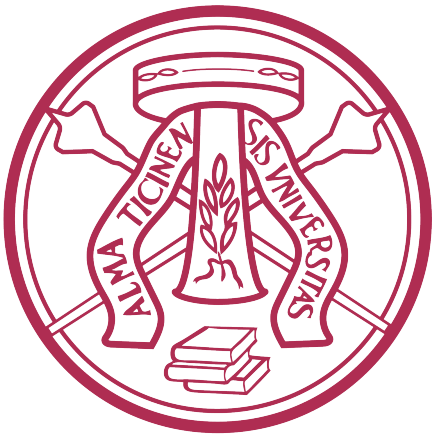


UNIVERSITÀ
DI PAVIA

Department of Physics

Master's Degree Course in Physical Sciences

Deep neural networks for the automatic segmentation of CT images of Head and Neck tumors: a preliminary study to enhance BNCT Treatment Planning System

Deep neural networks per la segmentazione automatica di CT di tumori testa-collo: studio preliminare per la personalizzazione del Piano di Trattamento per la BNCT

Master Thesis of:
Francesco Morosato

Supervisor:
Dott. Ian Postuma
Co-supervisor:
Dott.ssa Setareh Fatemi

A.A. 2021-2022

Abstract

Treatment Planning Systems (TPS) in Boron Neutron Capture Therapy (BNCT) are evolving more and more toward a precise and individualized treatment, and a more efficient calculation of in-patient dosimetry. However, organs and tumors manual segmentation is a very slow process and it introduces human-induced variability in the model reconstruction. In this thesis a preliminary study is carried out, aimed at using Deep Learning methods for Head and Neck (H&N) tumors automatic segmentation on CT images. This imaging modality and type of tumor where chosen because CT scans are the gold standard for treatment planning (TP) in radiotherapy and H&N tumor is the target most treated with BNCT nowadays. An automatic segmentation method would be helpful for radiologists in reducing their workload and the time they need to devote to manually segment images. On the other hand, especially in BNCT, it could also be helpful to researchers. In fact, it would allow comparing different treatment plans performed on different images of the same patient and made by different research groups. The compared dosimetry would be independent from the segmentation on which the specific TP is based, thus allowing to improve or test new BNCT TPS or new imaging systems quality. Moreover, fine tuning TPS parameters to achieve better performances needs a great amount of data for testing, hence the presence of a large database of images easily and quickly segmentable with an automatic tool would make the task much easier.

The first step was the creation of a standardized database composed only of those CT images useful for the study. Then the images were cropped so only the H&N area would remain. This reduces the computational weight and retains only the useful image information. Subsequently, the database was used to trained the neural network (NN). The neural network used was the nnU-Net, a deep learning-based segmentation method that automatically configures itself to the specific database. This DL method is the state-of-the-art in medical automatic segmentation. After the nnU-Net training on our database, its performance was tested and it provided excellent results. Therefore, a method for automatically segmenting ROIs in H&N cancer CT images was achieved, and it will be used as reference standard for the following studies. Moreover, a preliminary test was made on the dosimetry calculated in one patient using the true volume and the segmented one. Preliminary results are encouraging and future work will be devoted to deepen this aspect, using the DL tool developed in this thesis as an input.

Abstract

I Treatment Planning Systems (TPS) in Boron Neutron Capture Therapy (BNCT) si stanno evolvendo sempre più verso un trattamento preciso e individualizzato, e un calcolo più efficiente della dosimetria nel paziente. Tuttavia, la segmentazione manuale di organi e tumori è un processo molto lento e introduce una variabilità indotta dall'uomo nella ricostruzione del modello. In questa Tesi è realizzato uno studio preliminare, che mira all'uso di metodi di Deep Learning per la segmentazione automatica di tumore testa e collo (H&N) su immagini CT. La tipologia di immmagine diagnostica e di tumore sono scelti rispettivamnete perchè la CT è il golden standard per i piani di trattamento (TP) in radioterapia e i tumori H&N oggigiorno sono i più trattati con la BNCT. Un metodo di segmentazione automatica sarebbe utile ai radiologi per ridurre il loro carico di lavoro e il tempo che devono dedicare alla segmentazione manuale delle immagini. D'altra parte, soprattutto nella BNCT, potrebbe essere utile anche ai ricercatori. Infatti, permetterebbe di confrontare diversi piani di trattamento eseguiti su immagini diverse dello stesso paziente e realizzati da diversi gruppi di ricerca. La dosimetria confrontata sarebbe indipendente dalla segmentazione su cui si basa lo specifico TP, consentendo così di migliorare o testare la qualità di nuovi TPS per la BNCT o di nuovi sistemi di imaging. Inoltre, la messa a punto dei parametri del TPS per ottenere prestazioni migliori richiede una grande quantità di dati per i test, quindi la presenza di un ampio database di immagini facilmente e rapidamente segmentabili con uno strumento automatico renderebbe il compito molto più semplice. Il primo passo è stato la creazione di un database standardizzato composto solo dalle immagini CT utili per lo studio. Quindi le immagini sono state ritagliate in modo da mantenere solo l'area della testa e del collo. In questo modo si riduce il peso computazionale e si conservano solo le informazioni utili dell'immagine. Successivamente, il database è stato utilizzato per addestrare la rete neurale (NN). La rete neurale utilizzata è la nnU-Net, un metodo di segmentazione basato sul Deep Learning che si configura automaticamente in base al database specifico. La nnU-Net rappresenta lo stato dell'arte della segmentazione automatica in campo medico. Dopo l'addestramento della nnU-Net sul nostro database, le sue prestazioni sono state testate e hanno fornito risultati eccellenti. Pertanto, è stato ottenuto un metodo per la segmentazione automatica delle ROI nelle immagini CT del tumore H&N, che sarà utilizzato come standard di riferimento per gli studi successivi. Inoltre, è stato effettuato un test preliminare sulla dosimetria calcolata in un paziente utilizzando il volume vero e quello segmentato. I risultati preliminari sono incoraggianti e il lavoro futuro sarà dedicato ad approfondire questo aspetto, utilizzando come base lo strumento basato sul Deep Learning sviluppato in questa tesi.

Contents

Chapter 1

Introduction

1.1 Boron Neutron Capture Therapy (BNCT)

Boron Neutron Capture Therapy (BNCT) is a binary form of experimental hadron-therapy able to selectively destroy malignant cells, while sparing the normal tissue[1]. BNCT is based on the neutron capture that occurs when ^{10}B is irradiated with thermal neutrons ($E_n < 0.5eV$). An excited ^{11}B nucleus is created as a result of the neutron capture, and it decays almost instantly. The products of this reaction are an α particle and a recoiling ^{7}Li nucleus (**figure 1.1**), both have a high *Linear Energy Transfer* (LET). LET [KeVμm^{-1}] measures the amount of average energy deposited per unit length by the charged particle along their track. α particle's LET is approximately 150 KeVμm^{-1} and ^{7}Li nucleus's LET is approximately 175 KeVμm^{-1} [2]. This means that both particles are densely ionizing and able to cause non-reparable damage to cancer cells, i.e., complex or clustered damages to their DNA.

^{10}B is a stable non toxic element and has a high cross section for neutron capture (3837 barns at 0.025eV), which makes possible the use of this isotope for Neutron Capture Therapy (NCT). An important advantage of the use of ^{10}B is the small range in biological tissue of ^{7}Li nucleus and α particle, respectively of 4-5 μm and 9-10 μm, that are comparable with the diameter of a single cell. Thus, the damage caused by the reaction products is concentrated in the cells containing

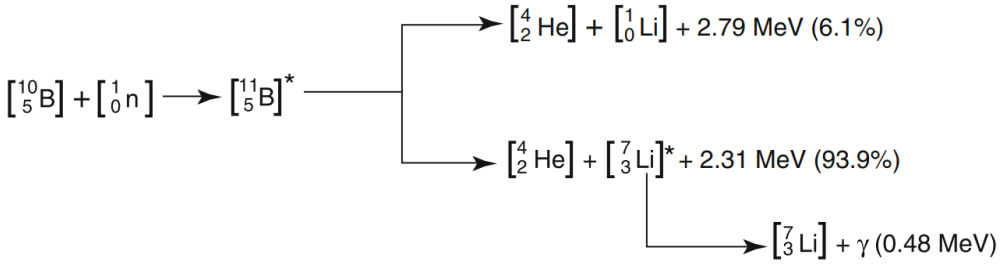


Figure 1.1: The reaction ^{10}B(n$_{th}$, α)^{7}Li is showed in figured, Inspired from [2].

^{10}B. This makes BNCT a selective therapy at the cellular level [1].

BNCT is performed in two steps. First a borated drug is administered to the patient through injection in blood, and after a time interval required for the drug to accumulate in the tumor, the second step can be carried out. That is, the patient is irradiated with a low-energy neutron beam, with spectrum optimized according to the depth of the tumor. Epithermal neutrons are used to treat deep tumors, because they lose energy by interacting with the superficial tissues, mainly via elastic scattering in hydrogen. In this way thermal neutrons can reach and interact with the boron in the malignant cells deep-seated in the body. Superficial tumors can be irradiated directly with a thermal neutron beam. The borated drug is designed to be incorporated by tumour cells, in concentrations 3-4 times higher than in healthy tissue, by exploiting various metabolic or immunological mechanisms [3]. The selectivity of BNCT is thus ensured by the preferential accumulation of ^{10}B in the tumor cells, triggering a localized dose deposition in the malignancy. In this way, only the cancer cells suffer lethal damage, while the healthy cells that have accumulated a lower concentration of ^{10}B can be spared. If a uniform thermal neutron field is obtained in the volume of the tumour-affected organ, the selectivity of the borated drug is exploited to the fullest, ensuring a tumour-selective dose deposition at a cellular level (**figure 1.2**).
Irradiating a patient with a neutron beam means that neutrons might also be captured by the hydrogen and the nitrogen atoms present in tissue, but the cross section of these interactions involved are at least 3 orders of magnitude lower. Therefore, it is possible to destroy tumor cells disseminated in the normal tissue, safeguarding the latter, if sufficient amounts of ^{10}B are delivered to the target volume. Thus, the most important requirements are: the presence of a sufficient concentration of 10^9 atoms/cell in the tumor, a sufficient tumor-to-normal tissue boron concentration ratio and an adequate thermal neutron flux obtained in the tumour volume [4].

Since BNCT effectiveness is based on the selective ^{10}B accumulation in tumor cells with a higher ratio than in normal cells, this therapy is a biologically rather than physically targeted type of radiation treatment, differently than photon radiotherapy or hadrontherapy. In photon radiotherapy, selectivity is ensured by the collimation of the X-ray beam and a treatment plan that allows the dose to be concentrated in the tumour with different irradiation directions. In hadrontherapy, selectivity is ensured by the specificity of the Bragg peak in depositing the dose at the tumour location and by the possibility of obtaining spread out Bragg peaks and moving the beams with magnetic fields [6]. Thus, radiotherapy or hadrontherapy cannot be used when the tumor location is difficult to point out with high precision or when there is a metastatic spread. On the contrary, BNCT could be a potential option in these cases because its sparing effect is guaranteed by the differential boron concentration obtained [4]. Therefore, BNCT might

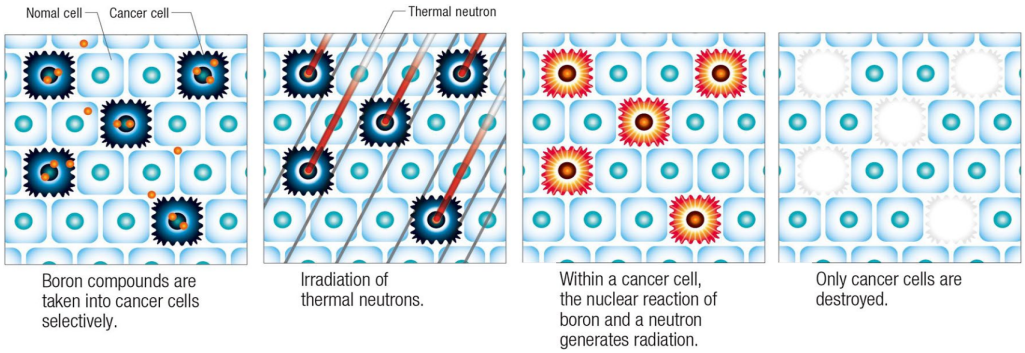


Figure 1.2: The concept of selectivity of BNCT. 1) A boron-containing drug is administered to the patient and it selectively accumulates in cancer cells. 2) The target is irradiated with a low-energy neutron beam, and the neutron captures occur in boron. 3) The thermal neutron capture by ^{10}B releases an α particle and a ^{7}Li nucleus in the cancer cell. 4) Tumor cells absorb a lethal dose while the healthy cells are spared. Figure taken from [5].

be a more effective treatment option for some tumor types, for example those that have spread or are infiltrating. Indeed, these kinds of tumor are frequently beyond the possibility of surgical removal and/or other forms of irradiation due to the target poor localization, proximity to a radiosensitive organ, or difficulty differentiating it from the surrounding normal tissue.
Another important BNCT advantage compared to low-LET radiotherapy and chemotherapy is its effectiveness independent from the presence of oxygen in the tumor cells or from the tumor cells proliferative cycle state [7], since α particles and ^{7}Li nucleus are both high LET particles. This means that the cell damage is in large percentage due to direct DNA damage and not to the formation of free radical as occurs in photon therapy.

The two tumor types more studied as a BNCT target are Glioblastoma Multiforme (GBM) and the Head and Neck (H&N) tumors, both in pre-clinical and in clinical research.

Glioblastoma Multiforme (GBM) is a neuroepithelial tumour originating from glial cells of the Central Neuron System (CNS). According to the WHO (World Health Organisation) classification, GBM or grade IV glioma is the most common malignant brain tumour among adults and is one of the most aggressive tumours. Although GBM rarely metastasize, it has a very infiltration-oriented development pattern [8]. This means that it needs to be handled as a whole-brain condition, instead of as a malignancy to be removed. Since the 1960s researchers are studying its treatment with BNCT, since high-grade glioma patients cannot be completely cured by chemotherapy or radiation, because these treatments are unable to effectively hit the micro-invasive tumor cells inside the healthy brain [4, 9, 10].

Another class of tumors which has been a target of BNCT is advanced or recurrent

Head and Neck cancers, such as Squamous Cell Carcinoma (HNSCC). This type of tumour poses difficult challenges, especially when the tumour is locally advanced or recurred. More details are explained in the section 1.4.

As mentioned above, a requirement for an effective BNCT treatment planning is the ability to correctly estimate the concentration of ^{10}B in the tumor and healthy tissues, to calculate a realistic dose absorbed by the target and the Organs at Risk (OARs). The calculation of the dose takes into account that neutrons interact not only with boron, but also with other elements in biological tissues, such as ^{1}H, ^{14}N, ^{12}C, ^{16}O. Of these, the ones that undergo neutron capture are ^{1}H ($\sigma \sim 0,332$ b) and ^{14}N ($\sigma \sim 1,8$ b). Even if the neutron capture cross sections are much smaller than the one of ^{10}B, the concentration of ^{1}H and ^{14}N atoms in the body is large compared to boron, hence the number of reactions is not negligible. Moreover, the epithermal neutrons get thermalized through elastic scattering with hydrogen nuclei in the tissue before reaching the tumor. The recoil hydrogen nucleus is a proton which delivers dose to the tissue, constituting a source of non-selective dose. Therefore, the neutron beam spectrum should not have very energetic components, and in fact the issue of neutron spectrum optimization is a very important part of the work when designing a clinical facility [11]. Finally, a neutron beam is always contaminated by a certain photon component that cannot be completely shielded, and that is responsible of unwanted dose deposition in the patient. Summarizing, the different contributions to the unspecific dose are:

- The structural gamma component present in the neutron beam.
- The 478 KeV photons of ^{10}B capture reactions.
- Scattering reactions ^{1}H(n, n')^{1}H', producing p^{+} with average energy equal to half the energy of the incoming neutron.
- ^{1}H(n, γ)^{2}H reaction releases a 2,2 MeV photon, with low LET but high penetration, so it can deposit its energy in distal organs.
- ^{14}N(n, p)^{14}C reaction releases a p^{+} with initial energy $E \approx 0.585$ MeV with high LET, and localized energy deposit.

Boron absorbed by normal tissue is another source of dose delivery that affects the normal cells and that must be taken into account.

The knowledge of ^{10}B concentration in the irradiated tissue, the neutron flux distribution and its spectrum in the volume of interest makes it possible to calculate the absorbed dose. Treatment planning simulates the patient irradiation using Monte Carlo methods for radiation transport in matter. The parameters of the irradiation are optimized in such a way as to maximise the dose in the tumour and minimise the dose absorbed by the healthy tissues. This is achieved by varying the position of the patient with respect to the beam port and by establishing how many different directions should be set-up. The irradiation time

is established prescribing the maximum dose tolerable by the most radiosensitive healthy tissue involved in the irradiation. The distribution of the absorbed dose enables the assessment of the effectiveness of treatment, for example through radiobiological models that link the dose to the clinical effect expected in patients [11, 12]. In order to apply these models it is necessary to translate BNCT dose into photon-equivalent units. In fact, the different BNCT radiation components cause different biological effects in cells [13]. The equivalence between BNCT dose and photon-equivalent units is obtained through experimental radiobiological data and appropriate models. These models are able to consider the different biological efficacy of all the radiation components of BNCT treatment. The traditional model was proposed by Coderre & Morris in 1999, and it is still used in clinical treatment [14]. The clinical dose expressed in photon-equivalent units is obtained by the moltiplication of the dose with Relative Biological Effectiveness (RBE) fixed factors or with Compound Biological Effectiveness (CBE) fixed factors for the dose from boron. In 2012 González and Santa Cruz developed a more refined model, the photon isoeffective dose[15], and further improved it in 2017 [16], that has shown greater predictive power on clinical data [17, 18] and on the outcome of the treatment.

As anticipated above, the two pillars of BNCT are the availability of a boron drug able to selectively enrich the tumor and the availability of a neutron beam with proper energy spectrum and sufficient intensity. Since BNCT effectiveness and selectivity is due to the distribution of ^{10}B, the role of the boron delivery agents is fundamental. Boron is electron-deficient and this makes boron atoms easily incorporated into chemical compounds, thus is possible to synthesize various boron carriers. [2, 8].
The ideal drug should respond to three requirements:

- Low systemic toxicity and low uptake in normal tissue, but high uptake in the tumor. In a pragmatic way the concentrations ratio of ^{10}B in tumor/normal tissue and tumor/blood should be $T > 3$.
- At least $\sim 20\ \mu g/g$ or $\sim 10^9$ atoms/cell of ^{10}B in the tumor.
- Rapid clearance from blood and normal tissue, but long retention of the drug in the tumor during BNCT.

At the moment, a drug that satisfies perfectly all criteria does not exists. The most important challenge is to develop a drug able to target only the tumor with minimal normal tissue toxicity and retention [4].
Only two compounds offer a sufficient tumor specificity and are approved for clinical use: the boron cluster sodium borocaptate (BSH) and the amino acid analogue boronophenylalanine (BPA) [2].
BSH has been used in BNCT treatments from more than 50 years. It was proposed by Soloway and Hatanaka in 1967 [19]. This compound was used especially

for cerebral tumors, because its selectivity is due to passive diffusion into the tumor across the damaged blood brain barrier [1, 20]. However, BSH hardly reaches the infiltrating tumor cells in the normal brain tissue, thus its effectiveness may be poor in the long range.
Boronophenylalanine (BPA) is an amino acid first obtained in 1958 by Snyder et al. [21] and thanks to its chemical structure similar to the melanin precursor, it was considered an effective boron carrier for melanoma treatment. Following further improvements, BPA was used to treat other kinds of cancer [22]. This compound carries only one boron atom, but it is actively transported inside of cells, more effectively if the tumor expresses the amino acid transporter LAT-1. In this case, although the number of boron atoms is lower, the effectiveness of BNCT may be higher due to the increased probability that the neutron capture products cross the DNA of the cells [1, 23, 24].

The other key factor for a successful BNCT is the neutron beam.

1.2 Neutron Sources

In the previous section it was underlined the key role of an adequate thermal flux in the tumor volume in BNCT treatment. This implies the importance of the neutron beams and of its optimization. If the tumor is superficial a thermal neutron beam can be used, otherwise a deep tumor needs to be treated with an epithermal neutron beam. When the epithermal neutron beam enters the tissue creates a thermal neutron flux deeper in the tissue. This shows the spare-skin effect, the epithermal neutrons get thermalized via elastic collisions with tissue hydrogens, and thermal neutrons arrive at the deep tumor's location. By raising the neutrons' average energy, the attainment of the maximum thermal neutron flux can be moved deeper. While a thermal neutron flux decrease exponentially after its entrance in the tissue (**figure 1.3**). The beam is characterized by its *intensity* and its *quality*. The first one, that is the number of neutrons per unit of time and area, is the main factor that determines the treatment time. The second one describes the kinds, energy, and relative intensities of all the radiations present in the beam [25].

So the ideal neutron source facility should be able to produce different kind of beams with different characteristics based on the treatment's needs, and this feature is also desirable for those facilities intended only for BNCT research.

1.2.1 Reactor-based BNCT facility

There are two options for neutron generation, reactor based and accelerator based facilities. In a reactor the neutrons are generated trough the following reaction:

$$n + {}^{235}_{92}\mathrm{U} \rightarrow {}^{140}_{56}\mathrm{Ba} + {}^{93}_{36}\mathrm{Kr} + 3n + Q \tag{1.1}$$

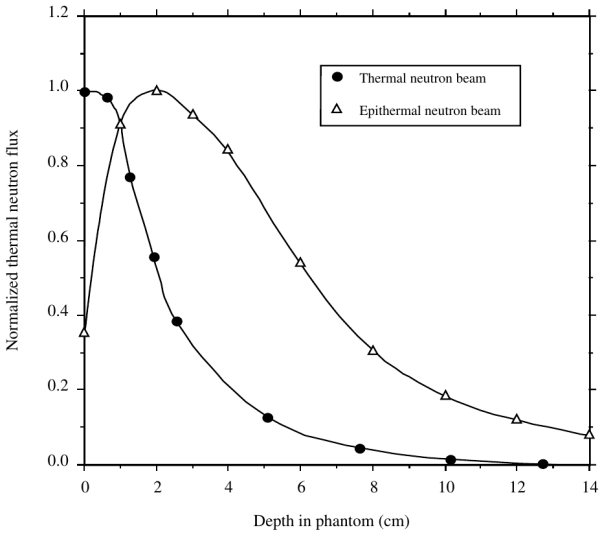


Figure 1.3: Comparison of flux-depth distributions for thermal and epithermal neutrons. Figure taken from [25].

But the spectrum of a reactor is composed mainly of prompt/fast neutrons, with a mean energy of 2 MeV and a most likely energy of 0,7 MeV, they exhibit a continuous energy distribution. So in order to use the neutrous in output from the reactor for BNCT treatment another step is needed. There are two ways to obtain a thermal/epithermal neutron flux from a reactor beam: *Spectrum Shifting* (SS) method and *filtering* method. The SS method is obtained through moderators, in this way the fast neutrons leaving the core are slowed down to the desired energy (thermal or epithermal energy). The second method uses filters, that block all the neutrons of different energy from the desired one. In some cases both methods can be used at the same time. The difference in the neutrons output obtained from the two methods is that the SS method is much more effective than the other one [25].

But there are some intrinsic problems in the use of reactors as BNCT neutron source facilities. Reactors are difficult to build inside of a hospital. There are problems about bureaucracy, safety and maintenance requiring experienced employees. Moreover in many countries as Italy the construction of new reactors can generate discontent in the population, which makes the construction more complicated. Therefore, BNCT researches have often used existing reactors, but they were general purpose built reactors, and not dedicated full time to the therapy. So usually the reactors were separated and far from the hospitals and the clinical treatment could interfere with the other reactor's tasks. Lastly it is not simple to modify a nuclear reactor to meet the demands of a healthcare procedure. Hence, in the last decade the BNCT community has focused on the construction and improvement of neutron source's facilities based on accelerators. This choice has more advantages than the reactors based facilities [25].

1.2.2 Accelerator-based BNCT facility

Accelerators are far smaller and simpler to install and maintain in healthcare settings compared to reactors. First accelerators are significantly more accepted by the general public. Second, there are typically less issues with license, responsibility, and waste disposal. Additionally, it may be turned on and off and it needs a substantially lower capital expense than a reactor [2, 25]. Many accelerator-based facilities for BNCT (AB-BNCT) were build around the world [26]. The availability of accelerators as BNCT neutron sources may result in a greater use of this therapy, possibly even in ordinary hospital procedures.

The are many nuclear reactions that produce neutrons and can be used in accelerators, but the two reactions mainly studied and used are ^{7}Li (p,n) ^{7}Be and ^{9}Be (p,n) ^{9}B. These reactions are endothermic, so the incoming proton needs a minimum threshold energy to make the reactions happen. When the proton has an energy bigger than the threshold but near it, the resulting neutron energy is very low [2].

The ^{7}Li (p,n) ^{7}Be reaction is more used than the other one, because it has better characteristics. Studying the reaction's cross section in function of the proton energy, it can be seen the energy threshold at 1.880 Mev and a big resonance at 2.25 MeV (580 mb). It is preferred to use a proton beam with 2.3 MeV energy, because it is the right compromise between a reaction's high cross section value and low maximum neutron energy (573 KeV). In this way neutrons slower than the ones generated trough reactors are obtained. Moreover in its spectrum there isn't the fast neutrons tail, that is hard to moderate. So without fast neutrons, neutrons can be slowed down by a smaller moderator and used in BNCT [2].

The ^{9}Be (p,n) ^{9}B reaction has the threshold energy at 2.06 MeV. The threshold value is similar to the previous one, but in order to obtain a yield similar to the ^{7}Li (p,n) ^{7}Be reaction at 2.3 Mev a proton beam more energetic is needed. The needed energy is about 4 MeV, this means that the generated neutrons have an higher energy (between 1.1 and 2.1 MeV) than the one generated from the previous reaction. Thus a more powerful and more expensive accelerator is required and a bigger moderator is needed for ^{9}Be (p,n) ^{9}B reaction [2].

In order to produce the wanted nuclear reaction what is needed is a high-current, low-energy accelerator and a suitable target. But the generated high flux of neutrons still needs to be moderated, and this operation is made by the Beam Shaping Assembly (BSA). The BSA is a custom-built structure composed of different materials, and it is designed to moderate, filtrate and collimate the neutron beam toward the patient.

In order to maximize the personalizing of the BNCT treatment, clinicians and medical physicists create a Treatment Planning System (TPS).

1.3 Treatment Planning (TP)

Delivering a therapeutic radiation dose to the target tumors, while reducing the risk of consequences to normal tissue is the aim of all radiation therapies. This is achieved through the *Treatment Planning*. TP is the process by which radiation oncologists, radiation therapists, and medical physicists plan the best radiotherapy treatment approach in order to satisfy the clinical needs.

The steps needed in the creation of a TP are:

- Acquisition of patient data (usually with CT, MRI or CT-PET images).
- Determination of tumour volumes and Organs At Risk (OAR).
- Creating a model of the patient and the radiation's beam(s).
- Calculate the dose absorbed by the tumor and healthy tissues in different beam and patient's position configurations and choose the best ones.
- Determining the correct positioning of the patient during treatment.

The diagnostic images used as foundation of the TP will be discussed in subsection 1.5.1.

1.3.1 Tumor's volumes definition

The tumour and organs contouring is manually made by expert radiologists through 3d tools on the diagnostic images. It is considered the golden standard about reliability and precision. Tumor and organs contouring/segmentation is explained in detail in subsection 1.5.4. The task of the tumor volumes delineation is made by different steps. In each step a volume is defined to include a different aspect that could affect negatively the results of the treatment, if not considered. All the volumes are defined in the ICRU Reports No. 50 [27] and 62 [28].

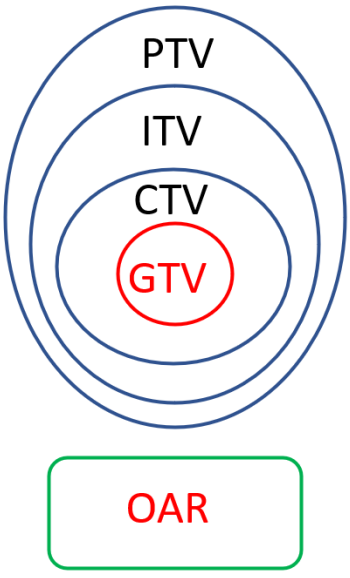


Figure 1.4: The tumor volumes defined by ICRU Reports No. 50 [27] and 62 [28].

Gross Tumor Volume (GTV): "The Gross Tumour Volume (GTV) is the gross palpable or visible/demonstrable extent and location of malignant growth" [27]. The GTV is delineated using data from diagnostic imaging as CT and MRI or from a combination of different modalities as CT-PET. Moreover information could be taken even from other diagnostic techniques such as histology and

pathology tests or other clinical evaluations. Sometime this volume is called primary Gross Tumor Volume (GTVp) [29].
Sometimes even the Lymph Node Gross tumor volume (GTVln) is defined. The GTVln is the volume of all the enlarged lymph nodes and the smaller ones too, if they are associated with high PET signals or any metastasis visible trough CT scans [30, 31].

Clinical Target Volume (CTV): "The clinical target volume (CTV) is the tissue volume that contains a demonstrable GTV and/or sub-clinical microscopic malignant disease, which has to be eliminated. This volume thus has to be treated adequately in order to achieve the aim of therapy, cure or palliation" [27]. This volume contains all the infiltrated tumor cell in the normal tissue, that cannot be seen through the diagnostic imaging, and other area considered at risk. Usually CTV is defined adding a variable or fixed margin to the GTV, this margin could be 1 cm, and in some case even 0 cm, so the CTV correspond to the GTV. All of this is determined by the radiation oncologist with consultation of pathologist and other specialist [29].

Internal Target Volume (ITV): it is defined as the CTV plus an *Internal Margin*. The internal margin is introduced to take in account any CTV's expected physiologic motion in relation to the Internal Reference Point and its corresponding Coordinate System (usually the bony anatomy is used as reference). There are many patient's involuntary movement that could deform and move the CTS's volume and position, as breathing, heartbeat and rectum and bladder filling [28].

Planning Target Volume (PTV): "The planning target volume (PTV) is a geometrical concept, and it is defined to select appropriate beam arrangements, taking into consideration the net effect of all possible geometrical variations, in order to ensure that the prescribed dose is actually absorbed in the CTV" [29]. So the PTV include the Internal Margin and a *Set-up Margin*, named *External Margin* too. The Set-up Margin takes in account uncertainties related to treatment set-up, machine tolerances and patient positioning.

Organ At Risk (OAR): "The OAR are normal tissues whose radiation sensitivity may significantly influence treatment planning and/or prescribed dose" [29]. If the tumor is near an OAR or if the OAR is a lot sensible to radiations, it could be necessary change the beam's entrance settings in the body.

1.3.2 Treatment Planning System (TPS)

When all the patient's body parts are delineated on the diagnostic images, it is possible to model the patient and the radiation's beam(s). The resulting model is used to calculate the absorbed dose by the involved body parts in different disposition. This will allow to find the best configuration of beam(s) and patient position that maximizes the absorbed dose by the tumor and minimizes the one

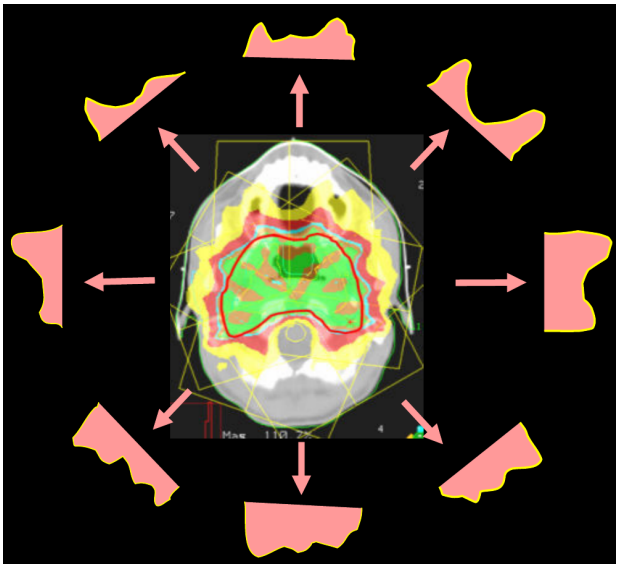


Figure 1.5: Example of reverse planning. It'is possible find the photon fluence profile for each beams based on a specified dose distribution in the tumor. Image taken from [32].

deliver to the healthy tissue. All of this process is made by the *Treatment Planning System* (TPS). TPS is a computer program, it receives in input the OAR's absorbed dose limit, the prescribed absorbed dose to the tumor's volume and other parameters, and then it shows the best beam configurations, energies, field widths, and fluence patterns to provide a dose distribution that is both safe and efficient. If the clinicians is not satisfied about the final treatment plan, they can change the input parameters and redo the process.

What just explained is how a photon radiotherapy TPS works. The TPS finds the best configuration through an *Inverse Planning*. The Inverse Planning is a mathematical problem, which consists in finding a set of optimal parameters (example radiation's fluence profile and beam's direction) from *a priori* specified tumor dose distribution and constraints that are the dose limiting OARs (**figure 1.5**). This is exactly what allowed the creation and advancement of Intensity Modulated RadioTherapy (IMRT).

Indeed, in photon radiotherapy, the software has to take in account only a type of particle (the photons), even if the photon's damage to the tumor tissue is delivered through the electrons. Thus, the algorithms used to solve the inverse problem can exploit experimental algorithms based on dose measures in water or even more advance algorithms. In this way the photon's TPS in order to calculate the dose distribution associated to a particular configuration set-up can use efficient and faster algorithms. On the other hand in BNCT TPs a complex mixed radiation's field is involved. As previously explained in section 1.1 there are many different contributions to the unspecific dose added to the boron dose, and the particles range from high to low LET. Moreover, the boron dose depends from the neutron

capture agent, its distribution in the body and the stochastic nature of neutron scattering. Also it's necessary to consider that different type of radiation at different energies have different biological efficacy and different tissues respond in different ways to different radiations. So each absorbed dose component's distribution varies and is influenced by the tissue composition, neutron, proton and photon fluence spectra, and other factors. In this situation there aren't any experimental algorithms, but the only way is using Monte Carlo simulations, even if it's very computationally expensive[2].

In BNCT the TPS is fundamentally a Monte Carlo software able to simulate the dosimetry of the mixed-field in patients and a radio-biological model to understand the radiations' biological effectiveness on different tissues. The radio-biological model converts the absorbed dose into photon-equivalent units (discussed in section 1.1), making it feasible to compare the results of a traditional treatment planning to the BNCT treatment planning, assuring the compliance of the prescribed tumor volume dose and dose limits of the OARs.

Thus the BNCT TPS consists in a software that first voxelizes the patient's segmented diagnostic images and it creates a model of the patient's geometry in the Monte Carlo software language. Then it sets the beam(s) configuration to get the prescribed absorbed dose in the tumor volume, while respecting the restraints on the OARs. This allows the neutron transport in the Monte Carlo code to simulate the dosimetry in the patient. The Monte Carlo codes mainly used in the BNCT TPS are MCNP [33] ,GEANT4 [34] and PHITS [35] or other written specifically for BNCT TP. The Monte Carlo simulation's results are analysed by the TPS and the latter is able to compute important radiological informations as isodose curves and Dose Volume Histograms (DVH) through the dose distribution. The DVH is a cumulative histogram that shows how much of the total examined volume takes a specific absorbed dose, the latter is plotted on the x-axis. The DVH gives information about the minimum and maximum dose and its uniformity in the volumes, but there is no information about the spatial distribution's absorbed dose in the volume. Usually in the TPS there is a Graphical User Interface (GUI), that allows the user to choose the parameters that describe irradiation condition (number of beams, beam direction, treatment time, define Regions of Interest (ROI).

At the moment there are alredy some BNCT TPS:

- NCTPlan [36, 37]
- JCDS [38]
- SERA [39]
- THORplan [40]
- TsukubaPlan [41]

- NeuMANTA [42]

The TPS used in this thesis is IT_STARTS.

1.3.3 IT_STARTS

Innovative Toolkit to Simulate neuTron cApture theRapy irradiaTion and doSimetry (IT_STARTS) is a BNCT TPS realised by the project IT_STARTS financed by the "Istituto Nazionale di Fisica Nucleare" (INFN). The TPS is born by the combined efforts of the Pavia's BNCT group and the BNCT argentine group. The argentine group is the same group that developed the photon isoeffective dose model [15, 16]. IT_STARTS is coded in Python and it will be shared with all BNCT researchers after further validation.

IT_STARTS receives in input the diagnostic image (in NIFTI format) alongside with the segmentation of the organs and Regions of Interest (ROI). Then it builds a patient's voxelized geometry in the languages of the most common Monte Carlo code (PHITS, GEANT4 e MCNP6). IT_STARTS builds the voxels using a *Multicell* approach implemented in [43]. Where there are big areas of the same material and homogeneous density the voxels will be bigger, on the other hand where there are big variability of material and/or density the voxels will be smaller to represent in the best way the body part. Thus there is a representative patient's model, but at the same time the total number of voxels is reduced. The patient's geometry is used in the Monte Carlo code to simulate different treatment set-ups, such as various beam(s) dispositions, and calculate the absorbed dose in each configuration after the particles' transport.
The organs and tumors' segmentation are fundamental for IT_STARTS, because it can calculate their spatial dose distributions only when the segmentations are provided in input. Otherwise there is no methods for the toolkit to discern the respective volumes. This toolkit in order to calculate organs or tumor's photon equivalent dose it uses radiobiological data. It makes the fit of the dose-effect curve, and the chosen dosimetric model's parameters are extracted through the fit and they are used to calculate the photon equivalent dose. IT_STARTS offers the classic RBE model and the innovative photon isoeffective dose model to calculate the photon equivalent dose. Also there is the possibility to calculate figures of merits about the clinical scenario of the treatment. Tumor Control Probability (TCP) and the Normal Tissue Complications Probability (NTCP) can be calculated through adequate models implemented in IT_STARTS. Lastly this toolkit can produce organs and tumor's isodose curves and DVHs.

This thesis' aim is to automatically segment the tumor's volume, so as a more efficient toolkit can be build by the fusion of IT_STARTS and this thesis results. In this work the tumor selected is the Head and Neck (H&N) cancers because it is one of the most studied cancer as a BNCT target.

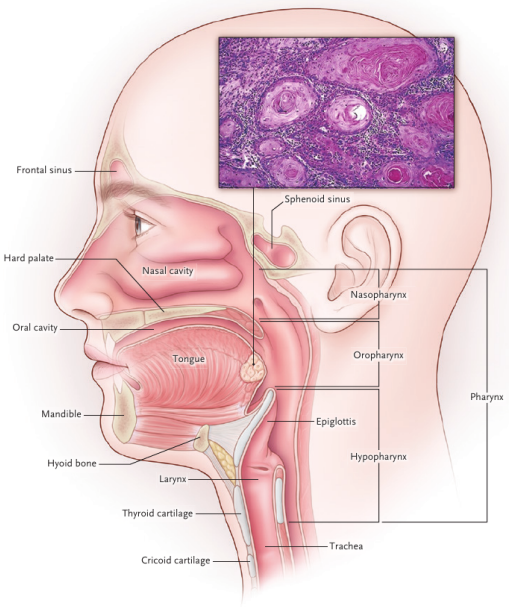


Figure 1.6: Major Anatomical Sites of Head and Neck Squamous-Cell Carcinoma. The inset displays the squamous-cell carcinoma's typical histologic characteristics that are present in head and neck cancer.[49]

1.4 The Head and Neck Cancer

Head and neck (H&N) cancers arise in an body area that starts at the base of the skull and extends to the clavicles. It includes the base of the skull, temporal bone, paranasal sinuses, nasopharynx, oropharynx, larynx, oral cavity, major and minor salivary glands, skin, and the neck [44] as shown in **figure 1.6**. They are a varied group of solid malignant tumors, that usually origin by squamous cells lining these tracts and cavities. In fact this type of cancer is called Head and Neck Squamous Cell Carcinoma (HNSCC) [44–46]. The head and neck region provides a therapeutic challenge since so many of its components are linked to essential basic physiologic systems including eating, breathing, communication, and speech. Tumors in this region may have severe, crippling symptoms depending on their exact location, size, and pattern of dissemination. Structure deformities and functional limitations may significantly impair social integration and quality of life. This happens because HNSCC develops near organs and tissues needed for essential physiologic functions. So the growing tumour as well as the treatment (surgery and radiation) could limit these functions, because with standard treatment procedure is difficult to avoid a negative impact on the sensitive organs in this body segment [44]. Moreover this type of tumor often recurs locally after surgery, radiotherapy or chemo-radiotherapy. Cancers that reappear locally are frequently regarded as inoperable, and re-irradiation is linked to significant toxicity, and the tumor may even show radio-resistance [47, 48].

Squamous Cell Carcinoma (SCC) represents about 95% of cancers that may de-

velop in this area, but even if the area in which the tumor arises is small, there is a big variability in the various types of histopathology, the tumor location, tumor size, and natural evolution of the cancer [44]. This largevariability could be an hindrance in making the study of tumor automatic, ad it is intended to be faced in this thesis.

H&N cancer was the seventh most common cancer in the world in 2018 (890,000 new cases and 450,000 deaths), the fifth among men, and the twelfth among women [50]. Tobacco and alcohol are still two important risk factors for head and neck cancer that influence majorly oral cavity, larynx, oropharynx and hypopharynx. Usually these risk components affect older patients that assumed tobacco and alcohol heavily, in fact thanks to the smoking habits slowly declining in the last decade, the number of new H&N cancers cases are also slowly decreasing [49][51]. More and more new cases of primary H&N tumor are located in the oropharynx, this is caused majorly by human papilloma virus (HPV) type 16. In the 2000s in U.S.A more than 73% of the oro-pharyngeal cancers were HPV-positive. In fact HPV induces oropharyngeal cancer. Young people are most affected by this type of cancer predominantly in North America and northern Europe, showing the occurrence of cancer 10 to 30 years after exposure to unprotected oral sex [44, 49]. This caused an increase of tumor number in tonsils and the base of the tongue [52, 53]. Moreover, other recognized risk factors are the exposure to ionizing radiation, specific product used in heavy industry and poor oral hygiene [54].

The three primary therapeutic modalities for H&N cancer are radiotherapy, surgery, and chemotherapy. The early stage disease (I or II stage) is curable with surgery alone or definitive radiotherapy alone. About 30% to 40% of patients present this characteristics. The treatment modality is chosen based on the accessibility of the tumor site, preservation of the functionality and minimization of the morbidity. Thanks to the treatment, 70% to 90% of early-stage patients have long-term survival rate. Unfortunately, more than of 60% of SCCs are diagnosed when they are a locally advanced (III or IV stage). This phase is distinguished by a large volume with marked local invasion or metastases in the near nodes. In this case, there are large probabilities of local tumor recurrence (15% to 40%) and metastasis in other part of the body. As a result, the 5 year survival is less than 50%. The size and anatomical location of the main malignancy, the disease stage, the patients' age, their preferences, their performance level, and any associated disorders all play a significant role in therapeutic decisions. In these cases both surgery and radiotherapy are applied, while chemotherapy is used as adjunctive or adjuvant treatment [49]. However, when the primary tumor is big, the surgery could drastically decrease the patient's quality of life, precisely because, as mentioned, this anatomical district has important functions, as the physiological one and as the cosmetic one, both very important for daily life. Patient affected by

inoperable tumors are treated with radiotherapy and chemotherapy [47].

Despite the continue progress in early diagnosis and treatment, more than 65% of HNSCCs develop locoregional tumor recurrence or distant metastasis caused by or the spread of the primary tumor before treatment or locoregional treatment resistant cancer cells. Both this cases are extremely difficult clinical scenarios [49, 55]. Recurrent tumors are often not operable, and in cases where the primary tumor has already been treated with radiation therapy, the possibility of new radiotherapy often implies high treatment toxicity, due to the previous dose absorbed by the healthy tissue surrounding the primary lesions. Moreover, recurrent SCCs could be radioresistant, and chemotherapy and immunotherapy are not very effective in the majority of patients [47, 48, 55]. This implies a dismal prognosis.

In this framework, BNCT is proving to be a valid treatment, its intrinsic characteristics respond to the two main problems in this clinical scenario. Thanks to the selectivity of the borated drug, in fact, the dose is mainly deposited in the tumor; in this way the treatment toxicity is lower and the functionality of the normal tissue can be preserved. Moreover, the tumor cells infiltrated in the healthy tissue surrounding the lesion will also absorb high dose, and so the probability of future recurrence decreases.

In 2001, Japanese researchers treated with BNCT a recurrent parotid gland tumor, originated from a primary tumor treated with conventional therapies, at the Kyoto University Research Reactor Institute (KURRI). It was the first treatment of this kind in the world [56]. The result was encouraging, the locoregional tumor control lasted for 7 years, then the patient died out of an inter-current disease. However, this success stimulated new BNCT clinical trials for H&N cancer both in Japan and Finland [57, 58].

The first set of 26 H&N patients receiving BNCT were published in 2004 by a team of Japanese researchers [59]. 19 SCC patients, 4 salivary gland carcinomas patients and 3 sarcomas patients, all with extremely advanced cancers who were not candidates for any other therapy entered the trial selection. 12 patients underwent complete remission, 10 partial regression and in 3 cases the treatment did not have any effect (one case was not evaluated). The average survival time was 13.6 months, but 24% of patients lived for 6 years or more.

Suzuki et al. [60] retrospectively reviewed the medical records of 62 patients treated with BNCT between 2001 and 2007 at Kyoto University. The tumors treated were unresectable or locally recurrent H&N cancers. For the 62 patients treated the overall response rate was 58% at 6 months, median survival was 10.1 months, and 2-year overall survival was 24.2% and the median follow-up was 18.7 months.

The Helsinki University Central Hospital used the neutron facility in Espoo, Finland, to perform one of the largest studies of BNCT for H&N cancers between 2003 and 2012. A total of 117 patients participated in the phase 1 and 2 trial, whose 79 cases were HNSCC cancers not treatable with surgery [48, 57]. Among the 79 patients, 69 treated patients were valuable for treatment response: follow up resulted in 25 total and 22 partial remissions, while in 17 patients the illness stabilized and in 5 the cancer continued to advance. The most common adverse effects were oral mucositis, oral pain and fatigue, all of which were clinically controllable. Three patients had 5.5, 7.8 and 10.3 years of disease-free survival. 10 months was the median lifespan, while 21% of patients survived for two years or more.[48].

1.5 Artificial Intelligence in cancer treatment

1.5.1 Diagnostic imaging

Computed Tomography scan (CT scan)

Computed Tomograpy scan (CT scan) is the natural evolution of an X-ray modality image. The latter shows only a 2d projection (axial image), but CT scan can reproduce and investigate the whole 3D patient's body. This has big repercussions on many medical fields, such as oncology and traumatology. This image modality is especially important for radiotherapy's TP, because it is used for planning the treatment as well as diagnostic and and follow-up purposes. Moreover at the moment it is the modality image most used as foundation of the TP. The CT scan is the golden standard of diagnostic imaging for radiotherapy and it is used to extract the Region Of Interest (ROI) segmentation. But this modality provides only anatomical information of the patient's body. Thus in order to have a metabolic information too, it is necessary to use another image modality: the Positron Emission Tomography scan (PET scan), that will be explained in the next subsection. So CT scan and PET scan are both necessary to have a complete information about cancer.

The process of the CT image's creation consists in the measurement of the transmitted X-ray (Ionizating radiation) trough the patient's body at many different angles (views). The quantity of transmitted X-ray depends on the linear attenuation coefficient ($\mu_{material}$) of the different body parts crossed by the X-ray. The $\mu_{material}$ depends on the photon energy, density and element composition of the crossed tissue. The initial X-ray intensity I_0 is known, the transmitted X-ray intensity I is measured and trough the relation $I = I_o e^{-\int_0^d \mu(x)\,dx}$ it is possible have information about $\int_0^d \mu(x)\,dx$, that is the sum of all μ of the different tissues crossed by the X-ray in the body thickness d. In this way the information in the transmission profiles of all the views are used in the filtered back-projection that reconstruct the CT images. Thanks to the filtered back-projection it is possible

to find the correct $\mu_{material}$ in each voxel of the 3d Images. What create the reconstructed body's image is the values of μ in each pixel. In particular it's a convention to change the $\mu_{material}$ values in the matrix with the Hounsfield Units (HU) defined by $HU_{material} = \frac{\mu_{material} - \mu_{H_2O}}{\mu_{H_2O}}$x1000. Then at every HU values a value in a grey scale is assigned to visualize the image. Usually the monitors that shows the CT scan can print only 256 grey values as grey scale. But the range of HU is much larger, and it's not possible show each HU values with a different grey value. So what is done is the *Windowing* operation, in which the user define the HU range that has to be showed with the gray scale, and the HU values greater o smaller of this range are showed respectively as white and black. Thus only by choosing the most suitable window range the user can see in the best way the tissues of its interest in the CT scan [61].

Positron Emission Tomograpy (PET)

In this image modality carriers labelled with positron-emitting radio-isotopes (example ^{18}F) are infused into the patient's blood. The carrier usually is glucose or other substance needed by the cell metabolism, and exploiting the accelerated tumor metabolism the radiotracers will tend to accumulate in the tumor volume. The most used tracer is ^{18}F-labbeled fluorodeoxyglucose ($_{18}$FDG). While the radionuclides decay, they emit a positron (e^+) that it will be annihilated in a range between a percentage of a millimeter and many millimeters based on its energy from its point of emission. The reaction between the e^+ and the atomic e^- will produce two 511 KeV photons with an angle of $\approx 180°$. What is detected are the 511 KeV photons in coincidence, thanks to the positron camera. Usually the positron camera is a ring composed of many single detectors that allows to measure the coincidence. The coincidence is made by detecting the two photons on the opposed patient's side. Each coincidence defines a lines in the space that links the two detectors on the opposed patient' s side. It's possible to cluster the coincidences in projections and then algorithms, such as the Filtered Back-Projections, are used to reconstruct the image that shows the radionuclide's distribution. As said in the previous subsection the PET scan gives metabolic information and not anatomical one, threfore it is perfect for finding tumor metastasis and so it's used together with CT scan image to have complete information both anatomical and metabolic. IN BNCT treatment the PET scan is used to find the boron distribution in the patient, because the ^{10}B carrier is labelled with ^{18}F [58]. Moreover it was demonstrated that the concentration of ^{10}B in the tumor's volume calculated trough PET was very similar with the experimental measures on surgical specimens. In particular it was analysed and validated the in vivo pharmacokinetics of fluorine-18-boronophenylalanine-fructose (L-18F-BPA) [62–64]. The primary benefit of the PET approach is the ability to objectively assess the boron tumour/healthy ratio, which allows for the individualization of therapy for each patient and a greater therapeutic impact.

However, PET imaging for BNCT has certain limitations because each boron carrier must be linked to the radioactive label, and until now this has only been done effectively for BPA. Moreover a CT scan is required as a reference, because the anatomical detail of PET imaging is extremely low, making it difficult to create the geometry of the tumor for the TPS. Thus to make a BNCT treatment is necessary running two ionizing radiation-based imaging scans. A possible solution could be the use of Magnetic Resonance Image (MRI).

Magnetic Resonance Image (MRI)

Magnetic Resonance Image (MRI) uses a strong external magnetic field $\mathbf{B}_0$ (range between $1.5 - 7$ T) to magnetize the patient's body in the $\mathbf{B}_0$'s direction. This is caused by the hydrogen's proton that precedes around $\mathbf{B}_0$. If a radio-frequency magnetic Field $\mathbf{B}_1$ is used with an oscillation $w_{RF} = w_0$, where w_0 is the angular velocity of the proton precession, it's possible to excite the protons and change their precession's axes, so the magnetization's direction changes. Different magnetic field gradient G_x, G_y, G_z are used to select the body's slice to reconstruct and obtain more data. Thanks to the use of this different magnetic fields in specific sequences, it is possible to receive an electrical signal from the hydrogen's proton of the water in the tissues. Because when $\mathbf{B}_1$ is turned off, the protons turn back to precede around $\mathbf{B}_0$ and doing so the magnetization turn back in its original direction aligned with $\mathbf{B}_0$. The change of direction of the magnetization while it turns back in it's original direction generates a signal in a receiving coil, and the latter can be measured. The signal contains the information about how the protons come back to precede around $\mathbf{B}_0$ after the specific sequence. The hydrogen's proton's response depends on which type of tissue is in the surrounding, in particular it depends on the composition of the tissue. Thus all the signals through the Filtered back-projection or other algorithms are converted in the body's image. An important MRI's characteristic is the better contrast of soft-tissue than CT image, thus MRI allows a better understanding of the brain and abdomen soft-tissue than CT. It's possible to make stand out or dampen a particular tissue through contrast agents, because they change the surrounding of protons, and so they can amplified or dampen the signal [65]. This is very important for BNCT treatment, because researchers are aiming to use ^{19}F labelled boron. ^{19}F is a contrast agent in MRI, and it could be used to study the ^{10}B concentration and distribution. Moreover MRI has good anatomical detail, so it could be directly used for the tumor and organ segmentation, specially for brain tumor as GBM. Although patients have not received BNCT treatment using the BNCT-MRI technology, the approach has been used on small animals with encouraging results [66–68].

1.5.2 The challenges of cancer treatment

Cancer is a complex disease, because it has a self-sustaining and adaptive nature that interacts dynamically with its microenvironment. Given this complexity, uncertainties arise in every step of the cancer treatment:

- Detection of early tumor lesions.
- Accurate distinction of pre-neoplastic and neoplastic lesions.
- During surgery, determining the margins of a tumor that has infiltrated.
- Monitoring the development of the tumor and potential treatment resistance.
- Prediction of tumor aggressiveness, where the metastasis will originate in the body, and an eventual recidivism.

The progress in medical imaging, and minimally invasive bio-markers, give hope in solving this complexity across the pipeline of cancer detection, treatment, and monitoring. Physicians have usually made use only of their personal knowledge and clinical experience while visiting patients and studying their medical history, but recent technological improvements produce a huge amount of data, hard to handle by human brain alone. This means that the decisions in every step in cancer treatment are made without considering all the possible information contained in data, because of the large amount of information or because of the complexity in extracting and correlating meaningful data. Clinical imaging is the most powerful method to evaluate the tumor response to the treatment and to study the characteristics of tumors, and the number of medical images per patient will only increase in time: the data available to physicians to take decisions will increase even more. This fact leads to a paradox: more data will result in a larger workload for the radiologist, who will be forced to spend less time examining the medical images, leading to an increase in errors in detection (failure to detect lesions) or misinterpretation (inability to properly diagnose a tumor). Hence, it is of utmost importance to overcome these hurdles, because the fundamental way to increase the patient survival rate is the early detection and accurate prognosis of cancer. Only this can enable a personalized and optimized treatment strategy for each patient. In fact, the risk of development of the disease, tumor recurrence, or death after treatment are strictly connected to the type of treatment and its effects. Thus, it is important to make the correct prognosis and to predict the correct outcome in response to a treatment, especially when the treatments are individualized or customized to patients [69–71].

At the moment it is difficult to personalize entirely a patient treatment, however digital and computational technique can be used to analyze a vast amount of data from many previous clinical cases to assess cancer prognosis and predict the disease development more accurately. Data can be collected from radiographic

images, genomics, pathology and electronic health records, and then analyzed and used together to obtain useful information about the task. The expertise of clinicians can be integrated with the automated capabilities of Artificial Intelligence (AI), and the provided assistance to clinicians will decrease their workload.

AI is "defined as a system's ability to correctly interpret external data, to learn from such data, and to use those learnings to achieve specific goals and tasks through flexible adaptation" [72]. So the concept of AI arises every time human cognitive functions are emulated by machines. Machine Learning (ML) is a subfield of AI. This area of research is concerned with providing computers the capacity to learn without being explicitly programmed, as the name suggests. In particular the algorithms use their "experience" (the past data used as training) to improve their performance or to make better predictions. This is possible because this study field is founded on statistics and probability theory. When the useful information in the data used as "experience" are extracted by the algorithms and not manually extracted by humans, it begins the Deep Learning (DL), a sub-filed of ML. The ML and DL concepts will be further explained in section 2.2.

For example, it can help with the interpretation of images, disease detection and determination of morphological characteristics. The AI can handle challenging tasks like the accurate volumetric contour of tumor size during its evolution over time, simultaneous finding of scattered lesions, using the tumor's visual characteristics (phenotype) it can infer implications on the tumor's genotype, and tumor development prediction using database comparisons of cases with similar characteristics.

These deep learning methods applied to clinical cases remain in the preclinical research domain at the moment. However, through the relentless improvement of such automatic methods there may be a systematic synergy in future between clinicians and AI to find a new approach in handling cancer [69].

1.5.3 Artificial Intelligence in cancer imaging

In the field of cancer imaging, artificial intelligence (AI) can be very helpful to clinicians. The main aid of AI in cancer imaging focus in tumor detection, characterization and monitoring, moreover it is important for the automatic segmentation of the organs of the body. This is possible because AI can identify complex patterns in images and may find quantitative features representing numerical information from images that cannot be detected by humans. This enables the work of image interpretation to progress from being qualitative and subjective to being measurable and easily repeatable, helping the clinicians in their decision making [69].

Tumor detection means the localization of neoplastic lesions in clinical images,

and the software AI based for detection are collectively known as computer-aided detection system(CADe). CADe systems aim to:

- Increase accuracy in diagnosis, while reducing observational oversights and failure to detect the tumor;
- Assist in early detection of cancer, helping find small lesions or lesions not visible to human eye;
- Reduce the time of the radiologist in exam evaluation.

Some practical example of CADe utilisation are identification of missed cancers in low-dose CT screening, localization of brain metastases in MRIs or micro-calcification clusters in mammography as signal of early breast carcinoma [69, 71].

Tumor characterization consists of segmentation, diagnosis, and staging. It can also go as far as to anticipate outcomes based on particular treatment modalities and prognostication based on a given illness.

1. The tumor segmentation is a fundamental step in the treatment. Segmenting means contouring the tumor to define its margins, in order to gather quantifiable data about its volume, morphology and textural patterns. Such information are meaningful for diagnostics as well as the creation of a radiation treatment planning system (TPS). Usually segmentation is a manual step made by radiologists; however its automatizing would reduce the subjective uncertainties and the time spent by radiologists.
2. Tumor diagnosis based on clinical image consists in identify suspicious lesions as benign or malignant. Normally this is a human task, physicians resolve this problem with their knowledge using subjective and qualitative tumor's features, in this way uncertainties are introduced. Computer-aided diagnosis (CADx) systems can be of help in this decision, because it utilizes quantitative tumor features (volume, morphology and textural patterns) to solve the task, thereby it can support physicians clinical decision-making.[69]
3. Tumor staging is classify the tumor in different categories that represents the expected cancer clinical course and treatment indication. The most widely used cancer staging system is the TNM classification, depending on the size of the tumor (T), whether the disease has migrated to neighboring lymph nodes (N), and whether it has affected other body areas (M, for metastasis), and there are alternative plans utilized for particular organs, such the central nervous system. [69][70].

Tumor monitoring is tracking changes in the tumor over time due to the natural course of the disease or treatment. Normally temporal monitoring of tumor is

done checking predefined metrics, as longest diameter measured, and quantifying tumor burden and evaluating therapy effectiveness through World Health Organization (WHO) criteria. But this procedure is criticized because it oversimplifies the complex tumor geometry and the criteria are not perfectly generalizable to all tumor types [69]. Thus AI can help because it grasps a big number of quantitative and discriminating features in the images over time through segmentation, much more than what humans can do.

1.5.4 Image segmentation

In this work we will concentrate on the automatizing of the tumor segmentation step, so now we will further analyze the segmentation problem, both of organs and tumors.

The purpose of image segmentation is to obtain quantitative information of a tumor lesion (Region Of Interest ROI) or organ about its volume, morphology and textural patterns. Such information are meaningful for diagnostic tasks as well as the creation of a radiation treatment planning system (TPS) [73]. The radiotherapy or hadrontherapy TPS starts with the segmentation (contouring) of the target volume (the tumor) and organs at risk (OAR) from CT, MRI, and PET scans. In this process, much precision of margin delineation is needed, as this will be the basis for radiation beam management to try to reduce the dose to healthy tissues, while maintaining the therapeutic dose to the tumor [74]. According to studies, radiation toxicity and tumor control are strongly connected with contouring accuracy. When there are steep dose gradient in intensity-modulated radiotherapy and proton beam therapy TPS, even minor contouring mistakes might cause tumor volumes to be missed or higher-than-intended doses can be delivered to healthy tissue [75]. In BNCT the ROIs and organs segmentations are used to calculate through simulations the received dose, so the precision is needed to avoid to prescribe more or less than the needed dose.

Generally, ROIs and OARs contouring is done by hand by experienced radiologists and is considered the golden standard, because it's judged as an accurate process, but it is a ***subjective*** task, because it is affected by:

- *Intra-observer variability* which can be seen when the same observer contours the same target multiple times, the discrepancies between the various contours thus obtained represent variability (subjectivity).
- *Inter-observer variability* which is quantified as the discrepancy between contours obtained by different observers on the same target, showing the variability [74].

The variability may be enchanted by the poor quality of the clinical images. In low dose CT the contrast is not so pronounced between soft tissue, which makes the observer prone to errors of intra/inter observer variability in the contour [75][76].

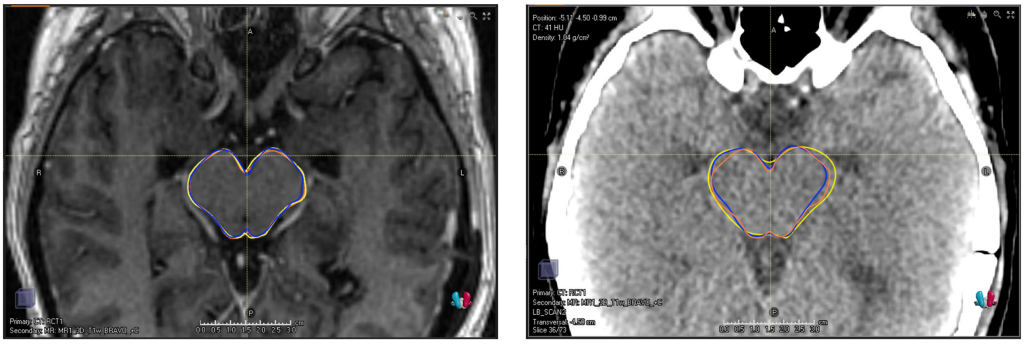


Figure 1.7: In this images it is clear the effect of low soft-tissue contrast on segmentation accurateness. In both images the same three physicians contour the brainstem, but on magnetic-resonance scan (left) the contrast between soft tissue is high and so the contours variability is low, otherwise on the CT scan (right) the decreased soft-tissue contrast increase the inter observe variability. Image taken from [75]

A example of inter-observer variability enchanted by the low quality of the clinical image is presented in **figure 1.7**.

In addition, manual contouring is a very ***time-consuming*** and ***tedious*** process for radiologists. This task can require several hours to contour manually complex cancer, and the time spent by physicians to manually segment, or to peer review, is taken away from the direct patient care. Moreover since the time spent on a single segmentation is a lot, the radiologist can become overburdened with this task alone. This can delay the start of the tumor treatment, leading to possibly less tumor control and survival probability [75].

An automatic segmentation system is needed to resolve, or at least alleviate, these problems. It should be fast and accurate, but it must contour in a comparable or in better way than the manual contouring done by clinicians and in less time, otherwise it will not be considered a useful tool. Another helpful and important system's feature may be the possibility of editing the automatic segmentation by a clinician, in this way the time management and quality assurance of segmentation would certainly be respected [75]. Because computer-aided detection systems are designed to support radiologists' diagnostic work rather than to replace it. In fact, the radiologist has the final word in the diagnosis process and can override descriptions produced by automated methods [73].

The segmentation problem is formed by two consecutive task both in manual and automatic delineation, but with different name: respectively object recognition/detection and object delineation/semantic segmentation. Object recognition/detection aim to identify the content or find the target location in the image, instead object delineation/semantic segmentation aims to delineate the boundaries of image components and then assign the resulting image segments to a category. As a result, each pixel in the image is labeled in categories that

indicate respective organs or tumors[73, 77].

In short, image segmentation is the process of dividing an image into numerous related pieces, or segments, using various similarity criteria, such as gray-scale, color, spatial texture, and geometric features. The objective of image segmentation is to identify and define the regions that represent relevant sections of the body for simpler analysis since segmenting medical pictures is considered to be a semantic segmentation problem. [78, 79].

Chapter 2

Deep Learning

Section 1.5.4 shows how manual segmentation is time consuming and has the intra- and inter- variability issues for every imaging modality. Deep learning can be a helping hand with the automatic segmentation by greatly reducing the time needed for the segmentation [80] and eliminating the human induced variability. The optimal automatic segmentatio tool should be robust, fast and accurate. The tool must lead to a comparable or better TP with respect to the one made only by manual segmentation. The aim is to aid the work of expert radiologist by standardising and reducing the time spent contouring. This will help to lessen the radiologists' workload allowing them to focus more on the single patient scan and at the same time increasing the total analysed scans number. Thus an automatic segmentation algorithm is useful for every image modality, but especially for multimodal MRI images, that are more challenging to read. In the case of BNCT, also preclinical research can earn a powerful tool, because this would allow comparing different treatment plans performed on different images of the same patient and made by different research groups. In fact, the compared dosimetry will be independent from the segmentation on which the specific TP is based thus allowing to improve or test new BNCT TPS or new imaging systems quality.

The contouring of organs and tumors is an *image semantic segmentation* problem in the Computer Vision field. The Computer Vision area aims at constructing a physical model of the world through the information obtained by images and videos. The constructed model is needed by AI system to take the adequate actions [81]. The term segmentation means the separation into parts (segments) of the image based on their features. Semantic means the classification of each image pixel in different classes or categories (organs or tumors). This is obtained by associating to each pixel a label that indicates the specific pixel category.

This thesis work aims at solving or lessening the mentioned segmentation problems, contributing to the construction of a valuable tool for BNCT researcher. The work is framed in a more comprehensive project named AI_MIGHT, funded

by National Institute of Nuclear Physics (INFN) in the scheme Young Researchers Grant in 2021.

2.1 AI_MIGHT

Artificial Intelligence methods applied to Medical ImaGes to enHance and personalize BNCT Treatment planning (AI_MIGHT) project intends to apply Deep Learning to create a segmentation method for a given tumor. AI_MIGHT first goal is to create a standardized dateset of public H&N and Glioblastoma CTs and MRIs. Then, the project will generate an automatic segmentation method for the tumors segmentation and finally it will obtain the automatic creation of the patient geometry based on the segmented image for the TPS input. Therefore, AI_MIGHT would be the perfect integration of the previuos IT_STARTS project discussed in subsection 1.3.3 and dedicated to the implementation of an innovative BNCT TPS. The availability of this tool in the TPS would speed up the preliminary evaluation performed by medical physicist, because they will be able to start to work on the TP of new patients without receiving all the clinicians completed segmentation. In this way the analysis of patient positioning, the irradiation time and the beam(s) configuration can be preliminarly set-up, while the clinicians produce the finished clinical information enabling the final tretament planning for the patient. Moreover, AI_MIGHT would be useful for BNCT researchers looking to improve their TPS and needing a great amount of segmented data to test their computational instruments. With this dual objective in mind the project has focused on the acquisition of a dataset for the two mentioned tumor types and both imaging modalities (CT and MRI).

In this thesis the focus will be on CT images and Head & Neck tumors and the first application of Deep Learning algorithms to segment those images.

Since deep Learning (DL) is used extensively in the AI_MIGHT project and in this thesis, in the next section the basis of DL will be described and the specific methods used in this thesis will be explained.

2.2 Introduction to Deep Learning

The goal of Artificial Intelligence (AI) is to create machines with the ability of resolve problems by autonomous decision making in conditions not seen before. The way of achieving this goal is emulating human cognitive function. Every time this approach is pursued, the field of investigation is AI. [72]. This ambitious goal cannot be fully achieved without a key requirement, that is the ability of learning from experience. The Machine Learning (ML) field is a sub-field of AI that wants to give the machines the ability to learn. In 1950 Alan Turing, the father of AI, argued that creating a "child" machine with the ability of learning would be simpler than creating an intelligent "adult" machine able of doing anything.

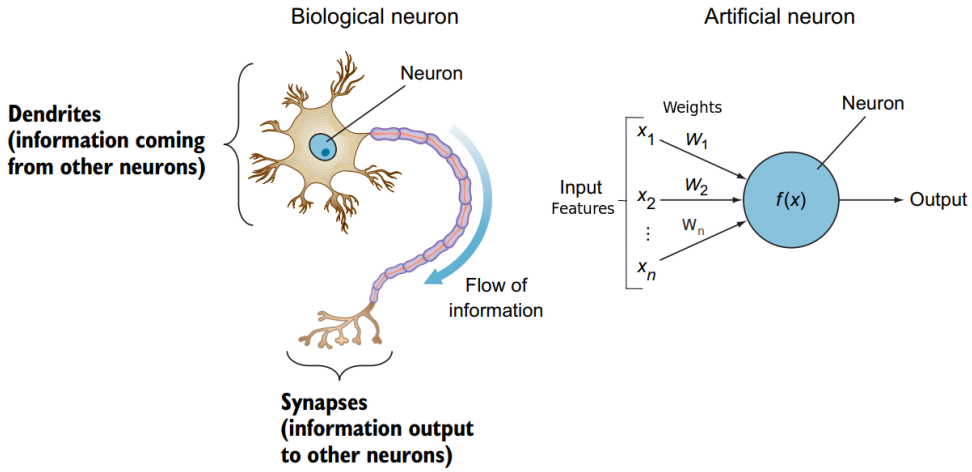


Figure 2.1: The resemblances between biological neuron and artificial one are blatant. The dendrites are emulated by the input connections, which collect input features instead of electrical signals. The function of the biological neuron central body is represented by the activation function and the activation state in the artificial neuron. Lastly the synapse is emulated by the output connection. Image taken from [81].

The development of this field started in 1956 by the efforts of Alan Touring and other scientists such as Arthur Samuel. The latter pioneer introduced the term Machine Learning in 1959, and wrote one of the first programs with the ability of learning. At that time, the fundamental idea was that intelligence arises from *connectionism*. Connectionists thought that natural intelligence was made by the interaction of a multitude of connected simple elements, such as the neurons in a brain. The idea of reproducing the human neural system generated the *Perceptron*, the model of a single neuron, and the first example of *Artificial Neural Network*, modeled as a collection of connected neurons.

2.2.1 The Perceptron

The Perceptron was designed to resemble a biological neuron. **Figure 2.1** shows the three main components of a natural neuron: the dendrites, that collects the transmitted information through the electrical signals, a central body that modulates the collected signals in different ways and transmits the signal through the third part, the synapse if the sum of all signals is higher than a certain threshold. The artificial neuron follows the same pattern, but it is adapted to computing process. The input information is not represented by electrical signals, but by *features* x_i. "A feature in machine learning is an individual measurable property or characteristic of an observed phenomenon" [81]. Thus in the imaging study a feature could be a color, a specific shape or a specific pixels pattern, moreover there are much more complex features. For example the presence of eyes, nose and mouth are features. These information are transported to the main body through the connections, that mimics the dendritis. The biological neuron

main body modulates the input signals, and to represent this process in the artificial neuron the connection are weighted. In particular a *connection weight* w_i is assigned to each connection, and this shows that not all the input features are equally important or useful, but the higher the weight, the more useful the information is in the decision making process and vice-versa. This represents the strength of the connection. The artificial neuron is modeled to have an *activation state* depending on the input features and their modulation. The intensity of the modulated inputs decides the value of the activation state. In the Perceptron this process is made by a simple calculus. *An activation function* is applied to the sum of the input features weighted by the the respective connection weight. In this model the activation function is a step function and defines the excitation state of the neuron. If the excitation passes a threshold, then the neuron is in activated state 1, otherwise the state remains 0. The activation function is what produces the decision making of the model. What is just explained is expressed in the following formula:

$$\text{Activation State} = \begin{cases} 1 & \text{if } \sum_{i=0}^{n-1} w_i x_i \geq \tau \\ 0 & \text{otherwise} \end{cases} \tag{2.1}$$

where x_i is the input feature, w_i is the connection weight of the corresponding connection and τ is the activation threshold. At the end of the calculus the neutron transmits its activation state through a connection that corresponds to the synapse. In this model the final output is the neuron activation state, that represents the solution to the proposed task described by the initial features. For example, the model predicts whether an objects with specific input features belongs to a class. The 1 state means yes and 0 means no.

The Perceptron model learns through *Supervised Learning*. This approach mimics the way children learn from their parents. The Perceptron first tries to guess the correct answer and then the correct answer is provided to it. The model changes the connection weights by comparing its predicted answer to the correct answer in order to give more importance to those input features that are useful to find the right solution. The process of learning is thus represented by the changing of the connection weights.

Unfortunately this algorithm has some drawbacks that do not allow a wide application. The step function used as the activation function, cannot be derived because it has a discontinuity: this implies the impossibility to use more advanced learning algorithms that use gradient-based optimization. Moreover, the model with only one neuron is able to discern only *linearly separable* classes, those classes perfectly separable by a straight line.

The first problem can be resolved with a small adjustment. The step function can be changed with other activation functions with better mathematical behavior.

The most common activation functions a(z) are:

$$
\begin{aligned}
a(z) &= \frac{1}{1+e^{-z}} && \text{(sigmoid)} \\
a(z) &= tan^{-1}(z) && \text{(arctangent)} \\
a(z) &= \frac{e^{z}-e^{-z}}{e^{z}+e^{-z}} && \text{(hyperbolic tangent)} \\
a(z) &= max(z,0) && \text{(ReLU)} \\
a(z) &= max(0.01 \cdot z, z) && \text{(Leaky ReLU)}
\end{aligned}
$$

where z is the total input to the neuron.
The result of these activation functions is a real number, that is the activation state value, and not anymore only 0 or 1. An important fact is that the activation function introduces the *non linearity*, because the activation state value is not proportional to the total input z. When only one neuron is used, this effect cannot be seen, but if many neurons are collected, this will produce a quality change. The sigmoid has range [0,1] and could be interpreted as probability, so it is used for binary classification problems, while the others for regression problems. The negative side of the sigmoid, arctangent and hyperbolic tangent is that they compress the output signals. Even if different large z values are used as input, the mentioned above activation functions will produce a very similar a(z) values; the same occurs for different very small z values. Moreover, if z is very large or small the gradient or derivative of a(z) will be close to zero slowing down the gradient-based learning algorithms. The RelU solves this problem with its linear behaviour, but it required the introduction of the Leaky RelU to mitigate the problem of the null gradient of the RelU when z<0. Despite the Leaky RelU generally works better than the RelU, the latter is wider used. The RelU and Leaky RelU are considered as the state of art of activation functions.

In the end the threshold τ is replaced by a *bias* b ($b = -\tau$), thus the mathematical model of the neuron is:

$$\text{Activation State} = a(z) = a(b + \sum_{i=0}^{n-1} x_i w_i) \tag{2.2}$$

The second problem is solved if a collection of connected neurons is used, the so called *Artificial Neural Network* (ANN) or simply *Neural Network* (NN), instead of a single neuron. This is possible because the interaction of different neurons with their non-linear activation state allows the description of much more complex patterns than the linear one. This starts the Deep Learning sub-field, characterized by ANNs. In particular, what will be explained in the next section is the MultiLayer Perceptron, and the fundamental concepts of DL that are the basis used in every NN.

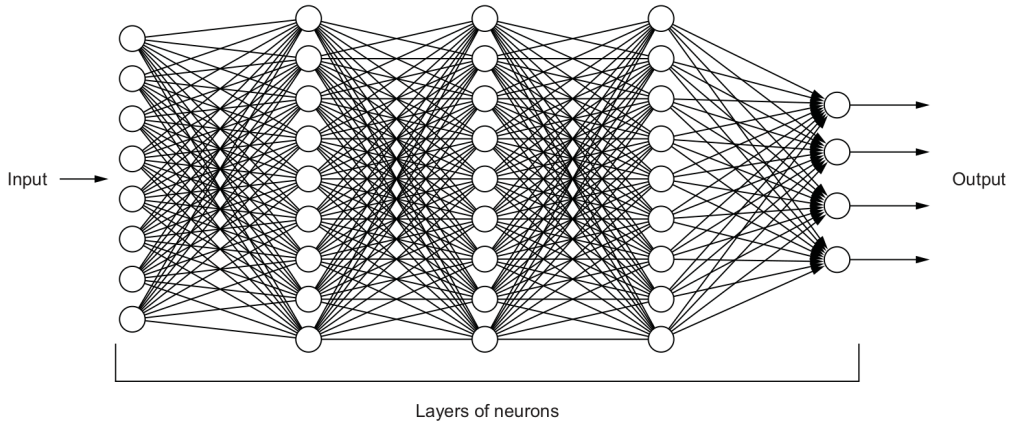


Figure 2.2: Example of MultiLayer Perceptron. Image taken from [81].

2.2.2 MultiLayer Perceptron

The MultiLayer Perceptron (MLP) is a *feedforword* ANN, thus the information goes in only one direction from the input to the output (directed) without loops (acyclic). There are only three types of neurons in the feedforward network: input, output and hidden ones.

- **Input Neurons** have no incoming connections and their activation state value is the respective input feature to the ANN. They are called *sensory neurons*, because their only task is to receive the input data.
- **Output Neurons** have no outgoing connections and their activation state is the output result calculated by the ANN.
- **Hidden Neurons** have incoming and outgoing connections. Their inputs are the activation states of the neurons in the previous layer, and their output is transmitted to the following layer of neurons. They are called "hidden", because the human user cannot "see" or control the input or output of these neurons.

In the MLP neurons are organized in layers and the neurons in a layer are ordered as a linear sequence. The first layer is formed only by input neurons, and the last is composed only by output neurons. All the layers between the first and last are formed uniquely by hidden neurons. MLP is a fully-connected or dense ANN, which corresponds to the fact that all the neurons inside a layer are connected to all the neurons in the previous layer. Another MLP characteristic is that only immediately subsequent layers can be connected (**figure 2.2**). The number of the connection groups between the layers defines the depth of the network. The neural network depth can be easily calculated as the number of the hidden layers plus 1. This is an index of the NN complexity. The activation function usually used in the input and hidden neurons is the ReLu, while the one used in the output neurons depends on the assigned task. Sigmoid or its multidimensional

version, the Softmax, are used for classification problems, while the arctangent or the hyperbolic tangent are used for regression problems.

The NNs driving force is the presence of the hidden neurons. The activation states of the hidden neurons are the new developed features for the next layer and so on. What happens is that the first hidden layer finds patterns in the input data (low level features), the second hidden layer finds more complex pattern in the first discovered patterns and so on. Thus, each hidden layer builds new more complex features (high level features) that will be analysed by the next hidden layer, till the completion of the task. So the deeper the NN is, the more information about the data it is able to extract. These features are not totally understood by humans and this the reason behind the name "hidden". This concerns all the NN and thus also the ones explained in the following sections. In particular this concept is showed by how the activation state of the i-th neuron of the $l > 0$ layer is calculated:

$$\text{Total Input} \qquad z_i^{(l)} = b_i^{(l)} + \sum_{r=0}^{n_l - 1} W_{ir}^{(l)} x_r^{l-1} \tag{2.3}$$

$$\text{Actication State} \qquad x_i^{(l)} = a(z_i^{(l)})$$

where $z_i^{(l)}$ is the total input to the neuron, $b_i^{(l)}$ is the bias and $W_{ir}^{(l)}$ is the weight of the connection with the r-th neurons of the layer $l - 1$.

This shows the first main difference between ML and DL. The ML algorithms need input features that are already the best ones and filled with useful information, so they need to be created or selected by humans. Instead DL algorithms, such as NN, do not require perfect features, because they build themselves the most useful ones (**figure 2.3**). The only requirements for DL methods are a large quantity of representative data about the task at hand and a significant computational capacity[81].

The learning process

In the previous section the Supervised Learning was introduced. The MLP and the NNs that will be discussed in the next section learn how to do a task trough this approach. It was already explained that the NN confronts its produced prediction and the correct answer to learn. How this is done is described below.

One of the fundamental cores of NN learning is the *Loss Function* $J(W, \mathbf{b})$, called also *Cost Function* or *Error Function*, which measures how big is the error in the NN prediction. So it measures how far is the prediction from the expected output. The smaller the Loss function value the better the NN does its task. Which Loss function to use is determined by the problem type and other aspects. The selection of Loss function is a critical aspect of the NN performance, because different functions will give different errors for the same prediction, and thus there is a significant impact on the NN performance.

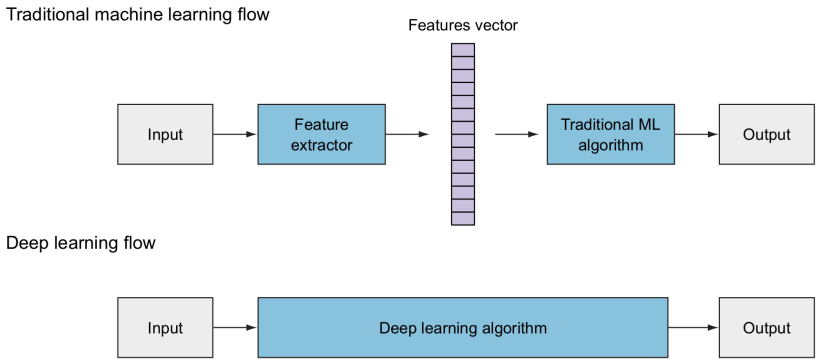


Figure 2.3: The traditional Machine Learning algorithms need already meaningful features as input. So it is necessary the work of another algorithm (feature extractor), that extracts "manually" the meaningful information (the feature vector) from the data. Instead the Deep Learning methods extract by them self the important features through the hidden layers. Image taken from [81]

The introduction of the Loss Function is an important turning point, because it allows to make the NN learning in an *Optimization Problem*, based on the minimization of the Loss function. The optimization is the change of the problem parameters to minimize (or maximize) an interested value.
Optimizing the Cost function in neural networks leads to modifying/updating the connections biases and weights, until the NN determines the ideal weight values to achieve the lowest error. Thus the NNs learn how to do a task by varying the connections weights. Following, two Loss Functions are explained as an example.

Cross Entropy quantifies the difference or divergence between two probability distributions. It is used in the classification problems, where the probability distributions are about classes or categories. In the case relevant for this thesis, the class distributions are the prediction of the NN (the automatic segmentation result) and the true distribution of the data (the ground-truth segmentation). The problem is yet more complex than this: the CT image is composed by N total voxels and for each i-th voxel the NN must calculate the predicted classes distribution. The number of classes C is the number of different structures, such as organs and tumor, that the NN needs to segment. Given the j-th training sample (a CT image) the Cross Entropy is calculated as:

$$L_{CE_j} = -\frac{1}{N} \sum_{c=0}^{C-1} \sum_{i=0}^{N-1} \overline{g}_{ic} \; log(\; \hat{s}_{ic} \;) \tag{2.4}$$

Where $\overline{g}_{ic}$ indicates the correct class distribution of the voxel i-th: if the class c is the correct one then $\overline{g}_{ic} = 1$ otherwise is 0. $\hat{s}_{ic}$ represents the corresponding predicted segmentation probability of the class c-th [82]. The Cross Entropy is zero if the two distributions are identical and its value increases at the increasing of differences. This quantity can be used in the Loss function definition for the

optimization problem, so the quantity to minimize is the *Cross Entropy Loss Function*:

$$J(W, \mathbf{b})_{CE} = -\frac{1}{m} \sum_{j=0}^{m-1} L_{CE_j} \tag{2.5}$$

where the sum runs on the j-th image and m is the number of the total samples.

The other Cost function is based on the *Dice Coefficient* (DC): a metric that quantifies the similarity between two volumes G and S. It is defined as:

$$DC = \frac{2 \cdot (G \cap S)}{G + S} \tag{2.6}$$

If there is perfect accordance then $DC = 1$, $DC = 0$ when there are no intersection between the volumes. Therefore the DC range is [0,1].
It is possible to build a Loss Function that directly maximizes this coefficient. The *Dice Loss Function* for the j-th CT image composed by N voxels and C different classes to segment is:

$$L_{DC_j} = 1 - \frac{\sum_{c=0}^{C-1} \sum_{i=0}^{N-1} g_{ic}\, s_{ic}}{\sum_{c=0}^{C-1} \sum_{i=0}^{N-1} g_{ic} + \sum_{c=0}^{C-1} \sum_{i=0}^{N-1} s_{ic}} \tag{2.7}$$

Where g_{ic} is the ground-truth segmentation and s_{ic} is the predicted segmentation of the i-th voxel for the c class. The Dice Loss Function of all the data is obtained through a mean of all L_{DC_j} such as was done for the Cross Entropy Loss function. There is a variant of this Dice Loss Function that is the *Square-Dice Loss Function*. The only difference is that both g_{ic} and S_{ic} at the denominator are squared [82]:

$$L_{DCsq_j} = 1 - \frac{\sum_{c=0}^{C-1} \sum_{i=0}^{N-1} g_{ic}\, s_{ic}}{\sum_{c=0}^{C-1} \sum_{i=0}^{N-1} g_{ic}^2 + \sum_{c=0}^{C-1} \sum_{i=0}^{N-1} s_{ic}^2} \tag{2.8}$$

The NN Cost function is a complicated function with local and global minima, and plateau (example in **figure 2.4**). Only some ML algorithms have a convex Cost Function with only one minimum. Hence an adequate optimization method is needed to obtain the global minima in the fastest and more precise way. The first learning gradient-based algorithm introduced is *Gradient descent*.

Gradient descent fundamental idea is that the gradient shows the steepest way to the function maximum through its direction in every point. Exploiting this fact, the steepest descent to the minimum is the opposing direction pointed out by the gradient. By taking the gradient of the loss function with respect to the connection weights, it is possible to update the connection weights making the NN learn and improving its performance.

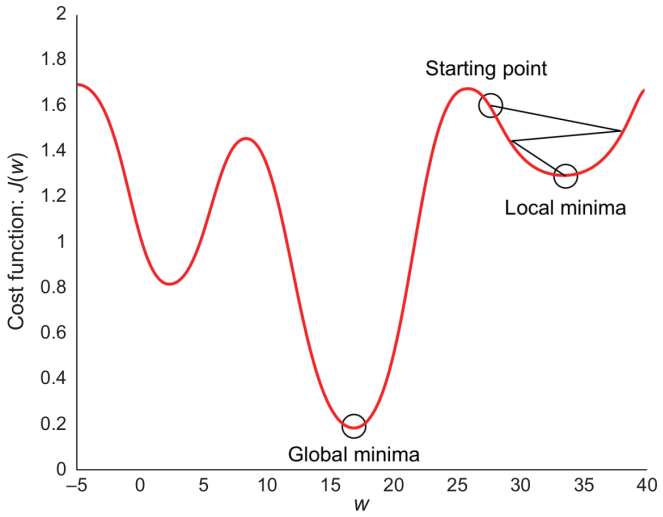


Figure 2.4: The NN Loss Function is a complex function with many minima, and there could be even plateau. This is an important problem, because if the learning algorithm is too simple, it could find a local minima as solution (like in this figure), or it could keep jumping between the walls of the minima without ever reach it. Image taken from [81].

Given some weights w_{old}, gradient computes the updated weights w_{new} as follow:

$$w_{new} = w_{old} - \eta \frac{\partial J(W, \mathbf{b})}{\partial w}$$

where η is the *Learning Rate*, that is the step size chosen as a prior.

This basic algorithm has some drawbacks. First, in order to update the weights one time it is necessary to calculate the Loss Function for all the samples, and this process could be done thousands of time. It is computational slow. Secondly it does not have a "memory" of the previous updating. For example, if η is too large, the gradient could jump around a minimum without ever entering it, or, if the Loss Function has a plateau, the gradient will be very small in this points, hence the updating towards the ideal weights will be very slow. In these cases the solution would be to adjust the length of the step based on the situations, instead of using a fixed value. If the gradient direction remains the same in many iterations would be better to take larger steps, on the other hand, if the gradient direction keeps changing, it would be better to take smaller steps.

The first problem is solved calculating the Loss Function only on a mini-batch of randomly extracted samples, instead of using all the data. So the weights are updated after the NN analyses a number of samples equal to the size of the mini-batch. The size of the mini-batch needs to be decided by the user, based on its computational capacity. If the mini-batch is 1 sample randomly taken, the the algorithm is called *Stochastic Gradient Descent* (SDG). The second problem is solved introducing the *Momentum* v, that works as a memory of the past steps. Instead of changing the parameters using only the gradient calculated in the last

iteration, it is a better idea using a parameter that depends on the gradients of all the past iterations, that is Momentum.

$$v_{new} = \alpha \cdot v_{old} + \frac{\partial J(W, \mathbf{b})}{\partial w}$$

Where v_{old} is the value of the momentum in the previous iteration and v_{new} is the new value in the present iteration. Thus if the sign of gradient keeps changing during the iterations, because the algorithm keeps jumping between the wall of a minimum, the value of the momentum tends to be zero. So if the momentum decides the length of the step for the updating of weights, the step will be smaller and smaller allowing the algorithm to descend the minimum. If the algorithm is on a plateau of the loss function, the calculated gradient will be almost zero. But in each iteration the gradient sign will be the same, and so the gradient will keep increasing the momentum. Thus if the momentum decides the length of the algorithm step, the length will keep increasing till the end of the plateau and the exit of the plateau will be reached in a faster way.

Thus adding Momentum to gradient descent means to apply the following algorithm:

$$\begin{aligned} v_{new} &= \alpha \cdot v_{old} + \frac{\partial J(W, \mathbf{b})}{\partial w} \\ w_{new} &= w_{old} - \eta \cdot v_{new} \end{aligned}$$

Where the coefficient $\alpha \in [0, 1)$ determines how much momentum is conserved between steps. If both of these adjustments are applied to basic learning algorithm, a more robust algorithm is obtained: *Mini-batch Gradient Descent with Momentum.*

Train, Test, Validation sets

As it was said before, an important requirement in the use of DL methods is the availability of a large number of data. These are needed for the *Training Process* in which the NN learns how to do a task through the gradient-based algorithm just explained. But not all the available data can be used for the training, because a part of it is necessary for the NN performance evaluation. It is important that the data used for testing the NN results are not seen/analysed by NN during the training, hence they should be completely independent from the model. Otherwise the test evaluation could be biased, because the NN could learn random pattern in the training data that are not necessarily present in other data. Moreover, there are some fixed parameters that are chosen *a priori* by the user and even those need to be tested in some way. Usually, they are about the architecture design choices or setting of the learning algorithm. They are called *hyperparameters* to distinguish them from the parameters (weights) learned by the NN. Usually multiple hyperparameters values are tested and then the best performing combination is chosen. So a measure of their performance is necessary,

but it is not possible using the data for the test and not even the training data because we could obtain biased evaluation even on the hyperparameters.

Therefore, when a DL method is used the data needs to be randomly splitted in three sets:

- **Training Set** is used during the model training. These data must be representative in the best way as possible of the real and complete phenomenon described by the dataset. The model performances are largely influenced by this aspect.
- **Validation Set** is used to evaluate the hyperparameters. After the model is trained with the chosen hyperparameters, it is evaluated on the validation set. This process is iterated for each combination and the best one is chosen.
- **Test Set** is composed of samples never used for training or validation. This set must be used only one time for the final evaluation of the trained model with the best hyperparameters combination.

If the data is scarce, it is possible to do a *k-fold validation*, instead of creating a validation set. The training set is divided randomly in k-subset. In this way k hyperparameters combinations can be valuated, because k models can be trained on different k-1 folds, and each model uses a different combination of hyperparameters. Then the trained models can be valuated on the remaining fold used as validation set, that will be different for each model. Then the best combination of hyperparameters can be chosen confronting their results. Hence the final training could be done using all the k-folds and the chosen combination of hyperparameters.

A useful definition is the number of *epochs*. This number shows how many times all the training set was analysed by the model during training.

MLP is a powerful DL methods, but it is not the best option when working with images. There are mainly two reasons: the problems are due its architecture and its connections. As said before, the MLP neurons are organized in a linear sequence inside the layer, but the images are at least a 2D matrix, so the images need to be flattened in a 1D vector to be analysed by the MLP. In this way, all the spatial information in the pixel pattern of the image are lost. Secondly, the MLP is a fully connected NN, hence there must be an input neuron for each input feature. This means that the number of connection weights are closely related with dimensions of the input features. But a small image is composed by 200x200 pixels and so 40000 neurons will be needed only for the first layer, and the number of connection weights would only increase exponentially with the increasing of the hidden layers. All of this is highly inefficient and computationally expensive. This leads to the development of the Convolutional Neural Networks.

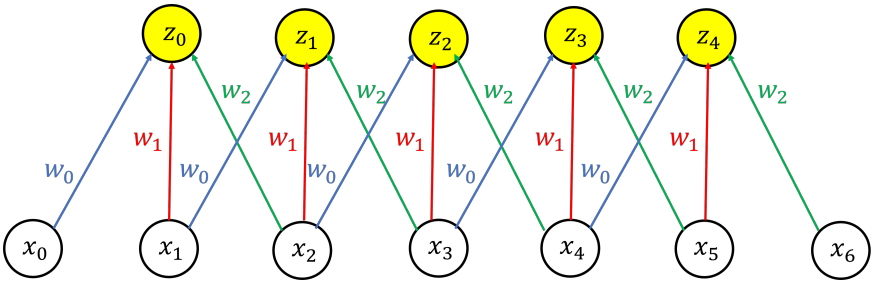


Figure 2.5: Convolution of a 1D signal **x** with a kernel **w** of size 3.

2.2.3 Convolutional Neural Networks

In order to overcome the MLP limits, it is necessary to look at the problem of image analysis in a new way. By exploiting the spatial information in the image it is possible to detect simple shapes by considering only a small group of neighboring pixels, that occupy only a small part of the image (*Locality Principle*). Therefore, it is better to start analysing only a small part of the image, instead of the whole image at the same time. This allows to reduce the number of input neurons keeping only the useful connections. Another intuition is that the same simple shape can be present in other parts of the image, therefore it would be useful to apply the group of connections used for the individuation of a specific pattern, on every part of the image (*Translation Equivariance*). This can be implemented in the DL methods by replacing the linear operator used for the total input z calculus in the MLP (equation 2.3) with *convolutions*. The resulting NN is a *Convolutional Neural Networks* (CCN).

Let's start looking at the 1D convolution to fully understand the characteristics of this new model, assuming a 1D image, resulting in $x_0, x_1, ..., x_n$ activation states of input or hidden neurons. The total input z_u of the u-th neuron of the following layer is calculated through a 1D convolution, and then the activation state x_u is computed through the activation function $a(z)$:

$$x_u = a(z_u) \tag{2.9}$$

$$z_u = \sum_{r=0}^{k-1} w_r \cdot x_{u+r} \tag{2.10}$$

where w_r is the r-th value of the connection weights vector $\mathbf{w} = (w_0, w_1, ..., w_{k-1})$ called *filter* or *kernel*. The u index tells us whose next layer neuron total input is computed and from which activation state it starts to compute the total input. It is helpful to imagine the filter translate or slide over the image, while it computes the total input in every position. In this way, the activation state of the next layer neurons are obtained. What is described in Eq. 2.10 is also shown in **figure 2.5**, which also displays the differences between MLP and CNN that are listed in the following:

- each total input z depends only on k neighboring input values,

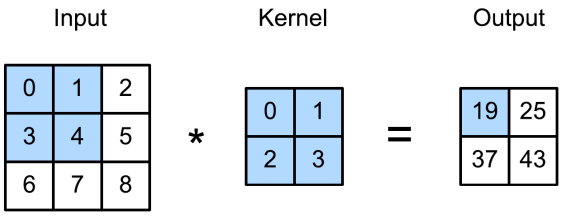


Figure 2.6: Example of 2D convolution. The light blue area on the input matrix shows the pixels used in the calculus of the total input of the light blue output neuron. Each pixel value is multiplied by the respective kernel value and the their sum is the total input. The kernel is moved by one step until all the total input is computed for each output neuron. Image taken from [83].

- the filter uses always the same k weights, thus it produces the same total input z where the same input values are found. This happens in every part of the image.
- the number of weights are independent from the size of the image.

All of these considerations lead to *sparse connection* idea opposing the MLP fully connected concept. What has been explained in this section justifies the use of the convolution in the signals and images studies.

In this first description, a 1D example was used, but the equation 2.10 can be easily adapted to 2D or 3D images, because it is just the 1D convolution application on each dimension.

Studying the 2D problem is helpful to highlight some aspects of the convolution operator that are important.
Let's start from the 2D convolution definition, that is quite similar to the 1D definition, but now the input image is a 2D matrix such as the filter, that is a $k \times k$ kernel:

$$z_{u,v} = \sum_{r=0}^{k-1} \sum_{t=0}^{k-1} w_{r,t} \cdot x_{u+r,v+t} \tag{2.11}$$

The convolution can be obtained as shown in **figure 2.6**. The weights of the kernel are overlaid to the image matrix elements and multiplied by the corresponding weights, the products are summed, and the result is taken as the total input to the first neuron ($z_{0,0}$). Then the kernel is moved by one step (one element in this case) to the right, and the second neuron total input is calculated. The process goes on until the kernel is positioned on each possible image location and all the neurons activation states are computed. This means that the convolution gets as input a 2D matrix and it produces in output another matrix, that is the matrix with the hidden neurons total input.

One drawback of the convolutions as defined in equation 2.11 is that the output image is smaller than the input. The convolution computed with a $k \times k$ filter

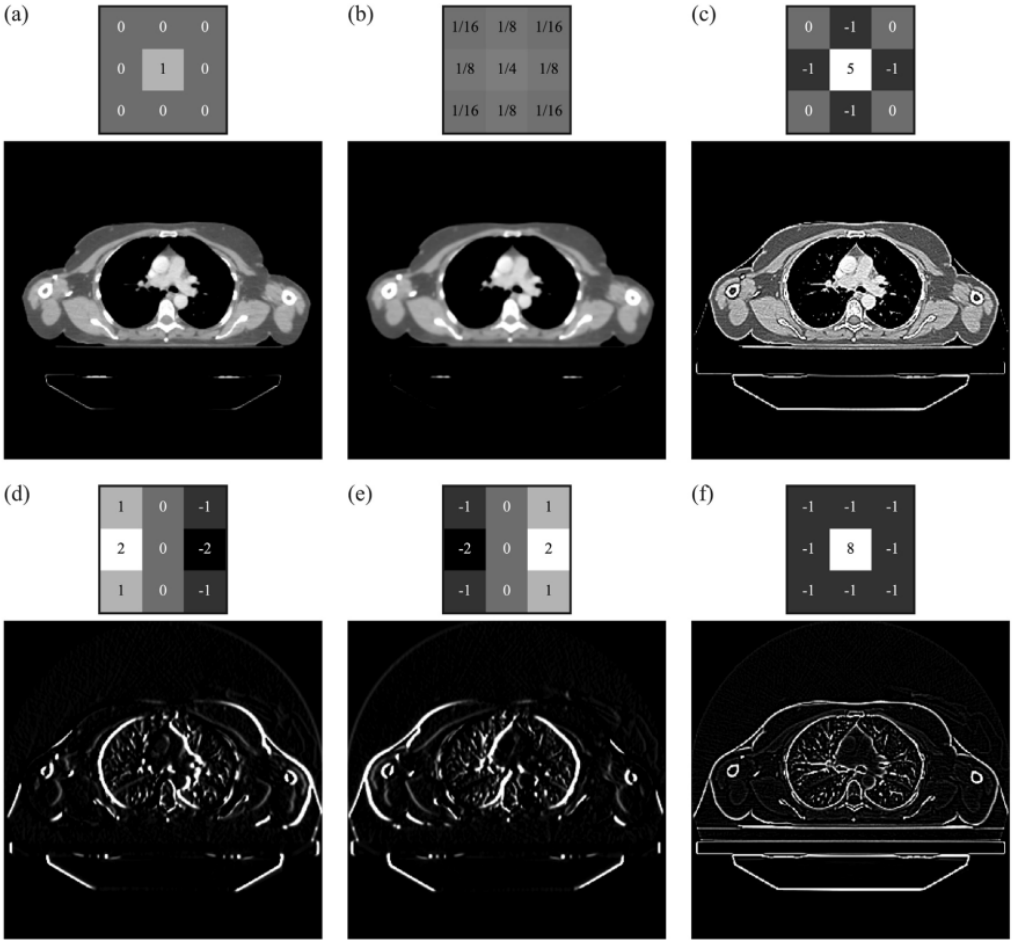


Figure 2.7: Examples of the application of 3×3 filters on a CT scan and their output. The filters are: (a) identity, which returns the original image; (b) Gaussian blur; (c) sharpening; (d) right-to-left gradient; (e) left-to-right gradient; (f) edge finding. Image taken from [75]

reduces the $h \times w$ input image to a $(h - k + 1) \times (w - k + 1)$ output images. So if multiple convolutions would be performed in succession, the image would disappear step by step. To control this phenomenon the *Padding* technique is commonly used. Padding consists in enlarging the input matrix adding a frame of elements, usually as zero. Thanks to this method the output dimensions may be chosen, the output can be the same size of the input image (*same size convolution*) or even enlarged (*full cunvolution*), or shrunk (*valid convolution*).

To recap, the filter is made of weights, that decide how much a pixel is important for the completion of the task. In particular the kernel output is a new image with new pixel values. This convoluted image is called *feature map* or *activation map*. Its name is due to the fact that the feature map indicates where the features, that were found by the kernel, are located on the input image. Each kernel searches a different feature from the other kernel (2.7). Some kernels find specific lines, shapes or even more complex patterns like a face. The filters can enhance or

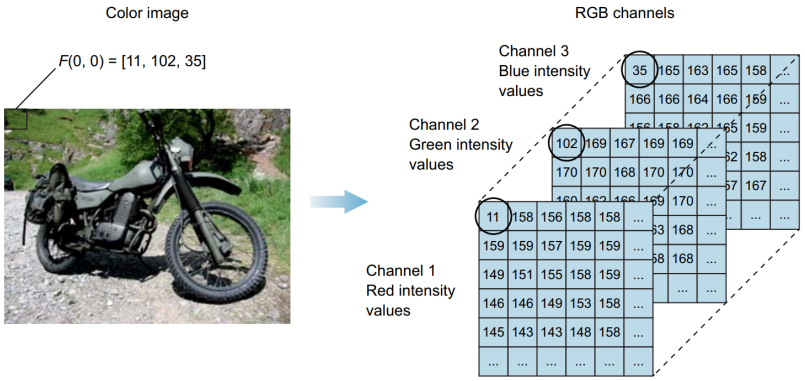


Figure 2.8: Example of a 3-channel image. The RGB image has a channel for each color (red, green and blue). The image can be seen as three 2D staked matrix, and each matrix element value describes the intensity of the respective color. Image taken from [81]

damp some features in the input images. The learning process of the weights of the kernel is the same as described in the previous section 2.2.2, so the NN learns autonomously the weights that make the kernel able to find a specific feature.

Based only on what explained until now, it could seem that for each layer only one kernel is applied. So from an input image there would be only one output image for every layer. This practice would be very inefficient and useless, due to its inability to extract complex information. Instead, on each layer multiple kernels need to be used on the input image, so the result is a set of feature maps. The next layer applies a new set of filters to the previous output feature maps, obtaining a new group of feature maps that indicate more complex patterns. This is necessary because only the combination of simpler patterns can allow finding more complex ones. Therefore, the output set of feature maps can be described by only one image that does not have a single value for each pixel, but a vector of values. This is the concept of *Multi-channel image*, because the image is now a 3D tensor. The image is now composed by multiple stacked 2D matrices and each matrix represents a *channel*. In each channel there is a different feature map with a different kind of information or pattern. An RGB image is an example of a multi-channel image. The colours are produced by 3 values in each pixel, quantifying how much red, blue and green form the color. So color images are multi-channel images with three channels and each color is contained in a matrix called channel (*figure 2.8*). The same concept is successfully applied to multi-channel 3D images. In our case, CT images have a single value in each pixel, that is the HU value, so the image has only one channel, like all the grey-scale images.

A $h \times k \times p \times n$ multi-channel filter used on a $h \times w \times n$ input n-channel image produces a $(h - k + 1) \times (w - k + 1) \times p$ output p-channel image through the 2D

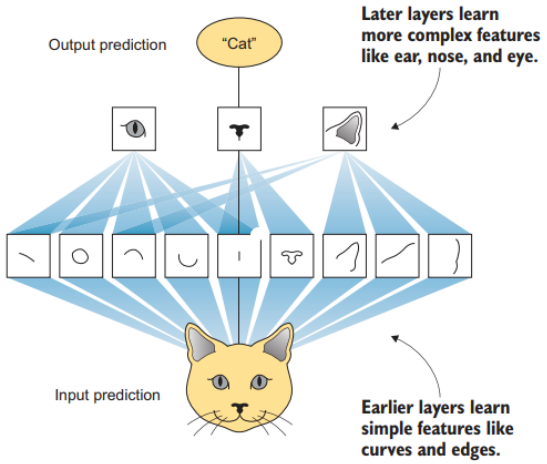


Figure 2.9: The figure explains what happen in a Convolution Neural Network. Given an image as input (cat photo), early hidden layers learn simple features as vertical lines and curved one. The next layers combine the simpler features in more complex ones, for example combining two diagonal lines an edge is obtained. This process continues until even more complex features are obtained like the eyes or the nose. Image taken from [81].

multi-channel convolution:

$$z_{u,v,d} = \sum_{c=0}^{n-1} \sum_{r=0}^{k-1} \sum_{t=0}^{k-1} w_{r,t,c,d} \cdot x_{u+r,v+t,c} \tag{2.12}$$

where (u, v) are the element indices of the output feature map of the channel d, (r, t) are the indices of the kernel weight and c indicates the input channel.

The multi-channel convolution allows to find easily patterns/features (lines, angles etc..) in the first layers and then it allows to combine the previous layer simpler patterns in more complex ones (corners... etc) and the process keeps going until it is able to do its task (**figure 2.9**). As the pattern complexity increases, also the number of channels increases, this is due to the presence of more complex features to be found comparing to the simple ones. So the number of channels is usually increasing with the depth of the NN to detect more patterns. The higher number of channels employed generates a problem from the computational point of view, because the parameters number increases exponentially. In order to solve this problem, the complexity of the model is compensated by decreasing the spatial resolution of the images. This means reducing the number of the activation states of the neurons in a feature map.

A way to reduce the image spatial resolution is including a *pooling* layer after some of the convolution ones. The pooling operation is done by a $s \times s$ kernel that takes a group of activation states and it replaces them with one of their statistic. So the the number of elements in a feature map is reduced by s^2. Usually the

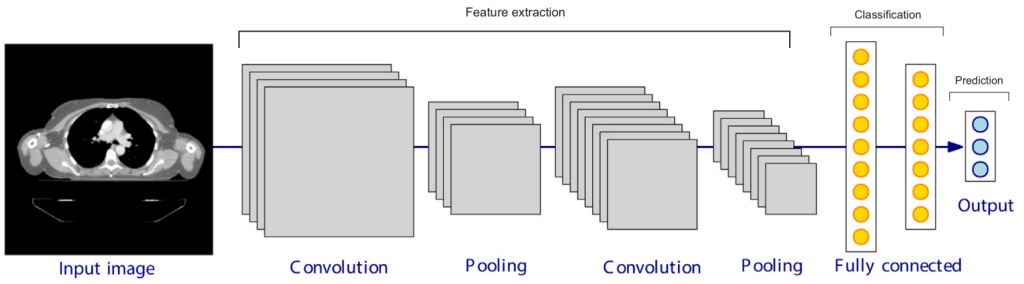


Figure 2.10: The use of a CNN in a classification problem. Given an image as input the convolutional layers extract the features, while the max pooling reduces the features' spatial resolution. When features full of semantic information are obtained, the classification step starts. The feature maps are flattened and go through fully connected layers, until the last layer is reached. Here a sigmoid (binary problem) or a softmax (multi-class problem) makes the final class prediction. Image taken from [75].

maximum value is taken (*max pooling*). The mean is also a good option (*mean pooling*).

Another option is the use of *strided convolutions* defined by:

$$z_{u,v,d} = \sum_{c=0}^{n-1} \sum_{r=0}^{k-1} \sum_{t=0}^{k-1} w_{r,t,c,d} \cdot x_{s \cdot u+r, s \cdot v+t, c} \tag{2.13}$$

Instead of moving the kernel of one pixel after the computation of a convolution, the filter is moved by s pixels. This produces a reduction in the output elements by a factor of s^2. Usually a strided convolution layer can be used every two-three same size convolution layers.

The idea behind this computational burden reduction strategy is that the exact location on the feature map is necessary only for the simpler patterns. When the complexity of the pattern increases, it is better to know that a pattern is present in the feature map than knowing its exact location. For example, when the task is to classify the animal in the image, being able to detect the distinctive complex features of the animal (the eyes, the ears, the nose, etc.), thanks to the lesser computational weight of the low resolution that allows the use of a deeper NN, is more important than knowing where are those features located on the original image.
This is the perfect solution for all the image classification problems. Because what is needed is only the probability to belong to the right class, and nothing else. In fact, in this kind of tasks, after the detection of the complex features, all the spatial information is useless, because only a global description of the image is needed. This is obtained by flattening the final convolution layer in a vector, and continuing the NN with fully-connected layers (**figure 2.10**). However, if the goal is to solve a semantic segmentation problem, this solution is not applicable. We need a *end-to-end* NN, that gets the input image and returns the segmented

images of the same size of the input image.

In the next section the NN that revolutionised the field of medical segmentation will be presented.

2.3 U-Net

In 2015 Ronneberger et al. proposed the U-Net [84] (**figure 2.11**). This NN has now become the backbone and standard in medical image segmentation thanks to its very good performances and the usage of few weights [85]. The U-Net is a *Fully Convolutional Network* (FCN), because it does not have any dense layer, but only convolutional ones. It is composed by two parts, the *Contracting path* or *Encoder* on the left and the *Expansive path* or *Decoder* on the right. Every path is the repetition of the same architecture module repeated with different characteristics.

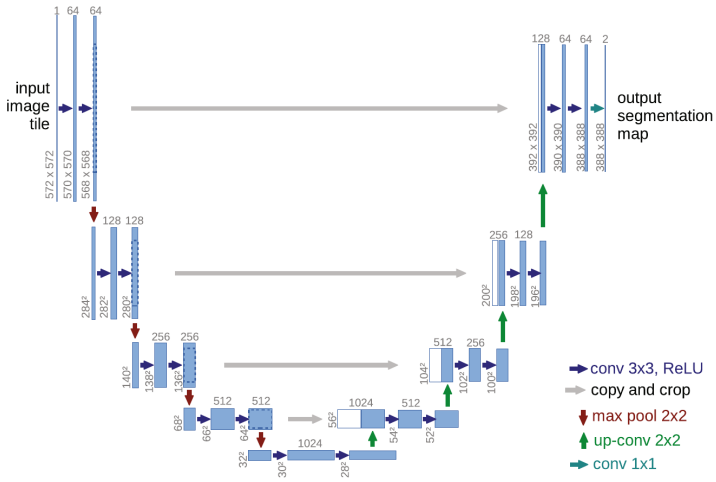


Figure 2.11: U-Net architecture (example for 32×32 pixels in the lowest resolution). Each blue box corresponds to a multi-channel image. The number of channels is denoted on top of the box. The x-y-size is provided at the lower left edge of the box. White boxes represent copied feature maps, that are concatenated to the feature maps represented by the blue box. The arrows denote the different operations. Taken from [84].

The first path serves to extract more and more complex features, in fact it has the architecture of a CNN. In the original article [84], the contracting module is composed by two convolutional layers (valid convolutions 3×3) with a RelU as activation function. The first layers double the number of channel, then a 2×2 max-pooling with stride 2 is applied to decrease the spatial resolution. Only the first contracting module increases the number of channels from 1 to 64, then it is repeated, as previously described, until it gets to the *Bottleneck* module, in which the max-pooling is not performed. The downsampling lowers the spatial resolution, so the image gets blurrier step by step. Even the organ boundaries get

blurred although they are especially important in segmentation problems. The bottleneck module is where the feature maps have the least spatial information, but they have the most semantic and complex features, for example in the original article of Ronneberger the image in the bottleneck has a size of 30×30, but 1024 channels.

The innovation of the U-net is how it retrieves the lost spatial information of the input image, that is the expanding path. The decoder part module is composed by an up-sampling, obtained through a 2×2 transposed convolution (it will be explained in the next subsection), a concatenation of the up-sampled feature maps with corresponding cropped feature maps of the contracting path (*skip connection*), and two convolutional layers (valid convolutions 3×3) with a RelU as activation function. The skip connection of the feature maps from the previous layers aims helping the NN with the localisation of the features in the maps and solving the issue of the blurred edges to get a more precise segmentation. The previous layers activation maps need to be cropped before the concatenation due to further image size reduction during the bottleneck module. The two convolutional layers are needed to learn how to use information from the feature maps, from the contracting path and the up-sampled ones to predict a more precise output. The expansive module is repeated until the desired image size is reached. The segmentation is obtained trough a 1×1 convolution layer with a sigmoid or softmax activation function, that gives the probability of each pixel to belong to a class [84, 85].

Other NN are born from the adaptation of the U-Net to 3D image segmentation problems, such as 3D U-Net [86] and V-Net [87].

Transposed Convolution

To fully understand the transposed convolution it is necessary to analyze the convolution in a matrix formalism.

Let's consider a 4×4 input image convoluted by a 3×3 kernel with zero padding (valid convolution) and stride $s = 1$, then the output image size will be 2×2. Instead of using the equation 2.11, let's flatten the input image in a vector of 16 elements and let's write the convolution operation action represented by a matrix. What it is obtained is a sparse matrix C where the weights of the kernel ($w_{r,t}$) are the only non-zero elements in the C matrix:

$$\begin{pmatrix} w_{0,0} & w_{0,1} & w_{0,2} & 0 & w_{1,0} & w_{1,1} & w_{1,2} & 0 & w_{2,0} & w_{2,1} & w_{2,2} & 0 & 0 & 0 & 0 & 0 \\ 0 & w_{0,0} & w_{0,1} & w_{0,2} & 0 & w_{1,0} & w_{1,1} & w_{1,2} & 0 & w_{2,0} & w_{2,1} & w_{2,2} & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & w_{0,0} & w_{0,1} & w_{0,2} & 0 & w_{1,0} & w_{1,1} & w_{1,2} & 0 & w_{2,0} & w_{2,1} & w_{2,2} & 0 \\ 0 & 0 & 0 & 0 & 0 & w_{0,0} & w_{0,1} & w_{0,2} & 0 & w_{1,0} & w_{1,1} & w_{1,2} & 0 & w_{2,0} & w_{2,1} & w_{2,2} \end{pmatrix}$$

In fact, the matrix product between the input vector and C produces a 4 elements vector. If the latter is reshaped, it is the 2×2 output matrix.

Therefore, the Transpose convolution is the operation described by C^T, and it

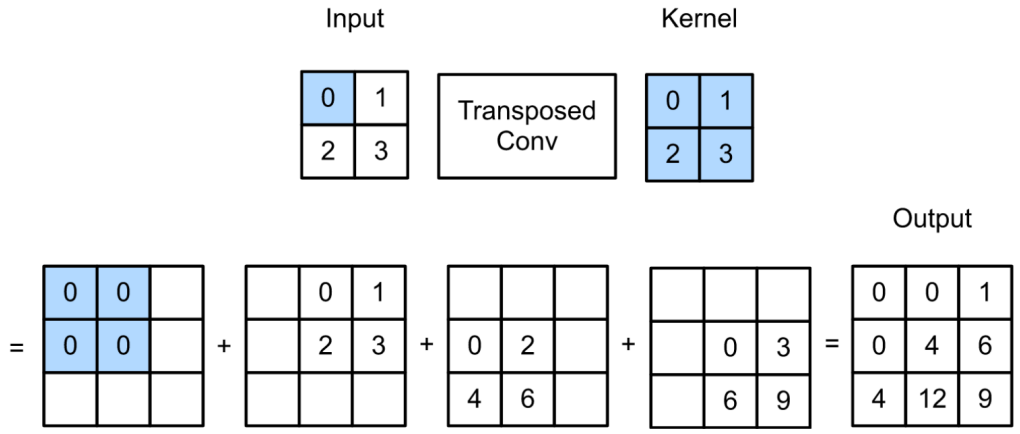


Figure 2.12: This is how the transpose convolution is computed. Given an input matrix, the first input element (light blue value) multiplies all the filter elements, the results are written on a intermediate matrix (light blue values). The second matrix element multiplies the kernel and results are written in another intermediate matrix, but a step on the right respect the previous results position. This process is done until all the input matrix elements used. Then the pixel values in the same position of the intermediate matrices are summed and the output image is obtained. This process is equal to write all the results directly on the output matrix and then sum the values in the same position. Image taken from [83].

permits to return to the 16 elements input vector, then reshaped in the 4×4 input image, if the matrix product between the 4 elements output vector and C^T is done.

Given a $w \times h$ input image a transposed convolution made through a $k \times k$ kernel produces a $(w + k - 1) \times (h + k - 1)$ output image. This can be obtained as showed in **figure 2.12**. The weights of the kernel are multiplied by the first input image element, and the corresponding results are written on the output image, then one step is made (stride = 1) to the right on the output image and the results obtained by the product of the kernel weights with the second input image elements are written in the respective locations and so on. Where some results are found on the same pixel, they get summed. In the end the output is the transposed convoluted image [83, 88].

In the next section a DL method for automatic segmentation will be presented. It is a state of the art method in the segmentation field, and it does not introduce new NN architectures, but it optimizes the construction and use of the U-net.

2.4 nnU-Net

The use of DL methods in semantic segmentation allows to analyse and quantify image features in many applications, but the model is specifically built based on the task, dataset properties and the computational capabilities. The model designing requires competence and experience, since its performance could be

compromised even by a small error. Moreover, the specificity of the model design decisions is so deep that the same model could perform very well on a dataset but worse on another one. The critical decisions to be made during the shaping and the training of the DL method are in every phase. The model architecture, how to do the training, the methods for data pre- and post- processing are just some examples of the decisions to take. Moreover it is necessary to set up the hyperparameters, such as the learning rate or the batch size. Therefore, the user can only go trough a manual and take on an experience-driven try and error process to find the optimal configuration, even if usually the used configuration is not the best one [89]. In 2021 Isensee et al. proposed No New U-Net (nnU-Net) [89] in order to face this issue. They created a tool that intends to make the whole model design process as objective as possible, leaving only a small space for manual decisions and without the task influence.

nnU-Net is a DL method for segmentation that autonomously configures itself on the whole segmentation task pipeline for any new datasets. This includes decisions about data preprocessing, NN architecture, training and data post-processing for any task. The driving force of the nnU-Net is not a revolutionary architecture of the NN or the use of a new Loss Function, but the optimization of the design decisions. In fact the name "no new U-net" underlines this aspect, because the NNs used are only 2D or 3D U-Net. Moreover nnU-Net is an out of the box tool, so it is easy to use and it does not require particular computational capacity out of the ordinary used for segmentation tasks.

This model is the state of the art in the segmentation field and it has shown better performances than other specialized NNs during the international biomedical segmentation competitions [89].

The nnU-Net's ability to configure the whole segmentation pipeline derives from the work of Isensee et al. of condensing the model design knowledge into parameters used during the configuration. This knowledge is based on a large group of development datasets and it produces robust design decisions even when it is applied on new datasets, allowing a powerful generalization. The key point is that the foundation of the design choices is not a single dataset with its characteristics, but many ones and each with different proprieties.

Fixed Parameters

Fixed parameters are composed by all of those design decisions that remain the same on every new dataset, because they proved to be the most robust ones in generalization when used on the development dataset by Isensee et al in the project developing. Some of the fixed parameters are about:

Architecture decisions, the blue-print model is a U-net like shape, and each computational module at a specific spatial resolution in Encoder and Decoder

is composed by two times the following operation sequence: convolution layer - instance normalization layer (a way to normalize the values in the feature maps) -Leake RelU activation layer. In the contracting path the decreasing of the spatial resolution is due to the first convolution being a convolution with stride > 1, instead in the expanding path the upsampling is done by a convolution transposed. The use of the skip connection remains identical to the original U-Net, but at the end of the decoding module (but no in the two lowest resolution modules), there is 1×1 convolution with a softmax in order to calculate another loss function. This is needed to inject gradients used for the learning algorithm deep into the NN, where they could become too small due to the mathematics behind the learning algorithm and slowing down or halt the learning process.

Architecture types, three different U-Net like architectures are designed and then the best one is automatically chosen: 2D U-Net, 3D full resolution U-Net (input images are at full spatial resolution) and 3D U-Net cascade. The latter is composed by a first U-Net, that works on coarse data with low resolution, then the predicted segmentation map is improved by a second U-Net that works at full resolution (this is configured for big images).

Training, the training ends after 1000 epochs, where 1 epoch is defined as 250 iterations on the training set. The learning algorithm is batch stochastic gradient descent with momentum ($\alpha = 0.99$) and initial learning rate of 0.01. The loss function is the sum of the Dice loss function (equation 2.7) and Cross entropy loss function (equation 2.4). This is decided empirically, because their combination improved the model performance. 5-fold cross validation is performed instead of creating a validation set.

Rule-based parameters

These parameters cannot be decided beforehand, because they are related to the specific dataset properties (dataset fingerprint). But the rule-based parameters are determined following specific heuristic rules applied to the dataset properties. This rules are distilled from the deep learning knowledge, and they are based on explicit relations between the *dataset fingerprint* and the *pipeline fingerprint*. The dataset fingerprint is a standardized representation of the dataset properties, thus the dataset characteristics are represented only by a small number of indexes. The latter are the median image size, voxel size (imaging spacing), the number of modalities (CT, MRI or PET) along with the number of classes to segment and the total size of the training set. The term pipeline fingerprint represents the final designed model with all the chosen parameters used for training. An example of rule-based parameters are about:

Network adaptation The downsampling obtained trough strided convolution ($s > 1$) is performed until the feature map size is small (minimum is $4 \times 4(\times 4)$).

Explaining the fixed parameters, it was said that the number of convolution in a contracting module is fixed, so the depth of the encoder and decoder are dependent from the input image size. The batch size is set to 2 in the mini-batch STG, but if the GPU capabilities allows it the batch size can be increased until the GPU memory is full.

Empirical parameters

This parameters cannot be decided only studying the dataset fingerprint, but they need to be chosen empirically after the evaluation of the validation results after the training phase. An example of empirical parameter is:

Model selection, knowing beforehand the performance of a model is not possible, because nnU-net generates three U-Net like models. In order to decide which model is the best one for the specific dataset and task, nnU-Net compares their performance only after the end of the 5-fold cross validation of each model [89].

Postprocessing, this phase is based on the idea that many times in the medical imaging field the target of segmentation is only one instance. Thus if during the validation removing all the smaller predicted/ segmented parts of a determinate class, but the largest, increases the Dice Coefficient, all the smaller parts of that class will be eliminated by all the predicted segmentation in the test set.

The nnU-Net is an out-of-the-box tool and it is the state of the art method in the automatic segmentation task. Therefore it was decided to use this DL algorithm to set the basic performance in our preliminary study of the H&N cancer automatic segmentation on CT images.

Chapter 3

Dataset creation and uses

In the previous chapter it was emphasized how the DL methods need large training sets. In this chapter it is discussed how our CT image dataset of H&N cancer was created for the training and test of the nnU-Net and how the images are implemented to improve the NN performance. The chapter ends with the explanation of what it can be done with the tumor predicted segmentation.

3.1 Dataset creation

The images used were downloaded from a public archive, The Cancer Imaging Archive (TCIA) [90]. It is possible to obtain many images about different kinds of cancer. Moreover all imaging modalities are available such as CT, PET and MRI, and some of them come with a corresponding manual segmentation image. Images are in DICOM format, that is the standard format type for medical imaging. DICOM files are made by two parts, the image (pixel image(s)) and the header. The header stores all kind of information about the patient, image itself and how the image was obtained. For example, the header reports the patient ID, the characteristics of the imaging machine, the way the image was obtained and many more information. Every information in the header is associated with a specific alphanumeric code, that makes the information unique and easily accessible.

One information of the DICOM file is the Modality that describes what kind of image is contained in the file. The CT images have a CT modality, and the same goes for the PET (PT) and MRI (MR), but the manual segmentation made by the clinicians could have different modalities,such as the Radiotherapy Structure Set (RTSTRUCT) modality or the Segmentation (SEG) one. In this preliminary study we only used the segmentation in files with a RTSTRUCT modality. Another modality that will be useful later on is the Radiotherapy Dose (RTDOSE), containing the dose distributions calculated by the TPS. There are many more modalities, but these are those relevant for this thesis.

The downloaded datasets with the patients number are the following:

- **Cetuximab** (111 patients)
- **CPTAC-HNSCC_V9** (64 patients)
- **HNSCC** (627 patients)
- **HNSCC_3DCT** (31 patients)
- **MRI-DIR** (9 patients)
- **OPC-Radiomics** (606 patients)
- **PET-CT-v2** (300 patients)
- **QIN** (279 patients)
- **RadiomicsHN1** (137 patients)

These datasets are organized all in the same way. Every patient has his own folder, with a sub-folder for each time the patient got hospitalized, treated or had a check-up. The sub-folders contain all the examinations that the patient received or the material necessary for the treatment. Because of the data organization, datasets need to be rationalized and more usable in a new database. Moreover, while downloading each dataset, a .csv file is attached. All the DICOM files of the dataset are listed in this .csv, moreover there are their properties and characteristics, including their path in the database. The file location is fundamental to been able to use and analyse the file itself, even if due to the presence of many sub-folder it's a complicated path to use. For all these reasons, the creation of a new large dataset is a must.

The first step is finding how many images with a specific modality there are in each dataset (**table 3.1**), and alongside how many and which images have each patient in each dataset. Hence a more detailed and close view on every patient is obtained.

Database name	CT	PET	MRI	RTSTRUCT	RTDOSE	other mod.
Centurimab	579	585		96	211	211
CPTAC-HNSCC	421		79			26
HNSCC	1297	1703		607	202	224
HNSCC 3DCT	93			185	93	
MRI-DIR	5		48	8		
OPC-Radiomics	591			613		
PET-CT-v2	503	298		831	497	531
QIN	906	924				1957
RadiomicsHN1	137	75		137		137

Table 3.1: The number of images for each modalities are showed. The empty space means the absence of the image modality in the dataset. The other modality column summarizes all the image modalities that are not useful in this thesis

Observing **table 3.1**, it seems that there are many RTSTRUCT files, but it should be noted that this number represents all the segmentations done on all image modalities, therefore, it is a larger number than the one used for this thesis work (segmentations on CT images). This is an important aspect to take into account for our purpose. **Table 3.2** shows the RTSTRUCT files associated to each image modality in the considered datasets. This information is obtainable because the RTSTRUCT file header contains the knowledge of the image modality used for the segmentation. The RTDOSE file similiarly has the same information in its header, therefore, it is possible to link the TPS's dose distribution output with the correct RTSTRUCT segmentation and with the diagnostic image modality. Thanks to this our final database is composed by 1934 CT images with usable RTSTRUCT, where 713 CT images have both the RTSTRUCT and the dose distribution. In the next step it will be explained how these numbers are obtained. All the RTDOSE files in the these datasets contain only dose distribution based on photon radiotherapy.

Database name	CT	PET	MRI
Centurimab	96		
HNSCC	606	1	
HNSCC 3DCT	185		
MRI-DIR			8
OPC-Radiomics	613		
PET-CT-v2	533	298	
RadiomicsHN1	137		

Table 3.2: Differentiation of the RTSTRUCT files of each dataset based on which diagnostic image the segmentation was created. The empty space means the absence of the RTSTRUCTs based on image modality in the dataset.

The second step is characterizing each diagnostic image and storing each image information. This is done creating a .csv file for each modality in the dataset and for each dataset. The image information are extrapolated directly from the image-header. To better understand the characteristic of the images available in the various datasets some image properties were analysed. The parameters considered were:

- Image modality,
- Image shape,
- Number of image slices,
- Slice thickness,
- Spacing between the slices,
- Image Field Of View (FOV) shape,

- Image Field Of View dimension,
- Methods of image reconstruction.

This is an important aspect for the results of this thesis, and also for the AI_MIGHT project. This thesis is focused on a preliminary study of the algorithm suing only those CTs that have an RTSTRUCT file without considering the other characteristics, however the instruments developed here will be used for the future improvement of the algorithm. Thus, it can be said that this work is the foundation for the achievement of the goals of the project.

The third step is the association of the CTs with the respective RTSTRUCT file. The issue in this phase is that there are some cases where there are more RTSTRUCT files associated to the same CT image. Comparing **table 3.1** and **table 3.2** the HNSCC 3DCT dataset is particularly interesting because it contains 93 CT images, but 185 RTSTRUCT files based on CTs. There could be different explanations of this situation: the clinician drew different segmentations or some files are simply repeated.

To solve this, let's halt the third step for the moment and move on with the fourth step. The nnU-Net needs the images in the NIFTI format file, and it is easier to use algorithms on this format than the DICOM one. So the Plastimatch [91] tool is used to convert all the CTs, RTSTRUCTs and RTDOSEs in NIFTI format.

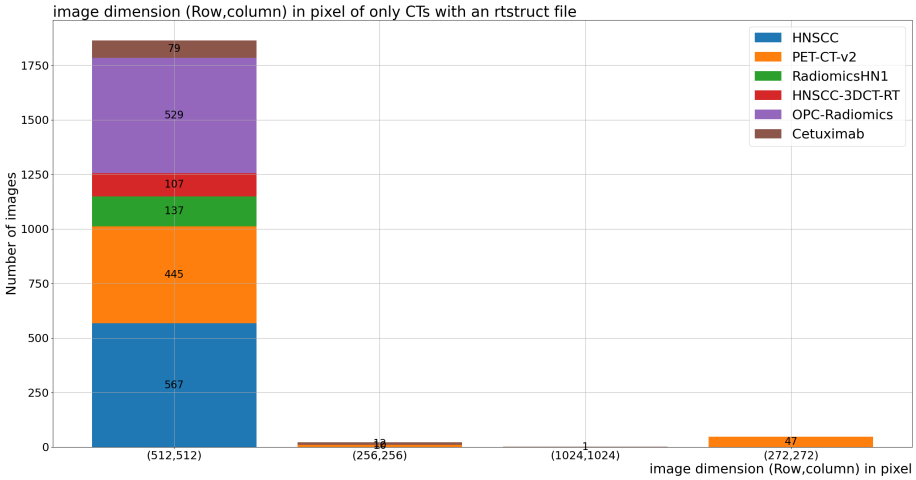
In the NIFTI files, there are the segmentations of many organs and many tumor volumes; the latter have been explained in subsection 1.3.1. Our goal is to train a NN able to localise the tumor in the CTs, thus the tumor volume has to be visible in the CTs in order to be identified. This means that only the GTV will be used as ground-truth during the training phase. But to take in account the worst case scenario and to use any information at our disposal, when the GTV_{ln} is present, it is fused with GTV creating only one mask. The other structures segmentations are not used in this thesis.

Now it is possible to go back to the third step. The GTV is a mask, so it is a tensor and its voxels are 0 when the tumor is not present and 1 otherwise. This property is used to check the presence of segmentation double files. In fact, two same size GTV masks named A and B, with the respective voxels $a_{i,j,k}$ and $b_{i,j,k}$ where i,j,k are the indices that describe the position in the tensor, are equal if

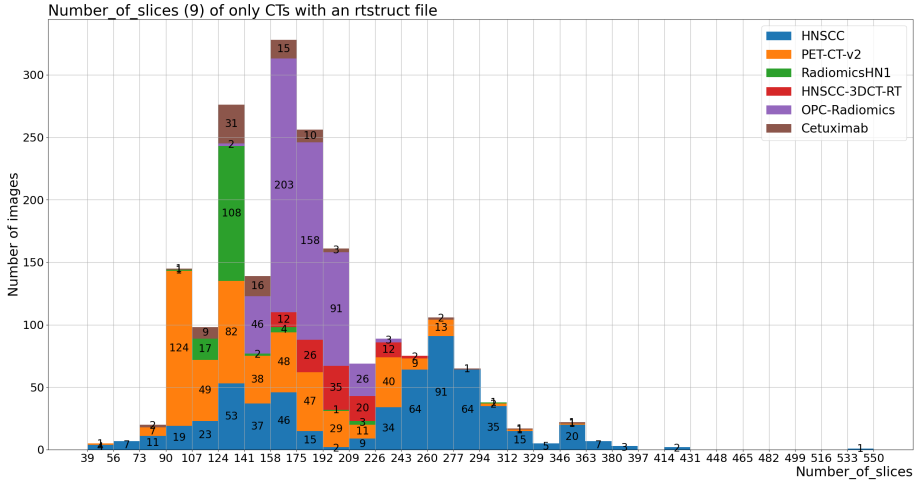
$$\sum_{i,j,k} a_{i,j,k} \cdot (1 - b_{i,j,k}) = 0$$

Using this relation, all the double files of patients are found and eliminated. When there are more GTV masks for the same CT image, they are kept. Those GTV masks without the respective CT are eliminated.

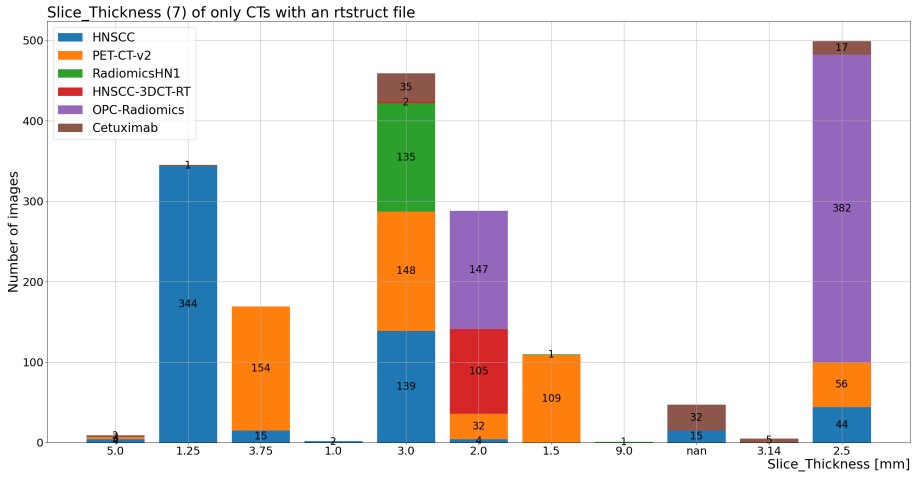
In the end our database composition is **1934 CT images with GTV mask**,



(a) *Image shape's distribution in our dataset*



(b) *Number of slice's distribution in our dataset*



(c) *Slice thickness' distribution in our dataset*

Figure 3.1: Characteristics of our database are showed, moreover the source of the images is reported.

where 713 CT images have both the GTV mask and the dose distribution. Some of the database properties are shown in the **figure 3.1**

3.1.1 Boundary Box (BB)

Now that the images are organized inside the dataset, it turns out that there are many information not useful. An example is reported in **figure 3.2 (a)**, there are body parts that we are not interested in, such as the torso and in some images even the legs, moreover there is the bed of the patient and all the background. These elements could be an hindrance to the leaning process of the NN, and they increase the computational weight and time without any reason.

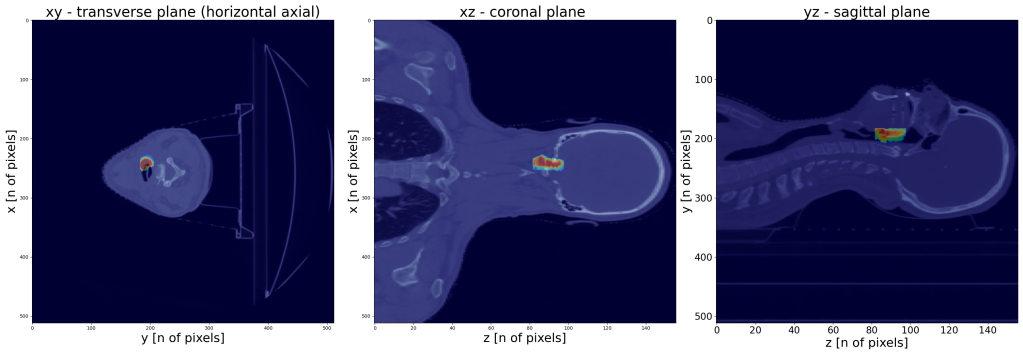
It is possible to decrease the computational cost, while increasing the NN performance, by cropping the images to maintain only the H&N region of the body. This is achieved through the use of a *Boundary Box* (BB). The first BB used was obtained through four different actions. The first is Windowing the images with an HU range of [-1000, 1000] (explained in section 1.5.1), this range improves the bones contrast. The second step is the Otsu threshold filter [92] application to the windowed images, the resulted images are no longer in grey-scale, but they are composed only by white or black pixels, separating the foreground from the background. Now a binary erosion is applied to the two color images: this operation removes single white pixels surrounded by the black ones, and then a binary dilation is performed. The latter enhances the white pixels increasing their number in the border between white and black pixels. In this way only the patient profile remains and it can be used to look for the two opposite points that describe the BB. The result of this process is showed in **figure 3.2 (b)**. In this way, the majority of the background and even the bed are eliminated, but some other unwanted body parts remain.

To select only the H&N region, some constraints are added on the previous BB. Fist it is established that on the x-axis a range of 20 cm centered on the middle of the BB's side along the x-axis is selected. The BB side along the y-axis is increased by 3cm, to avoid the deletion of the nose from the patient head. Lastly, the BB side along the z-axis is reduced, it starts from the top of the head and goes down for 24 cm, if the last pixel on z-axis of the GTV mask is too close or above this height, the new height limit is set as the position of the last pixel on z-axis of GTV mask +2 cm. The result of this rules is showed in **figure 3.2 (c)**

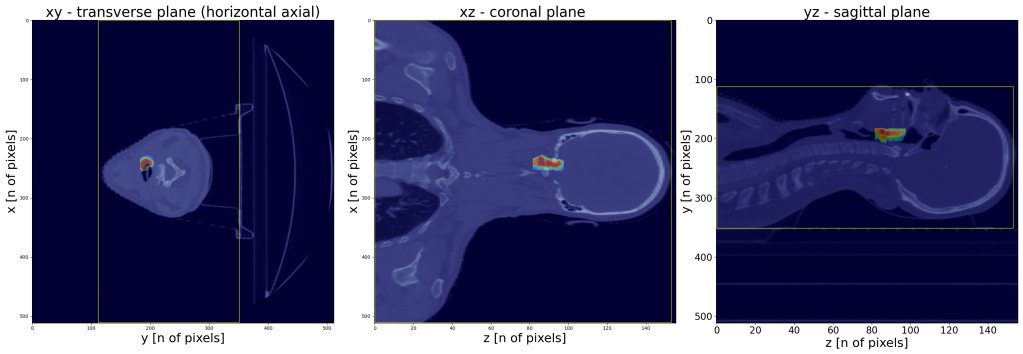
Finally, the images are cropped using the coordinates of the updated BB, thus the dataset is now complete and taylored for a more efficient NN application.

3.2 Dataset uses

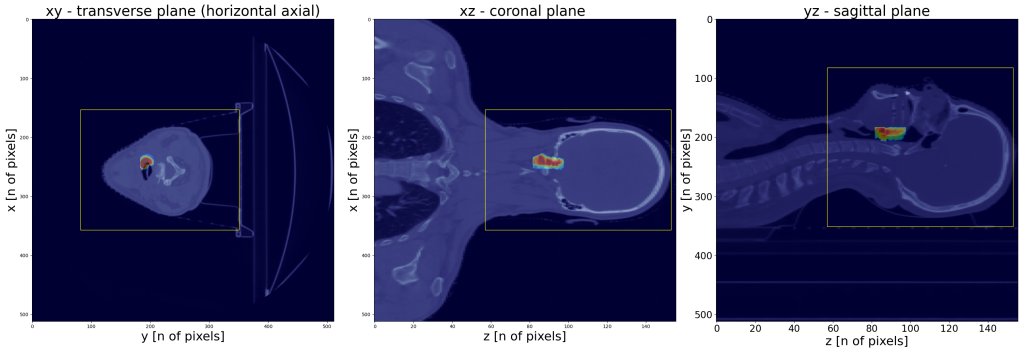
The completed dataset is used for the training of the NN, but as said before in section 2.2 it is important to have a test set to evaluate the NN performances



(a) *Initial Image*



(b) *Original Boundary Box*



(c) *Modified Boundary Box for the H&N region*

Figure 3.2: The steps in the image cropping are showed. The original image (a), the Boundary Box (BB) obtained with the first algorithm, represented by the yellow lines (b), and the adapted BB for the H&N region (c), represented by the yellow lines. Moreover, in every images the tumor is reported. For each view the tumor was projected respectively on the transverse plane, coronal plane and sagittal plane. The bluer is the color associated to a part of tumor, the less tumor volume was projected in that pixel, on the other hand the red corresponds to a larger tumor volume projected in the pixel.

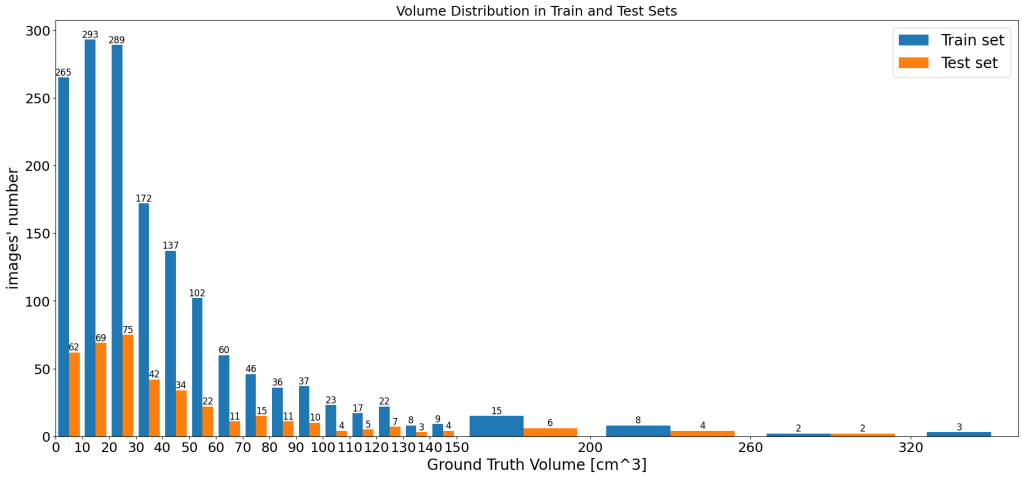


Figure 3.3: Tumor volume distribution in train and test set. This volume is the ground-truth volume used during training.

without bias. The train set is 80% (1547 images) of the dataset and the remaining 20% (387 images) is used as test test. Both sets are randomly generated, and the tumor volume distribution of both sets is plotted to avoid any bias between train and test set under this aspect. **Figure 3.3** shows how the tumor volume distribution is equal in both sets, proving the correct division of the data with respect to this parameter.

Three different training are scheduled:

- The first is the training used to find the minimum performance that we want to pursuit through the nnU-Net. This is obtained using the original images with all the useful and not useful information. Only one fold is used to train a 3D U-net at full resolution, because the time of training is around 12 days. This training is referred to as Train_A.
- The second training is done using the cropped images, using all the folds for each NN proposed by the nnU-Net. The computational saving is blatant, because training the NN on a fold takes only a mean of 20 hours. This training is referred to as Train_B.
- The third training is equal to the previous one, but the number of epochs is increased to 2000 for the best NN of the Train_B. This training is referred to as Train_C.

The NNs trainings are all performed on the calculation cluster EOS [93] of the Pavia University. It consists of a multiprocessor and multi-GPU cluster with parallel storage system. Its calculation capacity is dived in nodes, in particular among the nodes it has seven GPU nodes, each with 128GB RAM and 2 Nvidia V100 GPUs with 32GB RAM. Our trainings are done on a single GPU node.

Indexes are needed to evaluate the performance of the trained NN on the test set.

The indexes used in this thesis are defined next. Given a ground-truth volume G and the predicted volume S:

- **Dice Coefficient(DC)** [94], defined as $DC = \frac{2 \cdot (G \cap S)}{G+S}$.
  DC measures the similarity between two volumes, and its range is [0,1], where 0 is obtained when there is no similarity, on the other hand 1 means perfect superposition.

- **Geometrical Miss Index** (GMI) [95], defined as $GMI = \frac{G - (G \cap S)}{G}$.
  GMI is the fraction of the ground-truth volume G that is not predicted. It measures the under-contouring of the ground-truth volume. Its range is [0,1], the best case is 0, when all the ground-truth volume is predicted,the worst case is 1.

- **Discordance Index** (DI) [96], defined as $DI = \frac{S - (G \cap S)}{S}$.
  DI is the fraction of the predicted volume S that does not belong to ground-truth volume G, in particular it measures the over-contouring respect to the ground-truth. Its range is [0,1], the best case is 0, when all the predicted volume belongs to the ground-truth, while theworst case is 1.

- **Jaccard Coefficient** (JC) [97], defined as $JC = \frac{(G \cap S)}{G \cup S}$.
  It is similar to the DC, and it measures the concordance between the two volumes, and its range is [0,1], where 0 is obtained when there is no similarity, on the other hand 1 means perfect superposition.

- **Surface Dice Coefficient (SDC_τ)** [98], defined as $\frac{|S_G \cap B_{S,\tau}| + |c \cap B_{G,\tau}|}{|S_G| + |S_S|}$, where $S_G$and S_S are the surfaces of the volumes G and S, while $B_{i,\tau}$ is the border region of the surface i-th with thickness τ, that is used as tolerance parameter. Using this metric the performance is evaluated based on the fraction of the predicted surface S_S that needs to be redraw by the clinician in order to obtain S_G, with an acceptance of τ. Thus it measures the overlap of the surfaces. In this thesis $\tau = 2mm$ is used, and this index will be called Surface Coefficient Index with tolerance 2mm (SDC_2mm). Its range is [0,1], where 0 is obtained when there is no overlapping, on the other hand 1 means perfect overlapping.

Therefore, DC,JC and SDC_2mm measure the similarity between volumes or surfaces, while GMI and DI measure the dissimilarity between volumes.

At the end of the described process, 20 patients have been selected in the test set. The criterion of the selection is having two groups of 10 patients each (all with RTDOSE file).In the first set the NN performed very well, while in the second set the automatic segmentation performance was not as good. In the latter, a lower limit of the DC scored by the NN is adopted, that is $DC = 0.2$. This choice has been made because under this limit a clinician would not have a real advantage in modifying the automatic segmentation respect to starting a manual

segmentation from scratch. In both groups, patients have been chosen to assure a suitable variance of the ground-truth tumor size. These patients are used for further investigations. Both their ground-truth and predicted segmentations are used as input in the BNCT TPS IT_STARTS, the respective dose distributions are confronted and evaluated.

Chapter 4

The nnU-Net segmentation

4.1 Performance evaluation

4.1.1 Training

In section 2.4 how the nnU-Net works was explained. nnU-Net follows a precise pipeline in order to achieve the best results and adapt itself to any new database. The first step is the database fingerprint extraction, that is used with heuristic rules to choose the best rule-based parameters. This step produces the final pipeline fingerprint, that is the final designed model with all the chosen parameters used for training. The model is composed by the NN architectures that the nnU-net found suitable to perform the user task on the specific database.

Therefore the last parameters that have to be chosen are the empirical parameters: the best NN architecture and if the image post-processing has to be done. The empirical parameters can be chosen only after the models training. The Training is done trough a 5 fold-cross validation. So the NN is trained on each of the folds as explained in section 2.2.2, and for each fold the Dice Coefficient (DC) is calculated to understand the NN performance.

What happens is that the NN is trained on 4 folds, and the last fold is used as validation fold. So the NN predicts the tumor volume of each validation images, and the predictions are confronted with the respective ground-truth tumor volume to calculate the DC. Lastly the validation DC mean is calculated. Moreover, post-processing is applied to the predicted image, it means that the DC is calculated again, but everything except the largest foreground region is removed by the predicted segmentation. Then the DC mean value is taken. If the images has only two classes (in our case 0 bakground 1 tumor) the post-processing has only this option. But if there are more classes then another post-processing process is taken into account. The DC is calculated again, but in the predicted image all is removed but the largest connected component for each individual foreground class. Then the DC mean value is taken. Once the 5 fold cross-validation is

completed each fold has two or three DC values as evaluation.

When the 5 fold cross-validation has terminated, nnU-Net uses the 5 trained model from the cross-validation as an ensemble. Ensembling is implemented by averaging the softmax predictions. Then this model is the final model for the specific architecture proposed by the nnU-Net. If the post-processing is done on the test set and which post-processing is decided confronting the DC ensembles of the 5 folds. If the DC is higher without post-processing, then it will not be done in future steps, otherwise the post-processing will be done. Which kind of post-processing will be applied is always decided by confronting the DC values.

When this pipeline is done for all the NN architecture proposed by the nnU-net, the best model is decided confronting their DC values obtained during training. The best model will be the one tested on the test set.

The first training we did is going to be referred as Train_A, it was done on the original images without any modification. This step is important because it sets the minimum performance standard, that we want to obtain in the other Train sets after the image cropping. Due to the big image size only one configuration of the 5 fold cross-validation was used to train the 3D U-Net full resolution, and this is the only architecture trained in Train_A. So the DC values obtained on the only validation fold are used as final training DC values. Unfortunately using the original images means a big computational cost, in fact the training of only one 5 fold cross-validation configuration lasted more than 12 days. In Train_A the final DC values are $DC_{raw} = 0.58 \pm 0.18$ using the image prediction without post-processing and $DC_{post} = 0.54 \pm 0.20$ with the post-processing. Thus post-processing will not be applied to the new images.

The second training is Train_B. The training set used in this Train is composed by the cropped images. This allows us to to train all the NNs proposed by the nnU-Net. The architectures trained are:

- 2D U-Net. After the cross-validation the ensamble DC values are $DC_{raw} = 0.61 \pm 0.21$ and $DC_{post} = 0.56 \pm 0.20$.
- 3D U-Net full resolution. After the cross-validation the ensamble DC values are $DC_{raw} = 0.65 \pm 0.21$ and $DC_{post} = 0.60 \pm 0.21$.

The DC calculated after the post processing is smaller than the one calculated with the raw predicted images for both the architectures. Confronting the best DC of the NNs results that the best model is the 3D U-Net full resolution without post processing. A very important aspect is the significant reduction of the computational time thanks to the images cropping. Training one configuration of the cross-validation lasts 15 hours for the 2D U-Net and 25 hours for the 3D U-Net full resolution. This is a fundamental fact, because this will allow not only a reduction of time, but even an increasing of the performance.

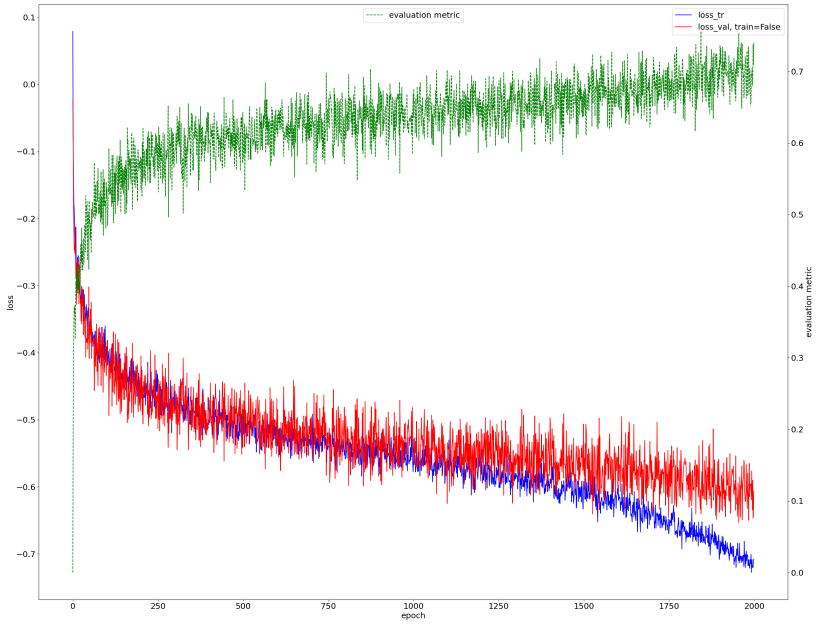


Figure 4.1: The learning process of the 3D U-Net full resolution using fold 3 as validation fold in Train_C. The training (blue) and validation (red) loss during training is showed. Moreover the green line is an approximation of the evaluation metric. This is just an approximation because it is calculated only using randomly extracted patches from the validation fold images, and then they are used to calculate the global dice as if the patches came all from the same volume.

Lastly, Train_C is a modification of Train_B. In the nnU-Net the number of epochs used during training is a fixed parameter and it is 1000. One epoch is defined as 250 iteration on the training set. But this parameter can be manually changed by the user, therefore in Train_C it was set a 2000 epochs. This is done to investigate if there is any room for further improvement or if increasing the epochs causes overfit. Overfit happens when the model starts to learn random patterns inside the training data and does not learn the meaningful ones. Thus, when new data are exposed to the model, the latter performance will be bad, even if the performance on the training data was very good. This is caused by the excessive complexity of the model for the task or because the training lasted too many epochs. So it was decided to train again with 2000 epochs (**figure 4.1**) the best Train_B model, 3D U-Net full resolution. This means doubling the computational time. Once the the cross-validation ended, the DC values obtained from the training were $DC_{raw} = 0.67 \pm 0.21$ and $DC_{post} = 0.62 \pm 0.22$. Also here the post-processing will not be operated on the test images.

When all the NNs are trained, what remains to do is evaluating their performances.

4.1.2 Evaluation

The performance evaluation of the trained NNs on the test set is fundamental. The indexes, used to measure the NN capability to predict the segmentation are explained in section 3.2. In the same section the different training with their characteristics are explained.

In order to confront in the best possible way the performance indices of the three Trains, the indices are plot in a box-violin plot. On the y-axes there is the range of the index, and it will be always $[0, 1]$ for the indexes used in this thesis. The index representation is on a specific value of the x-axes. Each index is represented by a Box-violin, and it shows different kinds of information. The light blue area is the index values distribution in the test set, thus where the area is larger, the bigger is the number of test samples that scored that specific index value. The green dot is the mean, instead the red line is the median. The black box centered on the median indicates the 25^{th} percentile (low median) and the 75^{th} percentile (high median), while the two lines (whisker), that terminates in the small horizontal segments indicate the interval $1.5 \cdot (q_n(75) - q_n(25))$. The black circumferences are outliers of the latter interval.

Dice Coefficient

The Dice Coefficient is the most important coefficient in this preliminary study, because the Loss Function used by the nnU-Net is based on this index. In **figure 4.2** and in **table 4.1** the results are reported. It is blatant the performance improvement from Train_A to Train_C. It is especially visible focusing on the changing of the DC values distributions. In Train_A the light blue area is more spread along all the index range with a thick tail in the DC range [0 , 0.4]. This indicates a significant number of test samples with low DC, in fact the 25% of the test set scored a $DC < 0.47$. Instead in Train_B and, and even more in Train_C, the distribution tail is slimmer, because much more samples scored high DC values. Moreover it is important to observe how in the values distribution a peak grows on the high DC values until it stops in Train_C on the range of $[0.70, 0.85]$. This is the proof of the improving capability of the NN to recognize the H&N tumor, and it is proved by the median value increasing and by the fact that 25% of the test set scored an $DC > 0.82$ in Train_C. What just said is even more expressed in **figure 4.3**, where the peak growth in the last two Train is well represented.

The performance increasing from Train_A to Train_B is due only to the cropping of the images, because not useful information was eliminated keeping only the meaningful one. Moreover this produced another important gain, the time necessary to train the NN was reduced significantly. Instead doubling the epochs number in the Train_C allows to improve further the performance on those tu-

mors already well segmented in Train_B. But this change does not influence much those tumors which have a very small DC in Train_B. This trend is also confirmed by the other indexes distributions.

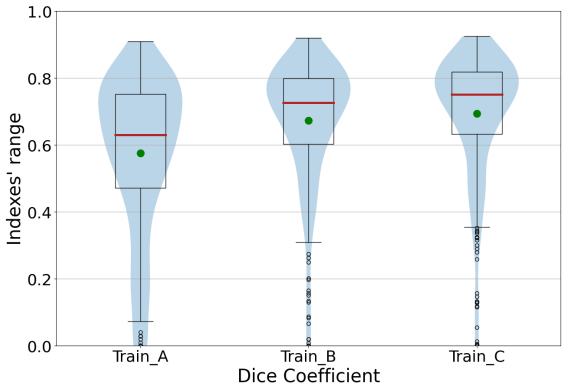
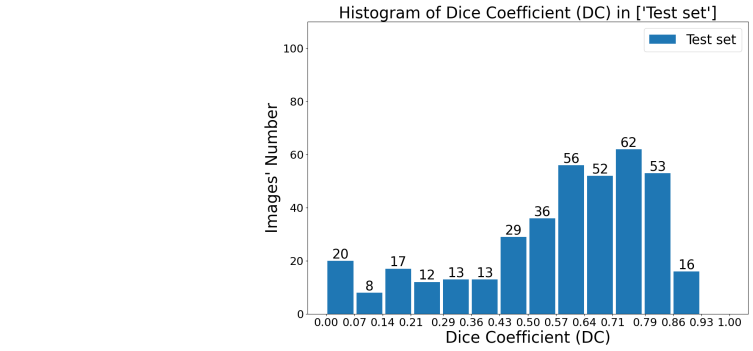


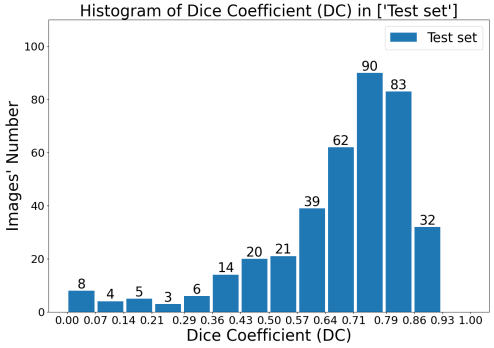
Figure 4.2: The box-violin plot of the Dice Coefficient (DC) is showed for each Train. The mean is the green dot and the median is the red line

Train ID	Mean	Median	Low Median	High Median
Train_A	0.58 ± 0.23	0.63	0.47	0.75
Train_B	0.67 ± 0.19	0.73	0.60	0.80
Train_C	0.70 ± 0.19	0.75	0.63	0.82

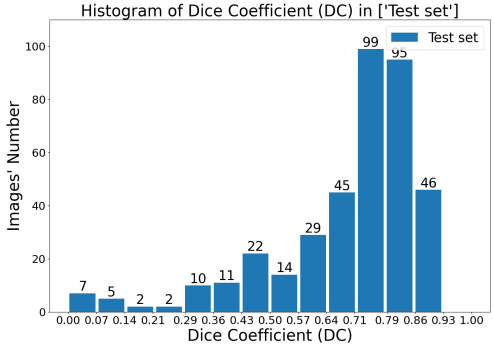
Table 4.1: In table the mean, median, low and high median of the Dice Coefficient (DC) are showed for each Train



(a) *Train_A*



(b) *Train_B*



(c) *Train_C*

Figure 4.3: The distribution of Dice Coefficient (DC) for each Train is showed.

Another important aspect is comparing the DC values obtained during training and testing for each Train. If the training DC value is much higher than the testing DC value, it means that the model is in overfitting, because it learns random patterns from the training data. On the other hand, if the training and test DC are similar, but they have a low value, then the model would be in underfitting, i.e., the model has not learnt enough information. But compsring our training and testing DC values (**table 4.2**), it emerges that our model is not in any of the above situations. The DC values are very similar and with high values, so the trained model *generalizes* well. This means that useful patterns are learnt and the model works properly even with unseen data.

Train ID	**Train DC**	**Test DC**
Train_A	0.61 ± 0.18	0.58 ± 0.23
Train_B	0.65 ± 0.21	0.67 ± 0.19
Train_C	0.67 ± 0.21	0.70 ± 0.19

Table 4.2: In table the DC values scored in train and test for each Train are showed

Geometrical Miss Index

The Geometrical Miss index is an important evaluator, because it can tell if the NN identifies the true tumor volume represented by the ground-truth volume or if it produces undercontouring. Observing **figure 4.4** it is possible to see an improvement of the second Train respected to the first and the third respect to the second. But it is just a slight change, not like the one in the DC. This is confirmed by the values reported in **table 4.3**, in fact they are quite similar. This means that cropping the images and doubling the epochs did not produce a qualitative change in the ability of segmenting the tumors, but just a small quantitative change, as show in **figure 4.5**. This could indicate the necessity of adding something else to reach a qualitative change. Maybe, in order to produce this change, could be useful perform some pre-processing on the image to make more visible the tumor on the scan. Moreover searching if a specific tumor characteristic makes the tumor hard to detect could be helpful. In this way a more selective NN could be trained.

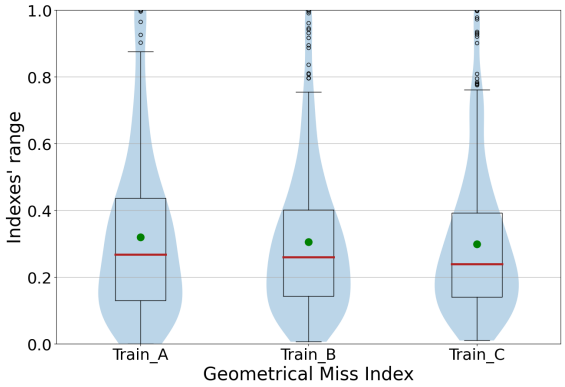
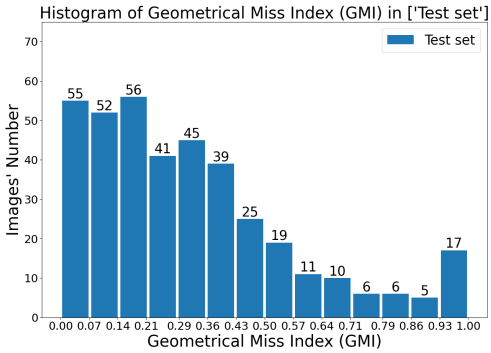


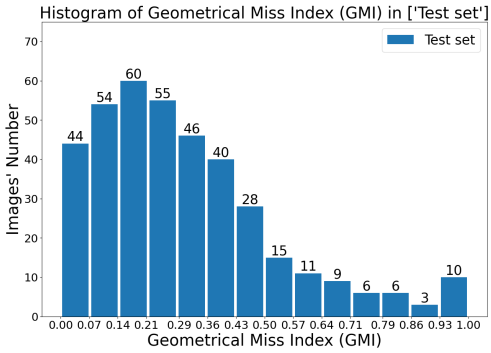
Figure 4.4: The box-violin plot of the Geometrical Miss Index (GMI) is showed for each Train. The mean is the green dot and the median is the red line.

Train ID	Mean	Median	Low Median	High Median
Train_A	0.32 ± 0.25	0.27	0.13	0.43
Train_B	0.30 ± 0.22	0.26	0.14	0.40
Train_C	0.30 ± 0.22	0.24	0.14	0.39

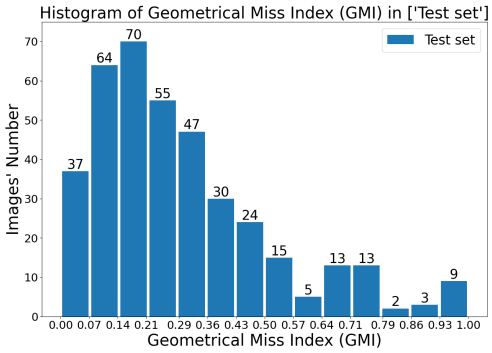
Table 4.3: In table the mean, median, low and high median of the Geometrical Miss Index(GMI) are showed for each Train.



(a) *Train_A*



(b) *Train_B*



(c) *Train_C*

Figure 4.5: The distribution of Geometrical Miss Index (GMI) for each Train is showed.

Discordance Index

The Discordance Index is complementary to the Geometrical Miss Index, because DI tells us how much of the predicted volume belongs to the ground-truth volume, so if the NN produces overcontouring. The best case is when DI=0 and the worst is when DI=1, thus **figure 4.6** represents a situation very similar to the DC distribution in **figure 4.2**. In Train_A the spread distribution with the thick tail reappeared, and a significant improvement occurs with the cropping of the images and the doubling of the epochs in Train_C. This is confirmed by **figure 4.7** and **table 4.4**. The DI and DC are correlated, and in this case the improving of the DC values are caused by the reduction of the overcontouring, i.e. the increasing of the DI values. In Train_B and Train_C the NN ability of recognizing what is not a tumor volume increased.

Train ID	Mean	Median	Low Median	High Median
Train_A	0.43 ± 0.27	0.38	0.21	0.60
Train_B	0.29 ± 0.21	0.23	0.13	0.37
Train_C	0.26 ± 0.20	0.20	0.11	0.32

Table 4.4: In table the mean, median, low and high median of the Discordance Index (DI) are showed for each Train

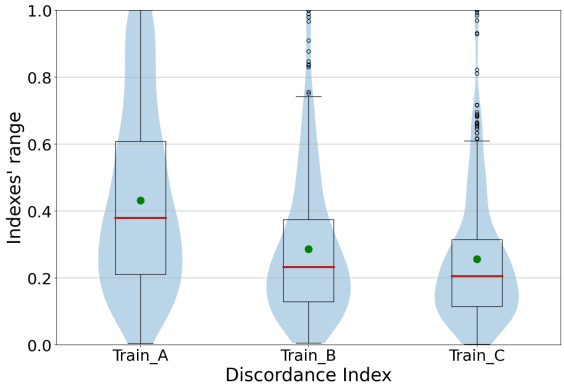
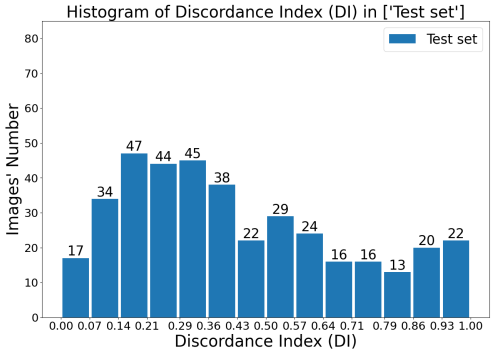
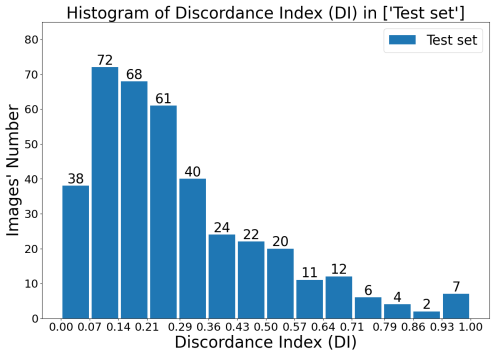


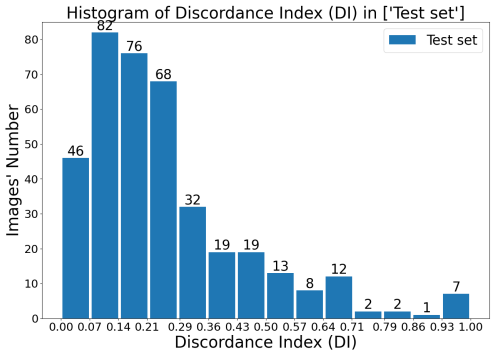
Figure 4.6: The box-violin plot of the Discordance Index(DI) is showed for each Train. The mean is the green dot and the median is the red line.



(a) *Train_A*



(b) *Train_B*



(c) *Train_C*

Figure 4.7: The distribution of Discordance Index (DI) for each Train is showed.

Jaccard Coefficient

The Jaccard Coefficient measures the accordance between two volumes, like the DC. Thus it is used to confirm the results obtained before through the DC distribution study. In **figures 4.8, 4.9** and **table 4.5** the same trend is found. Confirming the evaluation done with the DC index. The improvement between Train_A and Train_B is clearly visible and it is also possible to find a slight improvement by doubling the epochs from Train_B and Train_C.

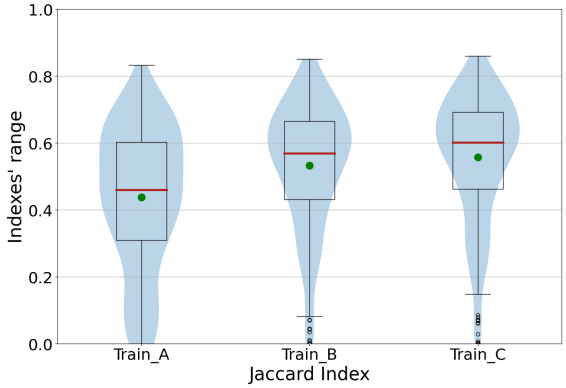
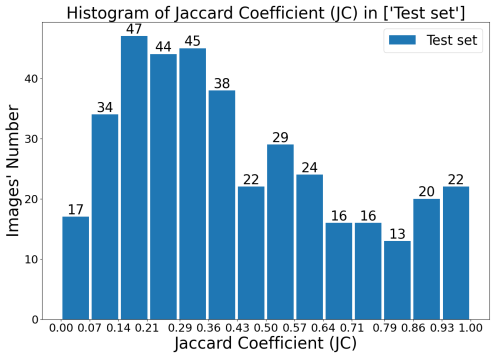


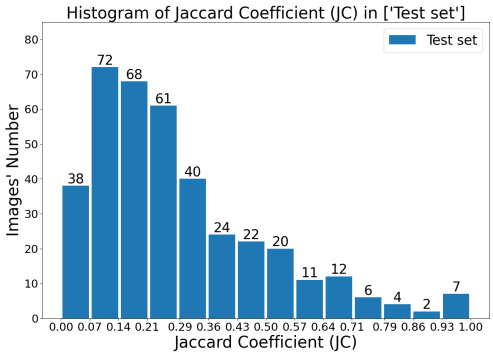
Figure 4.8: The box-violin plot of the Jaccard Index (JC) is showed for each Train. The mean is the green dot and the median is the red line.

Train ID	Mean	Median	Low Median	High Median
Train_A	0.44 ± 0.21	0.46	0.31	0.60
Train_B	0.53 ± 0.19	0.57	0.43	0.66
Train_C	0.56 ± 0.19	0.60	0.46	0.69

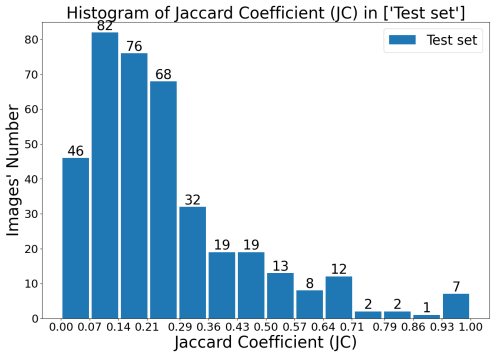
Table 4.5: In table the mean, median, low and high median of the Jaccard Index (JC) are showed for each Train.



(a) *Train_A*



(b) *Train_B*



(c) *Train_C*

Figure 4.9: The distribution of Jaccard Coefficient (JC) for each Train is showed.

Surface Coefficient Index (2mm)

It is interesting to analyse this index, because it takes in account the deviation magnitude and the number of predicted surface deviations from the ground-truth surface. Many small deviations will be corrected by the clinician in a longer time than a single deviation of a bigger magnitude, because in the former almost all the surfaces could be redrawn in many steps, instead in the latter only one edit could resolve the error. This aspect is not taken into account by the DC, because each dissimilarity between the two volumes are weighted in the same way[98]. In Train_A the DSC_2mm distribution has a bell shape centered on the middle of the index range. The peak shifts toward higher SDC_2mm values in the other two Trains, but the distribution keeps its bell shape. Thus the surface overlapping improvement is equally distributed on all the test sample, in contrast to the DC improvement that seems limited on those tumour volumes already well segmented.

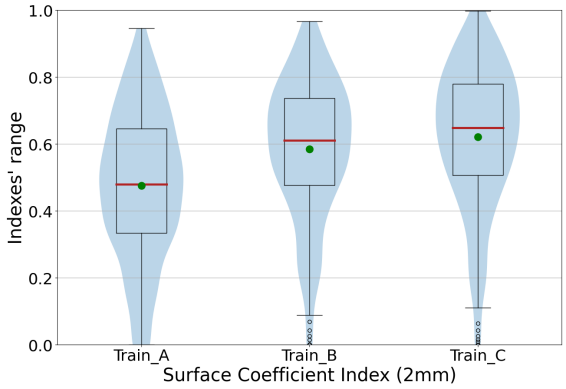
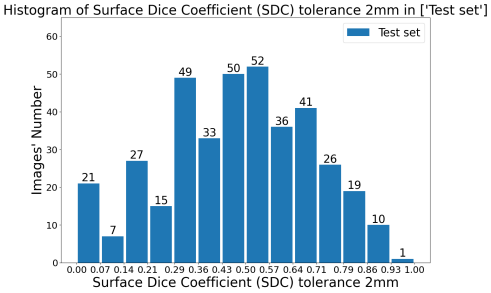


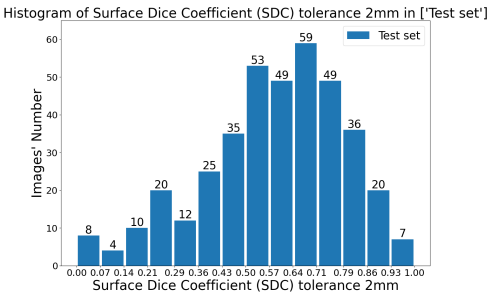
Figure 4.10: The box-violin plot of the Surface Coefficient Index (2mm) (SDC_2mm) is showed for each Train. The mean is the green dot and the median is the red line.

Train ID	Mean	Median	Low Median	High Median
Train_A	0.48 ± 0.22	0.48	0.33	0.65
Train_B	0.59 ± 0.20	0.61	0.47	0.74
Train_C	0.62 ± 0.21	0.65	0.51	0.78

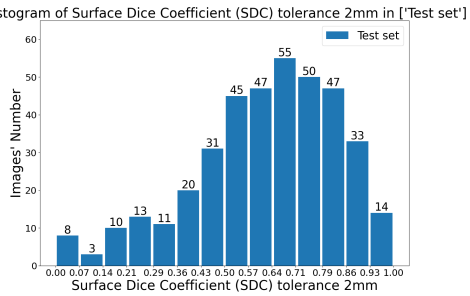
Table 4.6: In table the mean, median, low and high median of the Surface Coefficient Index (2mm) (SDC_2mm)) are showed for each Train.



(a) *Train_A*



(b) *Train_B*



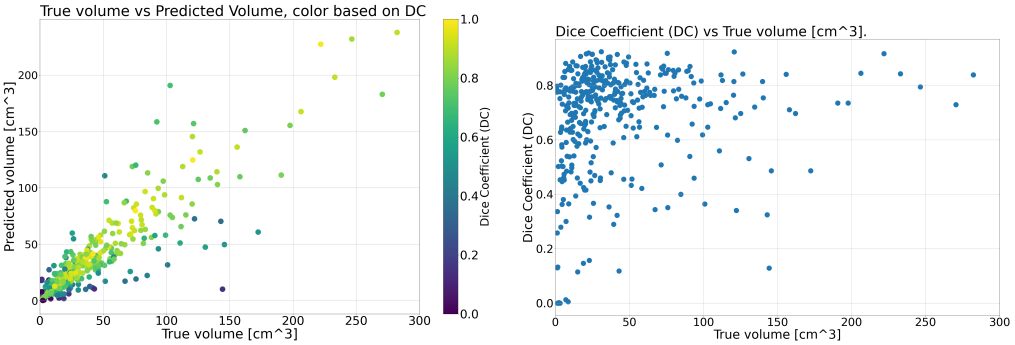
(c) *Train_C*

Figure 4.11: The distribution of Surface Dice Coefficient (2mm) (SDC_2mm) for each Train is showed.

Previous analysis of the DC distributions in the Train showed that some tumor volumes were not segmented despite the changes made to the training. The

Train_C indexes were further investigated to find if a possible explanation could be the tumor size. Two types of scatterplots are used to analyse the indexes. The first one shows the predicted tumor volume as a function of the ground-truth tumor volume, and each dot color is based on the index value (scatterplot (a)). The second one shows the index value plotted on the ground-truth volume (scatterplot (b)). The predicted and ground-truth volumes are in cm^3.

Let's start analysing the DC index. Focusing on the **figure 4.12**, the majority of all the DC small values are concentrated on the region of the small tumor size. In scatterplot (a) the darker blue dots represent samples that scored a very low DC value. These are mainly clustered when the ground-truth volume is small, instead at the increasing of the tumor size the dots get less blue and more yellow. In scatterplot (b) this trend is even more evident.

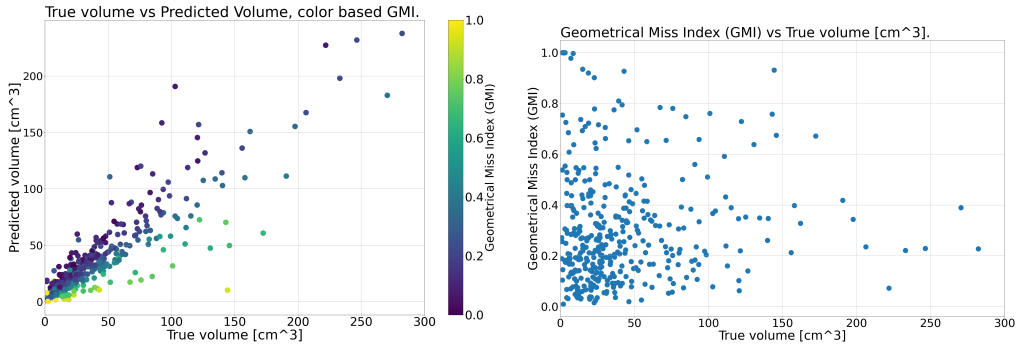


(a) (b)

Figure 4.12: Dice Coefficient (DC) scatterplots are showed.

This trend is present even on the analysis of the DI and GMI, although it is important to remember that for these indexes the best case is when they are close to zero.

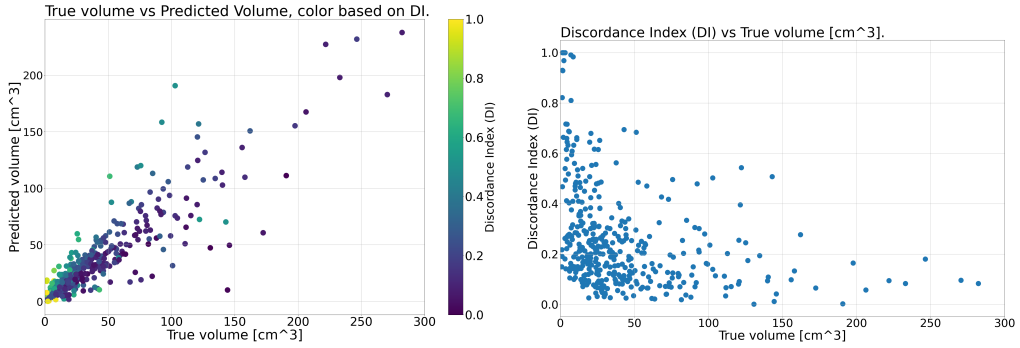
In **figure 4.13** that shows the GMI distribution, the yellow dots (worst case) are still clustered in the small tumor volume graphic area. And moving further in the crescent x-axes direction the dots gets darker and darker improving the GMI values (scatterplot (a)). Moreover it can be seen that in those samples, in which the predicted volume are larger than the ground-truth volume, the GMI values are better, and the other way around if the predicted tumor volume is smaller. Observing the scatterplot (b) the previous trend is still present, but it is less evident than the DC case.



(a) (b)

Figure 4.13: Geometrical Miss Index (GMI) scatterplots are showed.

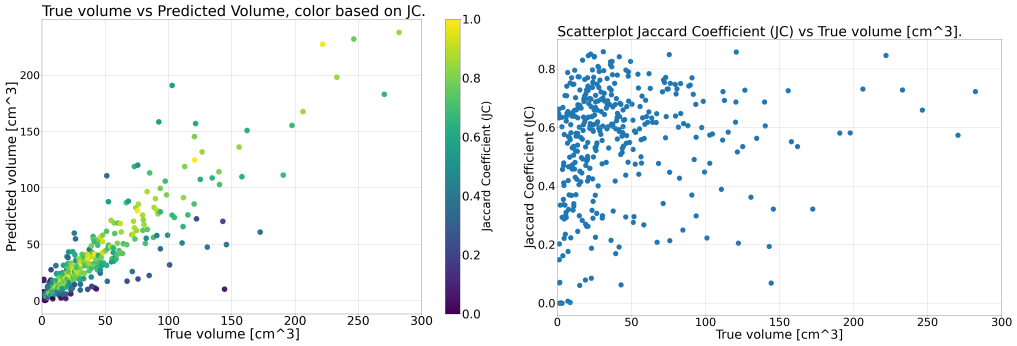
In **figure 4.14** the DI index presents the previous pattern associated with the tumor volume growth and in a more pronounced way than the GMI index, especially in scatterplot (b). However in scatterplot (a), it is possible to observe that the DI index value tends to be better when the predicted tumor volume is smaller than the ground-truth volume. Hence it is the opposite respect to the GMI index.



(a) (b)

Figure 4.14: Discordance Index (DI) scatterplots are showed.

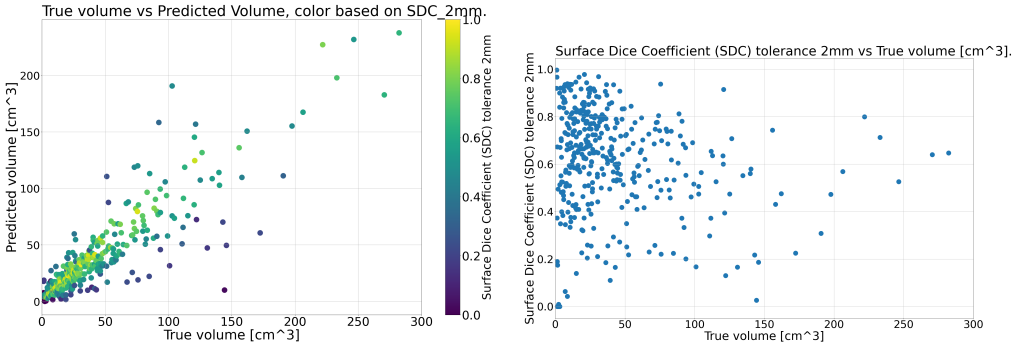
In **figure 4.15** the analysis of the Jaccard Coefficient is presented. As said before, this index is very similar to the DC, because they both measure the similarity between volumes as overlapping of the volumes. In fact all the considerations made on the DC scatterplots are confirmed by the JC scatterplots.



(a) (b)

Figure 4.15: Jaccard Coefficient (JC) scatterplots are showed.

The analysis of the Surface Dice Coefficient is quite interesting. Observing scatterplot (b) in **figure 4.16**, it is not so easy to say that increasing the tumor volume the SDC_2mm value increases. It is meaningful to analyse further this aspect to understand the relationship between the volume segmentation and this index that represents an evaluation on the surface segmentation. In future studies of the algorithm this parameter shall be checked in depth.



(a) (b)

Figure 4.16: Surface Dice Coefficient (SDC) tolerance 2mm scatterplots are showed.

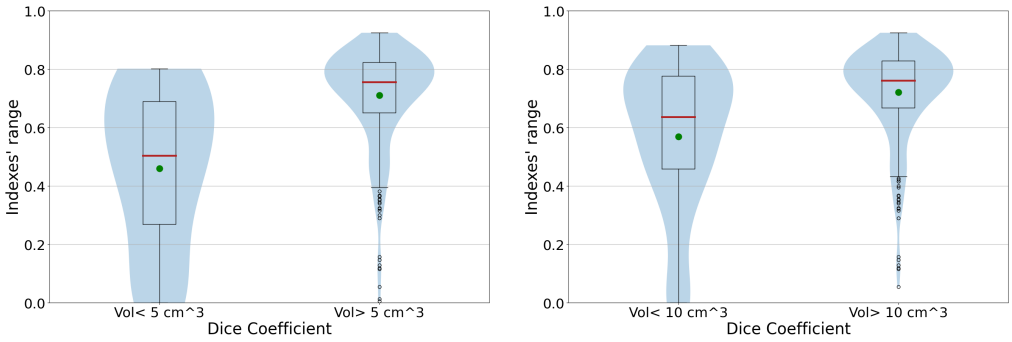
In order to investigate the pattern seen in the analysis of indexes distributions, the scored Train_C DC values were further studied. The idea is to use again the same box-violin plot type, but the DC values were divided in two groups based on the ground-truth volume size. And then, as done before, the distribution of these two groups is analysed to better understand the pattern. The two test samples batches are built choosing a particular volume size. The first group is formed by those samples that have a tumor volume inferior to the chosen threshold, the remaining ones compose the second batch. The chosen thresholds are 5 cm^3, 10

cm^3, 15 cm^3, 25 cm^3 and 50 cm^3. The results are proposed in **figure 4.17** and **table 4.7**.

Let's start with the smallest threshold of 5 cm^3 in figure 4.17 (a). The two distributions are not very similar, the group of the volumes under the threshold is spread along almost all the index range, in fact the light blue area is thick even in the distribution tail and the low median is 0.26 and the high median is 0.69. The DC values range from 0 to 0.8. Instead the volumes group over the threshold is much more clustered around an high DC value, its median is 0.76, while the low and high median are respective 0.65 and 0.82. If the threshold is increased to 10 cm^3 (b), bigger volumes enter the lower group, then the latter distribution changes. Now there is a higher peak on the high DC values, so the volumes entered scored with higher than low DC values. In fact the distribution tail gets thinner, and the best DC value in the group is bigger than 0.80, while the other distribution remains almost the same. When the threshold is 15 cm^3 (c) the smaller volume size distribution gets a higher median and low median than before, while the high median remains the same. This process goes on even when the threshold is 25 cm^3 (d) until the two distributions are almost the same, when the threshold is 50 cm^3 (e), because there are much more samples with bigger volume than 5 cm^3.

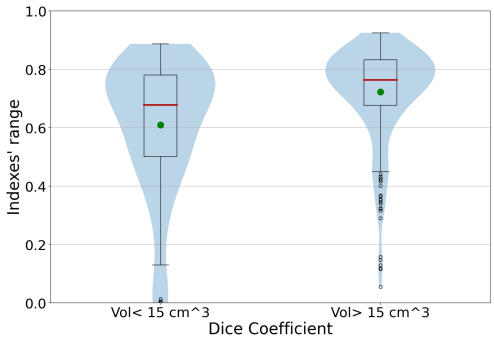
Volumes	**Images**	**Mean**	**Median**	**Low Median**	**High Median**
Vol $<$ **5cm**^{3}	27	0.46 ± 0.27	0.50	0.26	0.69
Vol $>$ **5cm**^{3}	360	0.71 ± 0.17	0.76	0.65	0.82
Vol $<$ **10cm**^{3}	71	0.57 ± 0.25	0.64	0.45	0.78
Vol $>$ **10cm**^{3}	316	0.72 ± 0.16	0.76	0.67	0.83
Vol $<$ **15cm**^{3}	100	0.61 ± 0.23	0.68	0.50	0.78
Vol $>$ **15cm**^{3}	287	0.72 ± 0.16	0.76	0.67	0.83
Vol $<$ **25cm**^{3}	175	0.65 ± 0.21	0.72	0.54	0.80
Vol $>$ **25cm**^{3}	212	0.73 ± 0.16	0.76	0.69	0.83
Vol $<$ **50cm**^{3}	282	0.69 ± 0.20	0.75	0.63	0.82
Vol $>$ **50cm**^{3}	105	0.71 ± 0.17	0.76	0.66	0.83

Table 4.7: Train_C Dice Coefficient (DC) mean, median, low and high median are showed for both volume groups for each volume threshold.

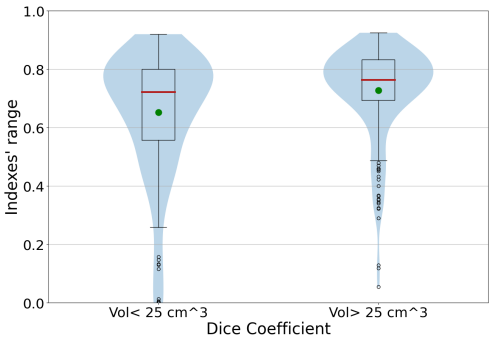


(a) *threshold 5 cm^3*

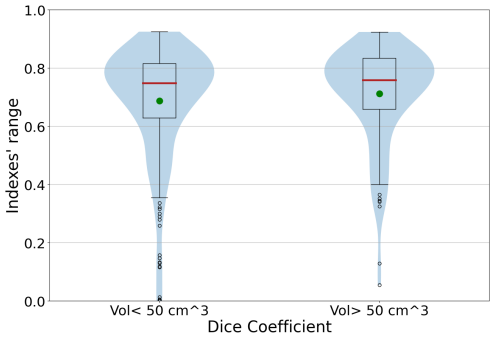
(b) *threshold 10 cm^3*



(c) *threshold 15 cm^3*



(d) *threshold 25 cm^3*



(e) *threshold 50 cm^3*

Figure 4.17: Dice Coefficient (DC) box-violin plots are showed for both volume groups for each volume threshold.

This means that the trained NN has a harder time segmenting small tumors than large ones, but not all the small tumors, only a part. Because as seen in figure 4.12 and 4.17 many small tumor volumes scored high DC. Hence there should be a characteristic in small tumors that determines if the volume will score a high or a small DC value. This parameter is interesting to improve the performances of the segmentation. By understanding the reason behind the poor segmentation of some of the smaller volumes it could be possible to add a segmentation step or refine the parameters to improve the nnU-Net ability to segment H&N tumours.

4.2 Segmented images

Terminated the indexes analysis, some examples of segmented images in the test set are shown. Given an input patient CT image, the same image segmented by the NN of Train_A, Train_B and Train_C is proposed.

Patient HN_P035_5

The first image (**figure 4.18**) belongs to the patient HN_P035_5. The figure shows the segmentation produced by the NN in Train_A. First of all this image is bigger than the other two (figure 4.19 and 4.19), because it is not cropped. Second, observing the green area that represents the predicted tumor it is evident that it is completely out of the red contour($DI = 1$), the ground-truth volume. So the NN contoured healthy tissue instead of the tumor ($GMI = 0$). In fact the similarity indexes values on this images are: $DC = 0$, $JC = 0$ and $SDC\_2mm = 0$, which means a complete failure.

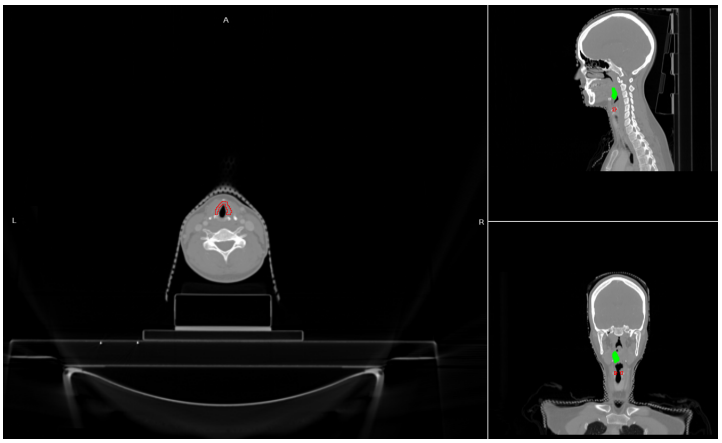


Figure 4.18: In the image is showed the segmented CT of the patient HN-P035_5 by the NN in Train_A. The red contour is the ground-truth volume, while the red area is the predicted volume. The segmentation indexes are $DC = 0$, $GMI = 1$, $DI = 1$ $JC = 0$ and $SDC\_2mm = 0$.

Instead, observing the same image segmentation produced by the Train_B (figure 4.19), there is a substantial improvement with respect to the previous one, now all the predicted volume is inside the ground-truth contour ($DI = 0.07$), but there is still margin of improvement, because there are still some areas not predicted ($GMI = 0.68$). The general similarity of the two volumes is described by: $DC = 0.58, JC = 0.33$ and $SDC\_2mm = 0.71$.

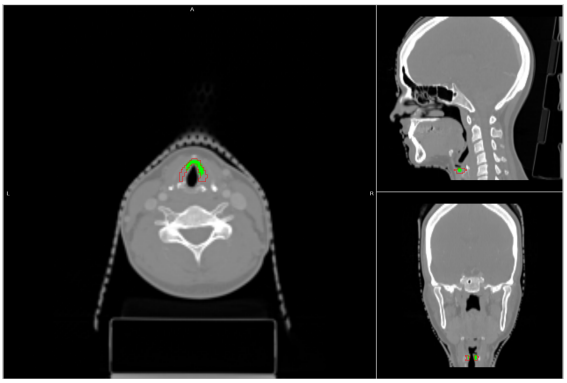


Figure 4.19: In the image is showed the segmented CT of the patient HN-P035_5 by the NN in Train_B. The red contour is the ground-truth volume, while the red area is the predicted volume. The segmentation indexes are $DC = 0.58$,$DI = 0.07$, $GMI = 0.68$, $JC = 0.33$ and $SDC\_2mm = 0.71$.

Lastly the tumor segmentation in **figure 4.20** is produced by the Train_C based on the same CT image. Now even more ground-truth volume is correctly predicted ($GMI = 0.26$), but increases the healthy tissue identified as malignancies ($DI = 0.19$), anyway the two volumes resemblance improved: $DC = 0.78$, $JC = 0.63$ and $SDC\_2mm = 0.91$.

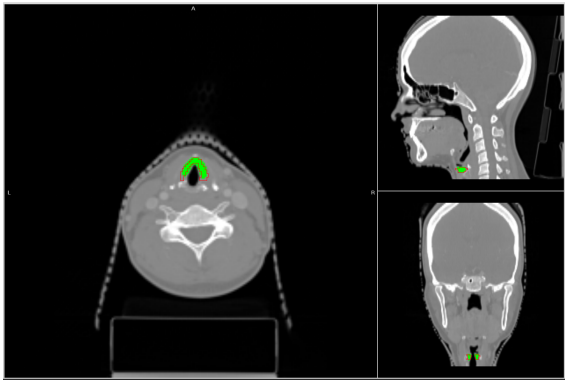


Figure 4.20: In the image is showed the segmented CT of the patient HN-P035_5 by the NN in Train_C. The red contour is the ground-truth volume, while the red area is the predicted volume. The segmentation indexes are $DC = 0.78$, $GMI = 0.26$, $DI = 0.19$, $JC = 0.63$ and $SDC\_2mm = 0.91$.

Patient HNSCC-01-0425_0

The patient image is a particular one, the **figure 4.21** has the original figure size and it did not need to be cropped. Maybe for this reason the NN trained by the Train_A obtained a great result. The majority of the predicted tumor volume (light blue area) is inside the red ground-truth volume contour ($DI = 0.17$), and also it recognizes a large part of the tumor ($GMI = 0.14$). The very good performance is also proved by the indexes values: $DC = 0.84$, $JC = 0.73$ and $SDC\_2mm = 0.83$.

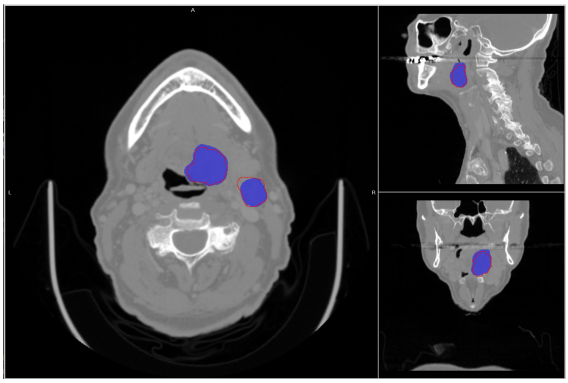


Figure 4.21: In the image is showed the segmented CT of the patient HNSCC-01-0425_0 by the NN in Train_A. The red contour is the ground-truth volume, while the red area is the predicted volume. The segmentation indexes are $DC = 0.84$, $GMI = 0.14$, $DI = 0.17$, $JC = 0.73$ and $SDC\_2mm = 0.83$.

The Train_B NN attains even a better results (**figure 4.22**), and it represents its best performance in all the tests set: $DC = 0.91$, $GMI = 0.05$, $DI = 0.10$, $JC = 0.85$ and $SDC\_2mm = 0.97$. This is an almost perfect segmentation.

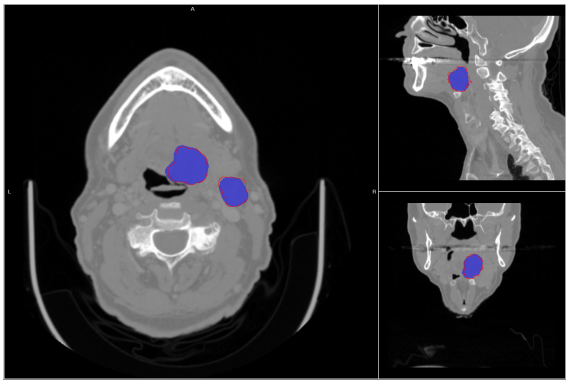


Figure 4.22: In the image is showed the segmented CT of the patient HNSCC-01-0425_0 by the NN in Train_B. The red contour is the ground-truth volume, while the red area is the predicted volume. The segmentation indexes are $DC = 0.91$, $GMI = 0.05$, $DI = 0.10$, $JC = 0.85$ and $SDC\_2mm = 0.97$.

Lastly this patient image (**figure 4.23**) is also where the Train_C NN reached its best performance on the entire test set. There is almost no overcontouring

($DI = 0.09$), and almost all the tumor volume is identified ($GMI = 0.05$). The similarity of the two volumes is proved by $DC = 0.92$, $JC = 0.85$ and $SDC\_2mm = 0.96$.

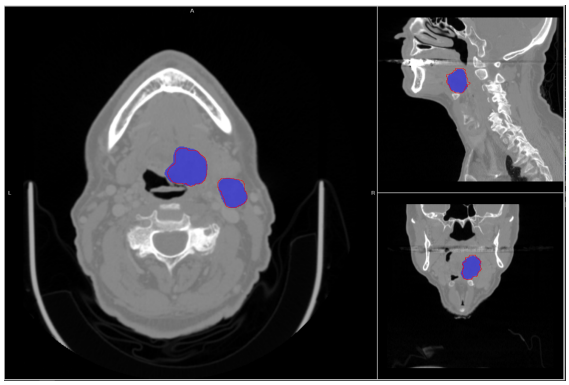


Figure 4.23: In the image is showed the segmented CT of the patient HNSCC-01-0425_0 by the NN in Train_C. The red contour is the ground-truth volume, while the red area is the predicted volume. The segmentation indexes are $DC = 0.92$, $GMI = 0.05$, $DI = 0.09$, $JC = 0.85$ and $SDC\_2mm = 0.96$.

If the automatic segmentation methods will reach this performance level, they will be a priceless tool for the clinicians and for the researchers, in BNCT and potentially in other types of radiotherapy.

4.3 Dosimetry: preliminary considerations

Up to this point we have evaluated the tumor mask inference from a geometrical point of view, by comparing it to the ground truth. Although this was not the objective of this thesis, we propose in this section to evaluate the most performing network (the one with the highest mean Dice Coefficient) from a dosimetric point of view. Our focus here is to generate from the CT image a geometry which can be implemented in a Monte Carlo simulation toolkit to perform a Boron Neutron Capture Therapy treatment plan. This process is not trivial and it consist in these steps:

Mask: Generating a mask of the CT image containing different tissue types.

Mesh: The 3D mask has to be transformed in a tetrahedral mesh.

TPS: For the Treatment Planning Simulation beam direction an patient position has to be optimized.

Dose: The treatment evaluation is performed by selecting the organ at risk and computing the dose delivered to the GTV.

One of the aims of the AI_MIGHT project is to develop an automatic procedure to perform those steps. This goal is out of the scope of the work presented in this manuscript. My contribution was to evaluate the critical aspects that prevent

automatising those steps. First of all, a mask is automatically generated from the CT scan by selecting materials based on HU values. To simplify this step we decided to select soft-tissue as all those voxels with HU in range (-200,200), bone as those voxels with HU greater than 200 and the rest as air, this is shown in Figure 4.24.

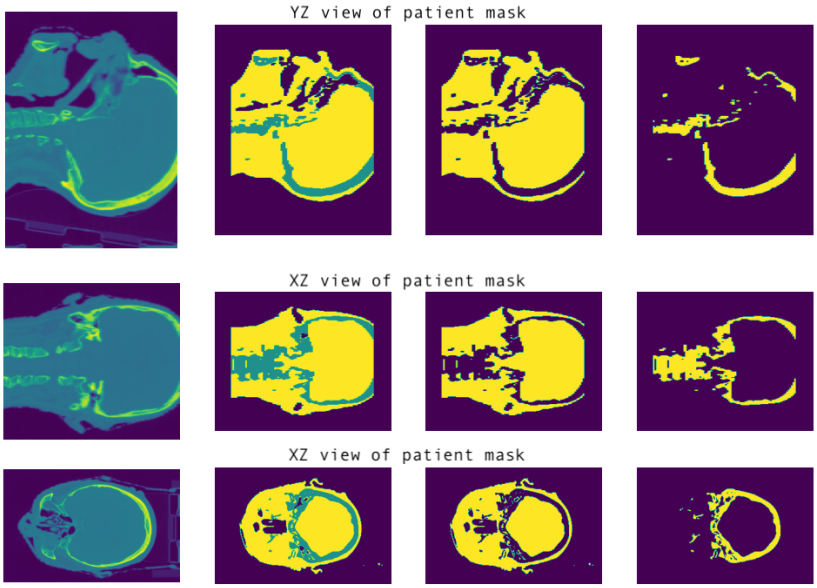
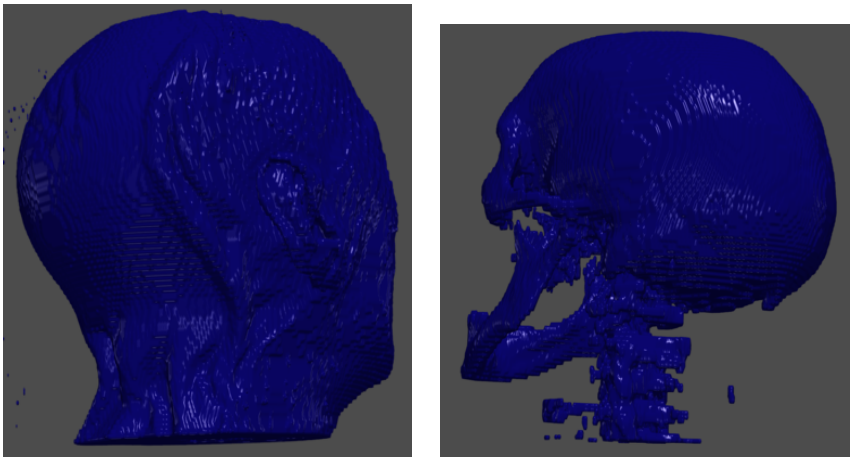


Figure 4.24: From left to right: CT image, masked image with bone and soft-tissue, soft tissue mask and bone mask. While from top to bottom the YZ, XZ and XY central slices are plotted.

In the second step the mask has to be manipulated and transformed into a tetrahedral mesh. This is a multi-step process which starts by applying the scikit-image marching cubes algorithm [99]. This algorithm is able to compute the iso-surface which contains all the voxels with a predefined value, the output of this process is shown in Figure 4.25. These surfaces can subsequently be transformed into tetrahedral mesh volumes by using the TetGen algorithm [100]. The image of the tetrahedrical reconstruction of the volume can not be shown in this thesis to be respectful of the patient anonymity but the reconstruction was performed successfully. Nonetheless, this process is error prone, since a slight overlap of the marching cube geometries might lead to problems during TetGen execution or later on during the Monte Carlo simulation of the BNCT treatment.

Unfortunately only one of the selected cases managed to pass the first two stages. Consequently, we decided to evaluate a simple treatment scenario where the patient is positioned in front of a realistic clinical neutron source, simulating the BNCT treatment with a single fraction. This is not a realistic situation, therefore, we decided to limit the observation of the minimum, mean and maximum absorbed dose rate ratio between predicted GTV volume and the ground truth GTV volume. The results are shown in Table 4.8, errors are not reported since



(a) (b)

Figure 4.25: The image shows the surface extracted by the marching cubes algorithm for soft tissue (left) and bone (right).

the simulations were performed to obtain a voxel value uncertainty of 1%. The result has no statistical significance, since it is computed on a single case. Anyhow, for this case the dosimetry difference of the predicted GTV volume is very low on the min absorbed dose value. Which is a promising result since tumor control probability significantly depends on the minimum dose to the tumor.

Table 4.8: Absorbed dose rate ratio comparison between ground truth and predicted GTV for the minimum, mean and maximum dose values.

$\frac{D_{min}(GTV_{predicted})}{D_{min}(GTV_{true})}$	$\frac{D_{mean}(GTV_{predicted})}{D_{mean}(GTV_{true})}$	$\frac{D_{max}(GTV_{predicted})}{D_{max}(GTV_{true})}$
0.98	0.92	0.86

Chapter 5

Conclusion

The aim of this thesis is using Deep Learning (DL) methods to automatically segment the Head and neck (H&N) tumors in CT images. The objective is to provide a more precise and more individualized BNCT treatment to the patients. The automatic segmented images are without human induced variability and could be produced in considerably lower time than the manual segmented ones. This tool has the potential to reduce the radiologist workload of manually contouring organs and tumors in the clinical application of BNCT. In the preclinical phase, it is of great help for the researchers comparing different TP, testing or improving BNCT TPS or evaluating the quality of new imaging systems thanks to the independence of the results from the segmentation used as basis in the workflow. To achieve the goal it is necessary a solid and standardized database, that is used to train the DL method.

A large quantity of medical images (DICOM format) are available on public databases of H&N cancer. However, there are different types of diagnostic scan such as MR, PET or CT, and other information used during the treatment planning, for example the image manual segmentation (RTSTRUCT file) or the dose distribution produced by the TPS (RTDOSE). All of this information needs to be rationalized and organized in order to create a solid database. The first step consisted in building a system for analysing the image characteristics: it extrapolated the image characteristics directly from the image header, and it linked the diagnostic images to the respective information produced during the TP pipeline.

In this preliminary study it was decided to select only the scan images which were CT and have a linked RTSTRUCT, without taking into account any other parameters. In particular, the Gross Tumor Volume (GTV) was the fundamental segmented part which was needed for the training of the DL methods. The final database was composed by 1934 CT images with GTV mask, where 713 CT images have both the GTV mask and the dose distribution. If the Lymph Node GTV was present, it was fused with the GTV, forming a single tumor volume. This decision was made to consider every possible useful information about the

volume that needed to be treated. The complete database was not only useful for this preliminary study, but also for the following ones.

The original images contained not only the H&N body area, but also many other information not useful for our goal, such as the background, the patient bed and body parts not relevant for our analysis. Using the image in its original size would have demanded excessive computational resources and an increase of computational time need for learning, without any gains. Therefore a Boundary Box (BB) algorithm was adapted to identify only the meaningful body parts for this project and to crop the unnecessary image parts. The use of cropped image not only sped up significantly the training time, but also made the DL method learning process easier.

The DL method used was the nnU-Net: a deep learning-based segmentation method that automatically configures itself to the specific database, that is the state-of-the-art in medical segmentation field. In order to train and to evaluate the DL method, the dataset was split in a train set (80% of samples) and in a test set (20% of the samples). Three training sessions were configured.
The first one (Train_A) used as training images the original size images, and its result was used as a performance benchmark for the other two sessions. The training of one of the 5 fold cross-validation lasted more than 12 days, so it was decided to train only one configuration for the 3D U-Net full resolution.
The second training section (Train_B) was done using the cropped images as training samples. Terminated the training, the best Neural Network (NN) architecture resulted the 3D U-Net full resolution. The computational time needed for the training of one cross-validation configuration was a mean of 20 hours.
In the last training session (Train_C) it was decided to increase the number of epochs from 1000 to 2000, using the best NN architecture of the second training session, i.e. 3D U-Net full resolution.

The performance of each training was evaluated on the test set. The Dice Coefficient (DC) was calculated for each session. Train_A obtained $DC = 0.6 \pm 02$, Train_B scored $DC = 0.7 \pm 02$ and Train_C $DC = 0.7 \pm 02$. Cropping the images produced a considerably improving of the NN performance. The DC distribution in the test set showed how the distribution shift toward the higher DC values in Train_B, and even more in Train_C, respect to Train_A. This also indicated how much the similarity between the predicted volume and the ground-truth tumor volume increased. However, the change of the Geometrical Miss Index (GMI) distribution in the three training section, only demonstrated a slightly improvement. This means that cropping the images and adding more epochs did not produce a qualitative change in the NN ability of segmenting the tumor. Instead, focusing on the Discordance Index (DI) distribution across the training sections, there was a consistent improvement with respect to Train_A. So the NN improved at recognizing what is not a tumor, and this, in turn, caused the improvement in the

DC values but not in the GMI values. Moreover, the change of the distribution of the Surface Dice Coefficient with tolerance 2mm (SDC_2mm) across the train section, proved an improvement. The increase of the SDC_2mm value seems to be uniform on all the test set samples, because the SDC_2mm distribution shifts toward higher values, while keeping its shape.

Investigating the relationship between the scored DC values and the ground-truth volume a pattern seems to emerge. The bigger the tumor volume gets, the bigger the DC value is scored. Moreover the majority of the low DC values are clustered in the small tumor volume area, even if many get a high DC value anyway. Thus the NN tendency is to segment in a better way the bigger tumor volumes, because a part of the small tumor volume have some characteristics, that prevent a good segmentation. This pattern was confirmed by all the other indexes, but SDC_2mm. It's distribution as function of the tumor volume neither confirmed or denied the pattern, but it created interest to study it further in the future.

In the next studies it is necessary to investigate why some small tumor volumes are not segmented well. Moreover it should be investigated why the GMI values do not increase by much even after cropping the images and doubling the epochs. Both of this hurdles need to be addressed searching the tumor characteristics that prevent the improvement of the NN segmentation. This could be useful in creating specific datasets on which the NN could be trained. Another option could be the use of image pre-processing trying to emphasise the tumor volume or further reduce aspects of the image not useful for the tumour segmentation. In this way the DL method performance could improve.

However, after the nnU-Net training on our database, the NN produced excellent results, as this preliminary work provided a first tool able to automatically segment H&N tumors with very good similarity with the ground-truth volume. In order to further study and evaluate the tool results, other analyses could be considered. For example, comparing the dosimetry produced by the TPS using as input both the automatic and manual segmentation, could give information about the grade of accordance needed between the two segmentation masks to obtain similar therapeutic outputs. Moreover, having the possibility to use the experience of a panel of radiologists could provide knowledge on the inter-variability of the segmentation in a clinical setting and open the possibility to compare the automatic segmentation results with a range in which the ground-truth could be identified. In the long range, the possibility to carry out retrospective studies using the same segmentation method could help in understanding better the relation dose-response in BNCT. Lastly, the possibility to compare and combine the dosimetric information obtained by the automatic segmentation and the application of BNCT and the dosimetry based on photon radiotherapy could give us insights on the improvement of BNCT and future steps in broadening the

therapeutic possibilities for Head and neck cancer and other tumours.

Bibliography

[1] Jeffrey A Coderre et al. "Boron neutron capture therapy: cellular targeting of high linear energy transfer radiation". In: *Technology in cancer research & treatment* 2.5 (2003), pages 355–375. DOI: 10.1177/153303460300200502 (cited on pages 1, 2, 6).

[2] Wolfgang AG Sauerwein et al. *Neutron capture therapy: principles and applications*. Springer Science & Business Media, 2012 (cited on pages 1, 5, 8, 12).

[3] Rolf F Barth, Peng Mi, and Weilian Yang. "Boron delivery agents for neutron capture therapy of cancer". In: *Cancer Communications* 38.1 (2018), pages 1–15. DOI: 10.1186/s40880-018-0299-7 (cited on page 2).

[4] Rolf F Barth et al. "Boron neutron capture therapy of cancer: current status and future prospects". In: *Clinical Cancer Research* 11.11 (2005), pages 3987–4002. DOI: 10.1158/1078-0432.CCR-05-0035 (cited on pages 2, 3, 5).

[5] Minoru Suzuki. "Boron neutron capture therapy (BNCT): A unique role in radiotherapy with a view to entering the accelerator-based BNCT era". In: *International journal of clinical oncology* 25.1 (2020), pages 43–50. DOI: 10.1007/s10147-019-01480-4 (cited on page 3).

[6] Sandro Rossi. "The national centre for oncological hadrontherapy (CNAO): status and perspectives". In: *Physica Medica* 31.4 (2015), pages 333–351. DOI: 10.1016/j.ejmp.2015.03.001 (cited on page 2).

[7] Song Wang et al. "Boron Neutron Capture Therapy: Current Status and Challenges". In: *Frontiers in Oncology* 12 (2022). DOI: 10.3389/fonc.2022.788770 (cited on page 3).

[8] Albert H Soloway et al. "The rationale and requirements for the development of boron neutron capture therapy of brain tumors". In: *Journal of Neuro-Oncology* 33.1 (1997), pages 09–18. DOI: 10.1023/A:1005753610355 (cited on pages 3, 5).

[9] Farina Hanif et al. "Glioblastoma multiforme: a review of its epidemiology and pathogenesis through clinical presentation and treatment". In: *Asian Pacific journal of cancer prevention: APJCP* 18.1 (2017), page 3. DOI: 10.22034/APJCP.2017.18.1.3 (cited on page 3).

[10] Shinji Kawabata et al. "Accelerator-based BNCT for patients with recurrent glioblastoma: a multicenter phase II study". In: *Neuro-oncology*

advances 3.1 (2021), vdab067. DOI: 10.1093/noajnl/vdab067 (cited on page 3).

[11] Ian Postuma et al. "A novel approach to design and evaluate BNCT neutron beams combining physical, radiobiological, and dosimetric figures of merit". In: *Biology* 10.3 (2021), page 174. DOI: 10.3390/biology10030174 (cited on pages 4, 5).

[12] L Provenzano et al. "The essential role of radiobiological figures of merit for the assessment and comparison of beam performances in boron neutron capture therapy". In: *Physica Medica* 67 (2019), pages 9–19. DOI: 10.1016/j.ejmp.2019.09.235 (cited on page 5).

[13] JA Coderre et al. "Boron neutron capture therapy of glioblastoma multiforme using the p-borophenylalanine-fructose complex and epithermal neutrons: tryal design and early clinical results". In: *J. Neuro-Oncol.* 33 (1997), pages 141–151. DOI: 10.1007/978-1-4757-9567-7_79 (cited on page 5).

[14] Jeffrey A Coderre and Gerard M Morris. "The radiation biology of boron neutron capture therapy". In: *Radiation research* 151.1 (1999), pages 1–18. DOI: 10.2307/3579742 (cited on page 5).

[15] Sara J González and Gustavo A Santa Cruz. "The photon-isoeffective dose in boron neutron capture therapy". In: *Radiation Research* 178.6 (2012), pages 609–621. DOI: 10.1667/RR2944.1 (cited on pages 5, 13).

[16] SJ González et al. "Photon iso-effective dose for cancer treatment with mixed field radiation based on dose-response assessment from human and an animal model: clinical application to boron neutron capture therapy for head and neck cancer". In: *Physics in Medicine & Biology* 62.20 (2017), page 7938. DOI: 10.1088/1361-6560/aa8986 (cited on pages 5, 13).

[17] Barbara Marcaccio. "From radiobiological experiments to treatment planning in patients: a BNCT dosimetry study". Supervisor: Silva Bortolussi, Co-Supervisor: Sara J. González and Ian Postuma. Master's thesis. University of Pavia, 2021 (cited on page 5).

[18] Erica Simeone. "Studi dosimetrici per la BNCT del Glioblastoma Multiforme con acceleratore". Supervisor: Silva Bortolussi, Co-Supervisor: Ian Postuma. Master's thesis. University of Pavia, 2022 (cited on page 5).

[19] A. H. Soloway, H. Hatanaka, and M. A. Davis. "Penetration of Brain and Brain Tumor. VII. Tumor-Binding Sulfhydryl Boron Compounds". In: *Journal of Medicinal Chemistry* 10.4 (1967). PMID: 6037065, pages 714–717. DOI: 10.1021/jm00316a042 (cited on page 5).

[20] Weilian Yang et al. "Boron neutron capture therapy of brain tumors: functional and neuropathologic effects of blood–brain barrier disruption and intracarotid injection of sodium borocaptate and boronophenylalanine". In: *Journal of neuro-oncology* 48.3 (2000), pages 179–190. DOI: 10.1023/A:1006410611067 (cited on page 6).

[21] HR Snyder, Albert J Reedy, and Wm J Lennarz. "Synthesis of aromatic boronic acids. aldehydo boronic acids and a boronic acid analog of tyro-

sine1". In: *Journal of the American Chemical Society* 80.4 (1958), pages 835–838. DOI: 10.1021/ja01537a021 (cited on page 6).

[22] Hiroshi Fukuda. "Response of Normal Tissues to Boron Neutron Capture Therapy (BNCT) with 10B-Borocaptate Sodium (BSH) and 10B-Paraboronophenylalanine (BPA)". In: *Cells* 10.11 (2021), page 2883. DOI: 10.3390/cells10112883 (cited on page 6).

[23] A Wittig et al. "Mechanisms of transport of p-borono-phenylalanine through the cell membrane in vitro". In: *Radiation research* 153.2 (2000), pages 173–180. DOI: 10.1667/0033-7587(2000)153[0173:MOTOPB]2.0.CO;2 (cited on page 6).

[24] Printip Wongthai et al. "Boronophenylalanine, a boron delivery agent for boron neutron capture therapy, is transported by ATB 0,+, LAT 1 and LAT 2". In: *Cancer science* 106.3 (2015), pages 279–286. DOI: 10.1111/cas.12602 (cited on page 6).

[25] *Current Status of Neutron Capture Therapy*. TECDOC Series 1223. Vienna: INTERNATIONAL ATOMIC ENERGY AGENCY, 2001. URL: https://www.iaea.org/publications/6168/current-status-of-neutron-capture-therapy (cited on pages 6–8).

[26] Andres Juan Kreiner et al. "Present status of accelerator-based BNCT". In: *Reports of Practical Oncology & Radiotherapy* 21.2 (2016), pages 95–101. DOI: 10.1016/j.rpor.2014.11.004 (cited on page 8).

[27] Douglas Jones. *ICRU report 50—prescribing, recording and reporting photon beam therapy*. 1994. DOI: 10.1118/1.597396 (cited on pages 9, 10).

[28] André Wambersie. "ICRU report 62, prescribing, recording and reporting photon beam therapy (supplement to ICRU Report 50)". In: *Icru News* (1999) (cited on pages 9, 10).

[29] WILLIAM Parker and HORACIO Patrocinio. "Clinical treatment planning in external photon beam radiotherapy". In: *Radiation oncology physics: A handbook for teachers and students. Vienna: IAEA* 219 (2005) (cited on page 10).

[30] Elizabeth A Eisenhauer et al. "New response evaluation criteria in solid tumours: revised RECIST guideline (version 1.1)". In: *European journal of cancer* 45.2 (2009), pages 228–247 (cited on page 10).

[31] Chun-Hung Chao et al. "Lymph node gross tumor volume detection in oncology imaging via relationship learning using graph neural network". In: *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer. 2020, pages 772–782 (cited on page 10).

[32] KY Cheung. "Intensity modulated radiotherapy: advantages, limitations and future developments". In: *Biomed Imaging Interv J* 2.1 (2006), e19. DOI: 10.2349/biij.2.1.e19 (cited on page 11).

[33] X-5 Monte Carlo Team. *MCNP—A General Monte Carlo N-Particle Transport Code, Version 5*. 2003 (cited on page 12).

[34] Sea Agostinelli et al. "GEANT4—a simulation toolkit". In: *Nuclear instruments and methods in physics research section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 506.3 (2003), pages 250–303 (cited on page 12).

[35] Tatsuhiko Sato et al. "Features of particle and heavy ion transport code system (PHITS) version 3.02". In: *Journal of Nuclear Science and Technology* 55.6 (2018), pages 684–690 (cited on page 12).

[36] R Zamenhof et al. "Monte Carlo-based treatment planning for boron neutron capture therapy using custom designed models automatically generated from CT data". In: *International Journal of Radiation Oncology* Biology* Physics* 35.2 (1996), pages 383–397 (cited on page 12).

[37] Sara J González et al. "Voxel model in BNCT treatment planning: performance analysis and improvements". In: *Physics in Medicine & Biology* 50.3 (2005), page 441 (cited on page 12).

[38] H Kumada et al. "Development of JCDS, a computational dosimetry system at JAEA for boron neutron capture therapy". In: *Journal of Physics: conference series*. Volume 74. 1. IOP Publishing. 2007, page 021010. DOI: 10.1088/1742-6596/74/1/021010 (cited on page 12).

[39] David W Nigg. "Computational dosimetry and treatment planning considerations for neutron capture therapy". In: *Journal of neuro-oncology* 62.1 (2003), pages 75–86 (cited on page 12).

[40] Tzung-Yi Lin and Yen-Wan Hsueh Liu. "Development and verification of THORplan—a BNCT treatment planning system for THOR". In: *Applied Radiation and Isotopes* 69.12 (2011), pages 1878–1881 (cited on page 12).

[41] Hiroaki Kumada et al. "Verification of nuclear data for the Tsukuba plan, a newly developed treatment planning system for boron neutron capture therapy". In: *Applied Radiation and Isotopes* 106 (2015), pages 111–115 (cited on page 12).

[42] Jiang Chen et al. "Development of Monte Carlo based treatment planning system for BNCT". In: *Journal of Physics: Conference Series*. Volume 2313. 1. IOP Publishing. 2022, page 012012 (cited on page 13).

[43] RO Farias and SJ González. "MultiCell model as an optimized strategy for BNCT treatment planning". In: *Proceedings of the 15th International Congress on Neutron Capture Therapy, Tsukuba, Japan*. 2012, pages 10–14 (cited on page 13).

[44] Loredana G Marcu et al. *Radiotherapy and clinical radiobiology of head and neck cancer*. CRC Press, 2018 (cited on pages 14, 15).

[45] Elisabete Weiderpass and Bernard W Stewart. "World Cancer Report". In: *Cancer research for cancer prevention* (2020) (cited on page 14).

[46] William M Mendenhall et al. "Squamous cell carcinoma metastatic to the neck from an unknown head and neck primary site". In: *American journal of otolaryngology* 22.4 (2001), pages 261–267. DOI: 10.1053/ajot.2001.24820 (cited on page 14).

[47] Teruhito Aihara and Norimasa Morita. "BNCT for advanced or recurrent head and neck cancer". In: *Neutron Capture Therapy*. Springer, 2012, pages 417–424. DOI: http://dx.doi.org/10.1016/j.apradiso.2014.04.007 (cited on pages 14, 16).

[48] Hanna Koivunoro et al. "Boron neutron capture therapy for locally recurrent head and neck squamous cell carcinoma: An analysis of dose response and survival". In: *Radiotherapy and Oncology* 137 (2019), pages 153–158. DOI: 10.1016/j.radonc.2019.04.033 (cited on pages 14, 16, 17).

[49] Laura QM Chow. "Head and neck cancer". In: *New England Journal of Medicine* 382.1 (2020), pages 60–72. DOI: 10.1056/NEJMra1715715 (cited on pages 14–16).

[50] Freddie Bray et al. "Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries". In: *CA: a cancer journal for clinicians* 68.6 (2018), pages 394–424. DOI: 10.3322/caac.21492 (cited on page 15).

[51] Sara Gandini et al. "Tobacco smoking and cancer: a meta-analysis". In: *International journal of cancer* 122.1 (2008), pages 155–164. DOI: 10.1002/ijc.23033 (cited on page 15).

[52] Anil K Chaturvedi et al. "Human papillomavirus and rising oropharyngeal cancer incidence in the United States". In: *Journal of clinical oncology* 29.32 (2011), page 4294. DOI: 10.1200/JCO.2011.36.4596 (cited on page 15).

[53] Per Attner et al. "The role of human papillomavirus in the increased incidence of base of tongue cancer". In: *International journal of cancer* 126.12 (2010), pages 2879–2884. DOI: 10.1002/ijc.24994 (cited on page 15).

[54] Neela Guha et al. "Oral health and risk of squamous cell carcinoma of the head and neck and esophagus: results of two multicentric case-control studies". In: *American journal of epidemiology* 166.10 (2007), pages 1159–1173. DOI: 10.1093/aje/kwm193 (cited on page 15).

[55] K Kian Ang. HM Mehanna. *Head and neck cancer recurrence: evidence-based, multidisciplinary managment*. Thieme, 2012 (cited on page 16).

[56] Itsuro Kato et al. "Effectiveness of BNCT for recurrent head and neck malignancies". In: *Applied Radiation and Isotopes* 61.5 (2004), pages 1069–1073. DOI: 10.1016/j.apradiso.2004.05.059 (cited on page 16).

[57] Leena Kankaanranta et al. "Boron neutron capture therapy in the treatment of locally recurred head-and-neck cancer: final analysis of a phase I/II trial". In: *International Journal of Radiation Oncology* Biology* Physics* 82.1 (2012), e67–e75. DOI: 10.1016/j.ijrobp.2010.09.057 (cited on pages 16, 17).

[58] Teruhito Aihara et al. "First clinical case of boron neutron capture therapy for head and neck malignancies using 18F-BPA PET". In: *Head & Neck: Journal for the Sciences and Specialties of the Head and Neck* 28.9 (2006), pages 850–855. DOI: 10.1002/hed.20418 (cited on pages 16, 18).

[59] Itsuro Kato et al. "Effectiveness of boron neutron capture therapy for recurrent head and neck malignancies". In: *Applied Radiation and Isotopes* 67.7-8 (2009), S37–S42. DOI: 10.1016/j.apradiso.2009.03.103 (cited on page 16).

[60] Minoru Suzuki et al. "Boron neutron capture therapy outcomes for advanced or recurrent head and neck cancer". In: *Journal of radiation research* 55.1 (2014), pages 146–153. DOI: 10.1093/jrr/rrt098 (cited on page 16).

[61] INTERNATIONAL ATOMIC ENERGY AGENCY. *Diagnostic radiology physics: a handbook for teachers and students*. International atomic energy agency, 2013 (cited on page 18).

[62] Yoshio Imahori et al. "Positron emission tomography-based boron neutron capture therapy using boronophenylalanine for high-grade gliomas: part I." In: *Clinical cancer research: an official journal of the American Association for Cancer Research* 4.8 (1998), pages 1825–1832 (cited on page 18).

[63] Yoshio Imahori et al. "Positron emission tomography-based boron neutron capture therapy using boronophenylalanine for high-grade gliomas: part II." In: *Clinical cancer research: an official journal of the American Association for Cancer Research* 4.8 (1998), pages 1833–1841 (cited on page 18).

[64] George W Kabalka et al. "Evaluation of fluorine-18-BPA-fructose for boron neutron capture treatment planning". In: *Journal of Nuclear Medicine* 38.11 (1997), pages 1762–1767 (cited on page 18).

[65] Robert W Brown et al. *Magnetic resonance imaging: physical principles and sequence design*. John Wiley & Sons, 2014 (cited on page 19).

[66] Diego Alberti et al. "Synthesis of a carborane-containing cholesterol derivative and evaluation as a potential dual agent for MRI/BNCT applications". In: *Organic & Biomolecular Chemistry* 12.15 (2014), pages 2457–2467 (cited on page 19).

[67] Yoshihide Hattori et al. "Study on the compounds containing 19F and 10B atoms in a single molecule for the application to MRI and BNCT". In: *Bioorganic & medicinal chemistry* 14.10 (2006), pages 3258–3262 (cited on page 19).

[68] Paola Porcari et al. "In vivo 19F MRI and 19F MRS of 19F labelled boronophenylalanine-fructose complex on a C6 rat glioma model to optimize boron neutron capture therapy (BNCT)". In: *Physics in Medicine & Biology* 53.23 (2008), page 6979 (cited on page 19).

[69] Wenya Linda Bi et al. "Artificial intelligence in cancer imaging: clinical challenges and applications". In: *CA: a cancer journal for clinicians* 69.2 (2019), pages 127–157. DOI: 10.3322/caac.21552 (cited on pages 20–23).

[70] Shigao Huang et al. "Artificial intelligence in cancer diagnosis and prognosis: Opportunities and challenges". In: *Cancer letters* 471 (2020), pages 61–71. DOI: 10.1016/j.canlet.2019.12.007 (cited on pages 20, 22).

[71] Macedo Firmino et al. "Computer-aided detection system for lung cancer in computed tomography scans: review and future prospects". In: *Biomedical engineering online* 13.1 (2014), pages 1–16. DOI: 10.1186/1475-925X-13-41 (cited on pages 20, 22).

[72] Andreas Kaplan and Michael Haenlein. "Siri, Siri, in my hand: Who's the fairest in the land? On the interpretations, illustrations, and implications of artificial intelligence". In: *Business Horizons* 62.1 (2019), pages 15–25. DOI: 10.1016/j.bushor.2018.08.004 (cited on pages 21, 27).

[73] Awais Mansoor et al. "Segmentation and image analysis of abnormal lungs at CT: current approaches, challenges, and future trends". In: *Radiographics* 35.4 (2015), page 1056. DOI: 10.1148/rg.2015140232 (cited on pages 23–25).

[74] Lin H et al. "Deep learning for automatic target volume segmentation in radiation therapy: a review." In: *Quant Imaging Med Surg* vol. 11(12) (2021), pages 4847–4858. DOI: 10.21037/qims-21-168 (cited on page 23).

[75] K. Harrison et al. "Machine Learning for Auto-Segmentation in Radiotherapy Planning". In: *Clinical Oncology* vol. 34 (2022), pages 74–88. DOI: 10.1016/j.clon.2021.12.003 (cited on pages 23, 24, 40, 43).

[76] Yabo Fu et al. "A review of deep learning based methods for medical image multi-organ segmentation". In: *Physica Medica* 85 (2021), pages 107–122. DOI: 10.1016/j.ejmp.2021.05.003 (cited on page 23).

[77] BigdataAILab. *What is Semantic Segmentation, Instance Segmentation, Panoramic segmentation?* https://becominghuman.ai/what-is-semantic-segmentation-instance-segmentation-panoramic-segmentation-3bbb03856c12. Accessed last time 19 Sept 2022. Apr. 2021 (cited on page 25).

[78] Lay Khoon Lee, Siau Chuin Liew, and Weng Jie Thong. "A review of image segmentation methodologies in medical image". In: *Advanced computer and communication engineering technology* (2015), pages 1069–1080. DOI: 10.1007/978-3-319-07674-4_99 (cited on page 25).

[79] Xiangbin Liu et al. "A review of deep-learning-based medical image segmentation methods". In: *Sustainability* 13.3 (2021), page 1224. DOI: 10.3390/su13031224 (cited on page 25).

[80] Ozan Oktay et al. "Evaluation of deep learning to augment image-guided radiotherapy for head and neck and prostate cancers". In: *JAMA network open* 3.11 (2020), e2027426–e2027426 (cited on page 26).

[81] MOHAMED ELGENDY. *Deep Learning for Vision Systems*. Manning, 2020 (cited on pages 26, 28, 31–33, 35, 41, 42).

[82] Jun Ma et al. "Loss odyssey in medical image segmentation". In: *Medical Image Analysis* vol. 71, 102035 (2021). ISSN: ISSN 1361-8415. DOI: 10.1016/j.media.2021.102035 (cited on pages 33, 34).

[83] Aston Zhang et al. *Dive into deep learning*. visited on 02/12/2022. URL: https://d2l.ai (cited on pages 39, 46).

[84] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. "U-Net: Convolutional Networks for Biomedical Image Segmentation". In: *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015* (2015), pages 234–241. DOI: 10.1007/978-3-319-24574-4_28 (cited on pages 44, 45).

[85] Huiyan Jiang, Zhaoshuo Diao, and Yu-Dong Yao. "Deep learning techniques for tumor segmentation:a review". In: *The Journal of Supercomputing* 78.3 (2022), pages 1807–1851. ISSN: 2071-1050. DOI: 10.1007/s11227-021-03901-6 (cited on pages 44, 45).

[86] Özgün Çiçek et al. "3D U-Net: learning dense volumetric segmentation from sparse annotation". In: *International conference on medical image computing and computer-assisted intervention*. Springer. 2016, pages 424–432 (cited on page 45).

[87] Fausto Milletari, Nassir Navab, and Seyed-Ahmad Ahmadi. "V-net: Fully convolutional neural networks for volumetric medical image segmentation". In: *2016 fourth international conference on 3D vision (3DV)*. IEEE. 2016, pages 565–571 (cited on page 45).

[88] Vincent Dumoulin and Francesco Visin. "A guide to convolution arithmetic for deep learning". In: *arXiv preprint arXiv:1603.07285* (2016) (cited on page 46).

[89] Fabian Isensee et al. "nnU-Net: a self-configuring method for deep learning-based biomedical image segmentation". In: *Nature methods* 18.2 (2021), pages 203–211 (cited on pages 47, 49).

[90] *The Cancer Imaging Archive (TCIA)*. https://www.cancerimagingarchive.net/ (cited on page 50).

[91] *Plastimatch*. http://plastimatch.org/ (cited on page 53).

[92] Nobuyuki Otsu. "A Threshold Selection Method from Gray-Level Histograms". In: *IEEE Transactions on Systems, Man, and Cybernetics* 9.1 (1979), pages 62–66. DOI: 10.1109/TSMC.1979.4310076 (cited on page 55).

[93] Calculation clusters Eos. URL: https://matematica.unipv.it/cluster-di-calcolo/ (cited on page 57).

[94] Lee R Dice. "Measures of the amount of ecologic association between species". In: *Ecology* 26.3 (1945), pages 297–302 (cited on page 58).

[95] Christina T Muijs et al. "Consequences of additional use of PET information for target volume delineation and radiotherapy dose distribution for esophageal cancer". In: *Radiotherapy and Oncology* 93.3 (2009), pages 447–453 (cited on page 58).

[96] Lucyna Kepka et al. "Delineation variation of lymph node stations for treatment planning in lung cancer radiotherapy". In: *Radiotherapy and Oncology* 85.3 (2007), pages 450–455 (cited on page 58).

[97] Michael V. Sherer et al. "Metrics to evaluate the performance of auto-segmentation for radiation treatment planning: A critical review". In: *Ra-*

diotherapy and Oncology 160 (2021), pages 185–191. DOI: 10.1016/j.radonc.2021.05.003 (cited on page 58).

[98] Stanislav Nikolov et al. "Deep learning to achieve clinically applicable segmentation of head and neck anatomy for radiotherapy". In: *arXiv preprint arXiv:1809.04430* (2018) (cited on pages 58, 70).

[99] Stefan Van der Walt et al. "scikit-image: image processing in Python". In: *PeerJ* 2 (2014), e453 (cited on page 81).

[100] Si Hang. "TetGen, a Delaunay-based quality tetrahedral mesh generator". In: *ACM Trans. Math. Softw* 41.2 (2015), page 11 (cited on page 81).

Ringraziamenti

In questa ultima sezione vorrei ringraziare tutte quelle persone che mi hanno aiutato e sostenuto in questo periodo complicato che è iniziato tempo fa alla triennale ed è finito in magistrale con questa tesi.

Ringrazio Ian, Setareh e Silva per avermi guidato e motivato in questo progetto, sia nei momenti di difficoltà di tutti i giorni sia negli imprevisti più inaspettati. Ringrazio i miei genitori Roberto e Simonetta per aver creduto in me e avermi dato ogni mezzo per poter raggiungere questo traguardo finale, a prescindere dai miei risultati ottenuti nel tragitto. Senza dimenticare la mia sorellina Maddalena che con la sua esperienza e le sue motivazioni mi ha dato forza nei momenti in cui la mia volontà vacillava.

Devo ringraziare tutte quelle persone che sono entrate nella mia vita a Pavia, accrescendola e apportando qualcosa di nuovo. I disperati, Erica, Gabriella e Matilde, che sono state fide compagne di studio in biblioteca, ma soprattutto di pause. Giorgia, con cui ho condiviso risate e bei momenti. Federico e Agnese, con cui ho riso e ballato alle feste.

Le costanti che rimangono. Elena e Giulia conosciute in periodi diversi, ma entrambe siete diventate come sorelle, una maggiore e una minore. Nonostante la distanza i rapporti non cambiano.

Devo ringraziare anche quelle persone che mi hanno aiutato a raggiungere una tappa intermedia, ma fondamentale, la laurea triennale.

I miei coinquilini Maria e Bartolo, con cui ho trascorso bellissimi momenti. Tutta la Balotta Tour con cui ho condiviso le sfide più dure del percorso. Veronica, con cui ho sempre affrontato discussioni profonde e bellissime gite in montagna, che durano tuttora.

E infine coloro che ci sono sempre quando torno a casa. Tutte le Giovani Marmotte con cui trascorro gite, vacanze e aperitivi, che rendono sempre una gioia ogni rientro a Verona. E le Teste di Miglio, che nonostante non ci vediamo più tanto come un tempo, rimangono sempre le solite belle e passionali Teste di Miglio.