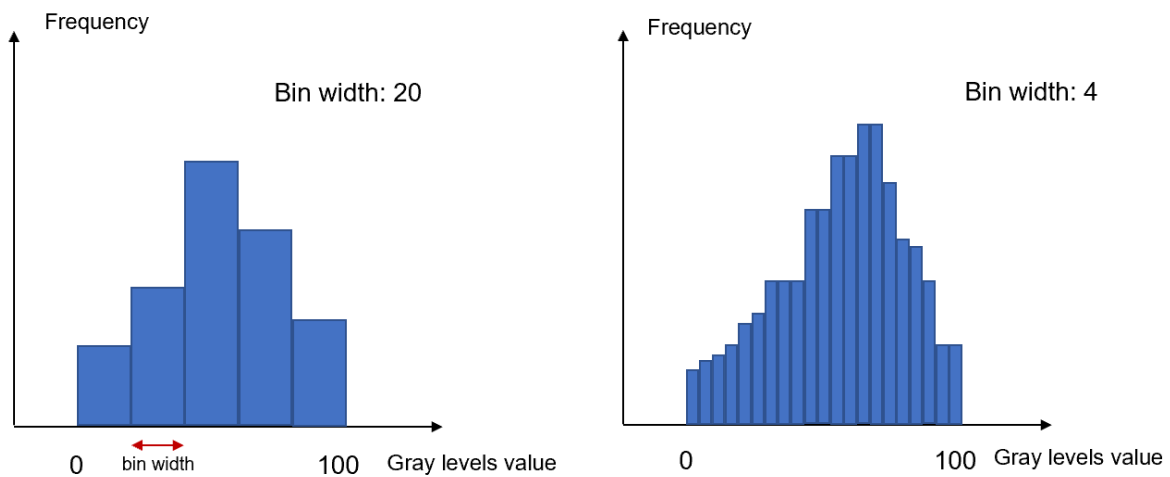
Figure 2. 6 – Different bin width discretization. At left side a greater bin width is set (wider bins), while at right side a smaller one is considered (narrower bins).

PyRadiomics allows the customization of both bin width and number of bins parameters, as there is no concrete evidence in favour of using one method over the other in all types of images. Since gray-level discretization affects the values of extracted features, different studies analysed its impact on PET [72], CT [73] and MRI [74] images, while, concerning dose maps, no researches address this kind of analysis.

In this thesis, particular attention was paid to the customization of number of bins and bin width since the final goal was to explore how different settings of parameters could influence the performance of trained model in predicting local recurrence. To determine the range of values to be assigned to the bin width parameter, the range of each kind of dose map was retrieved (reported in

Table 2. 3), while regarding the values assigned to the parameter bin count, a common practise is to consider powers of two [53]. In

Table 2. 4 fixed number of bins and fixed bin size which were explored are reported for each type of dose map.

Dose map	Fixed bin size	Fixed number of bins
Physical dose	0.1, 0.2, 0.3, 0.4, 0.5, 0.8, 1.0, 2.0, 3.0, 4.0, 5.0, 8.0, 10.0, 16.0	2, 3, 4, 5, 8, 10, 16, 32, 64, 128, 256, 512
RBE-weighted dose		
LET_d	0.01, 0.02, 0.03, 0.04, 0.05, 0.08, 0.1, 0.2, 0.3, 0.4, 0.5, 0.8, 1.0	

Table 2. 4 – Explored fixed bin size and fixed number of bins for each type of dose map.

From now on, the terms "binwidth" and "bincount" will refer to fixed bin width and fixed number of bins methods, respectively.

Features extraction was guided by the customization of binwidth or bincount configuration: after having set the value of the chosen parameter (binwidth/bincount), a *pxn* matrix was created, where *p* represents the number of samples (i.e. the number of patients) and *n* the number of extracted characteristics (i.e. *n* = 107). Considering bincount method, for all dose maps features extraction was repeated 12 times, one for each bincount configuration, while concerning binwidth method, the extraction was repeated 14 times for physical dose and RBE-weighted dose maps, while 13 times for LET_d.

A total of 107 features were retrieved for each type of dose map. Below are reported the number of extracted characteristics for each features type and their definitions.

Shape features

Shape features are referred to the geometry of the three-dimensional size and shape of the region of interest. A total of 14 shape features were retrieved: Elongation, Flatness, Least Axis Length, Major Axis Length, Maximum 2D Diameter Column, Maximum 2D Diameter Row, Maximum 2D Diameter Slice, Maximum 3D Diameter, Mesh Volume,

Minor Axis Length, Sphericity, Surface Area, Surface Volume Ratio and Voxel Volume.

First-order features

Concerning the first-order features, they describe the distribution of voxel intensities within the image region defined by a mask. Eighteen were the first-order extracted characteristics: 10 Percentile, 90 Percentile, Energy, Entropy, Interquartile Range, Kurtosis, Maximum, Mean Absolute Deviation, Mean, Median, Minimum, Range, Robust Mean Absolute Deviation, Root Mean Squared, Skewness, Total Energy, Uniformity and Variance.

Texture features (high-order features)

Texture features were not computed directly from the original image, but from different descriptive matrices encoding spatial relationships between voxels in the original image. Specifically:

- Gray Level Co-occurrence Matrix (GLCM) represents the number of times the combination of two intensity levels occurs in neighbouring voxels within a specific distance along a fixed direction;
- Gray Level Run Length Matrix (GLRLM) quantifies the gray level runs defined as the length of consecutive pixels or voxels sharing the same gray level value;
- Gray Level Size Zone Matrix (GLSZM) retrieves the number of connected voxels having the same gray level intensity;
- Gray Level Dependence Matrix (GLDM) quantifies gray level dependencies in an image, which is defined as the number of connected voxels within a specific distance that are dependent on the centre voxel;
- Neighbourhood Gray Tone Difference Matrix (NGTDM) quantifies the difference between a gray level value and the average value of its neighbours within a certain distance.

A total of 75 texture features were retrieved.

Details of Dosiomics features used in this thesis are reported in Appendix (A.1)

2.3. Features selection

To reduce data abundance and to prevent overfitting due to the high number of extracted features, a step of features selection was performed before training the model. Specifically, three different features selection methods have been applied:

- significance;
- intra-class correlation coefficient;
- correlation.

Significance

With the aim of accessing the significance of features in discriminating patients with a favourable treatment outcome (LC) from those with an adverse outcome (TR), statistical tests were applied after the step of features extraction. First of all, for each extracted feature, the distribution of its values has been tested applying Shapiro test: it is a normality test employed to verify the null hypothesis that data come from a normal distribution (significance level $\alpha = 0.05$). According to data distribution, indeed, it is considered either a parametric statistics, which assumes that the data are taken from a population normally distributed, or a non-parametric statistics, which makes no assumption on the distribution of population. In case of normal distribution, if patients who experienced a local recurrence and the ones who reached local control showed different variances, Welch's t-test for unpaired samples was applied to test whether the means of the two population are statistically equals (significance level $\alpha = 0.05$); on contrary in case of equal variances, t-test was considered to test both the equality of means and of variances (significance level $\alpha = 0.05$). Concerning non-parametric statistics, Mann-Whitney test was taken into account to test the null hypothesis that data come from independent distributions with equal medians (significance level $\alpha = 0.05$).

Intra-class Correlation Coefficient

Intra-class correlation coefficient (ICC) is an index used to measure reliability both in terms of correlation and agreement between measurements [75]. Reliability is mathematically defined as true variance over true variance plus error variance. In literature different types of ICC exist, leading to different results when applied to the same data. For this reason, it is important to choose the appropriate form of ICC considering the specific problem to deal with. To do that, "model", "type" and "definition" have to be taken into account [75]. Concerning the "model", it is referred to the set of raters (i.e. values which can assume binwidth/bincount parameter)

employed to measure all subjects (i.e. features): one-Way Random-Effects model is considered if each subject is rated by a different set of raters randomly chosen from a larger population of possible raters; in case of raters are chosen from a population showing similar characteristics (for example clinicians with the same years of experience to evaluate passive range of motion) two-Way Random-Effects Model is chosen; lastly, if the selected raters are the only raters of interest, two-Way Mixed-Effects Model is taken into account [75]. "Type" selection is related to the number of chosen raters: if it is greater than one, a "mean of k raters" type is considered (for example considering 4 raters, mean of k raters, with $k = 4$, is the selected type) [75]. "Definition" selection regards 2-way random and 2-way mixed-effects models: if the goal is to identify absolute differences between different raters, "absolute agreement" is taken into account, whereas if systematic differences between raters is under analysis, "consistency" is considered [75].

For this thesis:

- two-Way Mixed-Effects Model was selected as "model", indeed the same set of raters is considered for all subjects;
- for "type" definition, 12 was the considered number of raters in case of fixed number of bins for all the available dose maps, so a mean of 12 raters was the selected type. Instead, in case of fixed bin width, 14 raters were considered for physical dose and RBE-weighted dose maps while 13 raters in case of dose-averaged LET_d maps. A mean of 14 and 13 raters were the selected "types", respectively;
- consistency was the chosen "definition" since consistency between raters was under analysis.

ICC was applied as feature selection method aiming to the identification of robust features: for each extracted feature, the correspondent ICC value was calculated. To do this, a matrix of dimension pxm was constructed (where p represents the number of patients and m the number of raters) and it was given as input to *intraclass_corr()* function present in pingouin library (v0.4.0). This was repeated for each feature. It is suggested that [75]:

- ICC values less than 0.5 are indicative of poor reliability;
- ICC values between 0.5 and 0.75 indicate moderate reliability;
- ICC values between 0.75 and 0.95 indicate good reliability;
- ICC values greater than 0.95 are indicative of excellent reliability.

Correlation

Correlation is a statistical measure describing the dependency level between pairs of features and, in feature selection field, it belongs to filter methods. The level of correlation between couples of imaging biomarkers was identified based on the value assumed by Spearman correlation coefficient (SCC), which can vary between -1 and 1. Considering a couple of features, if the absolute value of SCC was found to be greater than 0.80, the two considered characteristics were labelled as highly correlated, so one of the two was deleted (since highly correlated features do not provide additional information to the model which will be trained, on contrary they will lead to overfitting phenomenon).

With the aim to reduce the total number of features to 10, two different approaches were adopted by combining the just explained different feature selection methods:

- **approach #1** – a specific ICC threshold value was found for each dose map type such that a maximum of 10 characteristics have been selected. This approach has been implemented considering only texture features with the aim to analyse how model performances vary by changing the value of binwidth and bincount parameters across each dose map type.
- **approach #2** – an ICC threshold value common to all dose map types was experimentally found for texture features with the aim to reduce to a maximum of 10 the total number of texture characteristics, while significance analysis was applied to shape and first order features to discard the non-significant ones. At this point, in case the total number of selected features (texture, shape and first order ones) exceeded 10, correlation method was applied until the total number of selected characteristics reached 10. The goal of this approach was to compare models' performance across different dose map types by fixing the value of binwidth or bincount parameter.

2.4. Modelling

In clinical trials, outcomes can have different statistical forms: they can be continuous, such as the size of a lesion, and can be analysed with linear regression, while others can assume the form of "either-or", e.g. if a patient has been diagnosed with cancer or not, so logistic regression can be employed. On the other hand, if a study is oriented towards to the time-to-event outcome, such as the time until a patient experiences a tumour's local recurrence, survival analysis is considered [76].

2.4.1. Survival analysis

Survival analysis refers to those researches where the goal is the investigation about the time at which a particular event occurs in a certain time window. In clinical field, time (also called survival time) can be defined as the time from the beginning of patient's follow-up period until an event occurs which depends on the type of the study (e.g. death, appearance of a disease, TR) [77]. Since the limited duration of the observation period, some of the involved patients may not experience the event of interest before the end of the study or they are dropped out before the occurrence of the interested event, thus the survival time is unknown. These subjects are called censored data and can not be ignored. To sum up, survival analysis is applied in those cases where the event of interest is analysed both in terms of its occurrence or non-occurrence during the period of time of the study and it provides tools to incorporate variables to predict the time of the event of interest.

Survival and hazard functions

In the context of survival analysis, different quantities of interest such as survival and hazard functions have to be defined. Survival function S(t) refers to the probability that the event of interest has not yet occurred by time t (Equation 2. 2) [78]. For example, if the interested event is "local recurrence of tumour", S(t) represents the "probability of not having tumour local recurrence beyond time t".

$$S(\text{t}) = \Pr\{T \geq t\} = 1 - F(t) = \int_{t}^{\infty} f(x)dx \qquad (2.2)$$

Equation 2. 2 – Definition of survival function S(t), where T is a random variable characterizing the time of the event, t refers to any specified time point, Pr { T ≥ t } is the probability of not experiencing the event up to a time t, F(t) represents the cumulative distribution function and f(t) the probability density function.

Concerning the hazard function h(t), it is defined as the rate at which the event of interest occurs instantaneously given that upon that point in time the event had not yet happened (Equation 2. 3) [78].

$$h(t) = \lim_{k \to 0} \frac{P(t \leq T < t + k \mid T \geq t)}{k} \tag{2.3}$$

Equation 2. 3 – Definition of hazard function h(t), where the numerator inside the limit function represents the conditional probability that the time T of event will occur in the interval [t, t+k) given that the event has not occurred before, while the denominator k represents the width of the interval. The ratio between the numerator and the denominator defines a rate of occurrence of event per unit of time. Passing to the limit, the instantaneous rate of occurrence is obtained.

Through some calculations it is possible to find the relationship between hazard and survival functions (reported in Equation 2. 4).

$$h(t) = \frac{f(t)}{S(t)} \tag{2.4}$$

Equation 2. 4 – Relationship between hazard and survival functions.

2.4.2. Cox proportional hazards model

Cox proportional hazards model is one of the most popular survival analysis techniques considered when dealing with censored data in medical studies [76] [79]. Particularly, it investigates the relationship between covariates (i.e. features) and time-to-event (e.g. local SBC recurrence) through the hazard function. In other words, the model evaluates how different factors influence the rate of the event happening at a particular point in time [76]. In Equation 2. 5 it is reported the hazard function for the Cox proportional hazards model: x_i is the i-th covariates, while b_i is the parameter to be estimated which represents the effect of the covariate on the outcome.

$$h(t|x) = h_0(t) \cdot e^{\sum_i b_i \cdot x_i} \tag{2.5}$$

Equation 2. 5 – Definition of hazard function $h(t|x)$ for a subject with a set of predictors x (i.e. features), where t refers to survival time, b_i represents the effect size of the covariate x_i, $h_0(t)$ is the baseline hazard (i.e. the value of the hazard function if all the covariates are equal to zero).

In this thesis, with the goal to develop a model for survival analysis, the interested event was identified as TR and the time-to-event as the time in which TR appeared in acquired images during follow-up, starting from the date of the last treatment. Concerning patients who experienced a favourable event (LC), they were labelled as censored data and the time was calculated starting from the date of last therapy treatment to the one of last follow-up.

2.4.3. Model's building

Cox proportional hazards model (r-Cox, scikit-survival v0.17.2) regularized with an elastic-net penalty was developed considering 80% of the dataset ($p = 40$) within a cross-validation procedure, while the remaining 20% of the data ($p = 10$) was exploited to evaluate the trained model on totally unseen samples. More in detail, the dataset was split randomly, but a stratification on the local control variable was applied to ensure an equal proportion of samples belonging to adverse and favourable local control in both sets (85% and 90% local control rate for train- and test-set, respectively). Before the training of the model, features were normalized with L2-norm, then survival model was built. During model's development, hyper-parameters tuning was performed through grid-search with a stratified 5-fold cross-validation routine repeated 10 times (stratification was performed considering LC). The best hyper-parameters' set was associated to the highest predictive performance in terms of Harrell Concordance index which will be discussed in paragraph 2.4.4. In Table 2. 5 possible values that each tuned hyperparameter could assume are reported. Specifically, the hyper-parameter "l1-ratio" refers to a trade-off between L1 and L2 penalization inside the elastic net parameter: if l1-ratio equals 0, then L2 penalization is employed, whereas for a value of 1, L1 penalty is considered. For $0 < l1\text{-ratio} < 1$, the penalty is a combination of L1 and L2. Concerning "penalty factor" hyperparameter, it is related to the penalty which can be applied to each feature, separately.

Hyper-parameter	Possible values
Penalty factor	0.1, 1
L1-ratio	1e-5, 1e-4, 0.001, 0.01, 0.2, 0.5, 0.8, 1

Table 2. 5 – Possible values that penalty factor and L1-ratio can assume during grid-search optimization.

Different survival models were built, specifically, for each value that binwidth and bincount parameter and for each dose map type, a Cox proportional hazards model was created.

2.4.4. Model's evaluation

With the purposes of both comparing models' performance across the same dose map type varying the binwidth/bincount parameter and across different dose map types fixing the value of binwidth/bincount parameter, a 5-fold cross-validation routine was randomly repeated 10 times. Harrell Concordance index (C-index) was the employed metric to evaluate models' predictive performance. C-index is defined as the proportion of all comparable patient pairs in which the predictions and outcomes are concordant [80]. As for the prediction of the time until local recurrence, C-index is calculated by considering all possible pairs of patients where at least one of whom experienced local recurrence. Two samples are labelled as "comparable" if both experienced the event at different times or if one sample had a shorter observed survival time, while the second one is known to have a survival time at least to the survival time of the first. In the latter case, the second sample is assumed to outlive the first. A pair of samples is defined as "not comparable" if the two experienced the event at the same time or if one of the two experienced the event and the other does not, but the second one has not been followed long enough to determine whether it will experience the event or not. The probability of concordance between the predicted and the observed outcomes is estimated by C-index [80]. This metric can assume values between 0 and 1: a value of 1 implies perfect concordance, while a value of 0 describes the opposite situation of perfect anti-concordance. C-index equal to 0.5 refers to completely random assignments [81].

To evaluate the performances of the built models, since a repeated cross-validation routine was employed, the median value and the interquartile range of C-indexes computed from the validation fold were calculated.

3 Results and discussion

In this chapter both the results related to features selection and the performances of developed models will be presented and discussed.

It must be pointed out that for the calculation of biologically effective dose, RBE value was considered constant in TPS (specifically equal to 1.1 for proton case). Since the RBE-weighted dose maps can be defined as the physical dose map weighted for the RBE factor and recalling the formula reported in Equation 1. 4, related to gray levels discretization method with fixed number of bins, it can be noticed that these kind of gray levels discretization is invariant to the product of the map by a constant factor. These leads to have equal features values for both RBE-weighted dose maps and physical dose maps, thus the analysis related to fixed number of bin method was performed only one time by considering physical dose maps (since the same results will be obtained for both physical and RBE maps).

Concerning the customization of the parameters related to intensity discretization, as regards the bin width parameter, ranges of the three dose maps types were calculated (19.36, 2.09 and 21.3 for physical dose, LET$_d$ and RBE-weighted dose maps, respectively), then values inside those ranges were chosen and assigned to the bin width parameter. Regarding the bincount parameter, powers of two were considered [53] with the addition of 3.0, 5.0 and 10.0 values. In Table 2. 4 are reported the set of chosen values related to intensity discretization for each dose map type.

3.1. Selected features

3.1.1. Selected features - Approach 1

In this paragraph the selected texture features for the three dose maps types using intra-class correlation coefficient are discussed. This approach was employed to identify robust characteristics. Specifically, considering a particular dose map type at a time, two ICC threshold values were experimentally found (one related to bincount method the other to binwidth one) aiming to select at maximum 10 texture features. Following this strategy, by fixing the dose map type, we perform a comparison of models' performance across different settings of binwidth/bincount values; in fact, the same 10 features will be obtained for all the settings of bincount/binwidth values.

In Table 3. 1, ICC threshold values are reported for physical dose, LET$_d$ and RBE-weighted dose maps respectively, divided by binwidth and bincount methods.

	ICC threshold	
Dose map type	**Binwidth**	**Bincount**
Physical dose	0.97	0.95
LET$_d$	0.98	0.96
RBE-weighted dose	0.98	0.95

Table 3. 1 - ICC threshold values for the three dose maps divided by binwidth and bincount methods.

Since for all the three dose maps types and for both binwidth and bincount method the ICC threshold value is greater or equals to 0.95, features that were selected and then used in building survival models are considered as robust. Details of the dosiomics selected features are reported in Appendix A.2593.2.1. It can be noticed that for all the dose maps, texture features related to gray level co-occurrence matrix were the most selected in both binwidth and bincount methods.

3.1.2. Selected features - Approach 2

To compare models' performance across different dose maps types by fixing the value of binwidth/bincount parameter, ICC, significance analysis and correlation were applied in features selection step. With the aim to reduce to a maximum of 10 the total number of texture characteristics, ICC threshold value was set to 0.95 for all types of dose maps (good robustness), instead, concerning significance analysis, it was employed to discard non-significant shape and first order features. Lastly, if the total number of selected characteristics (texture, shape and first order) exceeded 10, correlation was taken into account: a threshold of 0.8 was set to identify highly correlated features and among couples of features labelled as such, one of the two was discarded until the final total number of characteristics reached 10.

No-one of shape features turned out to be significant (p-value < 0.05) in all possible settings of binwidth/bincount values. Selected features for each dose map type and for each binwidth/bincount value are illustrated in Appendix A.2.2. Below bar plots reporting the number of selected features belonging to first-order and texture classes

are shown for each dose map type and for each value binwidth/bincount parameter can assume (Figure 3. 1 and Figure 3. 2).

Binwidth case

Due to the different grey levels ranges of the three dose maps types, different settings of binwidth values were employed to discretize grey levels values for LET$_d$ maps with respect to physical dose and RBE-weighted dose maps. In Figure 3. 1, considering the common set of binwidth values for all the three dose maps types, it can be noticed that the number of selected texture features is greater with respect to the one of the first-order for all dose maps types. Concerning the binwidth values belonging to the interval [0.01, 0.08], employed only in LET$_d$ map for intensity discretization, the same number of features belonging to first-order and to texture classes was selected for 0.02, 0.04 and 0.08 binwidth values, while for 0.01, 0.03 and 0.05 binwidth values, texture features prevailed. Lastly, considering binwidth values ranging from 2.0 to 16.0, employed only for physical and RBE-weighted dose maps, still a prevalence of texture features on those of the first-order was present.
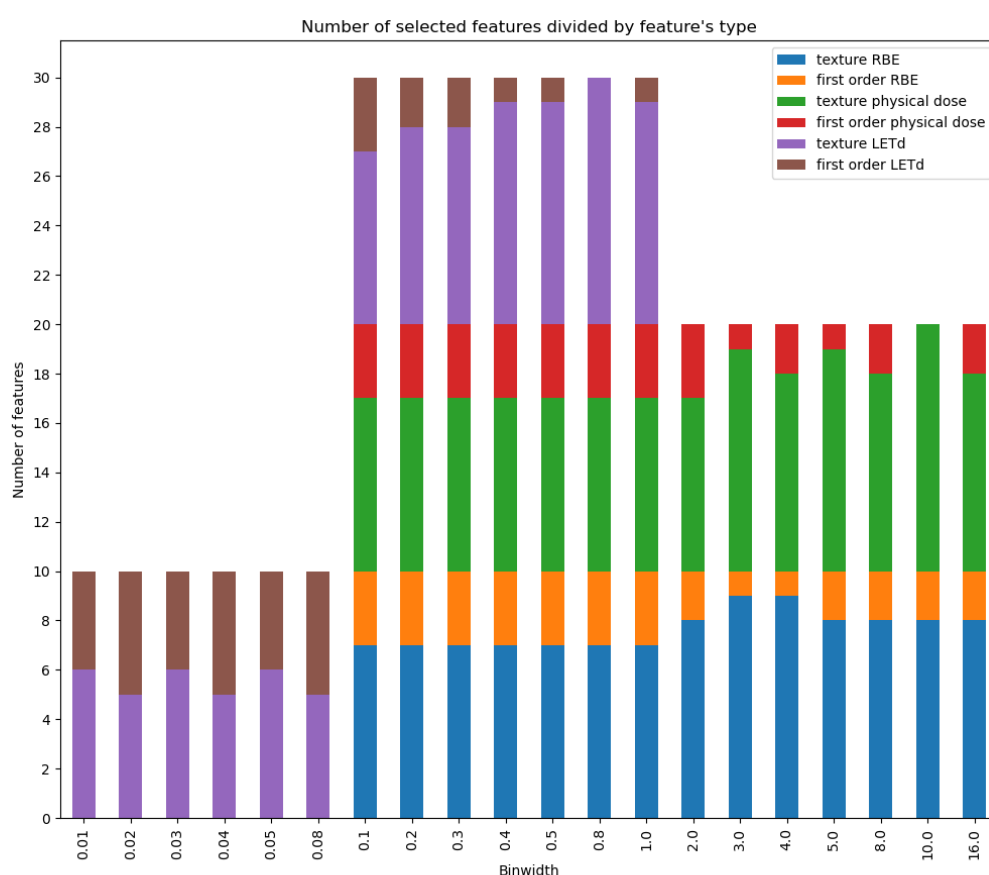
Figure 3. 1 - Bar plot grouped by binwidth value showing the number of selected features for physical dose, RBE-weighted dose and LET$_d$ maps.

Bincount case

In Figure 3. 2 a bar plot is reported showing the number of selected features belonging to texture and first-order for physical dose and LET$_d$ maps grouped by bincount parameter. It can be noticed that for each value the bincount parameter can assume, the number of selected texture features was greater than the one of first-order features for physical dose map, while considering LET$_d$ map, the same number of features belonging to the two features types was selected (except for bincount equals to 2.0 and to 3.0). Comparing the selected features for the two types of dose maps, a greater number of first-order features was taken into account for LET$_d$ map with respect to the one on physical dose.
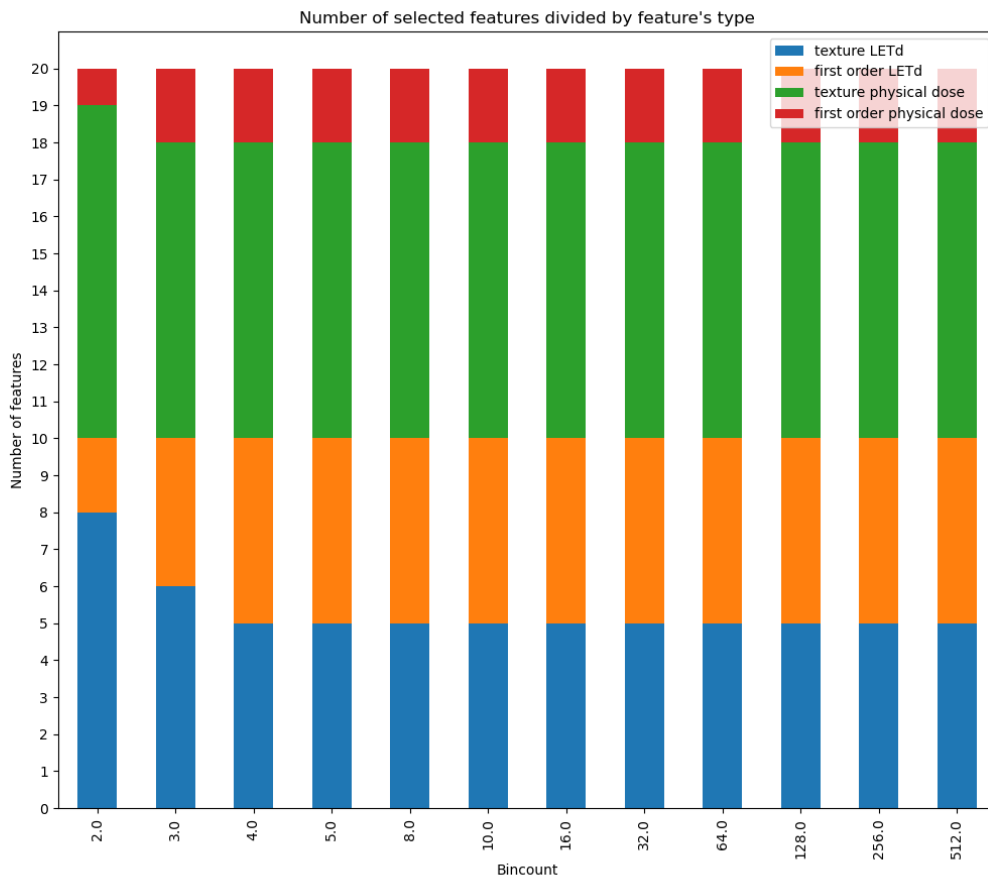
Figure 3. 2 – Bar plot grouped by bincount value showing the number of selected features for physical dose and LET$_d$ maps.

3.2. Models' performance

After features selection phase, features were normalized with L2-norm, then survival models were trained to discriminate patients who experienced TR from patients which reached LC. In this section, results in terms of models' performance are reported and discussed. Approach 1 and approach 2 refer to the two features selection strategies.

With the aim to compare models' performance across different settings of bincount/binwidth values fixing a dose map type (approach 1) and across different dose maps types by fixing the value of binwidth/bincount parameter (approach 2), different Cox proportional hazards models were trained. In particular:

- considering physical dose maps, a total of 12 models were built for bincount method, while 14 models for the binwidth one;
- 12 and 13 models were developed for bincount and binwidth methods respectively, feeding the algorithms with features extracted from LET_d maps;
- a total of 14 models were built for RBE-dose weighted maps in case of binwidth method.

Models' performance were accessed in terms of C-index, specifically, since a repeated cross-validation routine was employed, the median value and the interquartile range of C-indexes computed from the validation fold were retrieved and considered to evaluate the models (tables showing medians and interquartile ranges are reported in Appendix A.3).

3.2.1. Models' performance - Approach 1

Physical dose maps

In Figure 3. 3 the medians of C-indexes in function of the different bincount values are shown. The best model's performance (0.76/0.14 median/interquartile range) was retrieved for a discretization with a fixed number of bins equals to 256. Focusing on bincount values belonging to the interval [3.0, 16.0], even if C-indexes values resulted low, models showed stable performances (since the same C-index value is obtained), whereas increasing the bincount value stability was lost.

Concerning models' performance obtained by varying the binwidth parameter, from Figure 3. 4 it can be noticed that the best model's performance was retrieved for binwidth equals to 5.0 (0.76/0.19); stability was reached for binwidth values between 0.8 and 2.

Comparing the two discretization methods, bincount method showed better stability in models' performance with respect to binwidth one.
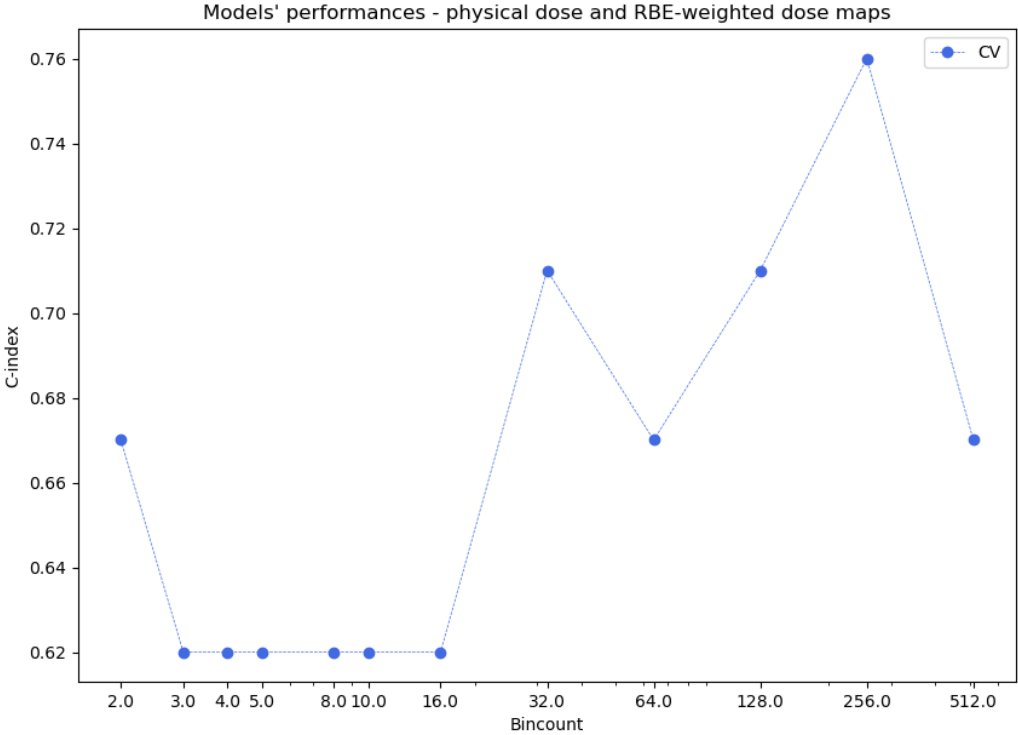

Figure 3. 3 - Medians of C-indexes versus bincount values for physical dose maps.
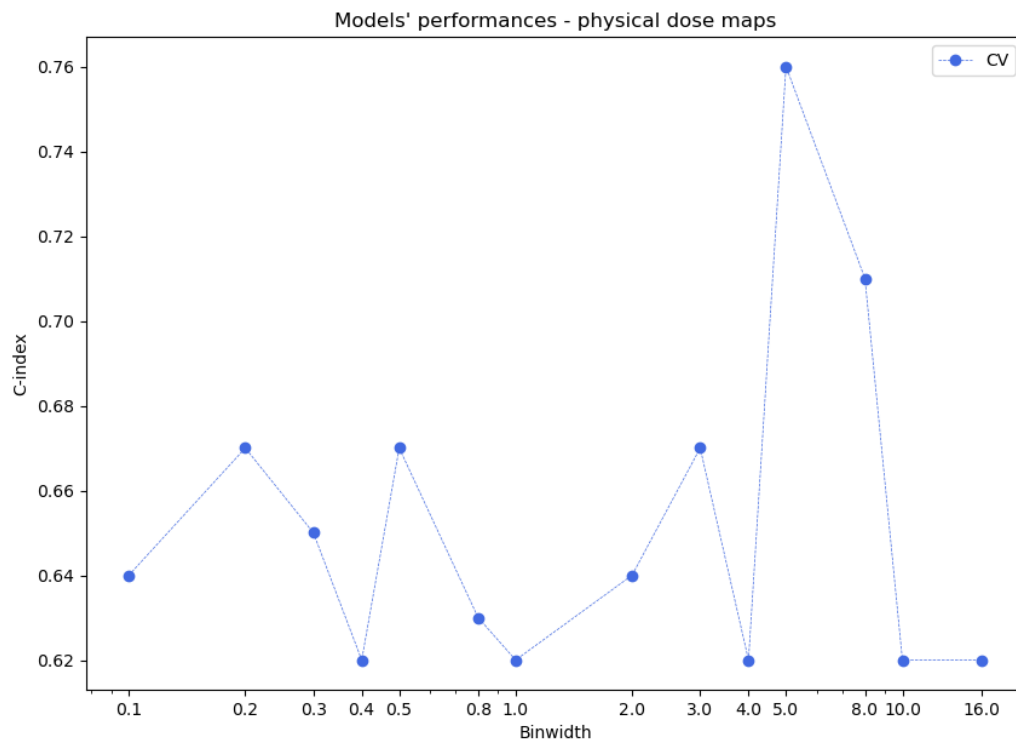
Figure 3. 4 - Medians of C-indexes versus binwidth values for physical dose maps.

LET$_d$ maps

In Figure 3. 5 it is possible to notice that the best value of C-index resulted from bincount value equals to 2 (0.64/0.24): this was an unexpected result since discretizing an image with only two grey levels intensities leads to a very poor differentiation between image intensities, so poor performance was foreseen. A more reliable result was obtained for bincount value equals to 16 (0.63/0.14). For bincount values ranging from 32 to 512, models' performance stability was reached, while decreasing the number of bins it was lost.

Concerning image discretization with binwidth method, it is possible to state that stability in models' performance was found in binwidth values between 0.01 and 0.04 (included), while the best performance was achieved with binwidth equals to 0.8 (0.69/0.18). Also in this latter case, intensity discretization performed by 0.8 as bin width leads to not enough differentiation in image intensities (since the very restricted range of LET$_d$ maps), so poor performance was expected. Figure 3. 6 shows the performances of the built models (median of C-indexes) in function of the binwidth value.

As in physical dose maps, a greater stability is reached in case of bincount as intensity discretization method with respect to binwidth one.
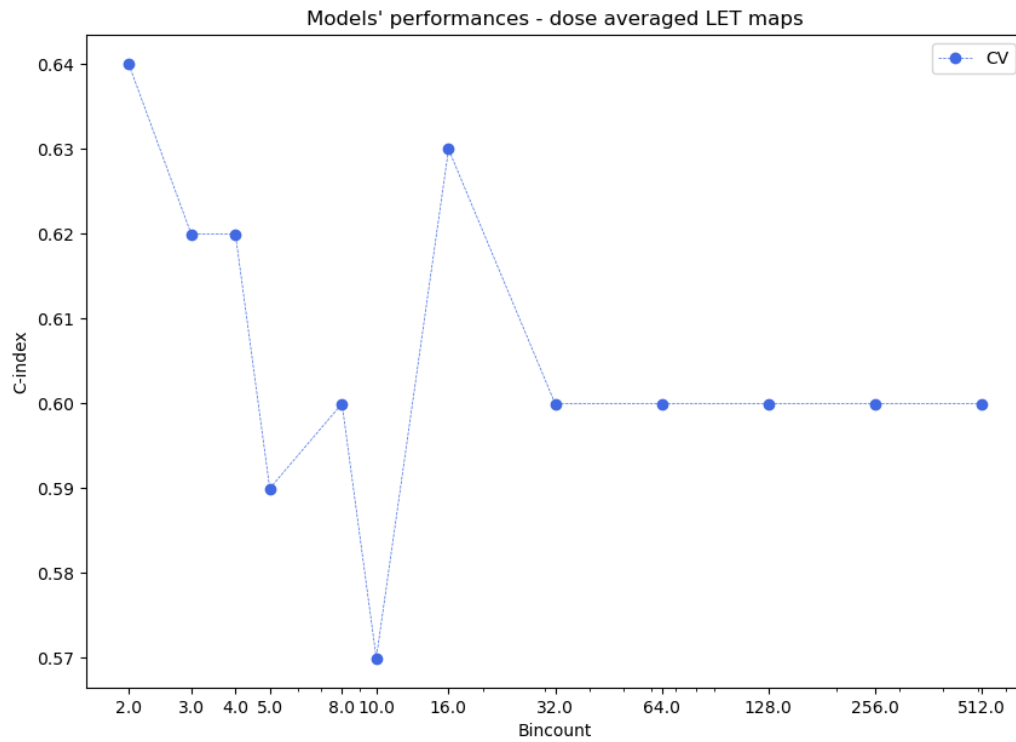
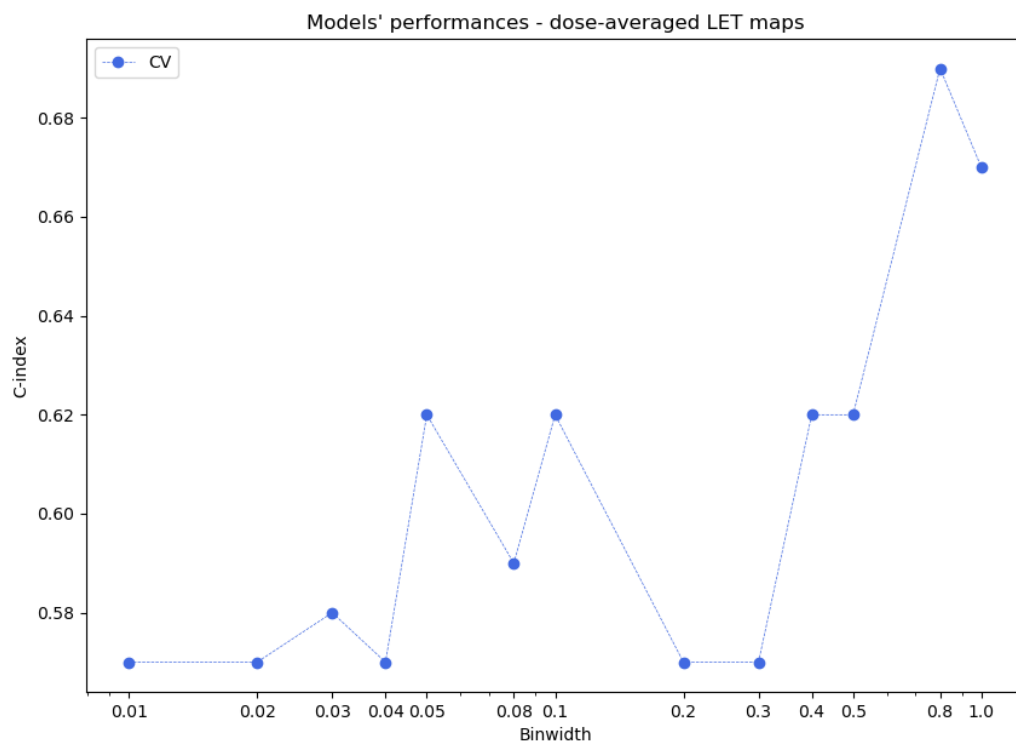Figure 3. 5 - Medians of C-indexes versus bincount values for LET$_d$ maps.

Figure 3. 6 - Medians of C-indexes versus binwidth values for LET$_d$ maps.

RBE-weighted dose maps

The best model's performance (0.71/0.14) was retrieved for images discretized with a bin width equals to 2. Focusing on binwidth values belonging to the interval [3.0, 16.0], even if C-indexes values resulted low, models showed stable performances; on contrary, decreasing the binwidth value, stability is lost. These results can be seen in the graph reported in Figure 3. 7Figure 3. 8.
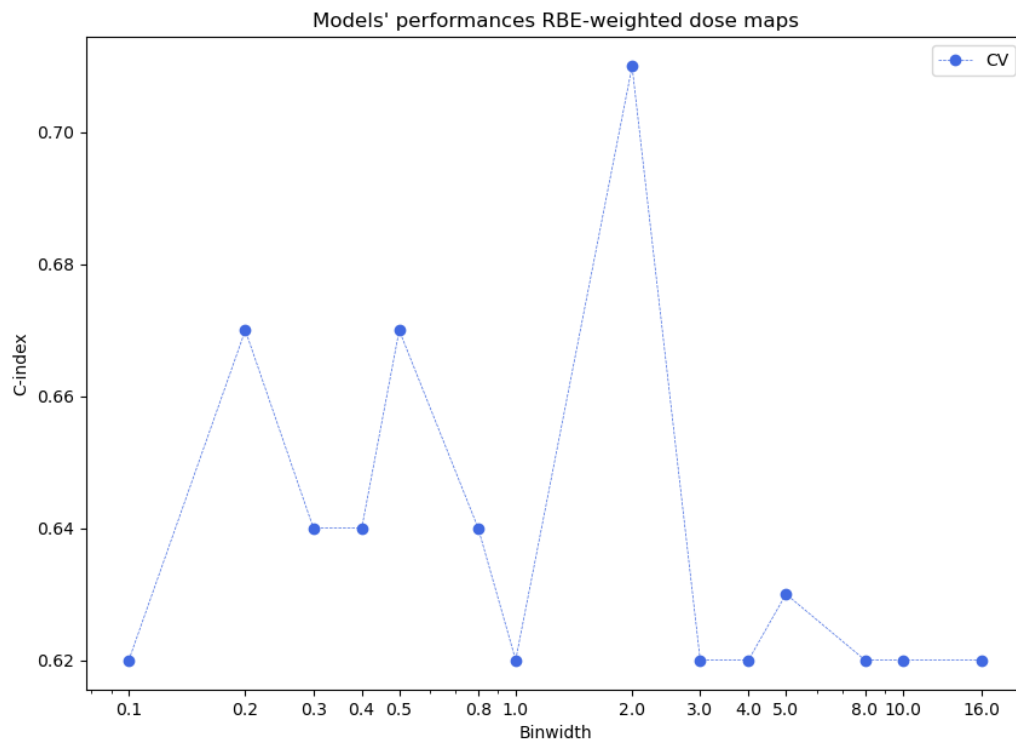
Figure 3. 7 - Medians of C-indexes versus binwidth values for RBE-weighted dose maps.

None of the two methods of intensity discretization turned out to be better regarding models' performance stability (as reported in paragraph 3, results obtained with bincount discretization method for physical dose map are the same of RBE-weighted dose map).

3.2.2. Models' performance - Approach 2

Binwidth case

Considering the bar plot shown in Figure 3. 9, it is possible to explore how models' performance vary between dose maps types by fixing the value of binwidth parameter. As already pointed out in paragraph 3.1.2, different settings of binwidth values were employed to discretize grey levels values for LET_d maps with respect to physical dose and RBE-weighted dose maps due to the different grey levels ranges of the three dose maps types.

A common value of the binwidth parameter showing the best model's performance for all dose maps types was not found, indeed concerning LET_d map type the highest C-index is associated to 0.03 and 1.0 as binwidth values, while for RBE-weighted and physical dose maps, 0.5 and 8.0 binwidth values were those related to best performances, respectively.

Focusing on possible values which binwidth parameter can assume in the interval [0.1, 1.0], performances related to LET_d map resulted lower with respect to the ones associated to physical and RBE-weighted dose maps (except for binwidth equals to 1).
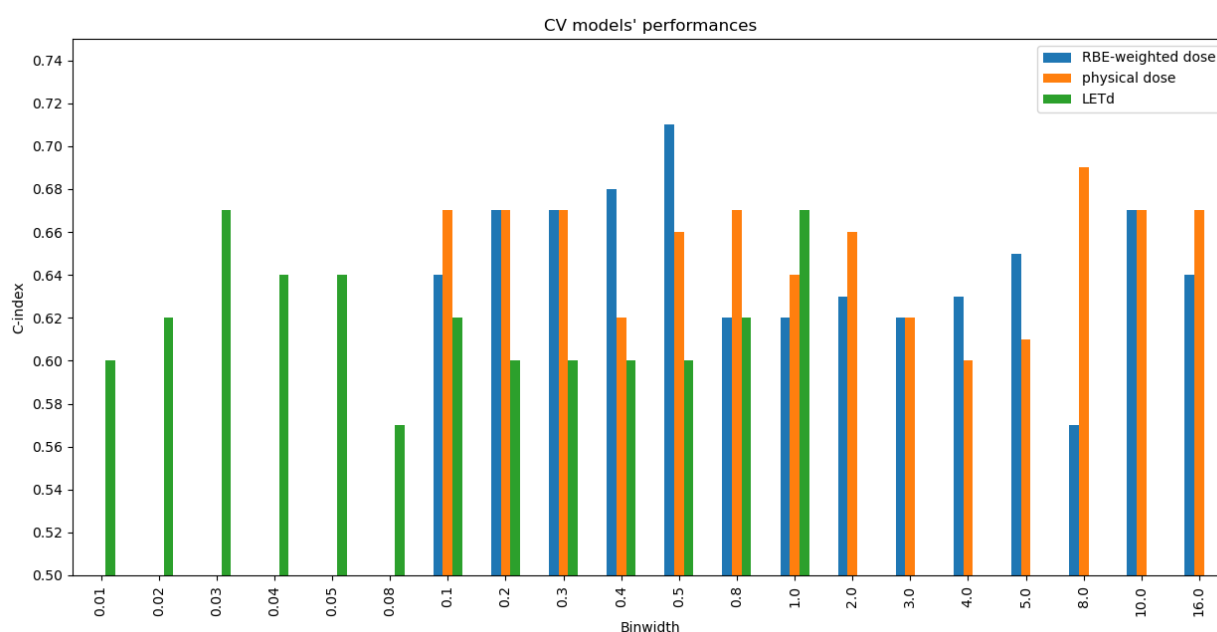

Figure 3. 9 - Median of C-indexes obtained through repeated cross-validation routine for physical dose, RBE-weighted dose and LET_d maps types grouped by binwidth values.

Bincount case

In Figure 3. 10 a bar plot grouped by bincount value reports models' performance related to physical dose and LET$_d$ maps types.

As in binwidth case, a common value of bincount parameter showing best models' performance for both the two dose maps type wasn't found (since 2 and 32 were the bincount values related to highest C-indexes for LET$_d$ and physical dose maps types, respectively).

Considering bincount values lower than 8 (with 8 included), by fixing the value of the parameter, comparable performances were obtained for LET$_d$ and physical dose maps; whereas, increasing the bincount value, a greater gap can be noticed between the performances of the two maps types (except for bincount equals to 16, 64 and 512).
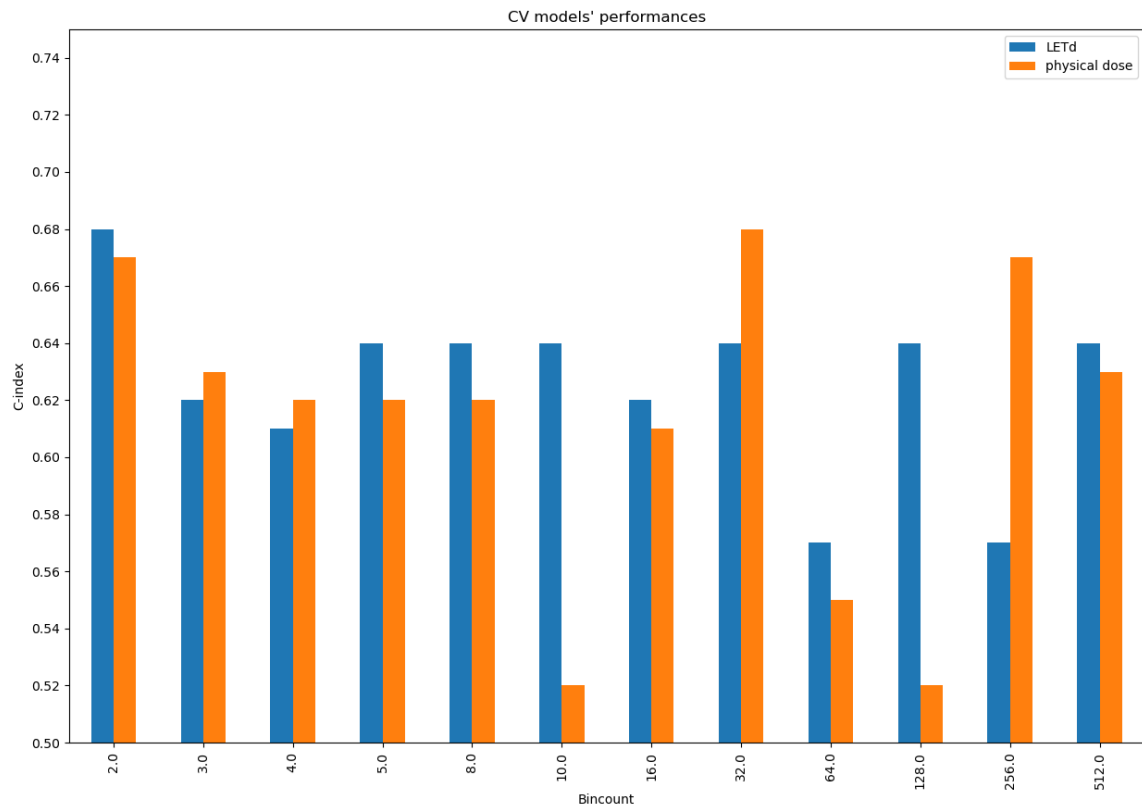

Figure 3. 10 - Median of C-indexes obtained through repeated cross-validation routine for physical dose and LET$_d$ maps types grouped by bincount values.

4 Conclusions

In this chapter the main results are summarized and possible future works are presented.

The purpose of this thesis was to explore how different features extraction settings can influence model's performance in predicting tumour recurrence for skull-base chordomas patients treated with proton therapy after a partial surgical resection.

To achieve this goal, starting from three available types of dose maps (physical dose, RBE-weighted dose and LET$_d$ maps) dosiomics features were extracted from CTV manually segmented by experts during treatment planning. Specifically, two different methods for image discretization were followed (fixed bin number and fixed bin width) and depending on those parameters, different values of dosiomics features were obtained. Then, after a features selection step, survival models (Cox proportional hazards model) were built. By the implementation of two different features selection strategies (approach 1 and approach 2), two analyses were conducted: a comparison between models' performance across the same dose map type by varying the value assumed by binwidth/bincount parameter and a comparison between models' performance across different dose maps types by fixing the value of binwidth/bincount parameter.

Considering the features selection – approach 1, the chosen ICC threshold values have brought to good robustness in selected texture features for all the dose maps types and for both binwidth/bincount intensity discretization methods, since ICC threshold values were equal or higher to 0.95 (Table 3. 1)[75]. Concerning models' performance, no high values of C-indexes were obtained: this could be explained by the fact that the usage of ICC as features selection method allows selecting robust characteristics, but these characteristics might not be the most informative. As regards stability of models' performance, it can be stated that a greater stability was obtained in models trained by varying the bincount parameter with respect to the binwidth parameter in case of physical dose. The same results were obtained for LET$_d$. Concerning RBE-weighted dose, no marked difference was present between the two intensity discretization methods (e.g. results obtained with bincount discretization method for physical dose map are the same of RBE-weighted dose map).

Concerning features selection – approach 2, since none of shape features turned out to be significant for all the dose maps and for both binwidth/bincount methods in discriminating patients who achieved LC from those who experienced TR, the conducted dosiomics analysis was not linked to contours (in contrast with the study of Buizza et colleagues [68], which found shape features to be promising); while as for first-order features a greater number of significant first-order features were found in LET_d for both the two intensity discretization methods with respect to physical and RBE-weighted dose maps.

Focusing on models' performance and considering the binwidth interval [0.1,1], performances related to physical and RBE-weighted dose maps resulted better with respect to the ones associated to LET_d in most cases. This could be linked with the range of image gray levels intensity, indeed physical and RBE-weighted dose maps showed greater ranges with respect to the one of LET_d maps, thus image discretization with bin width values belonging to the interval [0.1, 1] can lead to a not enough differentiation in case of LET_d maps compared to physical and RBE-weighted dose maps cases. Considering bincount case, by fixing the bincount value, in most cases better results in terms of models' performance were obtained for LET_d maps rather than physical dose maps: this is probabily related to the fact that a greater number of first order features were selected for LET_d maps.

From this study it is possible to state that by selecting only robust texture features, the discretization with fixed number of bins leads to better results for both physical dose and LET_d maps (considering values of parameter in the interval [3.0, 16.0] and [32, 512], respectively), while for RBE-weighted dose maps no marked difference is present between the two intensity discretization methods.

Concerning approach 2, the addition of significant first-order features in discriminating patients which experienced tumour recurrence from those who obtained local control did not lead to improvements in models' performance. Comparing the obtained performances with those already present in literature, Buizza et al. [68] conducted a radiomics and dosiomics study where SBCs patients treated with carbon ions were analysed: focusing on dosiomics model, Cox proportional hazards model was built and its performance resulted higher with respect to the ones obtained in this thesis. Higher performances were also obtained by Morelli et al. [69], where patients affected by sacral chordomas and treated with carbon ion radiotherapy were enrolled. Considering models' performance related to physical dose and to LET_d maps, better results were obtained for LET_d maps in bincount case, probably related to a greater number of selected first-order features.

The main limitation of this work is related to the sample size, which has affected the obtained results: only fifty patients were enrolled for the study, of which only 12% experienced TR. The reason why such a small dataset has been analysed is the fact that skull-base chordoma is a rare tumour and the chosen particle therapy approach was a peculiar one, so to conduct a survival analysis both the label related to LC and the time-to-event variable were needed.

Possible future developments may be related firstly to an increase in sample size by enrolling more patients for survival analysis since a limited number of patients' data was the one available for this study, then an exploration of other survival models and an improvement in already implemented one (Cox proportional hazards model). Finally, as regards the values assigned to fixed bin width and fixed number of bins discretization methods, different values from the ones considered in this thesis can be set to explore firstly whether models' performance remain stable inside and around the identified ranges, and if better performances can be achieved.

Bibliography

[1]   B. P. Walcott, B. V. Nahed, A. Mohyeldin, J.-V. Coumans, K. T. Kahle, and M. J. Ferreira, 'Chordoma: current concepts, management, and future directions', *Lancet Oncol.*, vol. 13, no. 2, pp. e69–e76, Feb. 2012, doi: 10.1016/S1470-2045(11)70337-0.

[2]   K.-S. Choi, M. J. Cohn, and B. D. Harfe, 'Identification of nucleus pulposus precursor cells and notochordal remnants in the mouse: Implications for disk degeneration and chordoma formation', *Dev. Dyn.*, vol. 237, no. 12, pp. 3953–3958, Dec. 2008, doi: 10.1002/dvdy.21805.

[3]   R. G. C. Santegoeds, Y. Temel, J. C. Beckervordersandforth, J. J. Van Overbeeke, and C. M. Hoeberigs, 'State-of-the-Art Imaging in Human Chordoma of the Skull Base', *Curr. Radiol. Rep.*, vol. 6, no. 5, p. 16, May 2018, doi: 10.1007/s40134-018-0275-7.

[4]   A. M. Frezza, L. Botta, A. Trama, A. P. Dei Tos, and S. Stacchiotti, 'Chordoma: update on disease, epidemiology, biology and medical therapies', *Curr. Opin. Oncol.*, vol. 31, no. 2, pp. 114–120, Mar. 2019, doi: 10.1097/CCO.0000000000000502.

[5]   Y. Fu *et al.*, 'Factors associated with artificial airway retention after skull base chordoma resection: A retrospective cohort study', *Front. Neurol.*, vol. 13, p. 992308, Sep. 2022, doi: 10.3389/fneur.2022.992308.

[6]   S. M. Barber *et al.*, 'Chordoma—Current Understanding and Modern Treatment Paradigms', *J. Clin. Med.*, vol. 10, no. 5, p. 1054, Mar. 2021, doi: 10.3390/jcm10051054.

[7]   E. Soule *et al.*, 'Current Management and Image Review of Skull Base Chordoma: What the Radiologist Needs to Know', *J. Clin. Imaging Sci.*, vol. 11, p. 46, Aug. 2021, doi: 10.25259/JCIS_139_2021.

[8]   C. Teng, Q. Yang, Z. Xiong, N. Ye, and X. Li, 'Multivariate Analysis and Validation of the Prognostic Factors for Skull Base Chordoma', *Front. Surg.*, vol. 8, p. 764329, Nov. 2021, doi: 10.3389/fsurg.2021.764329.

[9]   S. Stacchiotti and J. Sommer, 'Building a global consensus approach to chordoma: a position paper from the medical and patient community', *Lancet Oncol.*, vol. 16, no. 2, pp. e71–e83, Feb. 2015, doi: 10.1016/S1470-2045(14)71190-8.

[10]   M. Alahmari and Y. Temel, 'Skull base chordoma treated with proton therapy: A systematic review', *Surg. Neurol. Int.*, vol. 10, p. 96, Jun. 2019, doi: 10.25259/SNI-213-2019.

[11]   A. Carpentier, M. Polivka, A. Blanquet, G. Lot, and B. George, 'Suboccipital and cervical chordomas: the value of aggressive treatment at first presentation of the disease', *J. Neurosurg.*, vol. 97, no. 5, pp. 1070–1077, Nov. 2002, doi: 10.3171/jns.2002.97.5.1070.

[12]   J. Zhou, B. Yang, X. Wang, and Z. Jing, 'Comparison of the Effectiveness of Radiotherapy with Photons and Particles for Chordoma After Surgery: A Meta-Analysis', *World Neurosurg.*, vol. 117, pp. 46–53, Sep. 2018, doi: 10.1016/j.wneu.2018.05.209.

[13]   S. Stacchiotti *et al.*, 'Best practices for the management of local-regional recurrent chordoma: a position paper by the Chordoma Global Consensus Group', *Ann. Oncol.*, vol. 28, no. 6, pp. 1230–1242, Jun. 2017, doi: 10.1093/annonc/mdx054.

[14]   M. Uhl *et al.*, 'Highly effective treatment of skull base chordoma with carbon ion irradiation using a raster scan technique in 155 patients: First long-term results: Carbon Ion Irradiation in Skull Base Chordoma', *Cancer*, vol. 120, no. 21, pp. 3410–3417, Nov. 2014, doi: 10.1002/cncr.28877.

[15]   M. Durante and J. S. Loeffler, 'Charged particles in radiation oncology', *Nat. Rev. Clin. Oncol.*, vol. 7, no. 1, pp. 37–43, Jan. 2010, doi: 10.1038/nrclinonc.2009.183.

[16]   D. Eriksson and T. Stigbrand, 'Radiation-induced cell death mechanisms', *Tumor Biol.*, vol. 31, no. 4, pp. 363–372, Aug. 2010, doi: 10.1007/s13277-010-0042-8.

[17]   W. C. Dewey, C. C. Ling, and R. E. Meyn, 'Radiation-induced apoptosis: Relevance to radiotherapy', *Int. J. Radiat. Oncol.*, vol. 33, no. 4, pp. 781–796, Nov. 1995, doi: 10.1016/0360-3016(95)00214-8.

[18]   M. Oldham, 'Radiation physics and applications in therapeutic medicine', *Phys. Educ.*, vol. 36, no. 6, pp. 460–467, Nov. 2001, doi: 10.1088/0031-9120/36/6/303.

[19]   M. F. McNitt-Gray, 'AAPM/RSNA Physics Tutorial for Residents: Topics in CT: Radiation Dose in CT', *RadioGraphics*, vol. 22, no. 6, pp. 1541–1553, Nov. 2002, doi: 10.1148/rg.226025128.

[20]   D. Verellen, M. D. Ridder, and G. Storme, 'A (short) history of image-guided radiotherapy', *Radiother. Oncol.*, vol. 86, no. 1, pp. 4–13, Jan. 2008, doi: 10.1016/j.radonc.2007.11.023.

[21]   P. M. Evans, 'Anatomical imaging for radiotherapy', *Phys. Med. Biol.*, vol. 53, no. 12, pp. R151–R191, Jun. 2008, doi: 10.1088/0031-9155/53/12/R01.

[22]   D. De Ruysscher, U. Nestle, R. Jeraj, and M. MacManus, 'PET scans in radiotherapy planning of lung cancer', *Lung Cancer*, vol. 75, no. 2, pp. 141–145, Feb. 2012, doi: 10.1016/j.lungcan.2011.07.018.

[23]   P. Decazes, P. Hinault, O. Veresezan, S. Thureau, P. Gouel, and P. Vera, 'Trimodality PET/CT/MRI and Radiotherapy: A Mini-Review', *Front. Oncol.*, vol. 10, p. 614008, Feb. 2021, doi: 10.3389/fonc.2020.614008.

[24]   E. Huynh *et al.*, 'Artificial intelligence in radiation oncology', *Nat. Rev. Clin. Oncol.*, vol. 17, no. 12, pp. 771–781, Dec. 2020, doi: 10.1038/s41571-020-0417-8.

[25] International Commission on Radiation Units and Measurements, Ed., *Prescribing, recording, and reporting photon beam therapy*. in ICRU report, no. 50. Bethesda, Md: Internat. Comm. on Radiation Units and Measurements, 1993.

[26] T. S. Hong, W. A. Tomé, and P. M. Harari, 'Heterogeneity in head and neck IMRT target design and clinical practice', *Radiother. Oncol.*, vol. 103, no. 1, pp. 92–98, Apr. 2012, doi: 10.1016/j.radonc.2012.02.010.

[27] X. A. Li *et al.*, 'Variability of Target and Normal Structure Delineation for Breast Cancer Radiotherapy: An RTOG Multi-Institutional and Multiobserver Study', *Int. J. Radiat. Oncol.*, vol. 73, no. 3, pp. 944–951, Mar. 2009, doi: 10.1016/j.ijrobp.2008.10.034.

[28] C. E. Cardenas, J. Yang, B. M. Anderson, L. E. Court, and K. B. Brock, 'Advances in Auto-Segmentation', *Semin. Radiat. Oncol.*, vol. 29, no. 3, pp. 185–197, Jul. 2019, doi: 10.1016/j.semradonc.2019.02.001.

[29] N. Ebert, F. Tillner, and M. Baumann, 'Advances in Radiation Oncology', in *Reference Module in Biomedical Sciences*, Elsevier, 2018, p. B9780128012383654000. doi: 10.1016/B978-0-12-801238-3.65428-4.

[30] P. M. Putora, M. Fruh, and L. Kern, 'The place of radiotherapy in the palliative management of NSCLC', *Breathe*, vol. 8, no. 2, pp. 134–143, Dec. 2011, doi: 10.1183/20734735.012711.

[31] R. E. Drzymala *et al.*, 'Dose-volume histograms', *Int. J. Radiat. Oncol.*, vol. 21, no. 1, pp. 71–78, May 1991, doi: 10.1016/0360-3016(91)90168-4.

[32] B. A. Fraass, A. Eisbruch, and M. Feng, 'Intensity-Modulated and Image-Guided Radiation Therapy', in *Clinical Radiation Oncology*, Elsevier, 2016, pp. 294-324.e5. doi: 10.1016/B978-0-323-24098-7.00016-2.

[33] M. Durante, R. Orecchia, and J. S. Loeffler, 'Charged-particle therapy in cancer: clinical uses and future perspectives', *Nat. Rev. Clin. Oncol.*, vol. 14, no. 8, pp. 483–495, Aug. 2017, doi: 10.1038/nrclinonc.2017.30.

[34] M. D. Simoni *et al.*, 'FRED: a fast Monte Carlo code on GPU for quality control in Particle Therapy', *J. Phys. Conf. Ser.*, vol. 1548, no. 1, p. 012020, May 2020, doi: 10.1088/1742-6596/1548/1/012020.

[35] C. Grau, M. Durante, D. Georg, J. A. Langendijk, and D. C. Weber, 'Particle therapy in Europe', *Mol. Oncol.*, vol. 14, no. 7, pp. 1492–1499, Jul. 2020, doi: 10.1002/1878-0261.12677.

[36] L. Schaub, S. B. Harrabi, and J. Debus, 'Particle therapy in the future of precision therapy', *Br. J. Radiol.*, vol. 93, no. 1114, p. 20200183, Oct. 2020, doi: 10.1259/bjr.20200183.

[37] H. Murshed, 'Radiation Biology', in *Fundamentals of Radiation Oncology*, Elsevier, 2019, pp. 57–87. doi: 10.1016/B978-0-12-814128-1.00003-9.

[38]   D. A. Bluemke and S. Liu, 'Imaging in Clinical Trials', in *Principles and Practice of Clinical Research*, Elsevier, 2012, pp. 597–617. doi: 10.1016/B978-0-12-382167-6.00041-2.

[39]   F. Kalholm, L. Grzanka, E. Traneus, and N. Bassler, 'A systematic review on the usage of averaged LET in radiation biology for particle therapy', *Radiother. Oncol.*, vol. 161, pp. 211–221, Aug. 2021, doi: 10.1016/j.radonc.2021.04.007.

[40]   S. E. McGowan, N. G. Burnet, and A. J. Lomax, 'Treatment planning optimisation in proton therapy', *Br. J. Radiol.*, vol. 86, no. 1021, pp. 20120288–20120288, Jan. 2013, doi: 10.1259/bjr.20120288.

[41]   R. Mohan and D. Grosshans, 'Proton therapy – Present and future', *Adv. Drug Deliv. Rev.*, vol. 109, pp. 26–44, Jan. 2017, doi: 10.1016/j.addr.2016.11.006.

[42]   M. Beuve, 'Formalization and Theoretical Analysis of the Local Effect Model', *Radiat. Res.*, vol. 172, no. 3, pp. 394–402, Sep. 2009, doi: 10.1667/RR1544.1.

[43]   D. C. Weber *et al.*, 'Long term outcomes of patients with skull-base low-grade chondrosarcoma and chordoma patients treated with pencil beam scanning proton therapy', *Radiother. Oncol.*, vol. 120, no. 1, pp. 169–174, Jul. 2016, doi: 10.1016/j.radonc.2016.05.011.

[44]   A. Iannalfi *et al.*, 'Proton and carbon ion radiotherapy in skull base chordomas: a prospective study based on a dual particle and a patient-customized treatment strategy', *Neuro-Oncol.*, vol. 22, no. 9, pp. 1348–1358, Sep. 2020, doi: 10.1093/neuonc/noaa067.

[45]   M.-X. Zou, G.-H. Lv, Q.-S. Zhang, S.-F. Wang, J. Li, and X.-B. Wang, 'Prognostic Factors in Skull Base Chordoma: A Systematic Literature Review and Meta-Analysis', *World Neurosurg.*, vol. 109, pp. 307–327, Jan. 2018, doi: 10.1016/j.wneu.2017.10.010.

[46]   P. Lambin *et al.*, 'Radiomics: the bridge between medical imaging and personalized medicine', *Nat. Rev. Clin. Oncol.*, vol. 14, no. 12, pp. 749–762, Dec. 2017, doi: 10.1038/nrclinonc.2017.141.

[47]   R. J. Gillies, P. E. Kinahan, and H. Hricak, 'Radiomics: Images Are More than Pictures, They Are Data', *Radiology*, vol. 278, no. 2, pp. 563–577, Feb. 2016, doi: 10.1148/radiol.2015151169.

[48]   J. Wu, K. K. Tha, L. Xing, and R. Li, 'Radiomics and radiogenomics for precision radiotherapy', *J. Radiat. Res. (Tokyo)*, vol. 59, no. suppl_1, pp. i25–i31, Mar. 2018, doi: 10.1093/jrr/rrx102.

[49]   S. S. F. Yip and H. J. W. L. Aerts, 'Applications and limitations of radiomics', *Phys. Med. Biol.*, vol. 61, no. 13, pp. R150–R166, Jul. 2016, doi: 10.1088/0031-9155/61/13/R150.

[50] J. E. van Timmeren, D. Cester, S. Tanadini-Lang, H. Alkadhi, and B. Baessler, 'Radiomics in medical imaging—"how-to" guide and critical reflection', *Insights Imaging*, vol. 11, no. 1, p. 91, Dec. 2020, doi: 10.1186/s13244-020-00887-2.

[51] P. Lambin *et al.*, 'Radiomics: Extracting more information from medical images using advanced feature analysis', *Eur. J. Cancer*, vol. 48, no. 4, pp. 441–446, Mar. 2012, doi: 10.1016/j.ejca.2011.11.036.

[52] A. Zwanenburg *et al.*, 'The Image Biomarker Standardization Initiative: Standardized Quantitative Radiomics for High-Throughput Image-based Phenotyping', *Radiology*, vol. 295, no. 2, pp. 328–338, May 2020, doi: 10.1148/radiol.2020191145.

[53] P. Lohmann *et al.*, 'Radiomics in neuro-oncology: Basics, workflow, and applications', *Methods*, vol. 188, pp. 112–121, Apr. 2021, doi: 10.1016/j.ymeth.2020.06.003.

[54] A. Zwanenburg, 'Radiomics in nuclear medicine: robustness, reproducibility, standardization, and how to avoid data analysis traps and replication crisis', *Eur. J. Nucl. Med. Mol. Imaging*, vol. 46, no. 13, pp. 2638–2655, Dec. 2019, doi: 10.1007/s00259-019-04391-8.

[55] J. J. M. van Griethuysen *et al.*, 'Computational Radiomics System to Decode the Radiographic Phenotype', *Cancer Res.*, vol. 77, no. 21, pp. e104–e107, Nov. 2017, doi: 10.1158/0008-5472.CAN-17-0339.

[56] G. S. Collins, J. B. Reitsma, D. G. Altman, and K. Moons, 'Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): the TRIPOD Statement', *BMC Med.*, vol. 13, no. 1, p. 1, 2015, doi: 10.1186/s12916-014-0241-z.

[57] C. Vercellis, *Business intelligence: data mining and optimization for decision making*. Chichester, U.K: Wiley, 2009.

[58] Y. Jung and J. Hu, 'A *K* -fold averaging cross-validation procedure', *J. Nonparametric Stat.*, vol. 27, no. 2, pp. 167–179, Apr. 2015, doi: 10.1080/10485252.2015.1010532.

[59] J. Song, Y. Yin, H. Wang, Z. Chang, Z. Liu, and L. Cui, 'A review of original articles published in the emerging field of radiomics', *Eur. J. Radiol.*, vol. 127, p. 108991, Jun. 2020, doi: 10.1016/j.ejrad.2020.108991.

[60] H. J. W. L. Aerts *et al.*, 'Decoding tumour phenotype by noninvasive imaging using a quantitative radiomics approach', *Nat. Commun.*, vol. 5, no. 1, p. 4006, Jun. 2014, doi: 10.1038/ncomms5006.

[61] F. Urraro *et al.*, 'MRI Radiomics in Prostate Cancer: A Reliability Study', *Front. Oncol.*, vol. 11, p. 805137, Dec. 2021, doi: 10.3389/fonc.2021.805137.

[62]   X. Guan *et al.*, 'The preliminary results of proton and carbon ion therapy for chordoma and chondrosarcoma of the skull base and cervical spine', *Radiat. Oncol.*, vol. 14, no. 1, p. 206, Dec. 2019, doi: 10.1186/s13014-019-1407-9.

[63]   M. Dominietto *et al.*, 'Deep Learning Based on Radiomics Features Dataset to Predict the Outcome of Skull-Base Chordomas Patients Treated with Pencil Beam Scanning Proton Therapy', *Int. J. Radiat. Oncol.*, vol. 105, no. 1, p. E134, Sep. 2019, doi: 10.1016/j.ijrobp.2019.06.2168.

[64]   W.-C. Yang, F.-M. Hsu, and P.-C. Yang, 'Precision radiotherapy for non-small cell lung cancer', *J. Biomed. Sci.*, vol. 27, no. 1, p. 82, Dec. 2020, doi: 10.1186/s12929-020-00676-5.

[65]   X. Zhang *et al.*, 'Deep Learning With Radiomics for Disease Diagnosis and Treatment: Challenges and Potential', *Front. Oncol.*, vol. 12, p. 773840, Feb. 2022, doi: 10.3389/fonc.2022.773840.

[66]   A. Wu *et al.*, 'Dosiomics improves prediction of locoregional recurrence for intensity modulated radiotherapy treated head and neck cancer cases', *Oral Oncol.*, vol. 104, p. 104625, May 2020, doi: 10.1016/j.oraloncology.2020.104625.

[67]   Z. Dai *et al.*, 'PO-1811 Multi-modality fusion dosiomics for prediction of survival in nasopharyngeal carcinoma patients', *Radiother. Oncol.*, vol. 161, pp. S1540–S1542, Aug. 2021, doi: 10.1016/S0167-8140(21)08262-1.

[68]   G. Buizza *et al.*, 'Radiomics and Dosiomics for Predicting Local Control after Carbon-Ion Radiotherapy in Skull-Base Chordoma', *Cancers*, vol. 13, no. 2, p. 339, Jan. 2021, doi: 10.3390/cancers13020339.

[69]   L. Morelli *et al.*, 'A Dosiomics Analysis Based on Linear Energy Transfer and Biological Dose Maps to Predict Local Recurrence in Sacral Chordomas after Carbon-Ion Radiotherapy', *Cancers*, vol. 15, no. 1, p. 33, Dec. 2022, doi: 10.3390/cancers15010033.

[70]   S. Matsumoto *et al.*, 'Unresectable Chondrosarcomas Treated With Carbon Ion Radiotherapy: Relationship Between Dose-averaged Linear Energy Transfer and Local Recurrence', *Anticancer Res.*, vol. 40, no. 11, pp. 6429–6435, Nov. 2020, doi: 10.21873/anticanres.14664.

[71]   P. Sulewski, 'Equal-bin-width histogram versus equal-bin-count histogram', *J. Appl. Stat.*, vol. 48, no. 12, pp. 2092–2111, Sep. 2021, doi: 10.1080/02664763.2020.1784853.

[72]   R. T. H. Leijenaar *et al.*, 'The effect of SUV discretization in quantitative FDG-PET Radiomics: the need for standardized methodology in tumor texture analysis', *Sci. Rep.*, vol. 5, no. 1, p. 11075, Aug. 2015, doi: 10.1038/srep11075.

[73]   M. Shafiq-ul-Hassan *et al.*, 'Intrinsic dependencies of CT radiomic features on voxel size and number of gray levels', *Med. Phys.*, vol. 44, no. 3, pp. 1050–1062, Mar. 2017, doi: 10.1002/mp.12123.

[74]   L. Duron *et al.*, 'Gray-level discretization impacts reproducible MRI radiomics texture features', *PLOS ONE*, vol. 14, no. 3, p. e0213459, Mar. 2019, doi: 10.1371/journal.pone.0213459.

[75]   T. K. Koo and M. Y. Li, 'A Guideline of Selecting and Reporting Intraclass Correlation Coefficients for Reliability Research', *J. Chiropr. Med.*, vol. 15, no. 2, pp. 155–163, Jun. 2016, doi: 10.1016/j.jcm.2016.02.012.

[76]   B. George, S. Seals, and I. Aban, 'Survival analysis and regression models', *J. Nucl. Cardiol.*, vol. 21, no. 4, pp. 686–694, Aug. 2014, doi: 10.1007/s12350-014-9908-2.

[77]   S. Prinja, N. Gupta, and R. Verma, 'Censoring in clinical trials: Review of survival analysis techniques', *Indian J. Community Med.*, vol. 35, no. 2, p. 217, 2010, doi: 10.4103/0970-0218.66859.

[78]   J. In and D. K. Lee, 'Survival analysis: Part I — analysis of time-to-event', *Korean J. Anesthesiol.*, vol. 71, no. 3, pp. 182–191, Jun. 2018, doi: 10.4097/kja.d.18.00067.

[79]   L. D. Fisher and D. Y. Lin, 'TIME-DEPENDENT COVARIATES IN THE COX PROPORTIONAL-HAZARDS REGRESSION MODEL', *Annu. Rev. Public Health*, vol. 20, no. 1, pp. 145–157, May 1999, doi: 10.1146/annurev.publhealth.20.1.145.

[80]   F. E. Harrell, K. L. Lee, and D. B. Mark, 'MULTIVARIABLE PROGNOSTIC MODELS: ISSUES IN DEVELOPING MODELS, EVALUATING ASSUMPTIONS AND ADEQUACY, AND MEASURING AND REDUCING ERRORS', *Stat. Med.*, vol. 15, no. 4, pp. 361–387, Feb. 1996, doi: 10.1002/(SICI)1097-0258(19960229)15:4<361::AID-SIM168>3.0.CO;2-4.

[81]   F. E. Harrell and R. M. Califf, 'Evaluating the Yield of Medical Tests'.

Appendix

A.1 Features classes

Features were extracted from physical dose, RBE-weighted dose and LET$_d$ maps through PyRadiomics open-source python package (v.2.2.0). Features were extracted in 3D without the application of filters, normalization and resampling. Below are listed the extracted features divided by classes:

- **Shape**
  Elongation, Flatness, Least Axis Length, Major Axis Length, Maximum 2D Diameter Column, Maximum 2D Diameter Row, Maximum 2D Diameter Slice, Maximum 3D Diameter, Mesh Volume, Minor Axis Length, Sphericity, Surface Area, Surface Volume Ratio and Voxel Volume.

- **First-order**
  10 Percentile, 90 Percentile, Energy, Entropy, Interquartile Range, Kurtosis, Maximum, Mean Absolute Deviation, Mean, Median, Minimum, Range, Robust Mean Absolute Deviation, Root Mean Squared, Skewness, Total Energy, Uniformity and Variance

- **Texture:**
  - Gray Level Co-occurrence Matrix (GLCM)
    Autocorrelation, cluster prominence, cluster shade, cluster tendency, contrast, correlation, difference average, difference entropy, difference variance, inverse difference, inverse difference moment, inverse difference moment normalized, inverse difference normalized, information measure of correlation 1, information measure of correlation 2, inverse variance, joint average, joint entropy, joint energy, maximal correlation coefficient, maximum probability, sum average, sum entropy, sum squares;
  - Gray Level Run Length Matrix (GLRLM)
    GL non uniformity, GL non-uniformity normalized, GL variance, high GL run emphasis, long-run emphasis, long-run high GL emphasis, long run low GL emphasis, low GL emphasis, run entropy, run length non-uniformity, run length non-uniformity normalized, run percentage, run variance, short run emphasis, short run GL emphasis, short run low GL emphasis;
  - Gray Level Size Zone Matrix (GLSZM)

GL non uniformity, GL non uniformity normalized, GL variance, high GL zone emphasis, large area emphasis, large area high GL emphasis, large area low GL emphasis, low GL zone emphasis, size zone non uniformity, size zone non uniformity normalized, small area emphasis, small area high GL emphasis, small area low GL emphasis, zone entropy, zone percentage, zone variance;

- Gray Level Dependence Matrix (GLDM)
  Dependence entropy, dependence non uniformity, dependence non uniformity normalized, dependence variance, GL non uniformity, GL variance, high GL emphasis, large dependence emphasis, large dependence high GL emphasis, large dependence low GL emphasis, low GL emphasis, small dependence emphasis, small dependence high GL emphasis, small dependence low GL emphasis;

- Neighbourhood Gray Tone Difference Matrix (NGTDM)
  Busyness, coarseness, complexity, contrast, strength.

A.2   Selected features

A.2.1   Approach 1

In this section selected features related to approach 1 are reported in Table A. 1, Table A. 2 and Table A. 3for physical dose, LET$_d$ and RBE-weighted dose maps, respectively, divided by binwidth and bincount methods.

Physical dose map   -   selected texture features

Bincount method		Binwidth method	
GLCM	correlation, inverse difference, inverse difference moment, inverse difference moment normalized, inverse difference normalized	GLCM	difference entropy, inverse difference, inverse difference moment, inverse difference normalized, sum entropy
GLRLM	gray level non uniformity, run length non uniformity normalized	GLRLM	gray level non uniformity, run entropy
GLSZM	zone entropy	GLSZM	zone entropy

GLDM	dependence entropy, small dependence low gray level emphasis	GLDM	dependence entropy, small dependence low gray level emphasis

Table A. 1 - Selected features for physical dose map.

LET$_d$ map - selected texture features

Bincount method		Binwidth method	
GLCM	correlation, inverse difference, inverse difference moment, inverse difference moment normalized, inverse difference normalized	GLCM	difference entropy, inverse difference, inverse difference moment, inverse difference moment normalized, inverse different normalized, sum entropy
GLRLM	gray level non uniformity, run length non uniformity normalized, run entropy	GLRLM	run entropy
GLDM	small dependence low gray level emphasis	GLSZM	zone entropy
NGTDM	coarseness	GLDM	dependence entropy, small dependence low gray level emphasis

Table A. 2 - Selected features for LET$_d$ map.

RBE-weighted dose map – selected features

Binwidth	method
GLCM	difference entropy, inverse difference, inverse difference moment, inverse difference moment normalized, inverse different normalized, sum entropy
GLRLM	run entropy
GLSZM	zone entropy
GLDM	dependence entropy

Table A. 3 - Selected features for RBE-weighted dose map.

A.2.2   Approach 2

In this section selected features related to approach 2 are reported in the following tables (for each value the binwidth/bincount parameters can assume). From Table A. 4 to Table A. 29 results for physical dose are shown, from Table A. 30 to Table A. 54 the ones related to LET_d are reported and from Table A. 55 to Table A. 68 those referred to RBE-weighted dose.

Physical dose

Binwidth: 0.1	
GLCM	correlation, difference entropy, inverse difference, inverse difference moment
GLRLM	run length non uniformity normalized, gray level non uniformity
first-order	uniformity, minimum, range

Table A. 4

Binwidth: 0.2	
GLCM	correlation, difference entropy, inverse difference, inverse difference moment
GLCM	sum entropy
GLRLM	sum entropy, run entropy
first-order	entropy, minimum, range

Table A. 5

Binwidth: 0.3	
GLCM	correlation, inverse difference, inverse difference moment, joint entropy
GLRLM	run entropy, run length non uniformity normalized
GLSZM	gray level non uniformity
first-order	entropy, minimum, range

Table A. 6

Binwidth: 0.4	
GLCM	correlation, inverse difference, inverse difference moment, joint entropy
GLRLM	run length non uniformity normalized
GLSZM	gray level non uniformity, zone entropy
first-order	entropy, minimum, range

Table A. 7

Binwidth: 0.5	
GLCM	correlation, difference entropy, inverse difference, inverse difference moment, sum entropy
GLRLM	run entropy, run length non uniformity normalized
first-order	entropy, minimum, range

Table A. 8

Binwidth: 0.8	
GLCM	correlation, inverse difference, inverse difference moment, joint entropy
GLRLM	run length non uniformity normalized
GLSZM	gray level non uniformity, zone entropy
first-order	entropy, minimum, range

Table A. 9

Binwidth: 1	
GLCM	correlation, inverse difference, inverse difference moment, joint entropy, sum entropy
GLRLM	run length non uniformity normalized
GLSZM	gray level non uniformity

first-order	entropy, minimum, range

Table A. 10

Binwidth: 2	
GLCM	correlation, inverse difference, inverse difference moment, inverse difference normalized
GLRLM	run length non uniformity normalized
GLSZM	gray level non uniformity, zone entropy
first-order	entropy, minimum, range

Table A. 11

Binwidth: 3	
GLCM	correlation, inverse difference, inverse difference moment, inverse different moment normalized, inverse difference normalized, joint entropy, sum entropy
GLRLM	run length non uniformity normalized
GLSZM	gray level non uniformity
first-order	minimum

Table A. 12

Binwidth: 4	
GLCM	correlation, inverse difference, inverse difference moment, , inverse different moment normalized, inverse difference normalized, sum entropy
GLRLM	run length non uniformity normalized
GLSZM	gray level non uniformity
first-order	minimum, range

Table A. 13

Binwidth: 5	
GLCM	inverse difference, inverse difference normalized
GLRLM	gray level non uniformity, run length non uniformity, run length non uniformity normalized
GLSZM	gray level non uniformity, zone entropy, small dependence low gray level emphasis
first-order	minimum

Table A. 14

Binwidth: 8	
GLCM	correlation, inverse difference normalized
GLRLM	run length non uniformity, run length non uniformity normalized
GLSZM	gray level non uniformity, zone entropy, dependence entropy, small dependence low gray level emphasis
first-order	minimum, range

Table A. 15

Binwidth: 10	
GLCM	correlation, inverse difference normalized, joint entropy
GLRLM	gray level non uniformity, run length non uniformity, run length non uniformity normalized
GLSZM	gray level non uniformity, zone entropy, dependence entropy, small dependence low gray level emphasis

Table A. 16

Binwidth: 10	
GLCM	correlation
GLRLM	run entropy, run length non uniformity, run length non uniformity normalized

GLSZM	gray level non uniformity, zone entropy, dependence entropy, small dependence low gray level emphasis

Table A. 17

Bincount: 2	
GLCM	correlation, inverse difference, inverse difference moment, inverse difference moment normalized, inverse difference normalized
GLRLM	gray level non uniformity
GLSZM	zone entropy
GLDM	dependence entropy, small dependence low gray level emphasis
first-order	range

Table A. 18

Bincount: 3	
GLCM	correlation, inverse difference moment, inverse difference moment normalized, inverse difference normalized
GLRLM	gray level non uniformity
GLSZM	zone entropy
GLDM	dependence entropy, small dependence low gray level emphasis
first-order	minimum, range

Table A. 19

Bincount: 4	
GLCM	correlation, inverse difference moment normalized, inverse difference normalized
GLRLM	gray level non uniformity, run length non uniformity normalized
GLSZM	zone entropy
GLDM	dependence entropy, small dependence low gray level emphasis

first-order	minimum, range

Table A. 20

Bincount: 5	
GLCM	correlation, inverse difference moment normalized, inverse difference normalized
GLRLM	gray level non uniformity, run length non uniformity normalized
GLSZM	zone entropy
GLDM	dependence entropy, small dependence low gray level emphasis
first-order	minimum, range

Table A. 21

Bincount: 8	
GLCM	correlation, inverse difference moment normalized, inverse difference normalized
GLRLM	gray level non uniformity, run length non uniformity normalized
GLSZM	zone entropy
GLDM	dependence entropy, small dependence low gray level emphasis
first-order	minimum, range

Table A. 22

Bincount: 10	
GLCM	correlation, inverse difference moment, inverse difference moment normalized, inverse difference normalized
GLRLM	run length non uniformity normalized
GLSZM	zone entropy
GLDM	dependence entropy, small dependence low gray level emphasis
first-order	minimum, range

Table A. 23

Bincount: 16	
GLCM	correlation, inverse difference moment normalized, inverse difference normalized
GLRLM	gray level non uniformity, run length non uniformity normalized
GLSZM	zone entropy
GLDM	dependence entropy, small dependence low gray level emphasis
first-order	minimum, range

Table A. 24

Bincount: 32	
GLCM	correlation, inverse difference moment, inverse difference moment normalized, inverse difference normalized
GLRLM	gray level non uniformity
GLSZM	zone entropy
GLDM	dependence entropy, small dependence low gray level emphasis
first-order	minimum, range

Table A. 25

Bincount: 64	
GLCM	correlation, inverse difference, inverse different moment, inverse difference moment normalized, inverse difference normalized
GLRLM	run length non uniformity normalized
GLSZM	zone entropy
GLDM	dependence entropy
first-order	minimum, range

Table A. 26

Bincount: 128	
GLCM	correlation, inverse difference, inverse different moment, inverse difference moment normalized, inverse difference normalized
GLRLM	run length non uniformity normalized
GLSZM	zone entropy
GLDM	dependence entropy
first-order	minimum, range

Table A. 27

Bincount: 256	
GLCM	correlation, inverse difference moment, inverse difference moment normalized, inverse difference normalized
GLRLM	gray level non uniformity, run length non uniformity normalized
GLSZM	zone entropy
GLDM	dependence entropy
first-order	minimum, range

Table A. 28

Bincount: 512	
GLCM	correlation, inverse difference moment, inverse difference moment normalized, inverse difference normalized
GLRLM	gray level non uniformity, run length non uniformity normalized
GLSZM	zone entropy
GLDM	dependence entropy
first-order	minimum, range

Table A. 29

LET$_d$

Binwidth: 0.01	
GLCM	difference entropy, inverse difference moment
GLRLM	run length non uniformity normalized, run percentage
GLSZM	zone percentage
GLDM	small dependence emphasis
first-order	10 Percentile, mean, median, root mean squared

Table A. 30

Binwidth: 0.02	
GLCM	difference entropy
GLRLM	run length non uniformity normalized, run percentage
GLSZM	zone percentage
GLDM	small dependence emphasis
first-order	10 Percentile, mean, median, minimum, root mean squared

Table A. 31

Binwidth: 0.03	
GLCM	difference entropy, inverse difference moment
GLRLM	run length non uniformity normalized, run percentage
GLSZM	zone percentage
GLDM	small dependence emphasis
first-order	10 Percentile, mean, median, root mean squared

Table A. 32

Binwidth: 0.04	
GLCM	difference entropy, sum entropy
GLRLM	run percentage
GLSZM	zone percentage
GLDM	small dependence emphasis
first-order	10 Percentile, mean, median, minimum, root mean squared

Table A. 33

Binwidth: 0.05	
GLCM	difference entropy, sum entropy
GLRLM	run length non uniformity normalized, run percentage
GLSZM	zone percentage
GLDM	small dependence emphasis
first-order	10 Percentile, mean, median, root mean squared

Table A. 34

Binwidth: 0.08	
GLCM	difference entropy, joint entropy, sum entropy
GLSZM	zone percentage
GLDM	small dependence emphasis
first-order	10 Percentile, mean, median, minimum, root mean squared

Table A. 35

Binwidth: 0.1	
GLCM	difference entropy, inverse difference, inverse difference moment, joint entropy, sum entropy
GLSZM	zone percentage

GLDM	small dependence emphasis
first-order	mean, median, root mean squared

Table A. 36

Binwidth: 0.2	
GLCM	difference entropy, inverse difference, inverse difference moment, joint entropy
GLRLM	run percentage
GLSZM	gray level non uniformity, zone percentage
GLDM	small dependence emphasis
first-order	minimum, root mean squared

Table A. 37

Binwidth: 0.3	
GLCM	difference entropy, inverse difference, inverse difference moment, joint entropy
GLRLM	run length non uniformity normalized, run percentage
GLSZM	gray level non uniformity
GLDM	small dependence emphasis
first-order	mean, root mean squared

Table A. 38

Binwidth: 0.4	
GLCM	difference entropy, inverse difference, inverse difference moment, joint entropy, sum entropy
GLRLM	run length non uniformity normalized
GLSZM	gray level non uniformity, zone percentage
GLDM	small dependence emphasis

first-order	minimum

Table A. 39

Binwidth: 0.5	
GLCM	inverse difference, inverse difference moment, inverse difference normalized, sum entropy
GLRLM	run length non uniformity normalized, run percentage
GLSZM	gray level non uniformity, zone percentage
GLDM	small dependence emphasis
first-order	minimum

Table A. 40

Binwidth: 0.8	
GLCM	inverse difference moment normalized, sum entropy
GLRLM	run length non uniformity normalized, run percentage
GLSZM	gray level non uniformity, zone entropy, zone percentage
GLDM	dependence entropy, small dependence emphasis, small dependence low gray level emphasis

Table A. 41

Binwidth: 1	
GLCM	inverse difference moment normalized, sum entropy
GLRLM	run length non uniformity
GLSZM	gray level non uniformity, zone entropy, zone percentage
GLDM	dependency entropy, small dependence emphasis, small dependence low gray level emphasis
first-order	minimum

Table A. 42

Bincount: 2	
GLCM	correlation, inverse difference, inverse difference moment, inverse difference moment normalized, inverse difference normalized
GLSZM	zone entropy
GLDM	dependence entropy
first-order	minimum, root mean squared

Table A. 43

Bincount: 3	
GLCM	correlation, inverse difference, inverse difference moment, inverse difference normalized
GLSZM	zone entropy
GLDM	dependence entropy
first-order	mean, median, minimum, root mean squared

Table A. 44

Bincount: 4	
GLCM	correlation
GLSZM	zone entropy
GLDM	dependence entropy, small dependence low gray level emphasis
first-order	10 percentile, mean, median, minimum, root mean squared

Table A. 45

Bincount: 5	
GLCM	correlation
GLSZM	zone entropy
GLDM	dependence entropy, small dependence low gray level emphasis

NGTDM	coarseness
first-order	10 percentile, mean, median, minimum, root mean squared

Table A. 46

Bincount: 8	
GLCM	correlation
GLRLM	run percentage, short run emphasis
GLSZM	zone entropy
GLDM	dependence entropy
first-order	10 percentile, mean, median, minimum, root mean squared

Table A. 47

Bincount: 10	
GLCM	correlation, inverse difference normalized
GLRLM	run percentage, zone entropy
GLDM	dependence entropy
first-order	10 percentile, mean, median, minimum, root mean squared

Table A. 48

Bincount: 16	
GLCM	correlation, inverse difference, inverse difference moment
GLRLM	run entropy, zone entropy
first-order	10 percentile, mean, median, minimum, root mean squared

Table A. 49

Bincount: 32	
GLCM	correlation
GLRLM	run percentage, short run emphasis
GLSZM	zone entropy
GLDM	small dependence low gray level emphasis
first-order	10 percentile, mean, median, minimum, root mean squared

Table A. 50

Bincount: 64	
GLCM	correlation, inverse difference normalized
GLRLM	run entropy
GLSZM	zone entropy
GLDM	small dependence low gray level emphasis
first-order	10 percentile, mean, median, minimum, root mean squared

Table A. 51

Bincount: 128	
GLCM	correlation, inverse difference normalized
GLRLM	run entropy
GLSZM	zone entropy
GLDM	small dependence low gray level emphasis
first-order	10 percentile, mean, median, minimum, root mean squared

Table A. 52

Bincount: 256	
GLCM	correlation, inverse difference moment
GLRLM	run entropy
GLSZM	zone entropy
GLDM	small dependence low gray level emphasis
first-order	10 percentile, mean, median, minimum, root mean squared

Table A. 53

Bincount: 512	
GLCM	correlation, inverse difference normalized
GLRLM	run entropy
GLSZM	zone entropy
NGTDM	coarseness
first-order	10 percentile, mean, median, minimum, root mean squared

Table A. 54

RBE-weighted dose

Binwidth: 0.1	
GLCM	correlation, difference entropy, inverse difference, inverse difference moment, maximal correlation coefficient
GLSZM	zone percentage
GLDM	small dependence emphasis
first-order	minimum, range, uniformity

Table A. 55

Binwidth: 0.2	
GLCM	difference entropy, inverse difference, inverse difference moment, joint entropy
GLRLM	run percentage
GLSZM	zone percentage
GLDM	gray level non uniformity, small dependence emphasis
first-order	minimum, root mean squared

Table A. 56

Binwidth: 0.3	
GLCM	difference entropy, inverse difference, inverse difference moment, joint entropy
GLRLM	run length non uniformity normalized, run percentage
GLSZM	gray level non uniformity
GLDM	small dependence emphasis
first-order	mean, root mean squared

Table A. 57

Binwidth: 0.4	
GLCM	difference entropy, inverse difference, inverse difference moment, joint entropy, sum entropy
GLRLM	run length non uniformity normalized
GLSZM	gray level non uniformity, zone percentage
GLDM	small dependence emphasis
first-order	minimum

Table A. 58

Binwidth: 0.5	
GLCM	inverse difference, inverse difference moment, inverse difference normalized, sum entropy
GLRLM	run length non uniformity normalized, run percentage
GLSZM	gray level non uniformity, zone percentage
GLDM	small dependence emphasis
first-order	minimum

Table A. 59

Binwidth: 0.8	
GLCM	correlation, inverse difference normalized, joint entropy, maximal correlation coefficient, sum entropy
GLSZM	zone percentage
GLDM	small dependence emphasis
first-order	entropy, minimum, range

Table A. 60

Binwidth: 1	
GLCM	correlation, joint entropy, maximal correlation coefficient, sum entropy
GLSZM	zone entropy, zone percentage
GLDM	small dependence emphasis
first-order	entropy, minimum, range

Table A. 61

Binwidth: 2	
GLCM	inverse difference, inverse difference moment, maximal correlation coefficient
GLRLM	run length non uniformity normalized, run percentage

GLSZM	gray level non uniformity, zone percentage
GLDM	small dependence emphasis
first-order	minimum, range

Table A. 62

Binwidth: 3	
GLCM	correlation, inverse difference, inverse difference moment, inverse difference normalized, maximal correlation coefficient
GLRLM	run length non uniformity normalized
GLSZM	gray level non uniformity, zone percentage
GLDM	small dependence emphasis
first-order	minimum

Table A. 63

Binwidth: 4	
GLCM	inverse difference, inverse difference moment normalized, inverse difference normalized
GLRLM	run length non uniformity, run length non uniformity normalized
GLSZM	gray level non uniformity, zone percentage
GLDM	small dependence emphasis, small dependence low gray level emphasis
first-order	minimum

Table A. 64

Binwidth: 5	
GLCM	inverse difference moment normalized, inverse difference normalized
GLRLM	run length non uniformity normalized, run percentage
GLSZM	gray level non uniformity, zone entropy, zone percentage

GLDM	small dependence emphasis
first-order	Minimum, range

Table A. 65

Binwidth: 8	
GLCM	correlation, inverse difference moment normalized, maximal correlation coefficient
GLRLM	run percentage
GLSZM	gray level non uniformity, zone percentage
GLDM	small dependence emphasis, small dependence low gray level emphasis
first-order	minimum, range

Table A. 66

Binwidth: 10	
GLCM	correlation, maximal correlation coefficient
GLRLM	run entropy, run length non uniformity
GLSZM	gray level non uniformity, zone entropy
GLDM	dependence entropy, small dependence low gray level emphasis
first-order	minimum, range

Table A. 67

Binwidth: 10	
GLCM	correlation, inverse difference moment normalized, maximal correlation coefficient
GLRLM	run length non uniformity
GLSZM	gray level non uniformity, zone entropy
GLDM	dependence entropy, small dependence low gray level emphasis

first-order	minimum, range

<div align="center">Table A. 68</div>

A.3 Models' performance

A.3.1 Approach 1

In Table A. 69, Table A. 70, Table A. 71, Table A. 72 and Table A. 73 are reported both the values employed in features selection for the parameter bincount/binwidth and the corresponding model's performance in terms of C-index (median/interquartile range). Tables refer to physical dose, LET_d and RBE-weighted dose maps, respectively.

Bincount	2.0	3.0	4.0	5.0	8.0	10.0	16.0	32.0	64.0	128.0	256.0	512.0
C-index	0.67/0.19	0.62/0.14	0.62/0.23	0.62/0.14	0.62/0.19	0.62/0.18	0.62/0.19	0.71/0.1	0.67/0.1	0.71/0.14	0.76/0.14*	0.67/0.24

Table A. 69 – Bincount values employed for features extraction with relative model's performance (C-index) for physical dose maps. Results are shown in terms of median/IQR. The best value is highlighted with the symbol *.

Binwidth	0.1	0.2	0.3	0.4	0.5	0.8	1.0
C-index	0.64/0.14	0.67/0.14	0.65/0.14	0.62/0.14	0.67/.14	0.63/0.14	0.62/0.14

Binwidth	2.0	3.0	4.0	5.0	8.0	10.0	16.0
C-index	0.64/0.14	0.67/0.11	0.62/0.14	0.76/0.19*	0.71/0.21	0.62/0.14	0.62/0.19

Table A. 70 – Binwidth values employed for features extraction with relative model's performance (C-index) for physical dose maps. Results are shown in terms of median/IQR. The best value is highlighted with the symbol *.

Bincount	2.0	3.0	4.0	5.0	8.0	10.0	16.0	32.0	64.0	128.0	256.0	512.0
C-index	0.64/0.24*	0.62/0.19	0.62/0.19	0.59/0.18	0.6/0.21	0.57/0.13	0.63/0.14	0.6/0.15	0.6/0.14	0.6/0.17	0.6/0.15	0.6/0.17

Table A. 71 – Bincount values employed for features extraction with relative model's performance (C-index) for LET$_d$ maps. Results are shown in terms of median/IQR. The best value is highlighted with the symbol *.

Binwidth	0.01	0.02	0.03	0.04	0.05	0.08	0.1
C-index	0.57/0.16	0.57/0.14	0.58/0.14	0.57/0.14	0.62/0.17	0.59/0.14	0.62/0.14

Binwidth	0.2	0.3	0.4	0.5	0.8	1.0
C-index	0.57/0.1	0.57/0.14	0.62/0.16	0.62/0.14	0.69/0.18*	0.67/0.14

Table A. 72 – Binwidth values employed for features extraction with relative model's performance in terms (C-index) for LET$_d$ maps. Results are shown in terms of median/IQR. The best value is highlighted with the symbol *.

Binwidth	0.1	0.2	0.3	0.4	0.5	0.8	1.0
C-index	0.62/0.17	0.67/0.14	0.64/0.13	0.64/14	0.67/0.14	0.64/0.17	0.62/0.17

Binwidth	2.0	3.0	4.0	5.0	8.0	10.0	16.0
C-index	0.71/0.14*	0.62/0.21	0.62/0.1	0.63/0.22	0.62/0.14	0.62/0.1	0.62/0.14

Table A. 73 – Bin width values employed for features extraction with relative model's performance (C-index) for RBE-weighted dose maps. Results are shown in terms of median/IQR. The best value is highlighted with the symbol *.

A.3.2 Approach 2

In Table A. 74, Table A. 75, Table A. 76, Table A. 77 and Table A. 78 are reported both the values employed in features selection for the parameter bincount/binwidth and the corresponding model's performance in terms of C-index (median/interquartile range). Tables refer to physical dose, LETd and RBE-weighted dose maps, respectively.

Bincount	2.0	3.0	4.0	5.0	8.0	10.0	16.0	32.0	64.0	128.0	256.0	512.0
C-index	0.67/0.19	0.63/0.14	0.62/0.19	0.62/0.19	0.62/0.14	0.52/0.14	0.61/0.19	0.68/0.14*	0.55/0.14	0.52/0.17	0.67/0.17	0.63/0.24

Table A. 74 - Bincount values employed for features extraction with relative model's performance (C-index) for physical dose maps. Results are shown in terms of median/IQR. The best value is highlighted with the symbol *.

Binwidth	0.1	0.2	0.3	0.4	0.5	0.8	1.0
C-index	0.67/0.14	0.67/0.16	0.67/0.14	0.62/0.14	0.66/.16	0.67/0.19	0.64/0.18

Binwidth	2.0	3.0	4.0	5.0	8.0	10.0	16.0
C-index	0.66/0.17	0.62/0.14	0.6/0.19	0.61/0.14	069/0.17*	0.67/0.22	0.67/0.19

Table A. 75 - Binwidth values employed for features extraction with relative model's performance (C-index) for physical dose maps. Results are shown in terms of median/IQR. The best value is highlighted with the symbol *.

Bincount	2.0	3.0	4.0	5.0	8.0	10.0	16.0	32.0	64.0	128.0	256.0	512.0
C-index	0.68/0.18*	0.62/0.14	0.61/0.16	0.64/0.19	0.64/0.17	0.64/0.19	0.62/0.19	0.64/0.19	0.57/0.19	0.64/0.19	0.57/0.22	0.64/0.19

Table A. 76 - Bincount values employed for features extraction with relative model's performance (C-index) for LET$_d$ maps. Results are shown in terms of median/IQR. The best value is highlighted with the symbol *.

Binwidth	0.01	0.02	0.03	0.04	0.05	0.08	0.1
C-index	0.6/0.19	0.62/0.19	0.67/0.22*	0.62/0.21	0.64/0.19	0.57/0.21	0.62/0.18

Binwidth	0.2	0.3	0.4	0.5	0.8	1.0
C-index	0.6/0.19	0.6/0.22	0.6/0.17	0.6/0.21	0.62/0.19	0.67/0.22*

Table A. 77– Binwidth values employed for features extraction with relative model's performance in terms (C-index) for LET$_d$ maps. Results are shown in terms of median/IQR. The best value is highlighted with the symbol *.

Binwidth	0.1	0.2	0.3	0.4	0.5	0.8	1.0
C-index	0.64/0.19	0.67/0.18	0.67/0.14	0.68/19	0.71/0.14*	0.62/0.17	0.62/0.21

Binwidth	2.0	3.0	4.0	5.0	8.0	10.0	16.0
C-index	0.63/0.13	0.62/0.14	0.63/0.15	0.65/0.21	0.57/0.18	0.67/0.19	0.64/0.14

Table A. 78 – Bin width values employed for features extraction with relative model's performance (C-index) for RBE-weighted dose maps. Results are shown in terms of median/IQR. The best value is highlighted with the symbol *.

List of Figures

List of Tables

Acronyms

SBC — Skull-base chordoma

CT — Computed Tomography

MRI — Magnetic Resonance Imaging

EEA — Endoscopic Endonasal Approach

IMRT — Intensity Modulated Radiotherapy

PET — Positron Emission Tomography

GTV — Gross Tumour Volume

CTV — Clinical Tumour Volume

ITV — Internal Target Volume

PTV — Planning Target Volume

TPS — Treatment Planning System

IGRT — Image Guided Radiotherapy

CBCT — Cone-Beam CT

SOBP — Spread-Out Bragg Peak

RBE — Relative Biological Effectiveness

LET — Linear Energy Transfer

LET_t — track-averaging LET

LET_d — dose-averaging LET

CNAO — National Center of Oncological Hadrontherapy

VOI — Volume Of Interest

CV — Cross-Validation

LC — Local Control

TR — Tumour Progression/Local Recurrence

GLCM — Gray Level Co-occurrence Matrix

GLRLM — Gray Level Run Length Matrix

GLSZM — Gray Level Size Zone Matrix

GLDM   Gray Level Dependence Matrix

NGTDM Neighbourhood Gray Tone Difference Matrix

ICC   Intra-class Correlation Coefficient

C-index Harrell Concordance index

Acknowledgments

I would like to thank my advisor Chiara Paganelli and co-advisors Letizia and Giovanni, who helped me throughout the development of this project.

I don't know if I can find the right words to thank my parents, that supported me during these five years letting me to choose my way. As far as possible, I hope that the achievement of this goal can also be a reward for them, for the many sacrifices they have done. I also appreciate all the support I received from the rest of my family, who followed me in silence in all my way.

Thanks to all my friends for being always present during all the phases of my studies; a super thank you to "the friends of always", Alessia and Beatrice, for sharing the years of a beautiful and important path encouraging me and supporting me in every steps. Thanks to Giulia, my companion of tears and despair during exams sessions, but also a companion of adventures and laughter.

A special thank you to my boyfriend Alessandro, thank you for listening to my outbursts, for having scaled down my fears and worries in the darkest moments, but also thank you for all the carefree moments and for trying to convey to me that lightness and positivity which characterizes you during these years.

Thanks.
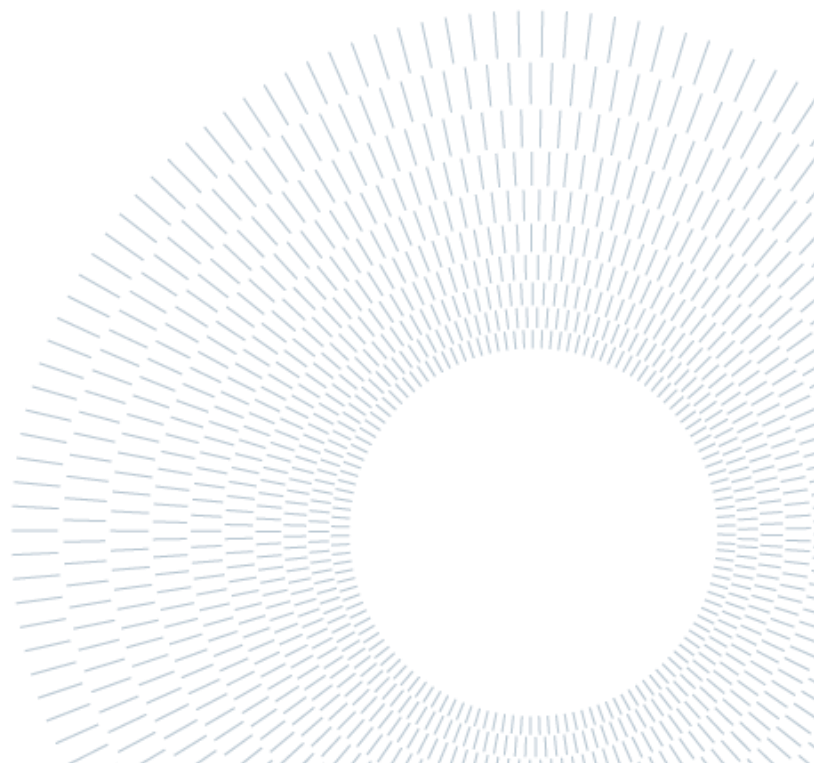
Ringraziamenti

A mia nonna,
che più di chiunque altro avrebbe voluto esserci.
Ai miei genitori,
per aver reso possibile il raggiungimento di questo traguardo.
Ad Alessandro,
compagno di momenti bui, ma soprattutto di momenti spensierati.
Ad Alessia e Beatrice,
le amiche di sempre.
A me stessa,
per aver creduto di non essere mai abbastanza, per non aver mai mollato.