Dipartimento di Fisica

Corso di Laurea Magistrale in Scienze Fisiche

Dosimetric Driven Evaluation of AI-based GTV Segmentation in BNCT Treatment Planning for Head and Neck Cancer

Tesi per la laurea di:

Carolina Ruzzon

Relatrice:

Prof.ssa Silva Bortolussi

Correlatrice:

Dott.ssa Setareh Fatemi

Anno accademico 2025 – 2026

Abstract

Boron Neutron Capture Therapy (BNCT) is a binary, cell-level-selective radiotherapy modality, increasingly investigated for recurrent and locally advanced head and neck (H&N) squamous cell carcinoma, for which conventional re-irradiation is often precluded by cumulative normal-tissue toxicity. Dose calculation in BNCT depends on many factors, one being the identification of the target volume. Errors in the delineation of the Gross Tumor Volume (GTV) propagate directly into errors in the estimated absorbed dose, motivating interest in automatic deep-learning-based segmentation as an aid to manual observer-dependent contouring.

Building on a precedent study on glioblastoma, this thesis extends the comparison between manual and automatic GTV contouring to the anatomically more complex H&N region: a patient cohort, selected to span the full range of achievable segmentation agreement, is processed through a Monte Carlo BNCT treatment-planning pipeline, comparing manual and automatic contours under identical conditions.

The analysis shows that automatic and manual contouring are dosimetrically comparable at the cohort level, but that this conceals substantial patient-specific discordance, driven largely by repositioning rather than by the segmentation mask alone. Segmentation accuracy correlates with dosimetric agreement, and the resulting radiobiological figures of merit require careful clinical interpretation.

These results indicate that automated contouring is a viable component of an H&N BNCT planning pipeline, but that its reliability is inseparable from the broader workflow in which it is embedded.

Contents

Abstract **i**

Introduction **1**

1 Boron Neutron Capture Therapy **5**

1.1 History of BNCT 5

1.1.1 Historical Development 5

1.1.2 BNCT in H&N Cancer 6

1.2 Fundamentals of BNCT 7

1.2.1 Nuclear Reactions 8

1.2.2 Boron Carriers 10

1.2.3 Neutron Beam and Sources 14

1.3 BNCT Dosimetry 18

1.3.1 Mixed Radiation Field and Dose Components 18

1.3.2 The Classical Fixed-Weighting Dose Model 21

1.3.3 The Photon-Isoeffective Dose Model 26

1.3.4 Tumor Control Probability and Normal Tissue Complication Probability 34

1.4 Clinical Experience of BNCT in H&N Cancer 36

1.4.1 Reactor-Based Clinical Trials 36

1.4.2 Accelerator-Based BNCT and Current Clinical Status 37

1.4.3 Dose-Response Relationships and Dose-Limiting Toxicity 39

2 Artificial Intelligence and Deep Learning in BNCT Treatment Planning **41**

2.1 History of Artificial Intelligence and Neural Networks 41

2.2 Artificial Neural Networks 43
2.2.1 The Perceptron and the Multilayer Perceptron 43
2.2.2 Network Depth and Feedforward Propagation 44
2.2.3 Training: Loss Functions and Gradient-Based Optimization . . . 47
2.2.4 Convolutional Neural Networks 50
2.3 Fully Convolutional Networks 51
2.3.1 The U-Net Architecture 52
2.3.2 The nnU-Net 54
2.4 Applications of AI in Cancer Imaging 57
2.4.1 Interobserver and Intraobserver Variability 58
2.4.2 Automated Segmentation Tools: TotalSegmentator and CERR . 60
2.5 The AI_MIGHT Project 62

3 Treatment Planning Systems for Boron Neutron Capture Therapy 64
3.1 General Framework of TPS 64
3.1.1 The ICRU Volume Definitions 65
3.1.2 Dose Reporting and Plan Evaluation 66
3.2 Treatment Planning System in BNCT 67
3.2.1 BNCT Dedicated Treatment Planning Paradigm 68
3.2.2 Patient Geometry Reconstruction 69
3.2.3 Tissue Composition and Boron Concentration Assignment . . . 74
3.2.4 Neutron Source Definition 75
3.2.5 Monte Carlo Dose Calculation and the KERMA Approximation 76
3.2.6 Patient Positioning for Fixed-Beam Facilities 78
3.2.7 Plan Evaluation and Optimization 79
3.2.8 Treatment Planning Systems Developed for BNCT 80
3.3 OpenPINT 82
3.3.1 Motivation and Scope 82
3.3.2 Software Architecture 83
3.3.3 From Segmented Geometry to MCNP Input 83
3.3.4 Automated Patient Positioning 84
3.3.5 Dose-Component Extraction and Biological Weighting 86
3.3.6 Radiobiological Model Database 87
3.3.7 Validation: Analytic and Voxelized Phantom Benchmarks 88

3.3.8 Limitations and Development 91

4 AI-Based Treatment Planning and Dosimetric Analyses for Head and Neck Patients 93

4.1 Patient Cohort 94
4.1.1 Data Provenance 94
4.1.2 Cohort Selection 95
4.1.3 Geometric Evaluation Metrics 96
4.2 Image Segmentation 99
4.2.1 Organ-at-Risk Segmentation 99
4.2.2 MCNP Geometry Generation 101
4.3 Treatment Simulation 103
4.3.1 Four-Arm Simulation Design 103
4.3.2 Patient Positioning 104
4.3.3 Voxel-Size Sensitivity Study 106
4.4 Dosimetric Analysis: Patient HN_0717 108
4.4.1 Two Radiobiological Dose Models 109
4.4.2 Dose Components 110
4.4.3 Dose-Volume Histograms 111
4.4.4 Mucosa-Limiting Irradiation Time 113
4.4.5 Tumor Control Probability and Normal Tissue Complication Probability 115
4.4.6 Configuration Comparison 116
4.4.7 Gamma-Index Analysis 117
4.5 Overall Results 122
4.5.1 Tumor Control Probability 122
4.5.2 RBE-Weighted and Photon-Isoeffective Dose Models 126
4.5.3 Mucosal Isoeffective Dose 129
4.5.4 Gamma-Index Analysis 130
4.5.5 Segmentation Accuracy and Dosimetric Discrepancy 133
4.5.6 Further Geometric and Dosimetric Analyses 134

Conclusions 141

Appendix A **142**

A.1 The Factor G_{ij} 142

Appendix B **144**

B.1 Loss Functions and Optimization Algorithms 144

B.1.1 Loss Functions 144

B.1.2 Gradient-Based Optimization Algorithms 147

B.2 The Backpropagation Algorithm 153

B.2.1 Why Backpropagation Is Needed 153

B.2.2 The Chain Rule 153

B.2.3 Formal Derivation 155

B.2.4 The Algorithm 157

Appendix C **158**

C.1 Derivation of the KERMA Approximation and the Charged-Particle-Equilibrium Condition 158

C.1.1 KERMA and Absorbed Dose 158

C.1.2 The CPE condition 159

C.1.3 Breakdown of CPE 161

C.2 Derivation of the Dose-Rate-to-Time Conversion Under a Dose-Volume Constraint 162

C.2.1 Dose Accumulation 162

C.2.2 From Dose-Volume Constraint to Maximum Irradiation Time . 163

C.2.3 Multiple Simultaneous Constraints 164

Appendix D **165**

D.1 Per-Patient Data 165

Bibliography **169**

Introduction

Head and neck (H&N) is a broad clinical term that encompasses a heterogeneous group of malignant cancers arising in the oral cavity, tongue, gums and lips, larynx, throat (nasopharynx, oropharynx and hypopharynx), salivary glands and paranasal sinuses [1]. Approximately 90% of H&N tumors are histologically classified as squamous cell carcinomas and represent one of the most frequent cancer sites worldwide, with roughly 890000 new cases and 450000 deaths in 2018 alone [2].

These tumors are particularly difficult to treat, not due to especially aggressive biology, but because of the anatomical complexity of the region. This represents one of the most complex challenges in modern oncology: irradiation with traditional techniques exposes the patient to severe and potentially fatal late complications, including temporal lobe necrosis, deep mucosal ulcers, pharyngeal fistulas, and carotid blowout syndrome (CBS) [3], particularly in patients who have already exhausted the dose tolerance of previously irradiated tissue.

In this clinical context, Boron Neutron Capture Therapy (BNCT) has found one of its most consistent areas of investigation [4, 5]. Its effectiveness relies on the selective targeting of tumor cells while minimizing damage to healthy tissues, a selectivity that can only be achieved provided a sufficiently favorable boron concentration ratio is established between tumor and healthy tissue. Because of the boron capture reaction, and the subsequent alpha-decay, the energy released is deposited almost entirely within the cell where the capture event occurs, granting BNCT its characteristic cellular-level selectivity.

Because BNCT dose depends jointly on tissue composition, boron uptake, and neutron penetration depth, the accurate delineation of the Gross Tumor Volume (GTV) and of the surrounding organs at risk (OARs) carries consequences that go beyond the purely geometric mismatch relevant to conventional radiotherapy: a contouring

error propagates directly into an error in the estimated absorbed dose. Manual contouring by an experienced radiation oncologist remains the clinical reference standard, but it is a time-consuming task subject to well-documented observer variability [6]. This has motivated growing interest in automatic segmentation based on convolutional neural networks, and in particular on architectures derived from U-Net [7] and its self-configuring extension nnU-Net [8], which have demonstrated segmentation performance comparable to expert observers across a wide range of H&N structures in conventional radiotherapy. Benchmarked without manual intervention across twenty-three public datasets drawn from international biomedical segmentation competitions, nnU-Net has been shown to match or outperform highly specialized purpose-built solutions in the great majority of tasks, and it has since become a common baseline, and often a production tool, in the medical imaging community. Its applicability to H&N anatomy specifically has been demonstrated at scale: a large collaborative study combining a custom three-dimensional U-Net architecture with a curated dataset of over six hundred clinical computed tomography scans showed that automatically generated OAR contours could match the geometric accuracy of experienced human observers across a wide panel of H&N structures, when validated against independent expert delineations [9].

A precedent directly relevant to the present work applied nnU-Net to the automatic segmentation of the GTV in glioblastoma multiforme (GBM) patients, and compared the resulting BNCT treatment plans against those obtained from manual contours on a subset of clinical cases [10].

Good dosimetric agreement was found for cases with high geometric overlap between the two segmentations, but the minimum target dose diverged systematically for lower-overlap cases, an effect attributed to the limited size of the training cohort and to the strong depth-dependence of the neutron dose profile. The authors identified the introduction of a clinically interpretable radiobiological figure of merit, such as the Tumor Control Probability (TCP), as the necessary next step to establish whether such dosimetric differences are clinically significant.

From this open question the present thesis takes its starting point. The aim of this work is to develop and critically evaluate an automated BNCT treatment planning pipeline for H&N cancer patients, extending the comparison between automatic and

manual contouring already applied to glioblastoma to a more anatomically complex region, while broadening the evaluation itself from geometric and dosimetric indices to radiobiological figures of merit of direct clinical relevance.

Computed tomography images of H&N patients, obtained from The Cancer Imaging Archive [11], are segmented automatically with nnU-Net and compared against expert manual contours using standard geometric indices. Each segmented anatomy is voxelized into a Monte Carlo geometry, and a treatment plan is computed twice per patient, once from manual and once from automatic contours, under identical irradiation conditions using the OpenPINT Treatment Planning System [12], which implements the photon isoeffective dose formalism [13, 14]. The comparison is then extended to the Tumor Control Probability (TCP) and Normal Tissue Complication Probability (NTCP) computed for each plan, in an attempt to determine whether residual segmentation discrepancies translate into clinically meaningful differences in expected outcome.

Chapter 1 lays the physical and radiobiological groundwork for BNCT. It begins with the nuclear reaction and the properties of the boron carriers and neutron sources currently used in the clinic, and then turns to BNCT dosimetry, presenting the models adopted to express the mixed neutron-photon dose in photon-equivalent terms.

Chapter 2 introduces the concepts of Deep Learning and Artificial Neural Networks required to follow the rest of the Chapter, then presents the U-Net architecture and the reasoning behind its use for medical image contouring, closing with a description of how it was trained and applied to delineate the GTV and the OAR in the H&N patients considered in this work.

Chapter 3 addresses Treatment Planning Systems for BNCT. It first compares them with the TPS used in conventional radiotherapy, then reviews the main features of the systems currently available for BNCT, and finally introduces OpenPINT, the system developed by INFN Unit of Pavia and adopted for the calculations of this work.

Chapter 4 presents the clinical cohort together with the simulations, results, and their analysis. Treatment plans obtained from manual and automatic contouring are compared in terms of the resulting dosimetry, and the geometric performance of the network is correlated with the observed dose differences. The Chapter closes by introducing the

Tumor Control Probability and the Normal Tissue Complication Probability as radiobiological figures of merit, used to assess whether the discrepancies between the two contouring approaches are clinically significant.

Chapter 1

Boron Neutron Capture Therapy

Boron Neutron Capture Therapy (BNCT) is a binary radiotherapy modality in which a non-radioactive ^{10}B compound and an external neutron field are combined to generate short-range high-linear energy transfer (high-LET) particles selectively within boron-loaded tissue. Neither component is therapeutically effective on its own: the boron carrier must first accumulate preferentially in tumor cells, and the neutron beam must then deliver an adequate thermal-neutron fluence to the target while limiting the non-selective dose to surrounding tissue, as illustrated schematically in Figure 1.1.

This Chapter lays the physical, radiobiological and clinical groundwork on which the rest of the thesis is built.

1.1 History of BNCT

1.1.1 Historical Development

The conceptual basis of BNCT was first articulated by Locher in 1936, only four years after Chadwick's discovery of the neutron: Locher proposed that selective absorption of neutrons by a non-toxic compound containing boron, lithium, gadolinium or gold introduced into a tumor, followed by irradiation with slow neutrons, could allow localized destruction of malignant tissue [15].

Clinical translation began in the early 1950s at Brookhaven National Laboratory and MIT, where patients with cerebral gliomas were treated with the boron compound sodium borate (Borax) and thermal neutrons from a nuclear reactor [16, 17]. These early trials used first-generation compounds with poor tumor selectivity and purely

thermal, low-penetration beams; results were disappointing and in some cases harmful, and clinical interest in BNCT waned for roughly three decades.

A resurgence began in the 1990s, driven by three parallel developments: the introduction of second-generation boron carriers with markedly improved tumor selectivity[1]; the development of epithermal neutron beams capable of penetrating several centimeters into tissue while minimizing superficial dose; and a more rigorous radiobiological and dosimetric framework for quantifying the mixed radiation field produced by BNCT [18, 19].

The transition from research reactors to compact accelerator-based neutron sources over the past decade and a half has substantially changed the clinical prospects of BNCT, allowing installation within or near hospitals rather than at dedicated nuclear facilities (Section 1.2.3). In 2020, the Japanese Pharmaceuticals and Medical Devices Agency designated and approved an accelerator-based BNCT system (the NeuCure treatment system and dose-calculation engine) together with borofalan (^{10}B, STEBORONINE), a second-generation BPA-based carrier, for unresectable, locally advanced or locally recurrent H&N carcinoma, on the basis of the JHN002 phase II trial [20, 21]: this is the first regulatory approval of BNCT as a routine, rather than purely investigational, cancer treatment worldwide.

1.1.2 BNCT in H&N Cancer

The majority of the modern clinical and dosimetric literature on BNCT has focused on high-grade glioma and glioblastoma multiforme, reflecting both the radioresistance of these tumors to conventional therapy and the historical origin of BNCT as a neuro-oncological treatment. The rationale for BNCT in H&N cancer, however, is at least as compelling.

Unlike glioma, most H&N squamous cell carcinomas are relatively superficial and accessible to epithermal neutron beams without requiring optimization for intracranial targets.

In H&N tumors, boron carriers such as BPA achieve tumor-to-blood and tumor-to-normal-tissue ratios that are comparable to, and in several series better than, those reported in glioma [22]. Moreover, because H&N recurrences after full-dose photon

[1] Sodium borocaptate (BSH) and boronophenylalanine (BPA), further explored in Section 1.2.2.

therapy represent a population for whom no further conventional radiotherapy is an option[2], BNCT could offer a viable re-treatment mechanism, rather than a competing first-line therapy. Nonetheless, ongoing clinical investigation is also exploring BNCT as a first-line treatment in selected cases judged likely to benefit from its dosimetric selectivity, rather than confining its role exclusively to salvage settings [25].

This clinical impasse, together with the biologically (rather than purely geometrically) targeted nature of BNCT, motivates its continued investigation for recurrent, inoperable or previously irradiated H&N cancer, and provides the clinical rationale for the patient-specific dosimetric pipeline developed in this thesis.

1.2 Fundamentals of BNCT

BNCT combines neutron-beam transport with chemically mediated selectivity. In photon therapy, selectivity is achieved predominantly through geometric conformation of the beam and fractionation. In BNCT, some physical selectivity from beam geometry is still required, but selectivity is predominantly determined by the spatial distribution of ^{10}B. The clinically relevant dose is consequently a complex quantity that depends on neutron energy, tissue composition, boron concentration, boron microdistribution, dose rate and irradiation time, features that make BNCT a particularly demanding modality from the perspective of dosimetry, radiobiology and treatment planning [13, 26, 27, 28].

A successful BNCT treatment requires several conditions to be satisfied simultaneously: a therapeutically relevant amount of ^{10}B must reach most clonogenic tumor cells; the boron compound must be minimally toxic and clear sufficiently from blood and dose-limiting normal tissue before irradiation; the neutron beam must deliver an adequate thermal-neutron fluence across the target; and non-selective fast-neutron and photon contamination must remain low enough to preserve a therapeutic advantage [18, 28]. These requirements are interconnected, rather than independent: a harder epithermal spectrum, for example, improves penetration but increases fast-neutron dose, so BNCT optimization cannot be reduced to maximizing a single physical quantity.

[2]Conventional re-irradiation is associated with severe late toxicity, including osteoradionecrosis and soft-tissue necrosis [23, 24].

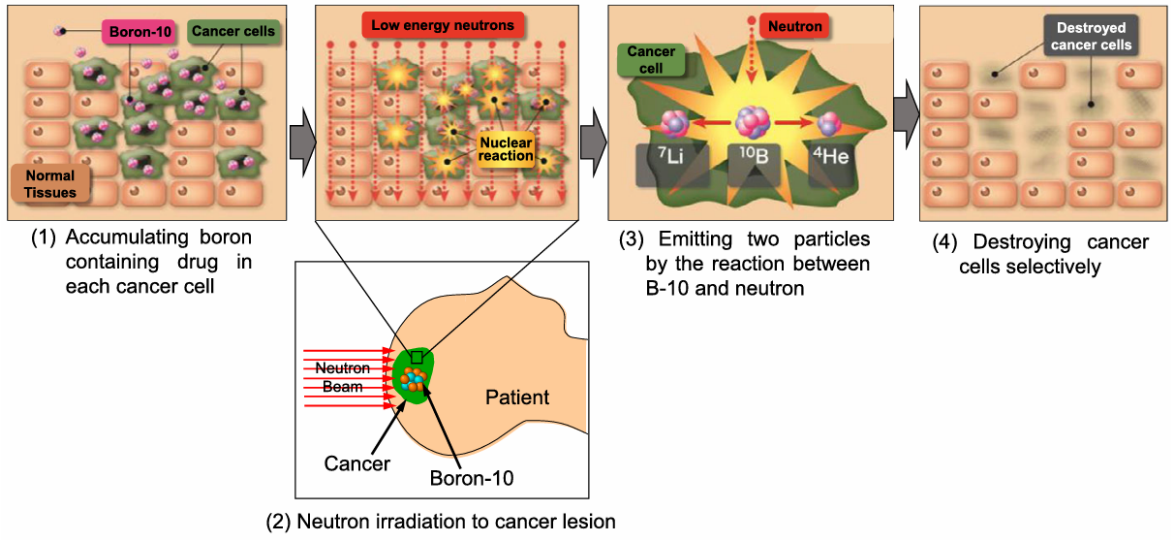


Figure 1.1: *Principle of Boron Neutron Capture Therapy* [35].

1.2.1 Nuclear Reactions

The physical foundation of BNCT is the thermal neutron-capture reaction of the stable isotope ^{10}B, which constitutes approximately 19.9% of natural boron. This isotope exhibits a large cross-section for thermal neutron capture, $\sigma_{\text{capture}} \approx 3837$ barn at a neutron energy of 0.025 eV, several orders of magnitude larger than the capture cross-sections of the major tissue constituents hydrogen ($\sigma \approx 0.33$ barn) and nitrogen ($\sigma \approx 1.7$ barn), as shown in Figure 1.2 [22].

Upon thermal neutron capture, the compound nucleus ^{11}B is formed in an excited state and decays essentially instantaneously via two branches:

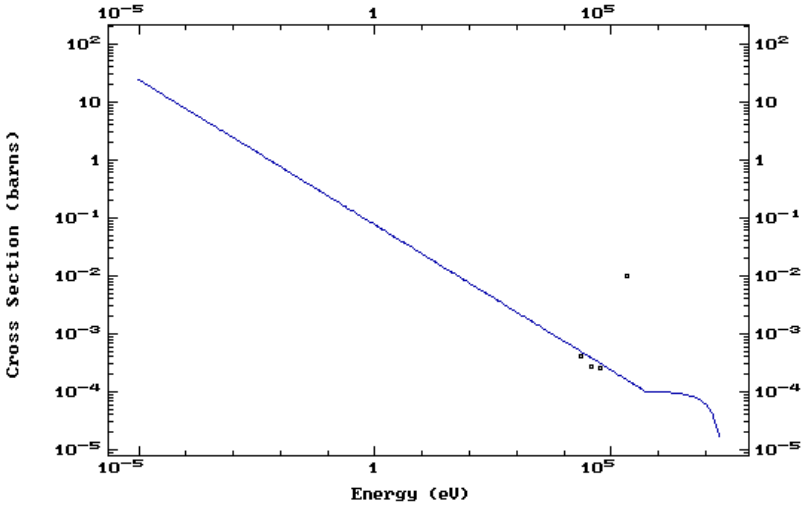
$$^{10}\text{B} + n \longrightarrow {}^{11}\text{B} \longrightarrow {}^{7}\text{Li}^{*} + {}^{4}\text{He} + 2.31\ \text{MeV} \qquad [\text{BR} = 93.9\%] \qquad (1.1)$$

$$^{7}\text{Li}^{*} \longrightarrow {}^{7}\text{Li} + \gamma + 0.478\ \text{MeV}$$

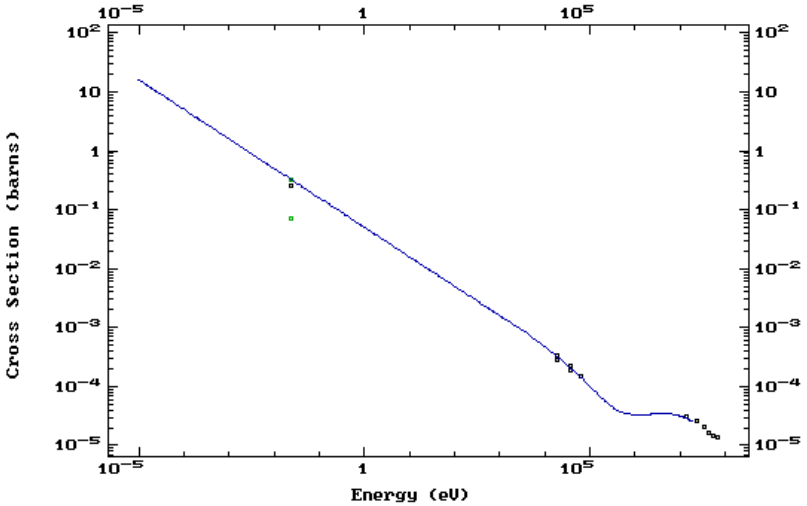
$$^{10}\text{B} + n \longrightarrow {}^{11}\text{B} \longrightarrow {}^{7}\text{Li} + {}^{4}\text{He} + 2.79\ \text{MeV} \qquad [\text{BR} = 6.1\%] \qquad (1.2)$$

In the dominant branch (93.9%), the reaction populates an excited state of ^{7}Li that decays with emission of a 0.478 MeV photon; the total reaction Q-value (2.31 MeV kinetic energy plus 0.478 MeV in the photon, $\approx$ 2.79 MeV) is therefore the same in both branches, consistent with mass-energy conservation.

The reaction products, α-particle and the recoiling ^{7}Li ion, are heavy charged particles with very short ranges in tissue, approximately 9 μm and 5 μm respectively, and



(a) *Neutron capture cross section on* ^{10}B.



(b) *Neutron capture cross section on* ^{1}H.

Figure 1.2: *Radiative neutron capture cross section on (a)* ^{10}B *and (b)* ^{1}H [36].

correspondingly high linear energy transfer, on the order of 150–175 keV/μm [19, 22]. Because these ranges are comparable to or smaller than a single cell diameter[3], essentially all of the kinetic energy released by the reaction is deposited within the cell in which the capture event occurred, or at most in its immediate neighbors. This is the physical basis of the selectivity of BNCT: cytotoxicity is confined to cells that have internalized a sufficient concentration of ^{10}B, while cells lacking boron uptake are relatively spared, independent of the macroscopic geometry of the neutron beam.

It is worth stressing that the reaction is not intrinsically tumor-specific: it is specific

[3]An oral cavity squamous epithelial cell measures approximately 80 μm.

to the spatial distribution of ^{10}B, not to malignant tissue as such. Normal cells that accumulate boron can receive a high local dose, while tumor cells with insufficient boron uptake may escape lethal damage. The microscopic heterogeneity of boron distribution is therefore as dosimetrically important as the macroscopic concentration used in treatment planning, a limitation discussed in Section 1.3.3.

Two further nuclear reactions contribute non-selectively to the total absorbed dose in any tissue exposed to the neutron field, independent of boron content: the radiative capture of neutrons on hydrogen, $^{1}\mathrm{H}(n,\gamma)^{2}\mathrm{H}$, producing a 2.2 MeV photon, and the thermal-neutron capture on nitrogen, $^{14}\mathrm{N}(n,p)^{14}\mathrm{C}$, producing a recoil proton. Together with elastic scattering of fast neutrons on hydrogen nuclei, $^{1}\mathrm{H}(n,n')^{1}\mathrm{H}$, which produces recoil protons of variable energy depending on the incident neutron spectrum, these four physical processes define the mixed radiation field characteristic of BNCT and are treated individually in the dosimetric formalism of Section 1.3. Figure 1.3 collects the relevant cross-section data for the nuclides involved.

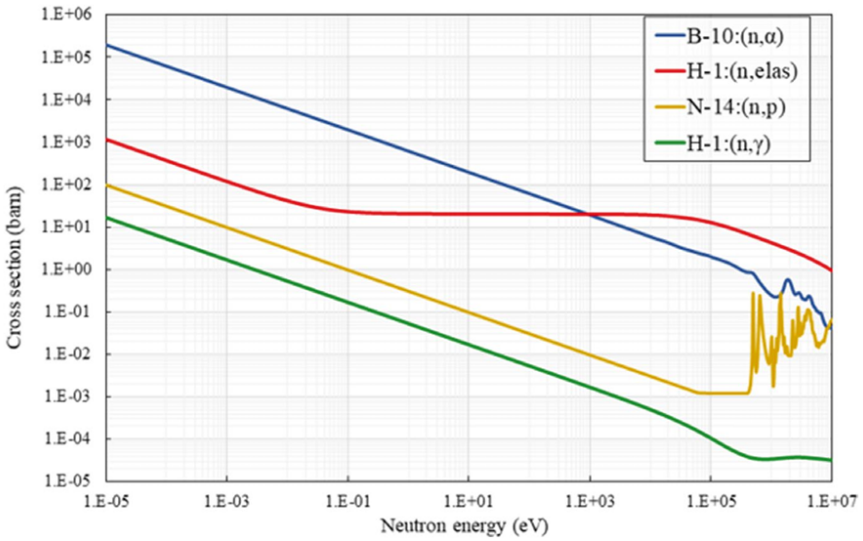


Figure 1.3: *Cross-section data of various nuclides related to the BNCT dose components* [37].

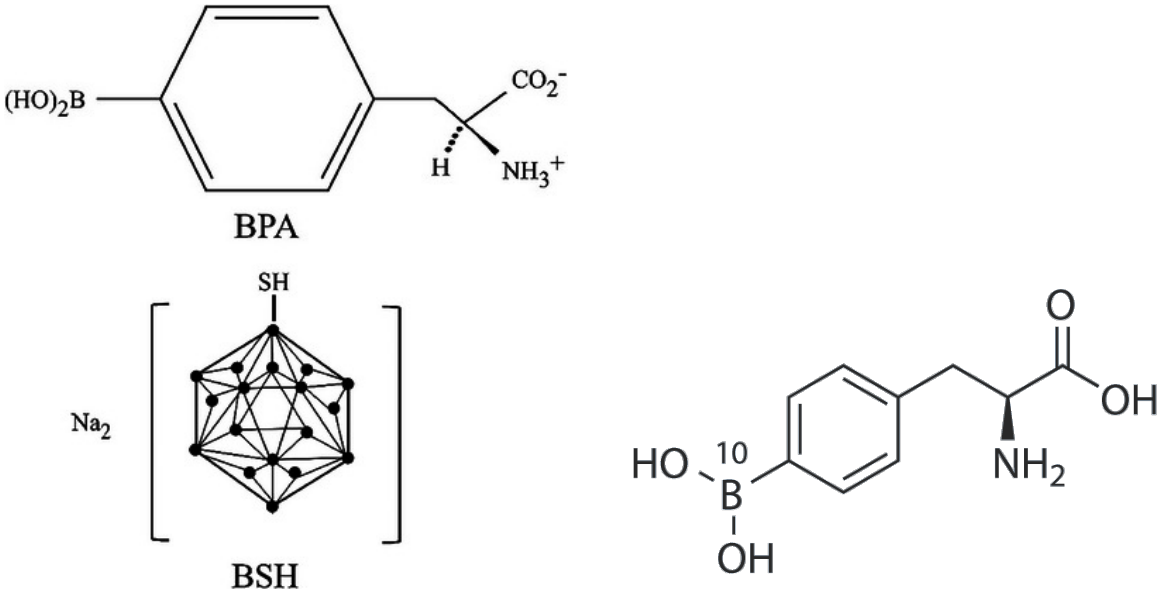
1.2.2 Boron Carriers

The therapeutic ratio achievable in BNCT is determined almost entirely by the tumor selectivity of the boron carrier. An ideal agent would combine a high absolute tumor boron concentration ($\gtrsim$ 20–35 μg/g, or approximately 10^9 atoms per cell, considered the minimum for therapeutic efficacy) with a high tumor-to-blood and tumor-to-normal-tissue concentration ratio, rapid clearance from blood and normal tissue, low

intrinsic toxicity, and retention in tumor cells for the duration of neutron irradiation [19, 33, 38].

First-generation compounds. Simple inorganic borates such as sodium borate, used in the 1950s trials, exhibited poor tumor selectivity (tumor-to-normal-tissue ratios close to unity) and are of historical interest only [33].

Second-generation compounds. Developed from the 1960s onward and still in clinical use today, these comprise two agents (Figure 1.4). Sodium borocaptate (*BSH*, $Na_2B_{12}H_{11}SH$) is a polyhedral borane cluster administered intravenously; it does not cross the intact blood–brain barrier and its uptake is mainly associated with disruption of that barrier and extracellular distribution, a property historically exploited in glioma BNCT. Boronophenylalanine (*BPA*, 4-borono-L-phenylalanine) is an amino-acid analogue actively transported into cells via the L-type amino acid transporter 1 (LAT1), which is characteristically upregulated in malignant tissue [22]. Because BPA has limited aqueous solubility, it is typically administered as the fructose complex (BPA-F); the fructose-complexed formulation developed in Japan, borofalan (^{10}B, development code SPM-011, trade name STEBORONINE), is more stable and easier to dissolve, and is the formulation used in the JHN002 trial [20, 23, 39].



(a) *Boronophenylalanine (BPA) and sodium borocaptate (BSH)* [40].

(b) *Borofalan* [21].

Figure 1.4: *Chemical structures of the principal boron carriers used in clinical BNCT.*

Third-generation agents. Still largely experimental, these include amino-acid analogues such as 3-borono-L-tyrosine (BTS), boronated peptides, nucleoside analogues, liposomal and nanoparticle-based carriers, and dodecaborate-conjugated compounds designed to improve subcellular targeting [22, 41].

Despite extensive preclinical work, clinical translation has been difficult: a carrier must simultaneously satisfy pharmacological, toxicological, manufacturing and regulatory requirements while remaining compatible with the timing of neutron irradiation [42]. Even though encouraging results on boron absorption and distribution have been obtained *in vitro* and *in vivo*, the translation to clinical use depends on extensive and expensive trials on human patients.

Pharmacokinetics in H&N cancer

In H&N tumors, BPA administration protocols differ between reactor-based and accelerator-based facilities. In the historical reactor-based protocol used in Finland, BPA is administered as a continuous intravenous infusion (typically 200–500 mg/kg body weight over 1–2 hours), followed, after an equilibration interval of 1–2 hours, by neutron irradiation delivered while boron blood concentration is monitored, typically by inductively coupled plasma atomic emission spectrometry (ICP-AES) on serial blood samples [23, 43]. In contrast, the accelerator-based protocols currently used in Japan and China maintain a lower-rate continuous BPA infusion throughout neutron irradiation itself (e.g., 200 mg/kg/h for the first 2 hours followed by 100 mg/kg/h during irradiation), in order to sustain a stable boron blood concentration for the duration of treatment [21, 34]. Pharmacokinetic modeling of the resulting blood boron time-course, and its relationship to tumor and normal-tissue concentration, has been described in detail for BPA-fructose administration [44].

Beyond blood-based pharmacokinetic monitoring, positron emission tomography with ^{18}F-labeled BPA (^{18}F-BPA-PET) provides a non-invasive, spatially resolved estimate of boron uptake prior to treatment, since the radiolabeled tracer shares the same LAT1-mediated transport mechanism as the therapeutic ^{10}B-carrier. In patients with unresectable, advanced or recurrent H&N squamous cell carcinoma, sequential ^{18}F-BPA-PET imaging performed during the pre-treatment work-up has been shown to track dynamic changes in tumor uptake, supporting its use as a patient selection tool to

identify lesions with sufficiently favorable, and sufficiently stable, boron accumulation before committing to irradiation [46]. This imaging-based selection step is complementary to, rather than a substitute for, the blood-sampling protocol described above: the latter quantifies the boron concentration actually present in blood at the time of irradiation, while ^{18}F-BPA-PET informs the *a priori* decision of whether a given tumor is a suitable candidate for BNCT at all. Beyond this selection role, ^{18}F-BPA-PET has recently been shown to carry dosimetrically relevant information that conventional blood-sampling protocols cannot capture: Teng et al. demonstrated, in a cohort of H&N BNCT treatments, that the oral mucosa exhibits substantial spatial heterogeneity in boron uptake, and that dose metrics derived from a PET-defined, uptake weighted mucosal subvolume predict oral mucositis substantially better than the uniform concentration assumption underlying both the Finnish and Japanese anatomical mucosa workflows [45]. This finding suggests that PET-derived boron heterogeneity may eventually inform not only patient selection but also normal-tissue dose-constraint modeling in H&N BNCT.

Tumor-to-blood ratios in H&N squamous cell carcinoma reported in the literature typically range from 2.5 to 4, comparable to or exceeding those achieved in glioma [22]; boron concentration in glioblastoma has similarly been shown to correlate with tumor cellularity [47]. A dosimetrically important feature specific to H&N BNCT is boron uptake in the oral and pharyngeal mucosa: owing to its high metabolic turnover, the mucosa accumulates BPA at concentrations intermediate between tumor and blood (mucosa-to-blood ratios of approximately 1.5–2 are typically assumed in clinical dose calculations). The mucosa is consequently treated as the principal dose-limiting organ at risk in H&N BNCT, in direct analogy to the role played by the vascular endothelium of the central nervous system in glioma BNCT [19, 43].

Because the compound-dependent biodistribution of boron between tumor, mucosa, skin and blood determines the tissue-specific biological weighting factors introduced in Section 1.3.2, an accurate model of BPA pharmacokinetics in H&N tissue is a prerequisite for any dosimetric treatment planning pipeline.

1.2.3 Neutron Beam and Sources

Neutron spectra

The choice of incident neutron energy critically determines the depth-dose profile achievable in tissue and is one of the central engineering problems in BNCT facility design. Neutrons are conventionally classified as thermal ($E < 0.5$ eV), epithermal (0.5 eV $< E < 10$ keV) and fast ($E > 10$ keV) [35], though the thermal/epithermal boundary varies slightly across sources (e.g. 0.4 eV in Table 1.2).

Purely thermal beams, as used in the earliest clinical trials, are rapidly attenuated in tissue and are only suitable for superficial targets; at depths beyond 2–3 cm the thermal-neutron fluence, and hence the boron-capture dose to the tumor, falls below therapeutically useful levels. Epithermal neutrons, by contrast, undergo moderation (inelastic and elastic scattering, principally with hydrogen) as they traverse tissue, building up a thermal-neutron flux that peaks at a depth of several centimetres before decaying, as shown in Figure 1.5; this makes epithermal beams the modality of choice for deep-seated or intermediate-depth targets, including glioma and the majority of H&N tumor sites [19]. Fast neutrons (> 10 keV) are undesirable in the therapeutic beam because they deposit a high-LET dose that increases with beam intensity irrespective of target depth (Section 1.3.1); their presence is minimized by the beam-shaping assembly (BSA) of the treatment facility.

Reactor-based sources

Until approximately 2010, clinically usable epithermal neutron beams for BNCT could only be generated by extracting and filtering the fission-neutron spectrum of a nuclear research reactor[4], typically using resonance-scattering filters (aluminum, sulfur, fluorine, or fluental) to remove fast neutrons, and bismuth or lead to attenuate accompanying gamma-ray contamination [19, 48]. Reactor-based facilities, while historically indispensable, present substantial practical limitations for routine clinical deployment: the requirement for a licensed nuclear reactor precludes installation in a conventional hospital setting, and several of the historically important research reactors used for BNCT have since been decommissioned [22, 48].

[4]Examples include the Kyoto University Research Reactor in Japan, the Finnish reactor FiR-1, the Tsing Hua Open-Pool Reactor in Taiwan, and the fission-converter facility at MIT, USA.

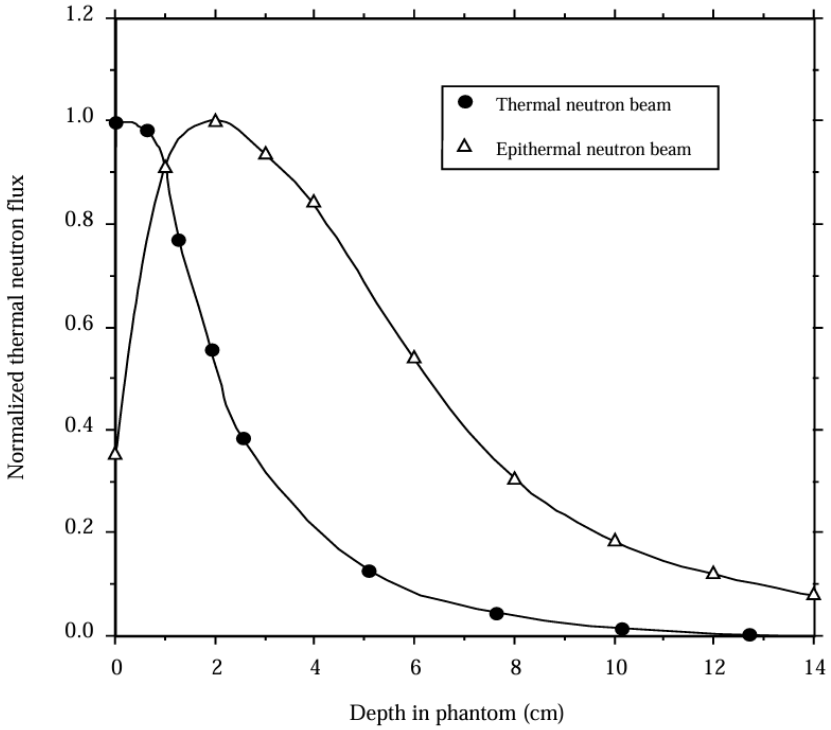


Figure 1.5: *Comparison of flux-depth distributions for thermal and epithermal neutrons* [28].

Accelerator-based sources. This limitation in the reactor-based approach has driven the development of accelerator-based BNCT (A-BNCT) systems, in which a proton beam (typically in the 2–10 MeV range) produced by a compact accelerator[5] is directed onto a neutron-producing target, most commonly $^{9}\mathrm{Be}(p,n)$ or $^{7}\mathrm{Li}(p,n)$, generating a fast-neutron spectrum (Figure 1.6). This spectrum is subsequently moderated and shaped by a dedicated BSA[6] into a clinically usable epithermal beam [35, 49, 50]. Lithium targets can produce a relatively low-energy neutron spectrum near threshold but present challenges related to heat removal and target lifetime; beryllium targets are mechanically more robust but require higher-energy protons and produce a broader spectrum, needing more extensive moderation.

A-BNCT systems can, in principle, be housed within a conventional hospital, dramatically improving accessibility. The first accelerator-based clinical BNCT device, using a cyclotron-based neutron source, entered routine clinical use in Japan in 2020 for recurrent or locally advanced H&N cancer, on the basis of the JHN002 phase II trial [20].

[5]Electrostatic, cyclotron, or radio-frequency-quadrupole (RFQ) linear accelerator.

[6]Typically MgF_2 or a similar moderator, combined with lead and borated-polyethylene shielding/reflector materials.

Comparable systems are now operating or under commissioning in Finland (Helsinki University Hospital), South Korea and Italy, and their specifics are shown in Table 1.1 [35, 49].

Table 1.1: Major A-BNCT devices under development worldwide, as reported by Kumada et al. [35], with status as of 2022. The "Facility status" column refers to the clinical readiness of the treatment facility as a whole, not to the technical completion of the accelerator hardware alone.

Country	Institute/Hospital	Accelerator type	Target material	Energy (MeV)	Current (mA)	Facility status
Japan	Southern Tohoku Hospital	Cyclotron	Be	30	1.0	Treatment
	Kansai BNCT Research Center	Cyclotron	Be	30	1.0	Treatment
	Kyoto University	Cyclotron	Be	30	1.0	Research
	National Cancer Center Hospital	Linac	Li	2.5	12	Clinical study
	Edogawa Hospital	Linac	Li	2.5	12	Clinical study
	University of Tsukuba	Linac	Be	8	2.1	Preparation of clinical study
	Nagoya University	Electrostatic	Li	2.8	15	Commissioned
	Shonan Kamakura Hospital	Electrostatic	Li	2.6	12	Commissioned
Finland	Helsinki University Hospital	Electrostatic	Li	2.6	12	Preparation of clinical study
China	Xiamen Humanity Hospital	Electrostatic	Li	2.5	10	Clinical study
	IHEP	Linac	Li	3.5, 2.8	2.9, 20	Commissioned
	China Institute of Atomic Energy	Cyclotron	Be	14	1	In planning
	Lanzhou University	Electrostatic	Li	2.6	15	In planning
Italy	CNAO	Electrostatic	Li	2.5	10	In planning
	INFN	Linac	Be	5	30	Under development
S. Korea	Gil Hospital	Linac	Be	10	8	Clinical study
U.K.	Birmingham University	Electrostatic	Li	2.6	12	Under development
Russia	Budker Institute	Electrostatic	Li	2.0–2.3	10	Research
Argentina	CNEA (deuteron)	Electrostatic	Be, C	1.45	30	Under development
Spain	University of Granada	Electrostatic	Li	2.1	30	In planning

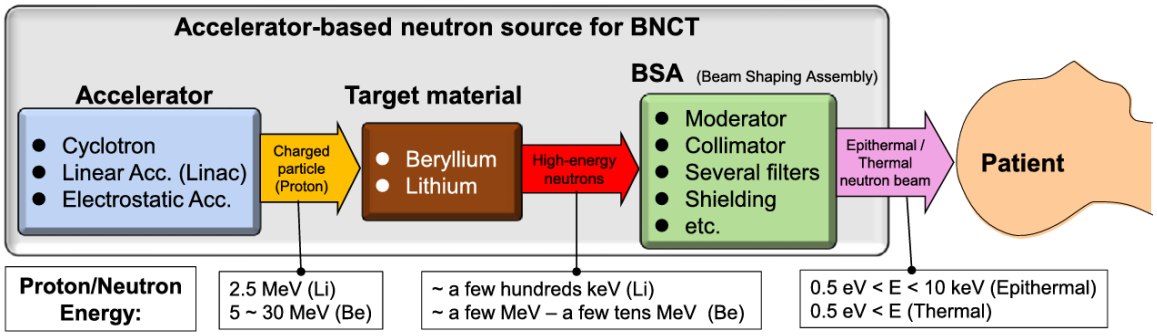


Figure 1.6: *Schematic diagram of an accelerator-based neutron source for BNCT* [35].

Beam quality criteria

The neutron beam quality criteria for clinical BNCT beams were established in detail by the IAEA in its foundational 2001 technical document [28] and subsequently updated in light of accelerator-based technology [18]. Desirable beam characteristics, summarized

in Table 1.2, include: an epithermal neutron fluence rate high enough to keep treatment times clinically practical (a beam of 10^9 n cm^{-2} s^{-1} requires an irradiation time of the order of one hour to deliver the required dose); a fast-neutron dose-to-epithermal-fluence ratio below approximately 2×10^{-13} Gy cm^2; a photon dose-to-epithermal-fluence ratio below approximately 2×10^{-13} Gy cm^2; and a high neutron current-to-flux ratio, indicative of a forwardly directed beam and hence of greater tissue penetrability [28].

These quantities describe beam quality in air; patient-specific performance also depends on anatomy, field size, beam incidence and target depth, and dedicated beam-filtration approaches (e.g. optional ^{6}Li filters) have been used to further improve dose targeting at fixed reactor facilities [51]. These criteria, originally formulated for reactor-based facilities, are now routinely met by both reactor-based and accelerator-based systems [18, 49].

More recently, an Italian–Argentinian collaboration proposed evaluating neutron beam suitability on the basis of in-patient radiobiological figures of merit – rather than the beam's physical in-air characteristics alone – introducing metrics such as the TCP, NTCP, and their combination into an Uncomplicated Tumor Control Probability (UTCP) [29, 30].

Table 1.2: Summary of the desirable characteristics for thermal and epithermal neutron beams for use in clinical studies of neutron capture therapy [19].

	Epithermal	**Thermal**
Neutron energy	0.4 eV $< E <$ 10 keV	$0 < E <$ 0.4 eV
Flux intensity (10^9 n cm^{-2} s^{-1})	> 2	> 0.5
Total contamination[a](10^{-13} Gy-w cm^2)	< 2.8	< 2.8
Advantage depth[b](cm)	> 9	> 4
Beam diameter (cm)	0–16	0–16
Collimation[c](J/Φ)	> 0.75	—

[a] Contamination refers to dose from the non-thermal or non-epithermal components of the beam. The contamination dose is expressed in weighted units, Gy-w, per thermal or epithermal neutron. The contamination dose is weighted by RBEs of 1.0 and 3.2 for photons and neutrons, respectively.

[b] Advantage depth is defined as the depth in tissue at which the dose to the tumor is equal to the maximum dose to normal tissue.

[c] Collimation: the ratio J/Φ, or the current to flux ratio, is a measure of the directionality of the beam.

1.3 BNCT Dosimetry

The radiobiology and dosimetry of BNCT are substantially more complex than those of conventional radiotherapy modalities. As discussed in Section 1.2.1, BNCT relies on the nuclear capture reaction between thermal neutrons and ^{10}B nuclei, whose products – the α-particle and the recoiling ^{7}Li ion – are responsible for the therapeutic dose. However, the ^{10}B(n,α)^{7}Li reaction is not the only nuclear process taking place during irradiation: neutrons also interact with other elements present in biological tissue, producing additional charged particles with distinct physical properties and biological effectiveness [28]. These particles deliver an unavoidable, non-negligible background dose to both tumor and healthy tissue alike. BNCT irradiation is therefore characterized by a mixed radiation field, and the dose contributions of its different components must be summed while properly accounting for their differing biological effectiveness.

1.3.1 Mixed Radiation Field and Dose Components

Regardless of the neutron source, the absorbed dose delivered to any point in tissue during BNCT irradiation is the sum of four physically and radiobiologically distinct contributions, corresponding to the reactions introduced in Section 1.2.1:

- the *boron dose*, D_B, arising from the $^{10}\mathrm{B}(n,\alpha)^{7}\mathrm{Li}$ reaction, proportional to the local ^{10}B concentration and the thermal-neutron fluence;
- the *nitrogen dose*, D_N, arising from the $^{14}\mathrm{N}(n,p)^{14}\mathrm{C}$ reaction in tissue nitrogen, essentially uniform across all tissues owing to the near-constant nitrogen content of soft tissue;
- the *fast-neutron (hydrogen) dose*, D_{fn}, arising from proton recoils produced by elastic scattering of fast, and to a lesser extent epithermal, neutrons on hydrogen, dependent on the fast-neutron contamination of the beam and decreasing with depth;
- the *photon (gamma) dose*, D_γ, comprising both the incident photon contamination of the beam itself and the secondary 2.2 MeV photons produced by the $^{1}\mathrm{H}(n,\gamma)^{2}\mathrm{H}$ reaction throughout the irradiated volume.

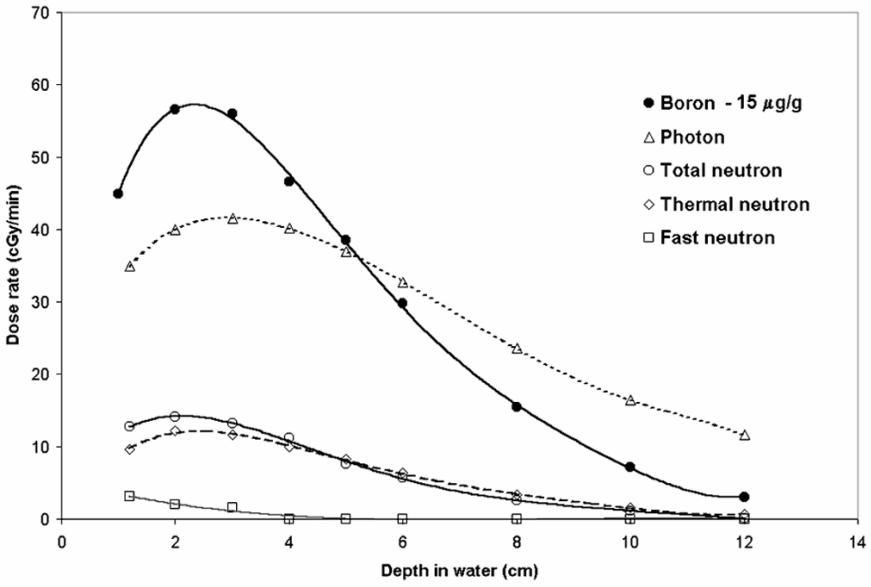


Figure 1.7: *Depth-dose profiles representative of an epithermal neutron beam at the MITR-II FCB Research Reactor in the United States* [19].

The measurement-based grouping shown in Figure 1.7 does not report a separate nitrogen dose curve, unlike the four-component classification adopted throughout the rest of this thesis. This is not a discrepancy in the underlying physics, but a direct consequence of how each quantity was obtained.

In-phantom measurements of this kind rely on instrumentation, typically twin ionization chambers, that cannot distinguish the proton produced by nitrogen capture from the recoil proton produced by fast-neutron scattering on hydrogen: both register as an indistinguishable high-LET proton signal, and are consequently reported as a single, combined "neutron" dose. Accordingly, the mixed radiation field could be classified by particle type and LET rather than by target nuclide (Figure 1.8):

- *low-LET gamma rays*, resulting primarily from the capture of thermal neutrons by hydrogen atoms in tissue, $^{1}\mathrm{H}(n,\gamma)^{2}\mathrm{H}$, but also present as incident gammas in the beam;

- *intermediate-LET protons*, either recoiling hydrogen nuclei due to collisions with fast neutrons (contaminants in the beam), $^{1}\mathrm{H}(n,n')^{1}\mathrm{H}$, or ejected due to capture of low-energy neutrons in nitrogen, $^{14}\mathrm{N}(n,p)^{14}\mathrm{C}$;

- *high-LET heavier charged particles* (^{4}He, ^{7}Li, ^{14}C), released as products of capture reactions in ^{10}B and ^{14}N, namely $^{10}\mathrm{B}(n,\alpha)^{7}\mathrm{Li}$ and $^{14}\mathrm{N}(n,p)^{14}\mathrm{C}$.

The nitrogen-capture reaction, in particular, contributes to two of the three categories simultaneously, since it emits both a high-LET proton and a recoiling ^{14}C nucleus [19].

Where nitrogen and hydrogen contributions *can* be resolved independently, as is the case of this thesis, the two are instead kept as distinct dose components, D_N and D_{fn}, for consistency with the treatment-planning results presented in Chapter 4.

The two conventions remain physically consistent with one another: summing D_N and D_{fn} recovers exactly the "total neutron" quantity reported in Figure 1.7.

Resolving these two contributions separately raises a natural question: why are nitrogen and hydrogen singled out rather than any other tissue constituent?

According to ICRU Report 44 [52], soft tissue is primarily composed of four elements, in the following mass fractions: oxygen (76.2%), carbon (11.1%), hydrogen (10.1%), and nitrogen (2.6%). For dose estimation in BNCT, the hydrogen and nitrogen dose components play a key role because of their comparatively high cross sections and KERMA factors, whereas the dose delivered by nuclear reactions involving carbon and oxygen is negligible [18].

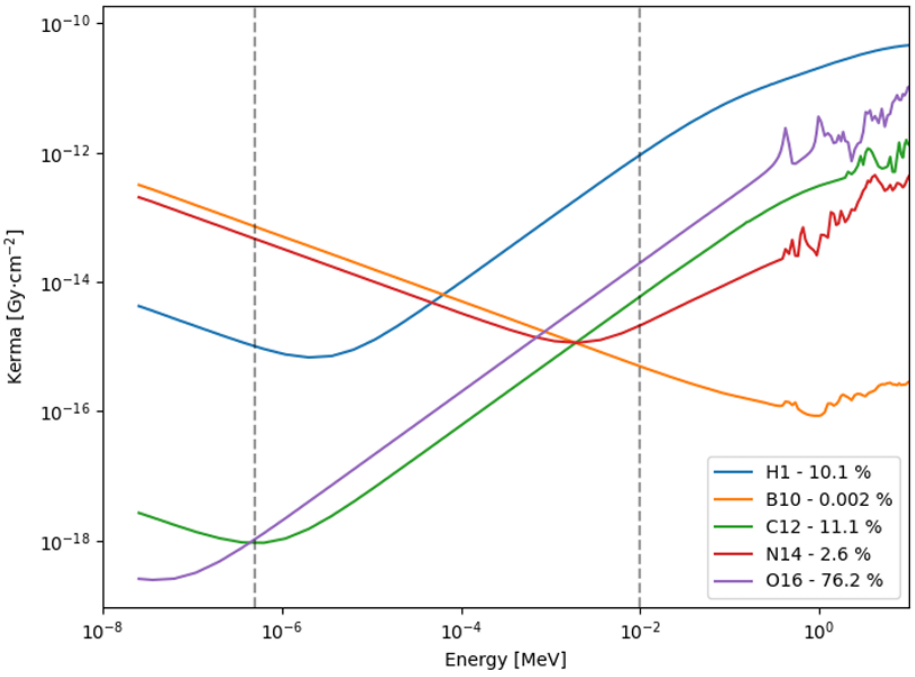


Figure 1.8: *Neutron cross section of main tissue elements* [53].

Elastic scattering between neutrons and hydrogen nuclei has a lower cross section than radiative capture; however, its dose contribution is substantial because of the sheer abundance of hydrogen atoms in tissue, and because the resulting recoil protons,

particularly those set off by neutrons above 10 keV, deposit a non-negligible dose. As epithermal and fast neutrons penetrate tissue, they are progressively thermalized through elastic scattering with hydrogen: in this $^{1}\mathrm{H}(n,n')p$ reaction, the incoming neutron loses on average half of its energy, releasing a high-LET recoil proton. Thermal neutrons, by contrast, are predominantly involved in radiative capture processes such as $^{1}\mathrm{H}(n,\gamma)^{2}\mathrm{H}$ and $^{14}\mathrm{N}(n,p)^{14}\mathrm{C}$. In the former case, a low-LET photon of energy $E_\gamma = 2.2$ MeV is produced, while the nitrogen-capture process releases an intermediate-LET proton of energy $E_p = 590$ keV [27].

Only the boron dose is selective for tumor tissue, to the extent that the boron carrier concentrates preferentially in malignant cells. The remaining three components are essentially independent of boron content and constitute an unavoidable, non-selective background dose common to tumor and normal tissue alike [19]. Because each of these components differs substantially in LET, and hence in relative biological effectiveness, they cannot simply be summed as absorbed doses when it is necessary to establish a dose-effect relationship; instead, a biologically weighted formalism is required to transform absorbed dose into photon-equivalent units. This is necessary because dose-response relationships for tumor and normal tissue are well established from conventional photon therapy, whereas dedicated TCP and NTCP models specific to each BNCT beam and boron carrier are not directly available. Figure 1.9 illustrates schematically the differing selectivity of the dose components for the tumor and for several normal tissues in a representative BNCT application.

1.3.2 The Classical Fixed-Weighting Dose Model

The radiobiological rationale for weighting the four absorbed dose components has traditionally rested on the classical definition of Relative Biological Effectiveness (RBE), the ratio of the absorbed dose of a low-LET reference radiation (conventionally 250 kVp X-rays or ^{60}Co γ-rays) to the absorbed dose of the test radiation required to produce an equal biological effect:

$$\mathrm{RBE} = \left.\frac{D_{\mathrm{reference}}}{D_{\mathrm{test}}}\right|_{\mathrm{iso\text{-}effect}} \tag{1.3}$$

High-LET particles such as the α-particle and ^{7}Li produce a denser pattern of ionization along their tracks than low-LET photons, resulting in a greater probability of complex,

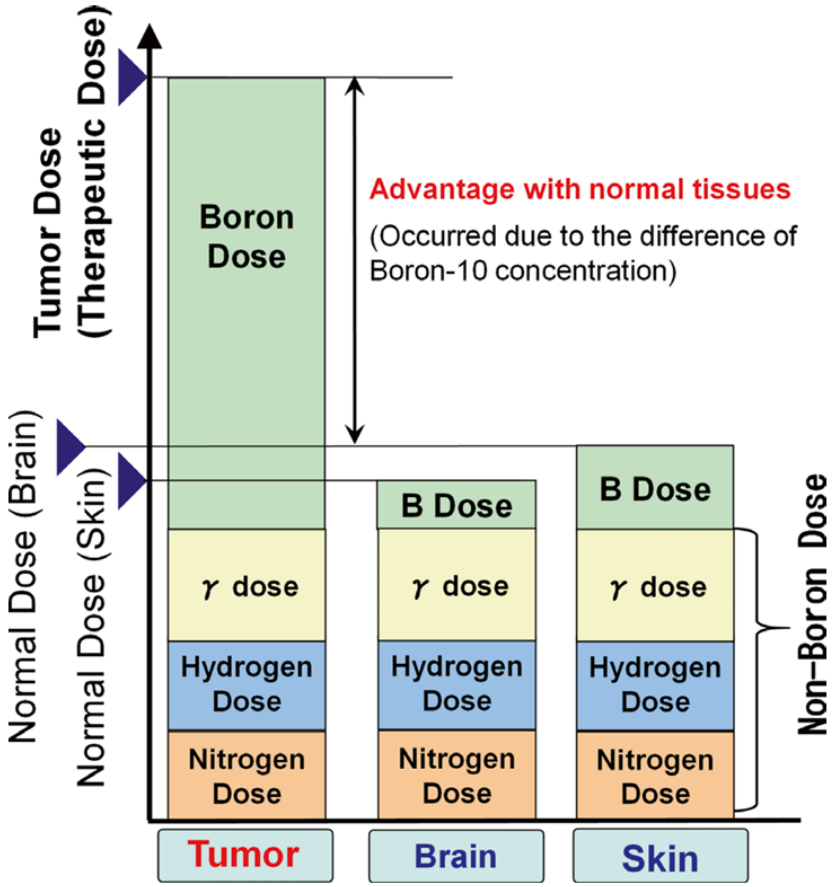


Figure 1.9: *Schema of the dose components for the tumor and for several tissues in BNCT dosimetry* [54].

unrepairable double-strand DNA breaks per unit absorbed dose, and hence RBE values substantially greater than unity [19]. However, the biological effect of the ^{10}B(n,α)^{7}Li reaction depends not only on the LET of the emitted particles, but also on the *microscopic distribution* of boron atoms relative to radiosensitive cellular structures. As a result, a single RBE value cannot adequately characterize the boron-dose component: two boron compounds with identical macroscopic boron concentration but different subcellular or intercellular distributions can produce substantially different biological effects. For this reason, the boron dose is instead weighted by a dedicated factor, the Compound Biological Effectiveness (CBE), which is kept formally distinct from the RBE used for the boron-independent high-LET dose components [19, 33].

Both RBE and CBE factors are derived experimentally from cell-survival or tissue-endpoint dose response curves by comparing the absorbed dose of a test radiation needed to reach a given biological endpoint with the reference dose needed to reach

that same endpoint.

The RBE of the high-LET components alone is measured using a neutron beam *without* boron present:

$$D_{X,\mathrm{ED}_{50}} = D_{n,\mathrm{ED}_{50}} \cdot \mathrm{RBE}_n, \tag{1.4}$$

where $D_{X,\mathrm{ED}_{50}}$ is the X-ray reference dose producing 50% incidence of the chosen biological endpoint and $D_{n,\mathrm{ED}_{50}}$ is the physical neutron-beam dose producing that same endpoint with the same incidence [19]. Rearranging Equation (1.4) gives

$$\mathrm{RBE}_{\mathrm{beam}} = \frac{D_{X,\mathrm{ED}_{50}}}{D_{\mathrm{beam},\mathrm{ED}_{50}}} \tag{1.5}$$

obtained entirely independently of boron.

Afterward, the same irradiation is repeated with the boron carrier administered beforehand. The total absorbed dose now includes both the beam component (weighted by the $\mathrm{RBE}_{\mathrm{beam}}$ already measured) and the boron-capture component (weighted by the unknown CBE). Solving for the CBE factor of the given compound and tissue yields:

$$\mathrm{CBE} = \frac{D_{X,\mathrm{ED}_{50}} - D_{\mathrm{beam},\mathrm{ED}_{50}}\,\mathrm{RBE}_{\mathrm{beam}}}{D_{^{10}\mathrm{B},\mathrm{ED}_{50}}}\,, \tag{1.6}$$

where $D_{^{10}\mathrm{B},\mathrm{ED}_{50}}$ is the absorbed dose from the $^{10}\mathrm{B}(\mathrm{n},\alpha)^{7}\mathrm{Li}$ reaction producing 50% incidence at the same endpoint [28].

CBE factors obtained in this way are strongly dependent on the boron carrier, the tissue, the biological endpoint, and the timing of irradiation relative to drug administration.[7] Table 1.3 compares the tissue-specific assumptions historically used in glioma BNCT trials [19] with those adopted in H&N BNCT protocols using BPA [19, 43]. The two protocols differ not only in the tissue set considered (brain being the dose-limiting normal tissue in glioma treatment, mucosa in H&N treatment), but also in the resulting CBE values themselves: the H&N mucosa CBE is nearly double the corresponding glioma brain CBE, reflecting the higher metabolic turnover, and hence higher BPA uptake of mucosal tissue relative to normal brain.

It is worth noting that dose fractionation, central to the therapeutic strategy of conventional photon radiotherapy, plays a comparatively minor role in BNCT: with tumor

[7] For example, CBE values for BPA in the rat central nervous system have been reported to range from 0.3 (short post-infusion interval, before blood–brain equilibration) to 1.5 (long infusion, near-equilibrium brain-to-blood boron ratio) [19].

Table 1.3: Comparison of tissue-specific boron concentration and CBE assumptions between the historical glioma-based BNCT protocol and the H&N-specific protocol. A fixed RBE of 3.2 is applied to the nitrogen-capture and fast-neutron dose components, and an RBE of 1.0 to the photon dose, in both protocols.

	Glioma		**H&N**	
Tissue	**Boron concentration**	**CBE**	**Boron concentration**	**CBE**
Blood	measured directly	—	measured directly	—
Brain	1.0× blood	1.3	—	—
Mucosa	—	—	∼1.5–2× blood	2.5
Scalp/skin	1.5× blood	2.5	—	2.5
Tumor	3.5× blood	3.8	—	3.8

boron concentrations of 20–30 μg/g, the great majority (85–90%) of the biologically weighted tumor dose arises from the high-LET boron and nitrogen-capture components, for which sub-lethal damage repair, and hence the therapeutic gain from fractionation, is minimal. At greater depth in normal tissue, the low-LET photon component can constitute 40–50% of the total dose, making fractionation comparatively more relevant for normal-tissue sparing [28].

Experimental studies in the rat spinal cord have shown, however, that the sparing effect of fractionation in BNCT is markedly smaller than would be predicted from the low-LET dose fraction alone, suggesting that any benefit of a multi-fraction H&N BNCT protocol (as used in the Finnish and Taiwanese trials, Section 1.4) may derive less from normal-tissue repair than from re-targeting of the boron carrier to a different subset of tumor cells between fractions [28].

Combining the four absorbed dose components with their respective weighting factors yields the standard clinical "RBE-weighted" or photon-equivalent dose, reported in units of Gray-equivalent, Gy-w:

$$D_w = \mathrm{RBE}_\gamma\, D_\gamma \;+\; \mathrm{RBE}_N\, D_N \;+\; \mathrm{RBE}_{fn}\, D_{fn} \;+\; \mathrm{CBE}\, D_B, \tag{1.7}$$

where each D_i denotes the physical absorbed dose of component i, computed from the corresponding neutron/photon fluence via the appropriate KERMA factors [13, 19].

The IAEA, in collaboration with the ICRU, has proposed a harmonized reporting framework for BNCT, explicitly modeled on ICRU Reports 50/62 for conventional photon-beam therapy [18, 28]. This framework retains the standard oncological definition of GTV, while introducing BNCT-specific extensions to account for sources of

geometric and dosimetric uncertainty peculiar to BNCT: fixed-beam patient positioning over extended irradiation times, spatio-temporal uncertainty in the boron distribution, and uncertainty in the biological weighting factors themselves [28]. The resulting total biologically weighted dose,

$$D_{bw} = w_\gamma D_\gamma + w_N D_N + w_{fn} D_{fn} + w_B D_B, \tag{1.8}$$

formally equivalent to Equation (1.7), is the quantity recommended for reporting and for comparison with conventional photon-therapy dose prescriptions [18, 28].[8]

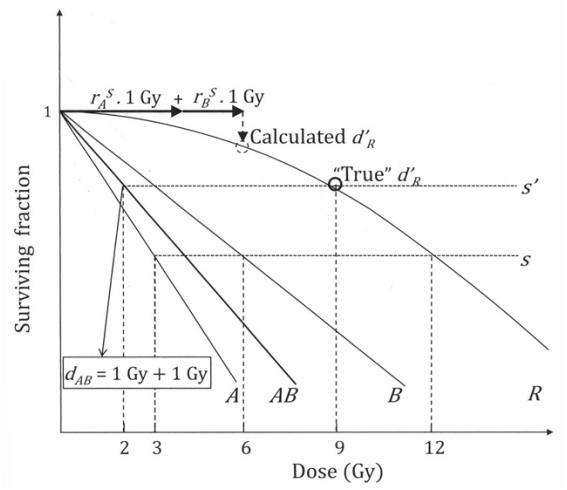
Significant limitations of this approach can be illustrated with a simple schematic example, restricted for clarity to only two radiation components A and B compared against a reference photon radiation R, as reported by González and Santa Cruz in 2012 [13]. Suppose that 3 Gy of radiation A, 6 Gy of radiation B and 12 Gy of the reference radiation R all produce the same level of biological effect s; the corresponding fixed RBE values at this level are then $r^s_A = 4$ and $r^s_B = 2$. If these same fixed factors are subsequently applied to calculate the RBE-weighted dose for a *different* combined irradiation, e.g., 1 Gy of A plus 1 Gy of B, Equation (1.7) yields $d'_R = r^s_A \cdot 1\,\text{Gy} + r^s_B \cdot 1\,\text{Gy} = 6$ Gy-w. However, the actual dose-effect curve for this particular combination shows that the true isoeffective photon dose, i.e., the photon dose producing the same survival level s' actually reached by the 2 Gy mixed irradiation, is 9 Gy (Figure 1.10a). The fixed-RBE method therefore *underestimates* the true isoeffective dose whenever it is applied at a survival level other than the one at which the RBE factors were originally derived.

The opposite bias can equally occur. Assume instead that 3 Gy of radiation A and 12 Gy of R produce the same effect level s_1, while 3 Gy of radiation B and 9 Gy of R produce a different effect level s_2, giving $r^{s_1}_A = 4$ and $r^{s_2}_B = 3$. Applying these two RBE values, each derived at a *different* survival level, to the combined irradiation of 3 Gy of A plus 3 Gy of B gives $d_R = r^{s_1}_A \cdot 3\,\text{Gy} + r^{s_2}_B \cdot 3\,\text{Gy} = 21$ Gy-w, whereas the true isoeffective dose for this 6 Gy combined irradiation is below 15 Gy (Figure 1.10b). In

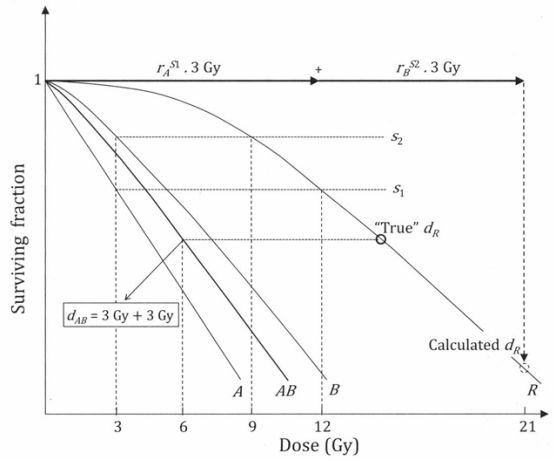
[8] The IAEA's original notation labels these same two components $w_n D_n$ and $w_p D_p$ — a "neutron dose" subscript n that in fact denotes the fast-neutron/hydrogen recoil-proton component, and a "proton dose" subscript p that denotes the nitrogen-capture component [28]. To avoid ambiguity with the D_N (nitrogen) and D_{fn} (fast neutron) notation used throughout this thesis (Section 1.3.1), the subscripts here have been relabeled accordingly; the equation itself is unchanged from the original source.

this case the fixed-RBE method *overestimates* the true isoeffective dose.

Taken together, the two examples demonstrate that the systematic bias introduced by fixed weighting factors is not merely a matter of imprecision in one direction: depending on which survival level the RBE factors are referenced to relative to the level actually reached by the mixed irradiation, the classical method can either underestimate or overestimate the biologically equivalent photon dose, with no way to know *a priori* which error applies in a given clinical calculation [13].



(a) *Mixed irradiation of 1 Gy of radiation A and 1 Gy of radiation B. Fixed RBE factors obtained for each individual radiation at level s are used to calculate the RBE-weighted dose for the combined 2 Gy total dose, underestimating the true isoeffective dose.*



(b) *Mixed irradiation of 3 Gy of radiation A and 3 Gy of radiation B. Fixed RBE factors obtained for radiations A and B at levels s_1 and s_2, respectively, are used to calculate the RBE-weighted dose for the combined 6 Gy total dose, overestimating the true isoeffective dose.*

Figure 1.10: *"True" vs. "calculated" RBE-weighted doses for two schematic mixed irradiations, illustrating the formal inconsistency of the fixed-RBE method* [13].

1.3.3 The Photon-Isoeffective Dose Model

The linear-quadratic model and dual radiation action

The starting point for treating mixed high-LET and low-LET irradiation is the linear-quadratic (LQ) survival model, in which the surviving fraction S of an irradiated cell population after a single-fraction dose D of a given radiation quality is

$$S(D) = e^{-(\alpha D + \beta D^2)} \tag{1.9}$$

where α (Gy^{-1}) and β (Gy^{-2}) are parameters relative to radiation and cell-line respectively, conventionally interpreted within the theory of dual radiation action (TDRA) as describing single-track (linear-in-dose) and two-track (quadratic-in-dose) lethal-lesion production [55, 56, 57, 58]. Within the TDRA, the quadratic term arises from the interaction of sub-lesions produced by independent energy-deposition events that would be individually insufficient to cause a lethal lesion, but combine to do so if formed sufficiently close in space and time. This microdosimetric interpretation provides the physical justification for treating the boron, nitrogen and fast-neutron dose components of BNCT as pure single-track ($\beta_i \approx 0$) contributors at clinically relevant doses, since the local energy density deposited by a single high-LET track already saturates the local lethal-lesion yield [57].

For protracted or fractionated irradiation, first-order repair of sub-lethal damage is incorporated via the generalized Lea–Catcheside time factor $G(\theta)$, originally formulated to describe the repair of sub-lethal chromosome damage [59], which modifies the quadratic term:

$$S(D,\theta) = e^{-(\alpha D + G(\theta)\,\beta D^2)} \tag{1.10}$$

with $G(\theta) \to 1$ for an acute exposure and $G(\theta) \to 0$ for irradiation protracted over a long time compared to the cellular sub-lethal repair half-time [13, 60]. In its general form, first-order repair proceeds with two characteristic kinetics, a fast and a slow component, so that G for a given radiation quality i is the weighted sum

$$G_i(\theta, t_{0f}, t_{0s}) = p_{if}\, G(\theta, t_{0f}) + p_{is}\, G(\theta, t_{0s}) \tag{1.11}$$

where p_{if} and p_{is} are the fractions of sub-lesions repaired with the fast and slow kinetics (with $p_{if} + p_{is} = 1$), t_{0f} and t_{0s} are the corresponding repair half-times, and

$$G(\theta, t_0) = \frac{2t_0}{\theta}\left[1 - \frac{t_0}{\theta}\left(1 - e^{-\theta/t_0}\right)\right] \tag{1.12}$$

is the single-kinetics form used in Equation (1.10), with t_0 the fast or slow characteristic repair time [13]. Repair is particularly important in BNCT because the irradiation time can be very long compared to the typical fractions of photon therapy, thus the number of cells able to repair during this time can significantly impact on the overall survival as a function of the dose.

Radiation synergism and the modified linear-quadratic formalism

Within the TDRA framework, sub-lethal lesions produced by different, simultaneously delivered radiation qualities coexisting in space and time within the same cell nucleus have a finite probability of interacting to produce an additional lethal lesion, not accounted for by summing the independent effects of each component: this phenomenon is termed *radiation synergism*, experimentally demonstrated for combinations of high-LET and low-LET radiation by multiple independent groups [13, 61].

Early evidence bearing on this question came from mixed X-ray/neutron irradiation studies on Chinese hamster cell proliferation [62], multicellular spheroids [63], and mouse gut epithelium [64]. In particular, split-dose experiments on V79 cell monolayers and spheroids demonstrated a qualitative asymmetry in sub-lethal damage repair kinetics: full recovery of the shoulder on the survival curve occurred only when both the initial and the challenge dose were X-rays, whereas an initial neutron dose left no detectable repair over several hours regardless of whether it was followed by X-rays or by further neutrons [63]. This finding is consistent with high-LET neutron damage being mechanistically distinct from low-LET X-ray damage, beside the former being less readily repairable than the latter.

For BNCT, in which four radiation qualities are delivered simultaneously rather than sequentially, the corresponding generalization of the synergism formalism to simultaneous multi-radiation exposure was derived by Suzuki and underlies the isoeffective-dose model described next [13, 65].

Independent-action approximation. Let $D_1, \ldots, D_4$ denote, respectively, the absorbed dose contributions from boron, nitrogen, fast neutrons and the γ component of the BNCT mixed field, and let D_R denote the dose delivered by the reference radiation R. The task is to determine $D_R = D_R(D_1, \ldots, D_4)$, referred to from here on as the *isoeffective dose*, such that it produces the same level of cell survival reached by the combination $D_1, \ldots, D_4$.

Denote by $S_i = S_i(D_i)$ the survival probability associated with dose component $i = 1, \ldots, 4$, and by $S = S(D_1, \ldots, D_4)$ the survival probability resulting from the joint action of all four radiations. Neglecting any synergistic interplay among the four components and treating each as acting independently of the others, the overall survival reduces to the plain product of the four individual survival probabilities, with

the quadratic term reserved for the photon component alone ($i = 4$):

$$S(D_1, \ldots, D_4) = \prod_{i=1}^{4} S_i(D_i), \qquad S_i(D_i) = \begin{cases} e^{-\alpha_i D_i} & i = 1, 2, 3 \\ e^{-(\alpha_i D_i + G_i(\theta)\, \beta_i D_i^2)} & i = 4, \end{cases} \tag{1.13}$$

where α_i and β_i are the parameters of the single-fraction linear-quadratic survival model for the corresponding radiation, and $G_4(\theta)$ is the generalized Lea–Catcheside time factor associated with the γ component of the beam.

Writing $S_R(D_R)$ for the survival probability of the reference radiation, the sought value $D_R = D_R(D_1, \ldots, D_4)$ must satisfy the isoeffect condition

$$S_R(D_R) = S(D_1, \ldots, D_4). \tag{1.14}$$

Substituting Equation (1.13) into Equation (1.14) gives

$$-\ln\big(S_R(D_R)\big) = \sum_{i=1}^{4} \alpha_i D_i + G_4(\theta)\, \beta_4 D_4^2. \tag{1.15}$$

Note that a quadratic dependence is allowed only for the low-LET component; sublethal-damage combination is thus permitted exclusively for this component. Equation (1.15) can equivalently be rewritten as

$$D_R(D_1, \ldots, D_4) = \sum_{i=1}^{4} r_i(D_R)\, D_i, \tag{1.16}$$

with

$$r_i(D_R) = \begin{cases} \dfrac{\alpha_i\, D_R}{-\ln\big(S_R(D_R)\big)} & i = 1, 2, 3 \\ \dfrac{\big(\alpha_i + G_i(\theta)\, \beta_i D_i\big)\, D_R}{-\ln\big(S_R(D_R)\big)} & i = 4, \end{cases} \tag{1.17}$$

where the r_i are the RBE factors expressed as a function of the reference dose D_R. Notice that the RBE factor for the γ component ($i = 4$) accounts both for sublethal-damage repair and for the temporal delivery pattern of the dose, through the factor G_4. The same RBE factors can alternatively be written as a function of the survival

probability $S = S(D_1, \ldots, D_4)$:

$$r_i(S) = \begin{cases} \dfrac{\alpha_i \, S_R^{-1}(S)}{-\ln(S)} & i = 1, 2, 3 \\ \dfrac{\left(\alpha_i + G_i(\theta)\, \beta_i D_i\right) S_R^{-1}(S)}{-\ln(S)} & i = 4, \end{cases} \tag{1.18}$$

where $S_R^{-1}(S)$ denotes the inverse of the reference-radiation survival function. Two properties of the RBE factors defined in Equations (1.17) and (1.18) deserve emphasis:

1. they depend on the reference dose, and
2. they must be evaluated at the reference dose that yields the *same* survival level S reached by the combination of the D_i, i.e., $D_R = D_R(D_1, \ldots, D_4)$.

Assume now that the survival of the reference dose follows the single-fraction linear-quadratic expression

$$-\ln\left(S_R(D_R)\right) = \alpha_R D_R + G_R(\theta')\, \beta_R D_R^2, \tag{1.19}$$

with α_R and β_R the LQ parameters and $G_R(\theta')$ the generalized Lea–Catcheside factor for the reference radiation. Inserting Equation (1.19) into expression (1.16) leads to

$$D_R(D_1, \ldots, D_4) = \sum_{i=1}^{3} \left(\frac{\alpha_i}{\alpha_R + G_R(\theta')\, \beta_R D_R} \right) D_i \; + \; \left(\frac{\alpha_4 + G_4(\theta)\, \beta_4 D_4}{\alpha_R + G_R(\theta')\, \beta_R D_R} \right) D_4. \tag{1.20}$$

Equation (1.20) is therefore the expression to be used for computing the photon-isoeffective dose in the mixed-LET BNCT field under the independent-action assumption.

When the reference survival data come from constant-dose-rate experiments, the irradiation time θ' needed to deliver D_R differs across measured points, so that Equation (1.20) generally has to be solved numerically.

If instead the survival experiment is carried out at constant irradiation time (varying the dose rate point by point), $G_R(\theta') = G_R$ is constant. If, further, the variation of $G_R(\theta')$ over the course of irradiation can be neglected – for instance when the irradiation time is much shorter than the characteristic repair time, so that $G_R \simeq 1$ –

Equation (1.20) collapses to a quadratic equation with the closed-form solution

$$D_R(D_1,\ldots,D_4) = \frac{1}{2}\frac{(\alpha/\beta)_R}{G_R}\left(\sqrt{1+\frac{4G_R}{\alpha_R(\alpha/\beta)_R}\left(\sum_{i=1}^{3}\alpha_i D_i + G_4(\theta)\,\beta_4 D_4^2\right)}\;-\;1\right). \tag{1.21}$$

Note that this expression makes no explicit use of the RBE factors from Equation (1.17) or (1.18); it depends instead directly on the LQ parameters of the BNCT components and of the reference radiation [13].

Synergistic-action formalism. Suppose now that the four components i produce their biological effect synergistically, i.e., sublesions generated by one radiation may combine with sublesions generated by any other radiation to yield a lethal lesion. The sublesion yield per unit dose for component i is taken to scale as $\sqrt{\beta_i}$. For each radiation considered on its own, the survival probability is

$$-\ln\big(S_i(D_i)\big) = \alpha_i D_i + G_i(\theta)\,\beta_i D_i^2, \qquad i=1,\ldots,4, \tag{1.22}$$

where $G_i(\theta)$ is the time factor for radiation i.

The expression derived from TDRA that describes the synergistic interaction between components i and j reads

$$-\ln\big(S_{ij}(D_i,D_j)\big) = G_{ij}(\theta)\,\sqrt{\beta_i\beta_j}\,D_iD_j, \qquad i\neq j=1,\ldots,4, \tag{1.23}$$

where $G_{ij}(\theta)$ is the time factor accounting for the first-order repair of sublesions from radiation i (radiation j) that lowers the probability of their interacting with sublesions from radiation j (radiation i) during irradiation. The survival probability for the combined action of the four radiations is then

$$-\ln\big(S(D_1,\ldots,D_4)\big) = \sum_{i=1}^{4}\alpha_i D_i \;+\; \sum_{i=1}^{4}\sum_{j=1}^{4} G_{ij}(\theta)\,\sqrt{\beta_i\beta_j}\,D_iD_j, \tag{1.24}$$

which recovers the quadratic term of Equation (1.22) when $i=j$. Assuming again that the reference-dose survival follows Equation (1.19) and equating it to Equation (1.24)

yields

$$\alpha_R D_R + G_R(\theta')\,\beta_R D_R^2 = \sum_{i=1}^{4} \alpha_i D_i + \sum_{i=1}^{4}\sum_{j=1}^{4} G_{ij}(\theta)\,\sqrt{\beta_i \beta_j}\,D_i D_j, \tag{1.25}$$

the general expression for the photon-isoeffective dose in the BNCT mixed field when synergistic interactions are retained.

When the variation of $G_R(\theta')$ for the reference data can be disregarded, i.e., $G_R(\theta') = G_R = \text{const.}$, Equation (1.25) admits the explicit solution

$$D_R(D_1, \ldots, D_4) = \frac{1}{2}\frac{(\alpha/\beta)_R}{G_R}\left(\sqrt{1 + \frac{4G_R}{\alpha_R(\alpha/\beta)_R}\left(\sum_{i=1}^{4}\alpha_i D_i + \sum_{i=1}^{4}\sum_{j=1}^{4} G_{ij}(\theta)\,\sqrt{\beta_i\beta_j}\,D_i D_j\right)} \; - \; 1\right). \tag{1.26}$$

This expression reduces to Equation (1.21) whenever the high-LET components depend on dose purely linearly, thereby excluding any synergistic contribution [13]. Figure 1.11 compares the resulting RBE-weighted and photon-isoeffective doses as a function of the total absorbed dose.

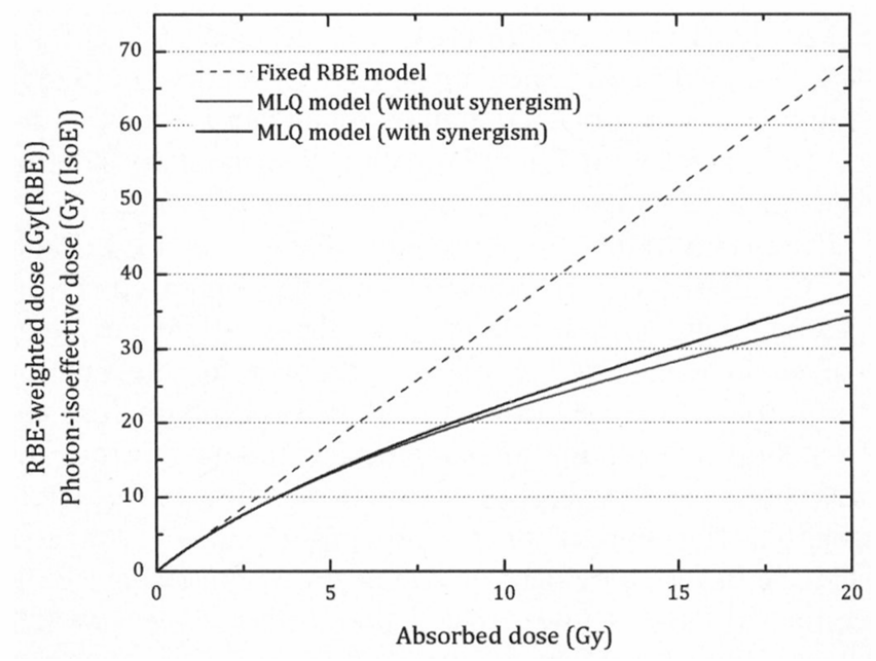


Figure 1.11: *RBE-weighted and photon-isoeffective doses as a function of the total absorbed dose computed with the fixed RBE model and the MLQ model, with and without synergism* [13].

Parameters of the model

As noted above, Equation (1.25) depends on a set of parameters, characteristic of the BNCT dose components and of the reference radiation, that enter both the independent-action and the synergistic-action versions of the LQ model. This section outlines a procedure for obtaining a consistent set of such parameters for a given cell line, starting from the cell-survival experiments typically performed in BNCT research.

Several groups have carried out cell-survival measurements under different experimental conditions with the purpose of establishing fixed RBE factors for various cell lines. Such experiments generally probe the dose response under three conditions: (1) irradiation with a reference photon beam; (2) irradiation with the neutron beam alone; and (3) irradiation with the neutron beam after administration of the boron carrier.

Starting from these three sets of dose-survival data, the model parameters can be obtained as follows. Equation (1.24) can be rewritten in terms of the total absorbed dose $D_T = \sum_{i=1}^{4} D_i$ and of the fractional contribution of each dose component, $f_i = D_i/D_T$:

$$-\ln\big(S(D_1,\ldots,D_4)\big) = \left(\sum_{i=1}^{4} f_i\alpha_i\right) D_T \;+\; G(\theta)\left(\sum_{i=1}^{4} f_i\sqrt{\beta_i}\right)^2 D_T^2, \qquad (1.27)$$

where the sixteen individual factors G_{ij} have all been replaced by a single common function G, a simplification justified by the considerations detailed in Appendix A.1. Throughout, the characteristic repair time t_0 entering the function G is treated as a known quantity, taken from the literature.

Determination of the reference-radiation parameters. Equation (1.19) is fitted to the photon-only survival data to extract the reference-radiation parameters α_R and β_R. Because survival experiments are typically carried out at a fixed dose rate, each experimental point corresponds to a different irradiation time; the resulting dependence of G_R on the irradiation time θ' is therefore built explicitly into the fit through Equation (1.12).

Determination of the BNCT radiation parameters. Equation (1.27) formally involves eight independent parameters describing the BNCT field: four associated with the neutron component, two with the boron component, and two with the total γ

component. To limit the number of free parameters, two simplifying assumptions can be introduced:

1. for the neutron components, $\alpha_2 = \alpha_3 = \alpha_n$ and $\beta_2 = \beta_3 = \beta_n$, based on the comparable response of biological systems to radiations with similar lineal-energy spectra;

2. for the photon components, $\alpha_4 = \alpha_R$ and $\beta_4 = \beta_R$, justified by the similarity between the lineal-energy spectrum of the reference photon source (^{60}Co, around 1 MeV) and that of the beam's γ contamination (typically around 2 MeV).

Under these assumptions, Equation (1.27) reduces to a four-parameter survival model, with α_n and β_n describing the neutron components and α_B and β_B describing the boron component. These four parameters are obtained simultaneously from a joint fit to the neutron-beam-only and neutron-plus-^{10}B-BPA survival data sets, with the dependence of the Lea–Catcheside factor on the irradiation time θ explicitly included for each experimental point.

It is worth stressing that, although the conventional BNCT practice is to derive radiobiological parameters *sequentially* – first fitting the beam-only data, and only afterward using the resulting values to extract the boron-specific parameters – the neutron-plus-^{10}B-BPA experiment in fact carries information on the neutron interaction with tissue beyond the ^{10}B(n,α)^{7}Li capture reaction itself. A joint fitting procedure that simultaneously minimizes both the beam-only and the neutron-plus-boron survival data is therefore preferable, as it makes full use of the available experimental information [13].

1.3.4 Tumor Control Probability and Normal Tissue Complication Probability

To translate absorbed or isoeffective dose distributions into clinically interpretable outcome predictions, BNCT dosimetry adopts Tumor Control Probability (TCP) and Normal Tissue Complication Probability (NTCP) models directly analogous to those used in conventional radiotherapy, with parameters re-derived from BNCT-specific radiobiological data. Within the MLQ formalism, the TCP for a tumor of a given cell number and radiosensitivity, at survival level S_T from Equation (1.15) evaluated for a

Table 1.4: Radiobiological parameters obtained for the 9L rat gliosarcoma *in vivo/in vitro* assay, with and without considering synergism, and corresponding RBE values for a cell surviving fraction $S = 0.01$ [13, 66].

Reference X rays	α (Gy^{-1})	β (Gy^{-2})
	0.2008	0.0078

	Model with synergism			**Model without synergism**		
	α (Gy^{-1})	β (Gy^{-2})	r_i^S	α (Gy^{-1})	β (Gy^{-2})	r_i^S
Beam γ photons	0.2008	0.0078	0.953	0.2008	0.0078	0.947
Neutrons	0.4972	0.088	2.786	0.844	0	2.689
Boron (BPA)	0.9091	0.0019	2.936	0.8896	0	2.835

single-fraction exposure, takes the general Poissonian form

$$\mathrm{TCP}_{\mathrm{MLQ}}(D) = \exp\Big[- C_1\, \phi\, C_2 \exp\big(- D(\alpha + \omega_2\omega_3 + \omega_2(1 - e^{-\omega_3 D})/D)\big)\Big], \quad (1.28)$$

with ω_2, ω_3 adjustable parameters fitted to clinical dose-response data and C_1, ϕ, C_2 accounting for the initial clonogenic cell density and tumor volume [13, 67].

For organs at risk, NTCP is generally modeled with a three-parameter empirical formula for inhomogeneous dose distributions across a reference tissue area or volume v,

$$p(D, v) = \exp\Big[- \frac{N_0}{v^k} \exp(-\gamma D)\Big], \quad (1.29)$$

where N_0, k, γ are coefficients fitted to single-fraction photon tolerance data for the endpoint of interest[9], and the equivalent uniform dose formalism can be used to reduce a heterogeneous OAR dose distribution to a single probability-equivalent uniform dose [68].

In H&N BNCT specifically, retrospective analyses have identified the minimum tumor dose delivered to the GTV, together with GTV size, as the strongest independent predictors of both regional control and overall survival: minimum GTV doses above approximately 15–18 Gy-w are associated with markedly improved outcome, while mucosal dose constraints (absorbed dose typically limited to approximately 6 Gy per fraction) are the principal factor limiting the achievable tumor dose in practice [43]. Figure 1.12 illustrates this dose-response relationship directly, together with the corre-

[9]For H&N BNCT these are typically moist desquamation of skin or grade $\geq$3 mucositis for the oral mucosa.

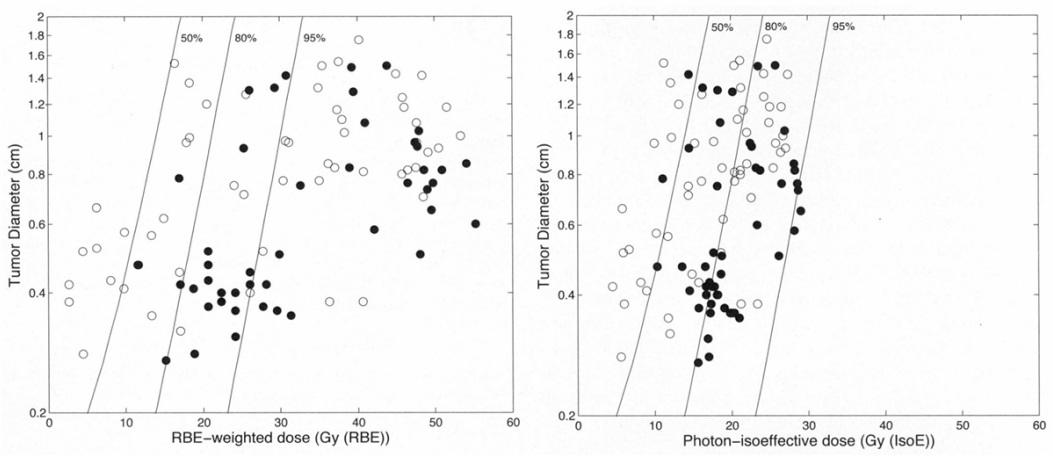


Figure 1.12: *Scatter plots showing the distribution of tumor diameters and minimum doses of the lesions with their response for fixed RBE-weighted (left) and photon iso-effective dose (right) models. Filled circles indicate complete response, and open circles indicate any other clinical outcome. Continuous lines represent the iso-tumor control probability curves computed with the TCP_{MLQ} for 50%, 80% and 95% of tumor control* [13].

sponding iso-tumor-control-probability curves.

1.4 Clinical Experience of BNCT in H&N Cancer

1.4.1 Reactor-Based Clinical Trials

The first attempt at BNCT specifically for H&N cancer was performed at the Kyoto University Research Reactor Institute in 2001, in a patient with a recurrent parotid gland tumor who had exhausted surgical, radiotherapeutic and chemotherapeutic options; regional control was achieved for seven years before the patient eventually died of recurrence [18]. This case prompted the systematic clinical investigation of BNCT in H&N cancer through the 2000s and 2010s at reactor facilities in Japan, Finland and Taiwan, extending the second-generation BPA-based protocols initially developed for glioma and melanoma to locally recurrent, inoperable, previously irradiated H&N squamous cell carcinoma [18, 69, 70], and subsequently to other H&N histology types including advanced salivary gland carcinoma [71]. The Kyoto University group reported outcomes for 62 H&N patients (49 recurrent, 13 unresectable) treated between 2001 and 2007 with BSH, BPA, or their combination, achieving a complete response rate of 29%, a partial response rate of 28%, and a two-year overall survival of 24.2% [18].

The Helsinki group (FiR-1 reactor, Finland) provided the most extensive reactor-

based experience in H&N BNCT. In the initial prospective phase I/II report, 12 patients with inoperable, locally recurred H&N cancer were treated with BPA-mediated BNCT; 10 patients responded, and the principal acute toxicities were mucositis, oral pain and fatigue [72]. In the subsequent final analysis of the same trial, 79 patients with locally recurrent H&N cancer, treated between 2003 and 2012 with two-fraction BPA-based BNCT and PET-guided target delineation, achieved a complete response rate of 36%, a partial response rate of 32%, a two-year locoregional progression-free survival of 38%, and a two-year overall survival of 21%; minimum GTV dose and GTV volume were identified as the strongest independent predictors of locoregional control and survival [18, 23]. Grade $\geq$3 toxicity, principally oral mucositis and oral pain, was observed in approximately half of patients, with osteoradionecrosis in roughly 20% of evaluable patients at long-term follow-up [23].

The same group also investigated recurrent laryngeal cancer specifically: in a nine-patient series, six of eight evaluable patients responded, and one patient remained disease-free with a functioning larynx at five years, demonstrating organ-preservation potential alongside the observation that durable control remained uncommon in this unselected recurrent population [73].

A comparable two-fraction reactor-based protocol was applied at Taipei Veterans General Hospital (using the Tsing Hua Open-Pool Reactor), Taiwan, in 17 patients (11 squamous-cell, 6 non-squamous-cell carcinoma) treated between 2010 and 2013, reporting complete and partial response rates of 35% each and a two-year overall survival of 47% [18, 39, 74]. The same Taiwanese program subsequently investigated fractionation strategies at the same facility, reporting clinically meaningful response rates alongside acute mucositis and other toxicities requiring careful management [75]; differences in neutron spectrum, fractionation scheme, boron administration protocol and dose reporting convention limit direct numerical comparison with the Finnish series.

1.4.2 Accelerator-Based BNCT and Current Clinical Status

The transition to accelerator-based neutron sources has substantially expanded clinical access to H&N BNCT over the past decade. An early accelerator-based series was reported by Southern Tohoku General Hospital, Japan, in 21 patients (13 squamous-cell, 8 non-squamous-cell carcinoma) treated between 2016 and 2019 with BPA-based BNCT, achieving a complete response rate of 24%, a partial response rate of 48%, a

two-year locoregional progression-free survival of 56%, and a notably high two-year overall survival of 85.3% [18].

The pivotal JHN002 trial, using a cyclotron-based epithermal neutron source and the carrier borofalan, reported a 71% response rate in patients with locally advanced or recurrent, unresectable H&N cancer [20], as shown in Figure 1.13. On the basis of these results, accelerator-based BNCT for recurrent H&N cancer was approved for reimbursement under the Japanese national health insurance system in 2020 — the first regulatory approval of BNCT as a routine, rather than purely investigational, cancer treatment worldwide [20].

Subsequent real-world experience accumulated under this insurance framework, and independent meta-analyses pooling reactor- and accelerator-based series, have reported objective response rates in the range of 70–80% and one-year overall survival rates of approximately 75% for recurrent H&N cancer treated with BNCT [3], broadly consistent with the earlier reactor-based experience while extending access beyond the small number of historically available research-reactor facilities. A retrospective analysis of the first 47 patients with recurrent H&N squamous cell carcinoma treated under the Japanese national health insurance system confirmed clinically relevant activity and identified prognostic factors for outcome in this post-marketing population [76]; these data remain non-randomized, and survival is influenced by tumor volume, site, prior treatment and subsequent therapy, but they demonstrate that accelerator-based BNCT can be delivered as routine medical care rather than only within a reactor-based research program.

As of 2025–2026, accelerator-based BNCT facilities for H&N indications are operational or in active clinical trial in Japan, Finland, South Korea, and China, and further systems are in commissioning or planning at additional centers, as shown in Table 1.1, reflecting the ongoing transition of BNCT from a research-reactor-dependent experimental therapy toward a more broadly deployable radiotherapy modality [18]. Multidisciplinary case-review procedures[10] are recommended by the IAEA as standard practice for approving individual BNCT indications in H&N cancer, reflecting the treatment's continued classification as a complex, multidisciplinary intervention even where nationally reimbursed [18].

[10]These typically involve a radiation oncologist, H&N surgeon, medical oncologist, pathologist, diagnostic/nuclear-medicine radiologist and palliative-care physician.

1.4.3 Dose-Response Relationships and Dose-Limiting Toxicity

Across both reactor-based and accelerator-based H&N series, a consistent dose-response relationship has been observed: higher minimum tumor dose is associated with improved locoregional control and survival, while treatment is limited principally by the tolerance of the oral and pharyngeal mucosa, which, because of its intermediate boron uptake relative to tumor and blood (Section 1.2.2), receives a substantial fraction of the prescribed tumor dose [23, 43, 45]. Acute toxicity commonly includes oral or pharyngeal mucositis, stomatitis, pain, fatigue and skin reactions; late effects may include xerostomia, dysphagia, soft-tissue necrosis and osteoradionecrosis, some of which reflect the cumulative effect of previous radiotherapy rather than BNCT alone, making attribution difficult.

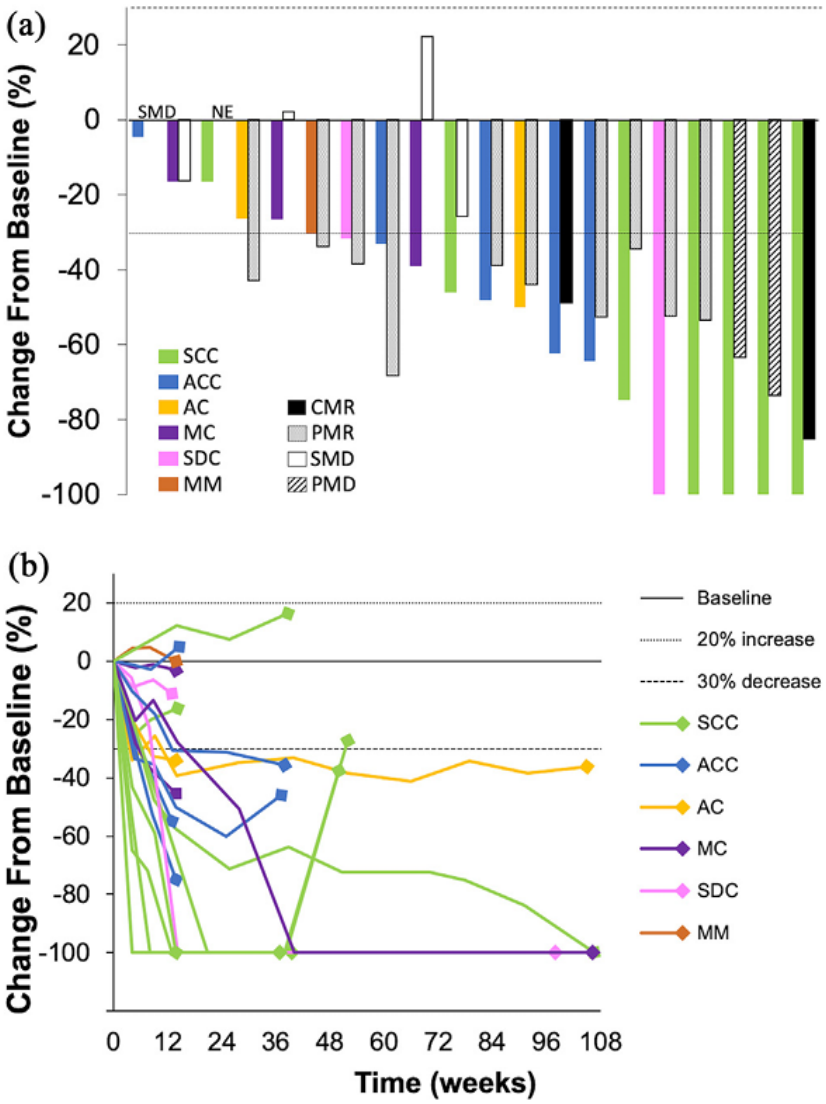


Figure 1.13: *Tumor responses observed in the JHN002 trial.*
(a) Best change in target-lesion size and in metabolic response for each patient within 90 days: for each patient, the left bar reports the maximum percentage reduction from baseline in the longest diameter of the target lesion, while the right bar reports the percentage change from baseline in the peak lean-body-mass-corrected standardized uptake value (SUL) at day 90. Bar color in the left-hand set indicates tumor histology; in the right-hand set, bar fill (black, dotted, white, striped) denotes the overall metabolic response category (CMR, PMR, SMD or PMD, respectively).
(b) Longitudinal change from baseline in target-lesion size over time, with each color again indicating tumor histology.
Abbreviations: *SUL, standardized uptake value corrected for lean body mass; R-SCC, recurrent squamous cell carcinoma; CMR, complete metabolic response; PMR, partial metabolic response; SMD, stable metabolic disease; PMD, progressive metabolic disease; RECIST, Response Evaluation Criteria in Solid Tumors; PERCIST, Positron Emission Tomography Response Criteria in Solid Tumors; NE, not evaluated; SCC, squamous cell carcinoma; ACC, adenoid cystic carcinoma; AC, acinic cell carcinoma; MC, mucoepidermoid carcinoma; SDC, salivary ductal carcinoma; MM, mucosal malignant melanoma.* [20].

Chapter 2

Artificial Intelligence and Deep Learning in BNCT Treatment Planning

Boron Neutron Capture Therapy shares with every other form of external or internal radiotherapy a structural dependence on the accurate geometric definition of the volumes that must receive dose and of the volumes that must be spared.

In current clinical practice, this geometric definition is still predominantly performed manually: an experienced radiologist or radiation oncologist reviews the patient's CT (and, when available, co-registered MRI) slice by slice and draws contours around each structure of interest. This procedure, while considered the reference standard, is time-consuming, requires substantial clinical expertise, and is affected by intrinsic variability that is difficult to eliminate, even among experienced clinicians [77, 78].

This Chapter introduces the family of Artificial Intelligence (AI) methods and of deep convolutional neural networks that make automated medical image segmentation possible.

2.1 History of Artificial Intelligence and Neural Networks

The idea of a machine capable of exhibiting intelligent behavior long predates modern computing, and its history has been extensively reviewed elsewhere from both a

technical and a philosophical standpoint [79].

The first mathematical model of a biological neuron capable of a rudimentary form of learning, the *Perceptron*, was proposed by Rosenblatt in 1958 [80] as a probabilistic model for information storage and pattern recognition in the brain. The Perceptron could learn to separate linearly separable classes of stimuli through iterative reinforcement of the connections (*weights*) between an input layer and a decision unit, and it generated considerable early enthusiasm for the connectionist research program in AI[1]. This enthusiasm was tempered by the well-known demonstration that a single-layer Perceptron cannot represent functions that are not linearly separable, which, together with the limited computational resources of the time, contributed to a prolonged reduction in funding and interest for connectionist approaches.

Interest in biologically inspired, layered neural architectures resurfaced with Fukushima's *Neocognitron*, presented in 1980 and formalized in 1982 [81]. The Neocognitron was explicitly designed to be tolerant to shifts in position and to moderate distortions of the input pattern[2]: it introduced, for the first time, the alternation of two functionally distinct cell types (a) simple, feature-extracting *S-cells* and (b) position-tolerant, pooling-like *C-cells*. This is directly recognizable as the ancestor of the convolution/pooling alternation used in every modern Convolutional Neural Network (CNN).

The subsequent development of the backpropagation algorithm for efficiently training multilayer networks, and the progressive increase in available computational power and labeled data, allowed increasingly deep architectures to be trained; the definitive turning point for computer vision is generally identified with the success of large CNNs on natural-image classification benchmarks in the early 2010s, which triggered the rapid adoption of deep learning across essentially every subfield of image analysis, including medical imaging [82].

[1] Opposing the rule-based paradigm that dominated AI research in the same decades, this idea was saying that intelligent behavior could emerge from networks of simple, neuron-like computing units rather than from explicitly programmed logical rules [79].

[2] The Neocognitron was built on the model of Hubel and Wiesel's studies of the visual cortex.

2.2 Artificial Neural Networks

2.2.1 The Perceptron and the Multilayer Perceptron

The Perceptron was designed to resemble the functional structure of a biological neuron, as shown in Figure 2.1. A biological neuron receives electrical signals as an input from its dendrites; these signals are then transmitted across its central body, where they are modulated, and the neuron fires an output signal through its synapses feeding another neuron, and so forth. This, however, happens only when the total amount of the input signals exceeds a certain threshold. The artificial Perceptron follows the same pattern.

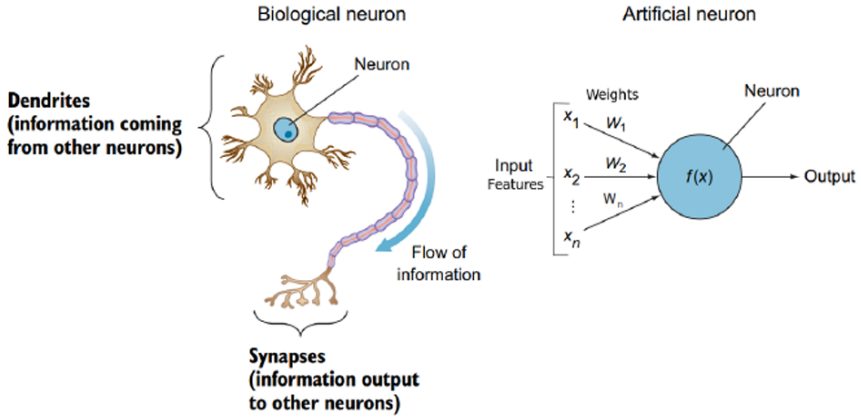


Figure 2.1: *Artificial neurons and biological ones share an analogous structure. The dendrites are emulated by the input connections, which collect input features instead of electrical signals. The function of the biological neuron central body is represented by the activation function and the activation state in the artificial neuron. Lastly, the synapse is reproduced by the output connection* [83].

A single artificial neuron, or Perceptron, receives a set of input features $x_1, x_2, \ldots, x_n$ — computationally defined as individual, measurable properties of an observed phenomenon — each weighted by a trainable coefficient w_i that reflects its relative importance. The neuron combines these weighted inputs into a scalar pre-activation,

$$z = \sum_{i=1}^{n} w_i x_i + b, \tag{2.1}$$

where b is a bias term, itself adjusted during training. As in its biological counterpart, the neuron does not simply output this raw value: it fires only once the total input exceeds a certain threshold, a behavior reproduced, in the artificial neuron, by a nonlinear *activation function* $a(z)$. This nonlinearity is essential, since a stack of purely linear layers would collapse algebraically into a single linear transformation, leaving the

network unable to represent anything but linearly separable functions [83]. Table 2.1 summarizes the activation functions in common use.

Table 2.1: Activation functions commonly used in feed-forward and convolutional neural networks [83, 84].

Name	Definition	Typical use
Linear	$a(z) = z$	rarely used in hidden layers
Sigmoid	$a(z) = \dfrac{1}{1+e^{-z}}$	output layer, binary classification
Hyperbolic tangent	$a(z) = \dfrac{e^{z}-e^{-z}}{e^{z}+e^{-z}}$	classical hidden-layer choice
ReLU	$a(z) = \max(0, z)$	standard hidden-layer choice [85]
Leaky ReLU	$a(z) = \max(0.01\,z,\, z)$	mitigates the "dying ReLU" problem [84]
Softmax	$a(z)_k = \dfrac{e^{z_k}}{\sum_j e^{z_j}}$	multi-class output / per-voxel class probability

The simplest of these — and the one originally adopted by Rosenblatt's Perceptron algorithm — is the step function, which produces a hard binary output,

$$\text{Activation Output} = \begin{cases} 0 & \text{if } z \leq \tau \\ 1 & \text{if } z > \tau \end{cases}, \tag{2.2}$$

where τ denotes the activation threshold [80]. This all-or-nothing rule reproduces, at the level of a single unit, the threshold behavior of the biological neuron; but because it is discontinuous and has zero gradient almost everywhere, it is unsuitable for gradient-based training of deeper networks, and was progressively replaced by the smooth activation functions of Table 2.1 once multilayer networks trained by backpropagation became the standard (Section 2.2.3).

A single layer of Perceptrons can only represent linear decision boundaries. Stacking several layers, with the output of each layer feeding the next, yields a *Multilayer Perceptron* (MLP), also called a fully connected or dense network because every neuron in one layer is connected to every neuron in the next, as illustrated in Figure 2.2.

2.2.2 Network Depth and Feedforward Propagation

Layers between the input and the output are termed *hidden layers*, precisely because neither their input nor their internal state is directly observed; each additional hidden layer allows the network, in principle, to represent progressively more abstract

combinations of the features extracted by the previous layer [83].

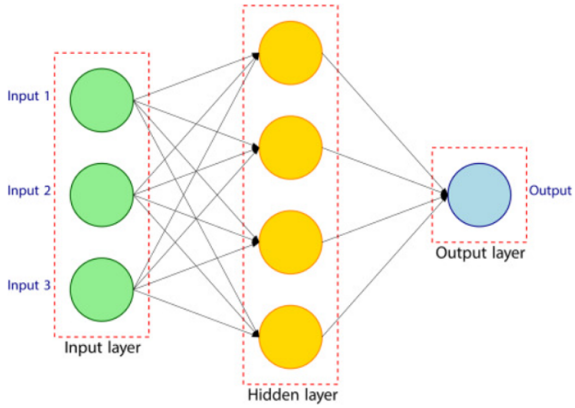
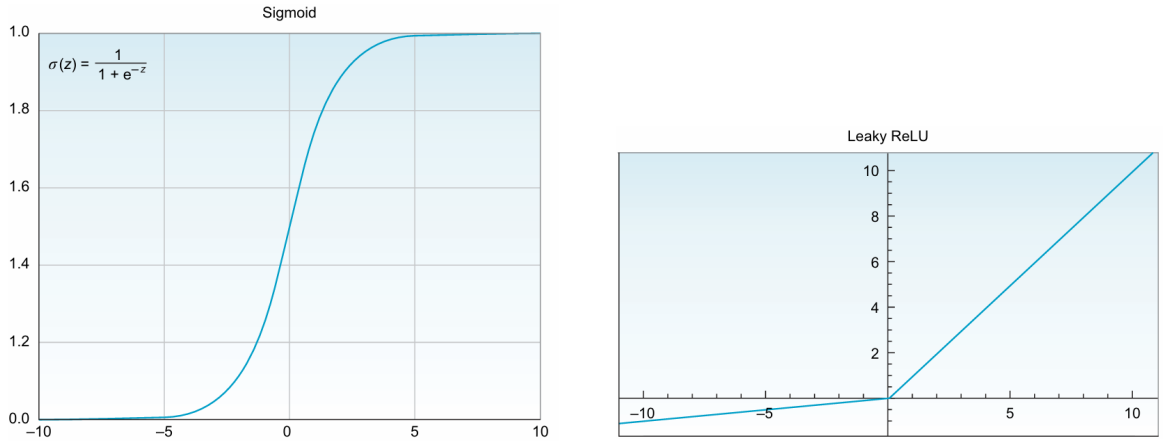


Figure 2.2: *Schematic of a Multilayer Perceptron with a single hidden layer. Input values are propagated forward through the network, with a weighted sum and an activation function computed at every unit, to produce the output prediction* [86].

More precisely, a hidden unit is never directly exposed either to the raw input data or to the final prediction: it receives as its only input the activation values produced by the layer immediately before it, and produces as its only output the activation values passed on to the layer immediately after it. Because the representation of each layer is built entirely on top of the one preceding it, depth allows a network to organize the features it learns hierarchically: an early hidden layer typically responds to simple, local patterns in the input, while progressively deeper layers recombine these into increasingly abstract, task-specific features [83]. Both the network depth (its number of hidden layers) and its width (the number of units per layer) are architectural choices that are, in principle, matched to the complexity of the task at hand: a deeper network can represent more complex functions of its input, at the cost of a correspondingly larger number of trainable parameters to optimize.

The choice of activation function (Section 2.2.1) also typically differs across a network layers (Figure 2.3). Hidden units commonly use the sigmoid or, particularly for image data, the leaky ReLU (Table 2.1); the output layer, by contrast, is chosen according to the task at hand. For multi-class classification, the standard choice is the softmax function, a direct generalization of the sigmoid to more than two classes, which converts the network raw outputs into a valid probability distribution over the candidate classes [83].

The process by which information flows through the network is called *feedforward propagation*, and consists of the repeated, layer-by-layer application of the weighted sum



(a) *Sigmoid function* **(b)** *Leaky ReLU*

Figure 2.3: *Shape of two of the activation functions of Table 2.1: (a) the sigmoid saturates for large $|z|$, whereas (b) the leaky ReLU retains a small, non-zero slope for $z \leq 0$, which is why it is preferred to mitigate the vanishing-gradient problem discussed in Section 2.2.3.*

of Equation (2.1) followed by an activation function. Denoting by $a_i^{(l)}$ the activation of the i-th unit in layer l, and by $w_{ij}^{(l)}$ the weight of the connection from unit j of layer $l-1$ to unit i of layer l, the activations of the first hidden layer are computed directly from the input features,

$$a_i^{(1)} = \sigma\left(\sum_j w_{ij}^{(1)} x_j + b_i^{(1)}\right), \tag{2.3}$$

and every subsequent layer activation is obtained in the same way from the activations of the layer immediately before it,

$$a_i^{(l)} = \sigma\left(\sum_j w_{ij}^{(l)} a_j^{(l-1)} + b_i^{(l)}\right), \qquad l = 2, \ldots, L, \tag{2.4}$$

where $\sigma(\cdot)$ denotes the activation function of the corresponding layer (not necessarily the sigmoid, despite the traditional notation), until the final L-th layer produces the network's output prediction $g = a^{(L)}$ (Figure 2.4).

Once a prediction g is obtained, its discrepancy from the corresponding, known ground-truth label can be quantified, and the network weights adjusted so as to reduce that discrepancy on subsequent predictions. This iterative, gradient-based adjustment of the weights is the process by which a neural network learns; it is described formally in Section 2.2.3.

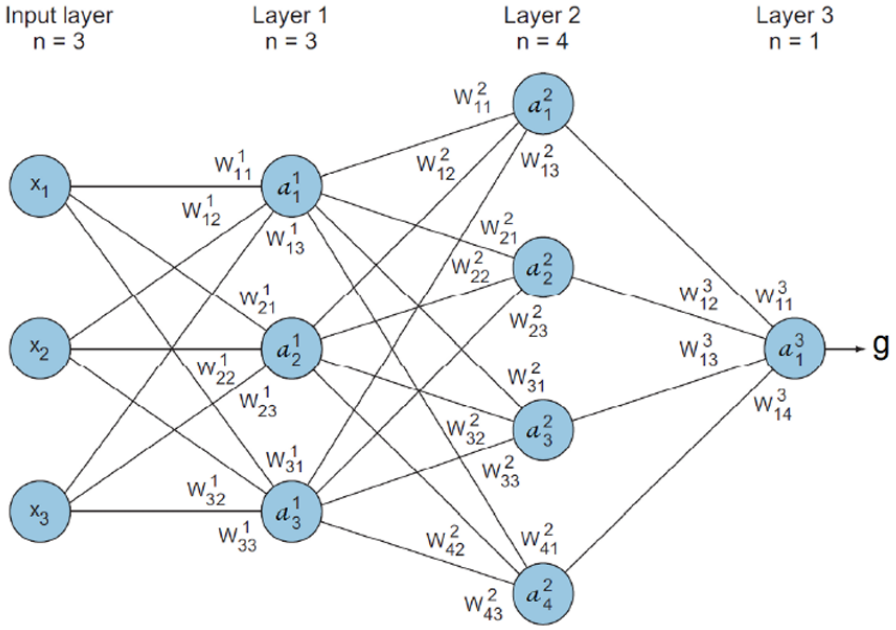


Figure 2.4: *Example of feedforward propagation through a multilayer network: the input features x_i are weighted and combined at each unit to produce the node activations $a_i^{(l)}$ of successive layers, up to the final output g* [86].

2.2.3 Training: Loss Functions and Gradient-Based Optimization

Training a neural network consists in finding the set of weights and biases that minimizes a scalar *loss function* (or cost function) $J(\theta)$, which quantifies the discrepancy between the prediction of the network and the ground truth for a given training example (Figure 2.5). For segmentation tasks, where the ground truth is itself a mask rather than a single label, the loss most directly relevant to this thesis is the *Dice loss*, built on the Dice Coefficient (DC). The DC measures the overlap between two volumes — the network predicted segmentation P and the ground-truth segmentation Q — and is the geometric metric used throughout this thesis to quantify segmentation agreement (see, e.g., Table 2.3). For two binary volumes P and Q, with $p_i, q_i \in \{0, 1\}$ indicating whether voxel i is, respectively, predicted or actually part of the structure of interest, the Dice Coefficient is

$$DC = \frac{2\sum_{i=1}^{N} p_i q_i}{\sum_{i=1}^{N} p_i^2 + \sum_{i=1}^{N} q_i^2}, \tag{2.5}$$

where the numerator, $2\sum_i p_i q_i$, counts twice the number of voxels on which prediction and ground truth agree (their intersection), and the denominator, $\sum_i p_i^2 + \sum_i q_i^2$, is simply the combined size of the two volumes, since $p_i^2 = p_i$ for a binary mask.

As a result, DC ranges from 0, when the two volumes do not overlap at all, to 1, when they coincide exactly — a scale that makes the coefficient directly interpretable regardless of the absolute size of the structure being segmented. The corresponding Dice Loss, $\mathcal{L}_{\text{Dice}} = 1 - DC$ (or a closely related multi-class generalization), is differentiable and can therefore be minimized directly by gradient descent, as first exploited by V-Net (Section 2.3.1) [87]. For a broader treatment of loss functions beyond Dice — including mean squared error and cross-entropy — see Appendix B.1.

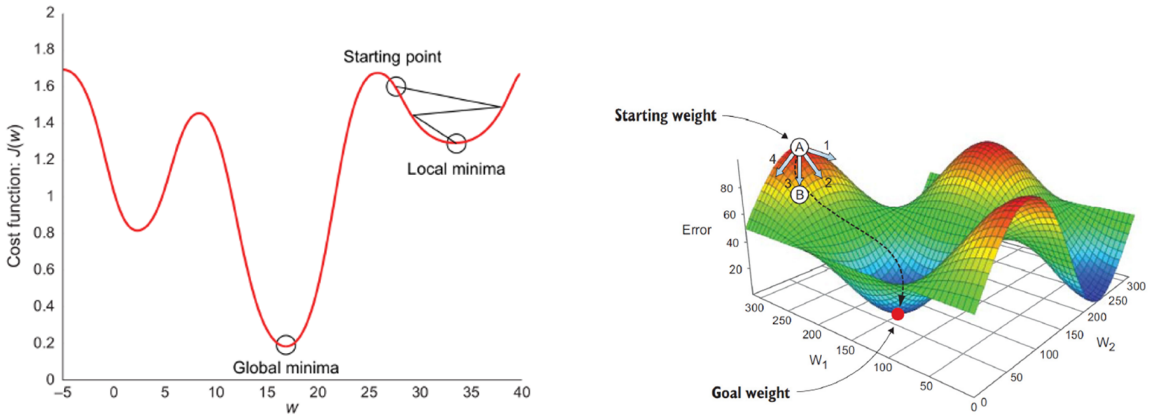
The dominant family of optimization algorithms for this minimization is *gradient descent*: the parameters θ are updated iteratively in the direction opposite to the gradient of the loss,

$$\theta \leftarrow \theta - \eta \, \nabla_\theta J(\theta), \tag{2.6}$$

where η is the learning rate [88]. Three variants of gradient descent differ in how much of the training data is used to compute a single gradient estimate: *batch* gradient descent uses the entire training set for every update, which is accurate but computationally prohibitive for large datasets; *stochastic* gradient descent (SGD) updates the parameters after every single training example, which is fast but produces a noisy, oscillating descent path; and *mini-batch* gradient descent, the standard choice in modern deep learning, computes the gradient over a small, randomly sampled subset (a mini-batch) of the training data at each step, trading off the accuracy of batch descent against the computational efficiency of SGD [83, 88].

Because the loss surface of a deep network is generally non-convex, with multiple local minima, saddle points, and plateaus, a range of adaptive refinements of SGD have been developed specifically to accelerate convergence and to reduce the sensitivity of training to the choice of learning rate [88]; these algorithms — Momentum, Adagrad, Adadelta, RMSprop, and Adam among others — are derived and compared in detail in Appendix B.1.

The gradient of the loss with respect to every weight in the network is computed efficiently, layer by layer, by the *backpropagation* algorithm, which applies the chain rule of differentiation backward from the output layer to the input layer; the full derivation is given in Appendix B.2. A full training pipeline further requires the available annotated data to be split into three distinct sets, following the general supervised-learning workflow shown in Figure 2.6:



(a) *2D representation* **(b)** *3D representation*

Figure 2.5: *The Cost Function with respect to its weights: (a) a single-weight Perceptron, represented with a 2D curve; (b) two weight factors, shown as a 3D surface* [83].

1. ***training dataset***, the sample of data used to tune weighting factors to train the model. These data must be representative in the best way possible of the real and complete phenomenon described.

2. ***validation dataset***, the sample of data used to provide an unbiased evaluation of a model fit on the training dataset while tuning model hyperparameters. These are distinguished from parameters as they deal with the architecture design of the ANN. Moreover, unlike weights and biases, hyperparameters are tuned manually before the model training phase and comprise the number of hidden layers, the chosen activation and loss functions, the optimization algorithm, as well as the learning rate, and the number of epochs [83].

3. ***test dataset***, the sample of data used to provide an unbiased evaluation of a final model fit on the training dataset.

The choice of activation function (Section 2.2.1) has direct consequences for how well this gradient-based procedure works in practice.

The sigmoid and the hyperbolic tangent are *saturating* functions: for large positive or negative values of z they flatten out almost completely (Table 2.1), so their slope, and therefore the gradient $\nabla_\theta J(\theta)$ propagated back through them, becomes vanishingly small. In a deep network this near-zero slope is compounded, layer after layer, by the chain rule, so the earliest layers of a deep network based on sigmoid or tanh receive an almost negligible learning signal and train extremely slowly: a phenomenon known as the *vanishing gradient problem*.

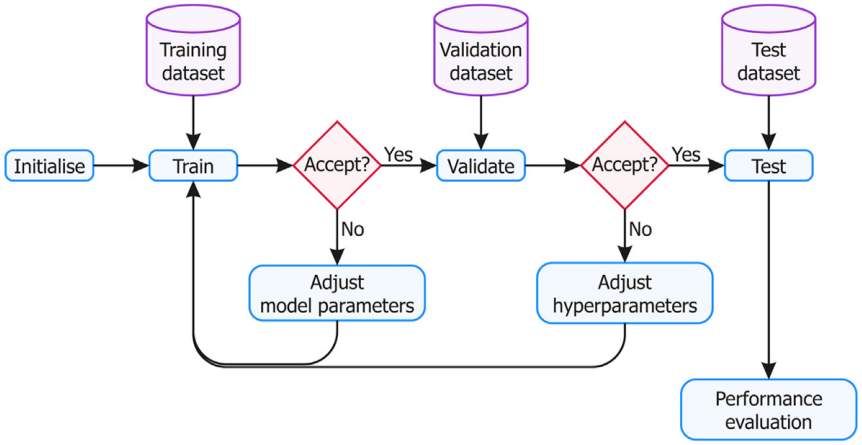


Figure 2.6: *Workflow for optimizing and evaluating the performance of a supervised-learning algorithm* [89].

Nair and Hinton showed that replacing sigmoidal units with ReLU units, whose slope is exactly 1 for any $z > 0$, substantially improves the training of deep networks for this reason [85]. The ReLU has its own weakness, however: because its slope is exactly zero for every negative input, a unit pushed into that region during training can become permanently inactive, since its gradient will remain zero regardless of any further update — a so-called "dead" neuron. The leaky ReLU addresses this by allowing a small, non-zero slope (0.01 in Table 2.1) also for negative inputs; Maas, Hannun and Ng confirmed, in the context of acoustic modeling, that this variant trains as reliably as the standard ReLU while remaining more robust to dead units [84].

2.2.4 Convolutional Neural Networks

A fully connected MLP is a poor architectural choice for image data for two related reasons. First, feeding a 2D (or 3D) image to an MLP requires flattening it into a 1D vector, which destroys the local spatial correlations between neighboring pixels or voxels that carry most of the relevant information in an image. Second, because every neuron in a dense layer is connected to every input, the number of trainable weights grows extremely quickly with image size, making fully connected networks computationally intractable for anything but small images [83].

The Convolutional Neural Network (CNN) addresses both problems by replacing dense connections with *convolutional filters* (kernels): small, spatially local weight matrices, typically 3×3 in 2D, or $3 \times 3 \times 3$ in 3D, that slide across the input, computing, at every position, the dot product between the filter weights and the corresponding

local patch of the image [83]. Because the same filter weights are reused at every spatial position (*weight sharing*), a convolutional layer both drastically reduces the number of trainable parameters relative to a dense layer of comparable size, and gives the network a degree of translation invariance directly inherited from the S-cell/C-cell architecture of the Neocognitron [81]. Different filters, learned during training, extract different features from the same input: edges, textures, or, in deeper layers, increasingly abstract and task-specific patterns (Figure 2.7).

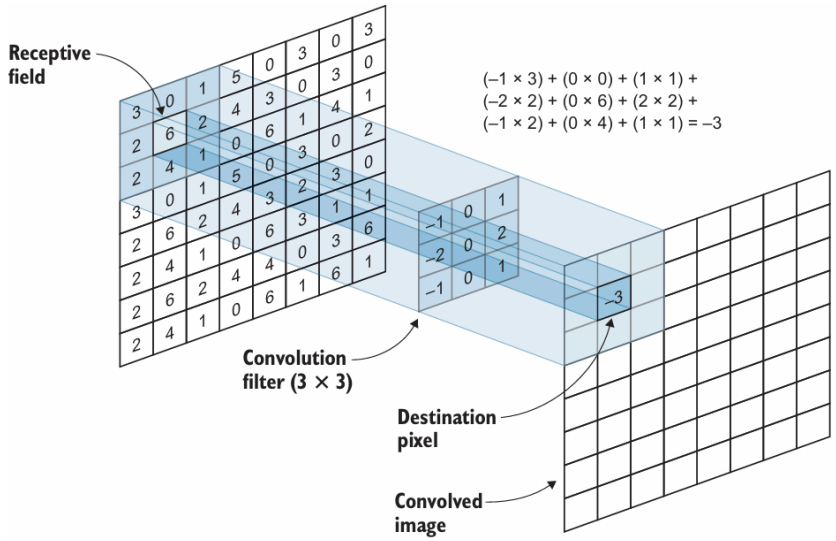


Figure 2.7: *A 3 × 3 convolutional filter is sliding over the input image: at each position the dot product between the kernel weights and the underlying image patch produces one pixel of the output feature map* [83].

Convolutional layers are typically alternated with *pooling* layers (most commonly max-pooling), which subsample the feature maps produced by the convolution — for instance, retaining only the maximum activation within each block — thereby progressively reducing the spatial resolution of the representation while increasing its receptive field and reducing the computational cost of subsequent layers [83] (Figure 2.8). A classification CNN typically ends with one or more fully connected layers that combine the final, low-resolution feature maps into a single class prediction via a softmax activation.

2.3 Fully Convolutional Networks

Image *segmentation* differs from image *classification* in that the desired output is not a single label for the whole image, but a label for every pixel (or voxel), assigning each of them to a class such as “GTV,” “mucosa,” or “background.” This prediction

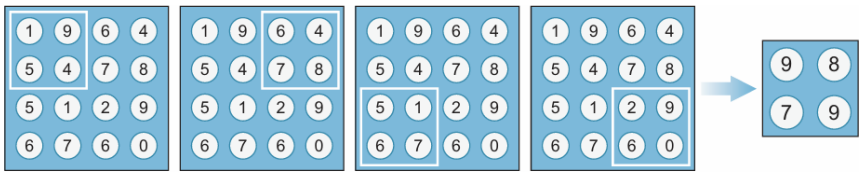


Figure 2.8: *A 2 × 2 pooling filter and strides of 2, reducing the feature map from 4 × 4 to 2 × 2* [83].

problem could be addressed by taking a standard classification CNN and replacing its final, fully connected layers with additional convolutional layers, producing a *Fully Convolutional Network* (FCN) capable of accepting an input of arbitrary size and returning a correspondingly sized, per-pixel classification map [90]. This architectural shift — from image-level to pixel-level prediction, achieved end-to-end within a single convolutional network — is the direct conceptual predecessor of every segmentation network used in this thesis.

2.3.1 The U-Net Architecture

The specific FCN variant that has become the *de facto* standard for biomedical image segmentation is the *U-Net*, introduced by Ronneberger, Fischer and Brox in 2015 [7]. The U-Net architecture is organized into two symmetric branches (Figure 2.9):

- a **contracting path** (encoder), structurally similar to a conventional CNN, which repeatedly applies pairs of 3×3 convolutions followed by a ReLU activation and a 2×2 max-pooling operation, progressively reducing the spatial resolution of the feature maps while increasing the number of feature channels — i.e., extracting increasingly abstract, but increasingly coarse-grained, semantic information;
- an **expansive path** (decoder), which mirrors the contracting path but replaces pooling with upsampling (transposed convolution) operations, progressively restoring the original spatial resolution.

The key architectural innovation of the U-Net, relative to a generic encoder-decoder FCN, is the presence of *skip connections* that concatenate the feature maps produced at each resolution level of the encoder directly to the corresponding level of the decoder [7]. Because the encoder alone discards spatial precision in exchange for semantic abstraction, and the decoder alone cannot recover information that was never propagated to it, the skip connections allow the network to combine the coarse, semantically

rich information from the bottom of the "U" with the fine-grained, spatially precise information available only in the shallow encoder layers, which is essential for accurately delineating structure boundaries. The final layer of the expansive path is a 1×1 convolution followed by a softmax activation, producing, for every voxel, the probability of belonging to each of the target classes.

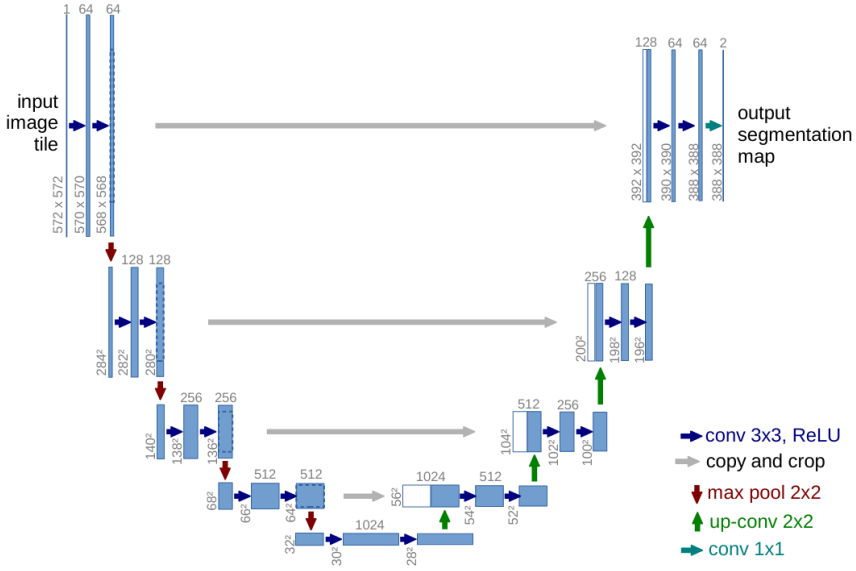
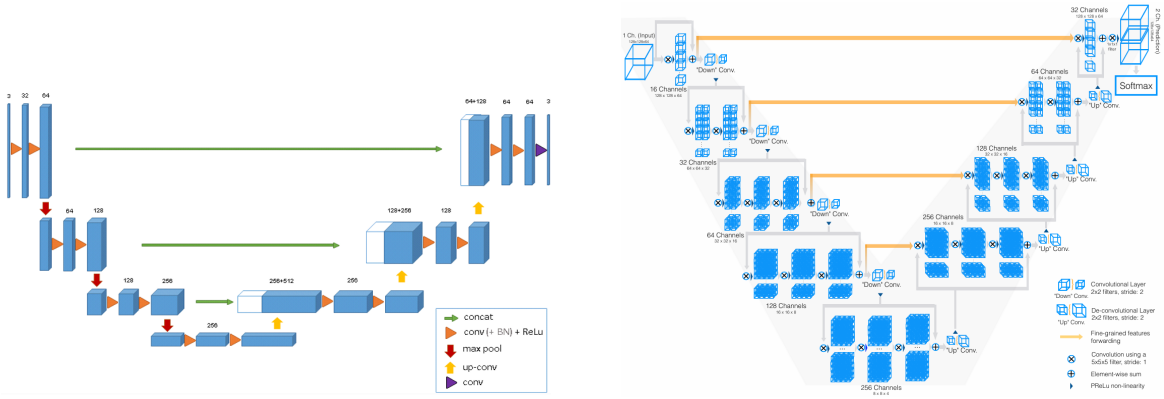


Figure 2.9: *The U-Net architecture (example for 32x32 pixels in the lowest resolution): a contracting path (encoder, left) progressively downsamples the input while extracting increasingly abstract features; an expansive path (decoder, right) progressively restores spatial resolution; gray arrows denote skip connections that concatenate encoder feature maps directly to the corresponding decoder level* [7].

The original U-Net operated on 2D image slices. Two extensions of direct relevance to volumetric CT data followed shortly after: *3D U-Net*, which replaces every 2D convolution and pooling operation with its 3D counterpart, allowing the network to exploit contextual information along all three spatial axes simultaneously rather than slice by slice [91]; and *V-Net*, which introduced a volumetric, fully convolutional architecture trained end-to-end with a novel objective function based directly on the DC rather than on a voxel-wise cross-entropy loss, defined and discussed in Section 2.2.3 [87]. Both architectures are shown in Figure 2.10.

The Dice Loss is particularly well suited to medical segmentation tasks in which the structure of interest occupies only a small fraction of the total volume — as is typically the case for a GTV relative to the full H&N CT volume — because a purely voxel-wise loss can instead be dominated by the background class and bias the network toward under-segmenting small structures. As discussed in Section 2.3.2, nnU-Net trains its

models using a combination of Dice Loss and cross-entropy loss.



(a) *3D U-Net* **(b)** *V-Net*

Figure 2.10: *Two volumetric extensions of the original U-Net architecture: (a) the 3D U-Net* [91]*; (b) the V-Net, trained end-to-end with a Dice-based loss* [87].

2.3.2 The nnU-Net

Although the U-Net architecture is conceptually simple, adapting it successfully to a specific new dataset requires a large number of interdependent design decisions — network depth, patch size, choice between 2D and 3D processing, normalization scheme, data augmentation strategy, loss function, learning rate schedule, and post-processing — that jointly determine the final segmentation performance to a much greater extent than any single architectural modification [92]. Isensee et al. observed that much of the published literature on segmentation network variants validates proposed architectural modifications against U-Net baselines that are themselves sub-optimally configured for the dataset at hand, making it difficult to determine whether a reported improvement is genuinely attributable to the proposed architecture or simply to a better-tuned baseline [92].

nnU-Net ("no-new-Net") was proposed as a direct response to this problem: rather than creating yet another architectural variant, it provides a framework that automatically and reproducibly configures a U-Net (2D, 3D, or a cascade of the two as shown in Figure 2.11) for an arbitrary new segmentation task, based entirely on a systematic analysis of the properties of the target dataset, without manual intervention [8, 92]. This automatic configuration is organized into three categories of decisions [8]:

1. **Fixed parameters**: design choices that do not require dataset-specific adap-

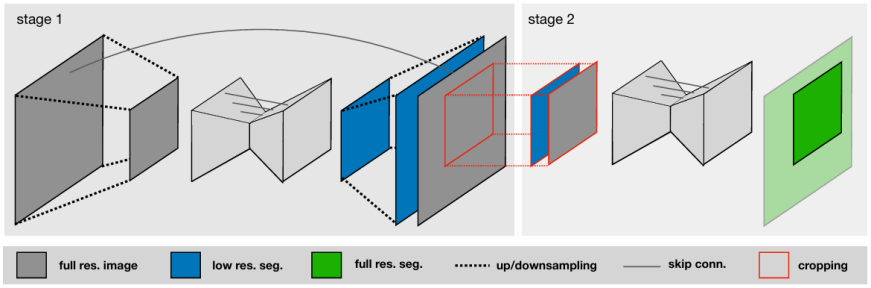


Figure 2.11: *U-Net Cascade: 3D U-Net is first trained on downsampled images (stage 1). The segmentation results of this U-Net are then upsampled to the original voxel spacing and passed as additional (one hot encoded) input channels to a second 3D U-Net, which is trained on patches at full resolution (stage 2)* [92].

tation and are therefore held constant across tasks. These include the overall architectural template (a close variant of the original 2D/3D U-Net, using leaky ReLU activations [92]), the training schedule (1000 epochs, each consisting of 250 iterations, by default), the optimizer (SGD with a fixed, decaying learning rate schedule in the original formulation), and the composite loss function, calculated as the sum of the Dice Loss and the cross-entropy loss.

2. **Rule-based parameters**: parameters that are set automatically according to heuristic rules derived from the properties of the specific dataset. Before training, nnU-Net extracts a *dataset fingerprint*, comprising image size, voxel spacing, number of image modalities, number of classes, and total dataset size; from this fingerprint, heuristic rules determine the network's patch size, batch size, target voxel spacing (whether to resample to isotropic resolution), and — critically for large volumes such as a full H&N CT — whether a cascaded, coarse-to-fine 3D U-Net is required because the target image cannot be processed at full resolution within the available GPU memory in a single forward pass.

3. **Empirical parameters**: parameters that cannot be derived from heuristics alone and must instead be selected empirically, after training, by comparing candidate configurations (2D U-Net, 3D U-Net, 3D U-Net cascade, or an ensemble thereof) on cross-validated performance; this stage also includes automated post-processing, such as removing all but the largest connected component of a predicted class when the training data indicate that the corresponding structure is anatomically expected to form a single connected region.

Figure 2.12 summarizes this three-stage configuration pipeline schematically.

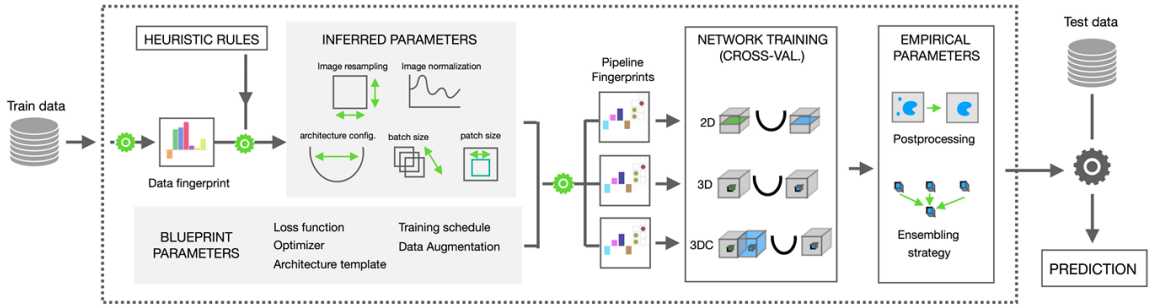


Figure 2.12: *Example of automated configuration by nnU-Net. Dataset properties are summarized in a "dataset fingerprint". A set of heuristic rules operates on this fingerprint to infer the data-dependent hyperparameters of the pipeline. These are completed by blueprint parameters, the data-independent design choices to form "pipeline fingerprints". Three architectures are trained based on these pipeline fingerprints in a 5-fold cross-validation. Finally, nnU-Net automatically selects the optimal ensemble of these architectures and performs postprocessing if required* [8].

Trained without any manual, task-specific tuning, nnU-Net achieved the highest mean DC across the majority of the tasks of the Medical Segmentation Decathlon, a challenge specifically designed to test generalization across ten heterogeneous segmentation problems (different organs, imaging modalities, and dataset sizes) without allowing task-specific manual adjustment. This method has been performed on 11 international biomedical image segmentation challenges comprising 23 different datasets and 53 segmentation tasks, testing a variety of organs, tumors and cellular structures in 2D as well as 3D images acquired by different medical imaging techniques, and some results are shown in Figure 2.13 [8, 92].

This combination of state-of-the-art accuracy and fully automatic, reproducible configuration is precisely what makes nnU-Net an appropriate computational backbone for the segmentation tools discussed in the next section, and, more generally, for a treatment-planning pipeline in which dozens of patients with heterogeneous anatomy must be processed under an identical, non-arbitrary protocol.

Table 2.2: Summary of the segmentation architectures discussed in this chapter and their relation to nnU-Net.

Architecture	Dimensionality	Key contribution
FCN	2D	first end-to-end, pixel-to-pixel dense prediction [90].
U-Net	2D	encoder-decoder with skip connections [7].
3D U-Net	3D	volumetric extension of U-Net [91].
V-Net	3D	Dice-loss training for volumetric segmentation [87].
nnU-Net	2D / 3D / cascade	self-configuring framework built on (close-to-vanilla) U-Net; backbone of TotalSegmentator [8, 92].

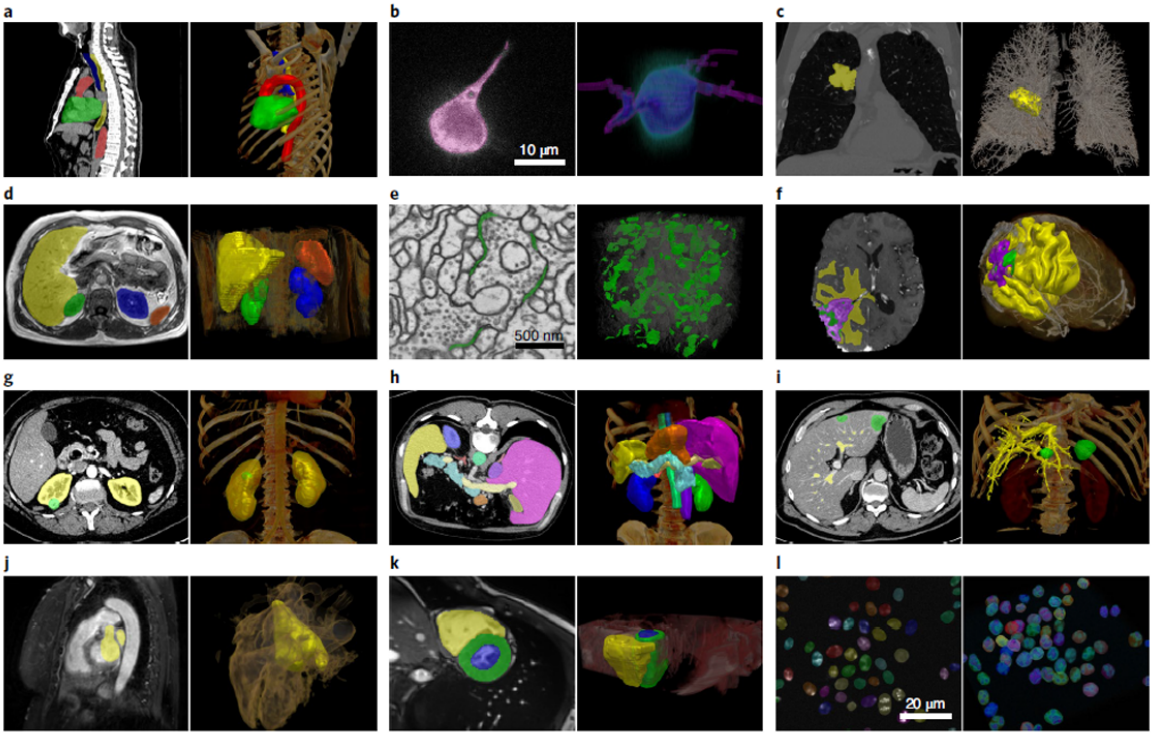


Figure 2.13: *A set of medical images segmented by the nnU-Net. Target structures for each dataset are shown in 2D projection (left) and in 3D (right). (a) Heart (green), aorta (red), trachea (blue) and esophagus (yellow) in CT images (dataset D18); (b) A549 lung cancer cells (purple) in FM (D22); (c) Lung nodules (yellow) in CT images (D6); (d) Liver (yellow), spleen (orange), left and right kidneys (blue and green, respectively) in T1 in phase MRI (D16); (e) Synaptic clefts (green) in EM scans (D19); (f) Edema (yellow), enhancing tumor (purple), necrosis (green) in MRI (T1, T1 with contrast agent, T2, FLAIR) (D1); (g) Kidneys (yellow) and kidney tumors (green) in CT images (D17); (h) Thirteen abdominal organs in CT images (D11); (i) Hepatic vessels (yellow) and liver tumors (green) in CT (D8); (j) Left ventricle (yellow) in MRI (D2); (k) Right ventricle (yellow), left ventricular cavity (blue) and myocardium of left ventricle (green) in cine MRI (D13); (l) HL60 cell nuclei in FM (D21)* [8].

2.4 Applications of AI in Cancer Imaging

Within oncological imaging, AI-based methods are typically grouped into three main clinical tasks: *detection*, *characterization*, and *monitoring* of tumors over time [93]. Detection, commonly referred to as Computer-Aided Detection (CADe) concerns the localization of suspicious findings within an image, and was among the earliest clinical applications of pattern-recognition software in radiology, predating deep learning by several decades [94]; CADe tools are designed to flag regions of interest for the radiologist's attention, functioning as a second reader intended to reduce oversight errors rather than to replace clinical judgment [93]. Characterization, the task most relevant to this thesis, includes segmentation, staging, and the quantitative description of a lesion's extent, and is the step on which subsequent dose-calculation and treatment-

planning tasks directly depend [93].

Auto-segmentation specifically has progressed through several distinct methodological generations: from intensity-based and shape-modelling techniques, through atlas-based methods, to early (non-deep) machine learning and, most recently, the deep-learning approaches described in Sections 2.3–2.3.2, a progression comprehensively reviewed by Harrison et al. [89].

2.4.1 Interobserver and Intraobserver Variability

In current clinical practice, GTVs and OARs are typically manually defined by experienced radiologists, involving clinicians drawing contours by hand to delineate anatomical regions of interest (ROIs). This practice is not only time-consuming, but is also intrinsically affected by variability between different observers (*interobserver variability*) and even between repeated delineations by the same observer (*intraobserver variability*) [95]. This variability is not a peripheral methodological detail: multiple studies across disease sites have linked inconsistent contouring directly to inferior disease control and increased toxicity, and cooperative-group trials have specifically identified contouring deviations as a contributor to inferior outcomes in H&N and gastrointestinal cancer protocols [95], an effect that is compounded when the imaging modality itself offers poor soft-tissue contrast (Figure 2.14).

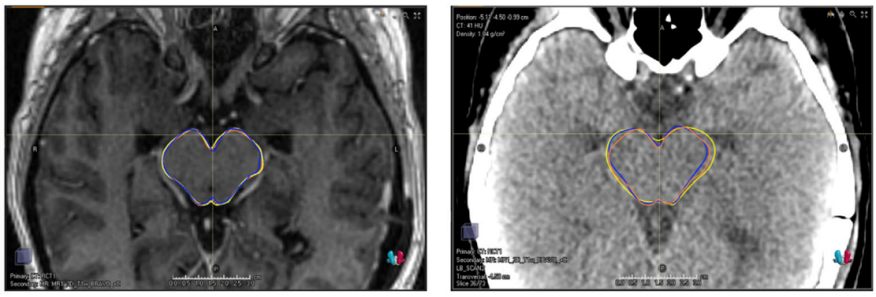


Figure 2.14: *Illustration of the effect of soft-tissue contrast on segmentation reliability. The brainstem was delineated with different colours by the same three clinicians on a magnetic resonance scan (left) and a CT scan (right). The decreased soft-tissue contrast of the CT scan leads to larger interobserver variability* [89].

Direct, quantitative evidence of this variability comes from studies comparing multiple expert delineations of the same structures. Teguh et al., evaluating atlas-based auto-segmentation of swallowing and mastication structures against multiple observers, is one of the earlier studies to formally quantify this problem in the H&N, as shown in

Figure 2.15 [77].

More recent work directly compares deep-learning-based auto-segmentation against a panel of multiple radiation oncologists rather than against a single reference observer, in order to determine whether automated contours fall within the range of variability that is already accepted between human experts, rather than requiring automated contours to match a single, inevitably somewhat arbitrary, "ground truth" [78]. Wong et al. found that, for OAR across several disease sites including H&N, deep-learning-based contours showed differences from expert contours comparable to the differences already observed between different expert observers, while requiring a small fraction of the contouring time (for example, on the order of tens of seconds rather than tens of minutes per case) [78]. This is the central empirical argument, beyond mere convenience, for treating an appropriately validated automated segmentation as methodologically defensible: the alternative, a single manual delineation, is not a variability-free "ground truth" but merely one sample from an underlying distribution of equally plausible expert contours.

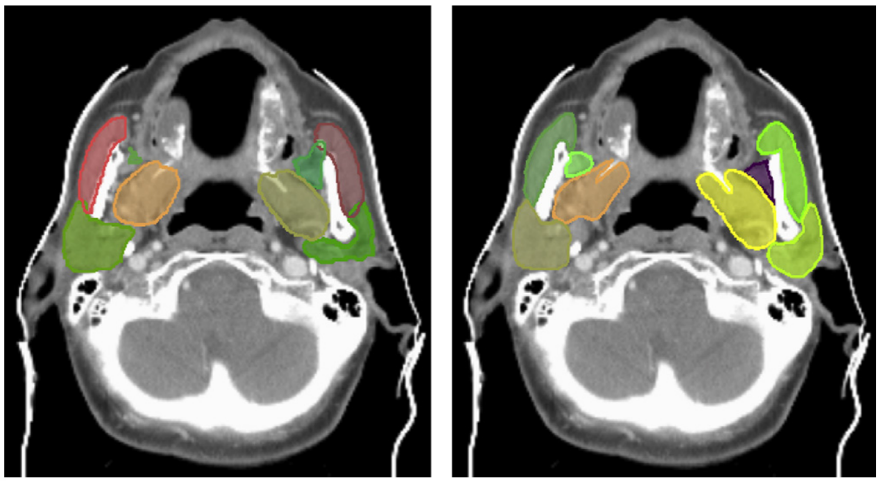


Figure 2.15: *Autocontouring of OARs: mastication muscles and parotid glands (left) vs. reference contours (right)* [77].

Sherer et al. provide a broader taxonomy of metrics used to evaluate auto-segmentation performance against this backdrop, distinguishing geometric metrics — such as the DC and Hausdorff distance, already introduced in Equation (2.5) and Section 2.2.3 — from dosimetric metrics, which capture the downstream effect of a segmentation difference on the delivered dose distribution and are arguably of most direct relevance to a Monte Carlo dosimetric pipeline such as the one developed in this thesis, as well as time-based metrics and qualitative clinical-acceptability scoring [95].

Table 2.3: Representative DC values reported in the literature for deep-learning-based auto-segmentation of H&N structures, illustrating typical agreement between automated and expert-derived contours.

Structure(s)	DSC	Comparison basis
Masseters	0.87	retrospective test set [96]
Medial pterygoids	0.80	retrospective test set [96]
Larynx	0.81	retrospective test set [96]
Pharyngeal constrictor	0.69	retrospective test set [96]
Whole body divided in 104 structures	0.943 (mean)	independent test set, multiple pathologies [97]
CNS/H&N/prostate OARs & CTVs	comparable to inter-observer range	against multiple expert observers [78]

2.4.2 Automated Segmentation Tools: TotalSegmentator and CERR

The pipeline developed in this thesis (Chapter 4) combines two independent, AI-based segmentation sources: *TotalSegmentator*, a standalone nnU-Net–based tool, and a library of validated deep-learning segmentation models distributed through the *CERR* platform.

TotalSegmentator

TotalSegmentator is an open-source tool for the automatic segmentation of anatomical structures in CT (and, in a more recent extension, MRI) images, developed by Wasserthal et al. and trained, via the nnU-Net framework described in Section 2.3.2, on a dataset of 1204 CT examinations sampled from routine clinical practice at the University Hospital Basel [97]. The CT version segments 104 anatomical structures — 27 organs, 59 bones, 10 muscles, and 8 vessels — and reported a DC of 0.943 on an independent test set spanning a wide range of clinical pathologies, scanners, and acquisition protocols, outperforming another publicly available multi-organ segmentation model on the same comparison dataset (0.943 vs. 0.871) [97].

Two aspects of the TotalSegmentator methodology are directly relevant to the way it is used in this thesis:

- **Training-data heterogeneity and robustness.** Unlike earlier multi-organ segmentation models trained on comparatively small and homogeneous datasets (a single scanner, a single acquisition protocol), TotalSegmentator was deliberately trained on real-world, heterogeneous clinical data — with and without

contrast agent, from different scanners and acquisition settings [97].

- **Iterative, semi-automated annotation workflow.** The 104-structure ground truth itself was generated through an iterative bootstrapping procedure: an initial nnU-Net model, trained on a small number of manually segmented cases, was used to pre-segment additional cases, whose predictions were then manually corrected by the annotating physicians and fed back into the next round of training [97]. This is methodologically analogous to (though independent of) the AI_MIGHT annotation strategy discussed in Section 2.5.

An MRI-compatible extension of the same tool, *TotalSegmentator MRI*, trained jointly on MRI and CT data to segment 80 structures independently of MRI sequence, has since been released [98]; while the present thesis works exclusively from CT-derived geometry, this extension is mentioned here as a natural direction for future multimodal work within the same pipeline.

CERR

The Computational Environment for Radiotherapy Research (CERR) is an open-source, MATLAB-based software platform, originally introduced by Deasy, Blanco and Clark, for the import, visualization, and quantitative analysis of radiotherapy treatment planning data [99]. Beyond DICOM import and dose-volume/radiomic feature extraction, CERR distributes a library of validated, deep-learning-based auto-segmentation models — deployed via containerized environments and callable directly from the CERR interface [100].

In this thesis, CERR-distributed models are used as a second, independent AI-based segmentation source alongside TotalSegmentator, providing structures absent from the TotalSegmentator task set used in this pipeline: the brainstem, the oral cavity, and the laryngeal soft tissue (as opposed to the laryngeal airway alone, which TotalSegmentator does segment) are contoured exclusively via CERR-distributed models and incorporated directly into the combined patient geometry. One of the validated models distributed through this library is a prospectively evaluated network for the larynx, pharyngeal constrictor muscle, masseters, and medial pterygoids, trained on 242 H&N cancer patients and benchmarked against expert inter-observer variability [96], a structure set that closely matches several of the CERR-derived masks used in this pipeline.

2.5 The AI_MIGHT Project

AI_MIGHT (*Artificial Intelligence methods applied to Medical ImaGes to enHance and personalize BNCT Treatment planning*) is a project developed at the University of Pavia, funded by the Istituto Nazionale di Fisica Nucleare (INFN) under its Young Researchers Grants scheme in 2021, concerned with the implementation of a deep-learning tool for the automatic segmentation of tumor volumes from CT and MRI examinations. Its stated aim is to speed up the preliminary clinical evaluation of medical images for BNCT treatment planning: manual GTV contouring, as discussed in Section 2.4.1, is time-consuming, and delays in its completion can in turn delay the start of patient positioning, beam configuration, and Monte Carlo dose calculation, all of which are themselves computationally demanding steps in the treatment planning workflow. By providing an automatically generated tumor contour as soon as imaging is available, AI_MIGHT allows medical physicists to begin this downstream work without first waiting for a clinician to complete manual segmentation; the automatically generated contour is then reviewed, approved, or corrected by the treating clinicians before being used in the final dose calculation. Beyond individual patient workflows, a fast, automated tumor segmentation tool is also valuable for BNCT methodological research more broadly, since evaluating or improving a treatment planning system typically requires a much larger volume of segmented cases than can realistically be manually contoured; AI_MIGHT was explicitly conceived to be coupled with the OpenPINT [12] treatment planning system (Section 3.3) for exactly this purpose.

The AI_MIGHT segmentation model is an nnU-Net (Section 2.3.2), trained on a standardized dataset of publicly available H&N and brain-tumor CT and MRI examinations [8, 92], comprising 1934 H&N and 230 brain-tumor cases with corresponding, physician-drawn RTSTRUCT[3] files used as training ground truth. Prior to training, the images were cropped to remove anatomically irrelevant background and reduce training time, and the dataset was subsequently split into training, validation, and test sets following nnU-Net's standard automated procedure (Section 2.2.3), with segmentation performance evaluated using the geometric metrics discussed in Section 2.4.1. This model was applied, in a prior study conducted within the same research group,

[3]RTSTRUCT (RT Structure Set) is a DICOM-RT object used to store contoured regions of interest, such as the GTV or an OAR, as a set of 2D polygonal contours drawn slice-by-slice on a reference CT or MRI study, together with metadata linking them back to that study.

to a cohort of glioblastoma patients, comparing the BNCT dosimetry obtained from physician-drawn ground-truth GTV contours against that obtained from the corresponding AI_MIGHT-generated segmentation [10].

In the pipeline developed for this thesis, AI_MIGHT is used in an analogous role: it is the tool responsible for generating the GTV contour from each patient's H&N CT. This is a distinct task from, and not an alternative to, the organ and tissue segmentation role played by TotalSegmentator and CERR (Section 2.4.2), which are instead used exclusively for the surrounding OARs and normal-tissue structures. The three tools therefore operate on non-overlapping anatomical targets within the same geometry-generation pipeline: AI_MIGHT delineates the tumor, while TotalSegmentator and CERR jointly delineate everything else.

Chapter 3

Treatment Planning Systems for Boron Neutron Capture Therapy

This chapter describes the computational infrastructure, the Treatment Planning System (TPS), that turns a segmented patient geometry, a neutron source description, and a set of radiobiological weighting assumptions into a predicted dose distribution that can be used for clinical decision-making.

3.1 General Framework of TPS

A Treatment Planning System is the computational tool that translates a clinical prescription into a spatial dose distribution and a set of deliverable machine or irradiation parameters. In external-beam radiotherapy, the treatment planning process is conventionally organized into a sequence of steps: identification of the tumor and of the OARs through medical imaging; selection of a reproducible patient positioning and immobilization strategy; selection and optimization of the beam arrangement; evaluation of the resulting dose distribution over the volumes of interest; and, finally, translation of the prescribed dose into deliverable machine settings. The entire process is carried out under the joint responsibility of the medical physicist, who is responsible for the dosimetric accuracy of the plan, and the radiation oncologist, who prescribes the treatment and approves the plan before delivery [108, 109].

3.1.1 The ICRU Volume Definitions

A prerequisite for any quantitative treatment planning, in BNCT as in conventional radiotherapy, is a shared vocabulary for describing the anatomical volumes involved. This vocabulary was formalized by the International Commission on Radiation Units and Measurements (ICRU) in Report 50 and its supplement, Report 62 [110, 111], and has since been extended to intensity-modulated, stereotactic, particle, and MRI-guided radiotherapy by later reports, without altering its core definitions [108]. The framework distinguishes:

- the **Gross Tumor Volume (GTV)**, the palpable or imageable extent of the malignant growth, identified through clinical examination and imaging (CT, MRI, PET) and, when applicable, separately reported for the primary tumor (GTV-T) and for any nodal involvement (GTV-N) [108, 111]; used directly, together with the CTV and PTV defined below, by essentially every computerized TPS for automatic margin generation, segmentation-based beam geometry definition, and DVH-based plan evaluation [109];

- the **Clinical Target Volume (CTV)**, the GTV plus a margin accounting for suspected, sub-clinical microscopic disease that cannot be resolved by imaging but must nonetheless be treated to achieve local control [111];

- the **Internal Target Volume (ITV)**, introduced in ICRU Report 62, which adds to the CTV a margin accounting for physiological, patient-frame-relative motion of the target (e.g., breathing, swallowing) [111];

- the **Planning Target Volume (PTV)**, a purely geometric construct that adds to the ITV a further margin for set-up uncertainty and machine tolerances, ensuring that the prescribed dose is delivered to the CTV despite unavoidable positioning and delivery uncertainties [111];

- the **Organs at Risk (OARs)**, non-target structures whose radiosensitivity constrains the achievable treatment, each generally associated with a tolerance dose beyond which the probability of a clinically significant complication becomes unacceptable [111, 112].

Figure 3.1 illustrates the resulting hierarchy of nested volumes graphically.

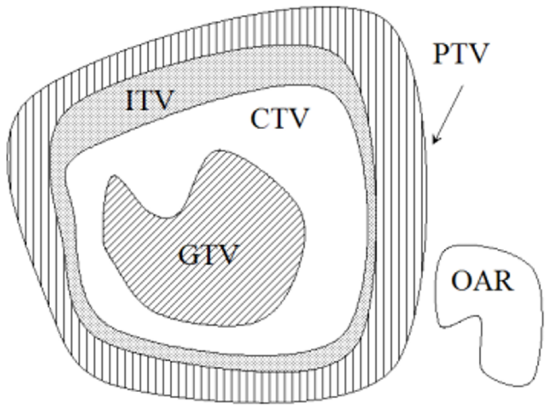


Figure 3.1: *Graphical representation of the volumes-of-interest, as defined by the ICRU 50 and 62 reports* [109].

This framework was devised for photon beams that can be geometrically shaped and conformed to a target; as discussed in Section 3.2.1, several of these concepts require reinterpretation, rather than literal transposition, when applied to BNCT.

3.1.2 Dose Reporting and Plan Evaluation

Once target and OAR volumes are defined, a treatment plan is evaluated through standard dosimetric quantities. Isodose curves, superimposed on the anatomical images, display the spatial distribution of the absorbed (or weighted) dose, typically normalized to a reference or maximum value. Dose-volume histograms (DVHs) summarize the same information in a volume-independent form, plotting the cumulative fraction of the relevant volume receiving at least a given dose; a DVH curve shifted toward higher doses at the same cumulative volume indicates better target coverage or, for an OAR, poorer sparing [109]. Successive ICRU reports have progressively standardized which DVH-derived metrics must be reported for a given technique, as summarized in Table 3.1.

As will become apparent in Section 3.2.1, BNCT planning reports a formally analogous set of quantities ($D_{2\%}$, $D_{50\%}$, $D_{98\%}$, D_{mean}, expressed in RBE- or isoeffective-weighted units), but the constraint that ultimately determines the treatment time is, in essentially every published protocol, an OAR dose-volume limit rather than a target coverage requirement [43].

Despite this formal analogy, no internationally agreed standard for BNCT dose reporting currently exists, in marked contrast to the ICRU-report lineage of Table 3.1

Table 3.1: Dose-reporting parameters recommended by successive ICRU reports for external-beam radiotherapy, summarized from [108]. $D_{x\%}$ denotes the dose received by $x\%$ of the structure volume.

ICRU Report	Reported quantities
50 & 62	Dose at the ICRU reference point; minimum ($D_{\min}$) and maximum ($D_{\max}$) dose to the PTV
83	Full dose–volume histograms; $D_{2\%}$, $D_{50\%}$, $D_{98\%}$, and mean dose (D_{mean}) to the PTV; D_{mean} and $D_{2\%}$ to OARs and to the planning organ-at-risk volume
91	Same quantities as Report 83, together with dose homogeneity and conformity indices
93	RBE-weighted D_{mean} and $D_{\max}$ for targets and OARs; linear energy transfer distribution, where available
97	Daily adaptive-plan parameters, deformable registration and intrafraction motion metrics

for conventional radiotherapy. Because the weighting factors that convert each of the four physical dose components into a weighted or isoeffective dose (Section 1.3.2) are specific to the neutron spectrum, boron carrier, and target tissue considered at a given facility, a single weighted-dose value computed at one center cannot in general be transferred to, or reproduced at, another center without full knowledge of the underlying component doses and weighting factors used to obtain it [113]. This is not merely a matter of incomplete standardization effort: the CBE used to weight the boron-dose component alone has been reported to carry an uncertainty of 16–36%, depending on the ^{10}B carrier [113], so that reporting only a single total weighted-dose figure, without the four underlying physical dose components, discards precisely the information a different center would need to reinterpret that figure under its own weighting convention. For this reason, the component-wise dose maps preserved throughout the OpenPINT pipeline (Section 3.3.5), rather than only their biologically weighted combination, are directly relevant to the broader reproducibility concerns discussed in this Chapter.

3.2 Treatment Planning System in BNCT

Building a BNCT dose calculation requires a sequence of interdependent computational choices: how the patient anatomy is reconstructed for Monte Carlo transport, how tissue and boron composition are assigned to that geometry, how the neutron source is defined, how the resulting dose is actually calculated, how the patient is positioned relative to the (typically fixed) beam, and, finally, how the resulting plan is evaluated. This Section addresses these steps in turn, and closes with a survey of the TPSs that

have implemented them at different BNCT facilities worldwide.

3.2.1 BNCT Dedicated Treatment Planning Paradigm

Reusing the conventional-radiotherapy TPS machinery for BNCT is not possible, for reasons that are physical rather than merely technical.

Chemical Selectivity

In photon, electron, or proton therapy, tumor selectivity is obtained primarily by shaping and conforming the beam to the target volume; the ICRU volume hierarchy of Section 3.1.1 is built around this geometric logic. In BNCT, as discussed at length in Section 1.2, selectivity is instead obtained biochemically, through the preferential uptake of the boron carrier in tumor cells. The neutron beam itself is comparatively unshaped: epithermal beams are not highly collimated and lose directionality rapidly with depth in tissue as neutrons thermalize and diffuse [114]. A BNCT TPS must therefore model the spatial and temporal distribution of the boron carrier, and its pharmacokinetics relative to the timing of irradiation, in addition to the neutron and photon transport itself [114].

A Mixed Radiation Field

As established in Section 1.3.1, the absorbed dose at any point in tissue during BNCT is the sum of four physically and radiobiologically distinct components (boron, nitrogen, fast-neutron, and photon), each with a different depth and material dependence. Photon-therapy dose engines, whether based on empirical pencil-beam kernels or model-based algorithms, exploit the comparatively simple depth-dose behavior of a single low-LET radiation quality and can therefore rely on partially analytical approximations. No such simplification is available for the strongly scatter-dominated, multi-component neutron/photon field of BNCT: essentially every BNCT TPS developed to date relies on general-purpose Monte Carlo transport codes, most commonly MCNP, to compute the four dose components with adequate accuracy [114, 115, 116]. This has direct practical consequences: Monte Carlo calculations are substantially more computationally demanding than the deterministic or semi-analytical algorithms of conventional TPS, and planning throughput in BNCT has historically been limited

by the available computing resources rather than by algorithmic considerations alone [114].

Dose Prescription to Normal Tissue

Because the dose-effect relationship of the BNCT mixed field on tumor tissue is not established with the precision available for photon therapy, BNCT protocols do not prescribe a target dose in the way conventional radiotherapy does. Instead, treatment time is set by a tolerance-dose constraint on the most radiosensitive dose-limiting normal tissue, historically the brain in glioma protocols and the oral/pharyngeal mucosa in H&N protocols (Section 1.3.4), while the tumor dose is simply reported as a consequence of that constraint rather than imposed as a planning goal [23, 43]. A BNCT TPS must therefore be built around dose-volume constraint evaluation on OARs as the primary planning driver, with target coverage reported for outcome interpretation rather than for direct optimization.

A Rigid Beam Geometry

Section 3.2.6 discusses this point in detail: at both reactor-based and accelerator-based facilities, the neutron beam port is spatially fixed, and beam-angle optimization, which is routine in conventional radiotherapy, is replaced by patient repositioning relative to a small number of achievable beam orientations.

Taken together, these four factors explain why every BNCT TPS surveyed in Section 3.2.8 converges on the same basic architecture, namely automated patient-specific geometry construction, full Monte Carlo transport, and dose-volume-constraint-based plan evaluation, regardless of the specific facility or institution that developed it.

3.2.2 Patient Geometry Reconstruction

The first computational step of any BNCT treatment plan is the construction, from the patient's medical images, of a geometric model suitable as input to a Monte Carlo transport code (Figure 3.2). Three main reconstruction strategies have been used historically, with a fourth, hybrid strategy developed more recently to combine their respective advantages.

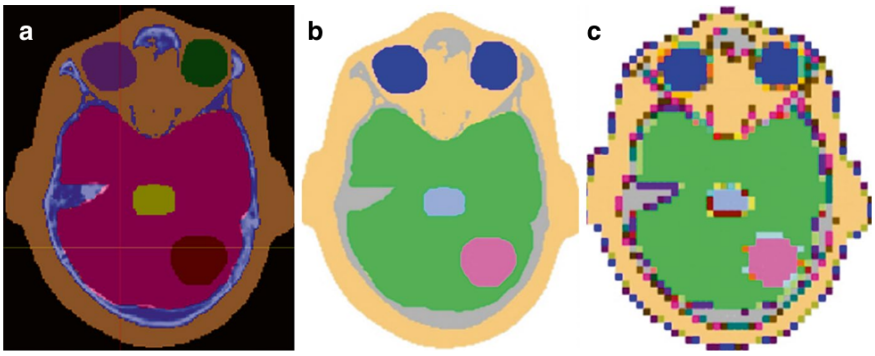


Figure 3.2: *Example of constructing computational models used for dose calculation in the treatment planning process* [117].

Voxel models

The voxel approach superimposes a three-dimensional grid of rectangular parallelepipeds on the CT volume; each voxel, generally coarser than the native CT pixel, is assigned a single material and density based on the Hounsfield units it contains [118]. Voxel geometries are straightforward to generate automatically from segmented images and are the most widely adopted reconstruction strategy in BNCT TPS, including NCTPlan and its predecessor MacNCTPlan [118, 119], as illustrated by the MiMMC voxelization example of Figure 3.3. Their principal drawback is a resolution trade-off: coarser voxels reduce the number of surfaces the transport code must track, and hence the computation time, but approximate curved anatomical boundaries with a staircase pattern that can bias the dose at shallow depths and near surfaces, most importantly for skin, one of the classical dose-limiting OARs in BNCT [118]. A systematic analysis of the standard 1 cm voxel resolution used in NCTPlan and MacNCTPlan found this resolution adequate for most reported benchmark and phantom comparisons, while identifying degraded accuracy at very shallow depths that could be mitigated through an improved surface-voxel material-assignment strategy [118].

Univel models

The univel (uniform-volume-element) approach, adopted by the SERA system, is geometrically similar to the voxel approach but assigns a material label to each individual image pixel/voxel rather than to a coarser superimposed grid, preserving the full spatial resolution of the source medical images [120], as illustrated in Figure 3.4. Fast, largely integer-arithmetic line-rasterization algorithms allow particle tracks to be advanced

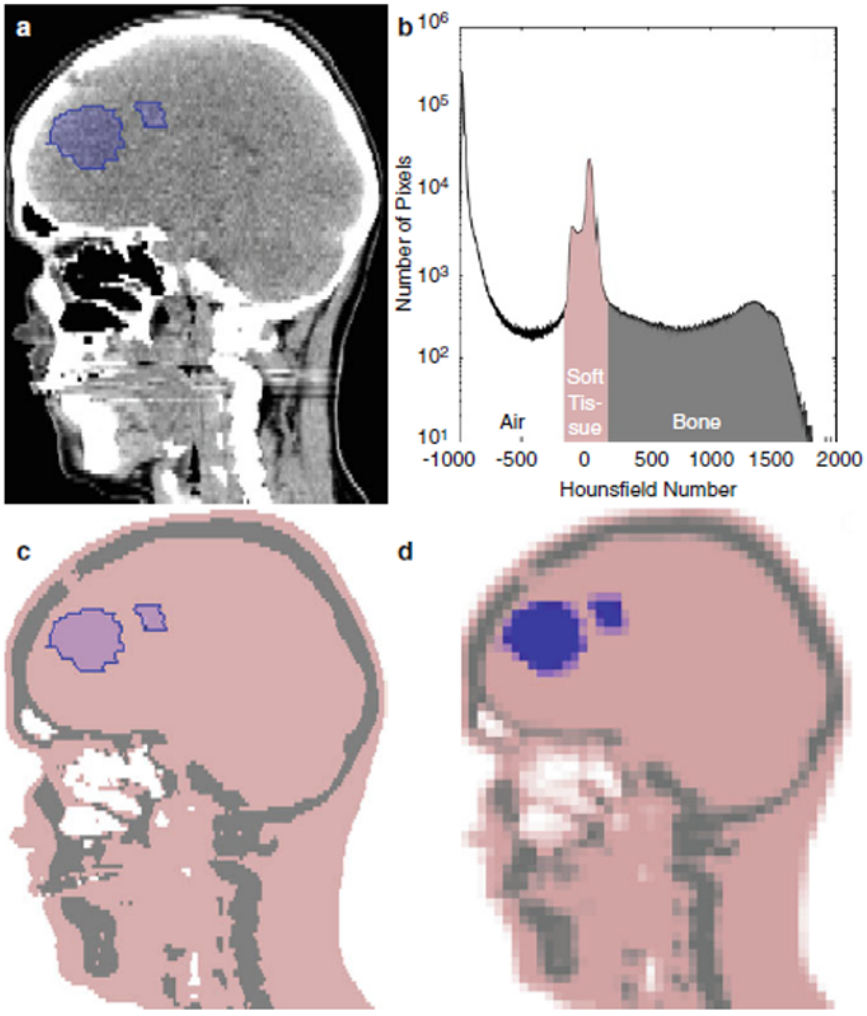


Figure 3.3: *Construction of a voxel model from CT image data in the MiMMC treatment planning system.* (a) *Sagittal CT image slice with contoured target volumes shown in blue.* (b) *Image intensity histogram showing the range of Hounsfield numbers for the selected tissues, i.e. air, soft tissue, and bone and* (c) *the resulting thresholded sagittal image.* (d) *The corresponding slice through a 4-mm voxel model* [54].

efficiently through a univel geometry that may comprise several million elements, at an execution-time cost reported to be five to ten times lower than the NURBS-based reconstruction it replaced, while retaining comparable statistical accuracy [120]. The price of this fidelity is the manual or semi-automated labor required to assign a material to every pixel of every relevant structure, a task that can become time-consuming for anatomically complex regions [120].

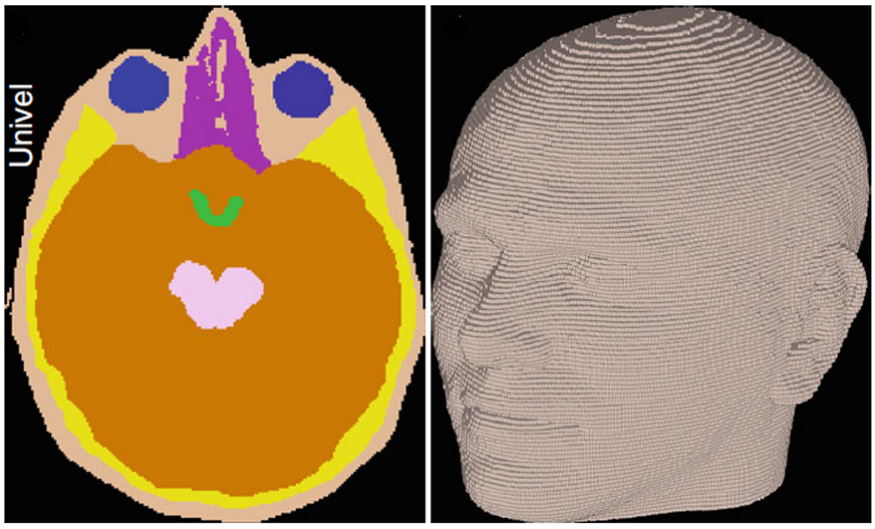


Figure 3.4: *2D slice Univel model where air, skin, skull, brain, brain stem, eyes, optic chiasm, and sinuses are defined. Each univel corresponds to a pixel in the MR images used to derive the model, which have a pixel size of 1 × 1 × 2 mm3 (left). Right: the corresponding 3D representation* [54].

NURBS models

Non-Uniform Rational B-Spline (NURBS) models represent anatomical surfaces as continuous, free-form curves and surfaces defined by a set of control points placed manually or through edge-detection algorithms on the medical images, as used in the early BNCT_Rtpe system that preceded SERA [120] and shown in Figure 3.5. NURBS reconstruction preserves smooth external anatomy with high fidelity but can suffer from poorly defined internal structures (e.g., sinuses, skull architecture) and from the risk of overlapping surfaces if control points are placed inconsistently, which can cause particles entering the overlap region to be lost from the simulation.

Multi-resolution and MultiCell models

A fourth strategy, intermediate between the accuracy of a fine voxel grid and the computational economy of a coarse one, allows the voxel size to vary locally as a function of anatomical heterogeneity: finer voxels along material interfaces, coarser voxels in homogeneous regions. This multi-resolution philosophy underlies the MultiCell model developed by the Argentine Comisión Nacional de Energía Atómica group, in which the patient volume is reconstructed as a combination of variable-sized parallelepiped cells suitable for direct import into MCNP, keeping the total cell count, and hence the computational burden, low while preserving high geometric precision at material boundaries [121]. The same group subsequently developed the photon-isoeffective dose

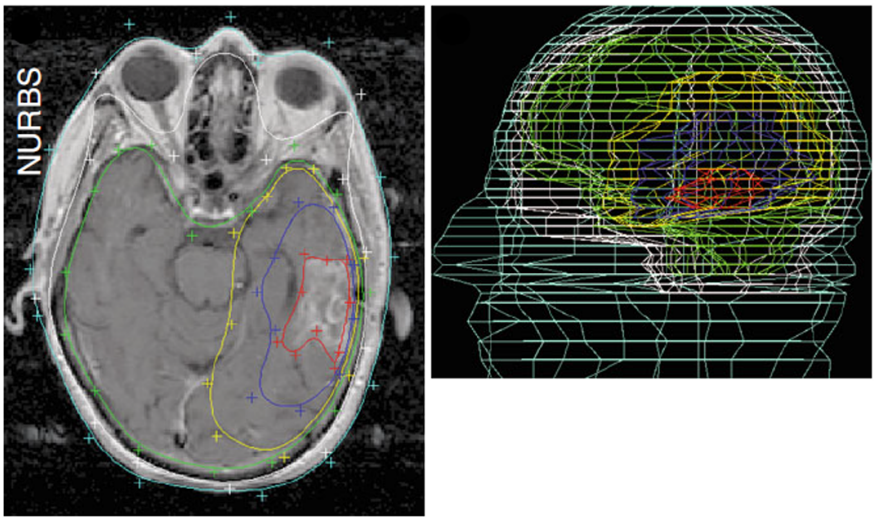


Figure 3.5: *Left: control points indicated by plus signs guide the placement of the spline in the NURBS model, which is overlaid on an MR image. Right: wire-frame rendering of the NURBS model shows the outer boundaries of skin (cyan), skull (white), brain (green), CTV (yellow), edema (blue), and GTV (red)* [54].

formalism used throughout this thesis (Section 1.3.3), and its MultiCell-based TPS has been applied beyond cranial BNCT, including thoracic geometries for non-small-cell lung cancer [122].

Table 3.2 summarizes the four reconstruction strategies described above. The pipeline developed for this thesis (detailed in Chapter 4) adopts a fixed-resolution voxel approach, consistent with the automated, batch-oriented segmentation workflow of Chapter 2: TotalSegmentator and CERR derived organ masks and the AI_MIGHT-derived GTV mask are combined into a single labeled NIfTI volume and rasterized directly into an MCNP-compatible lattice, following the same voxel philosophy as NCTPlan but built on an automated, rather than manually contoured, segmentation source.

Table 3.2: Comparison of patient geometry reconstruction strategies used in BNCT treatment planning.

	Resolution	**Automation**	**Main weakness**	**Representative TPS**
Voxel	Coarser than native image (typ. ≥1 cm)	High	Staircase approximation of curved boundaries, shallow-depth bias	NCTPlan / MacNCTPlan [118, 119]
Univel	Native image resolution	Low–moderate (manual labelling)	Labor-intensive material assignment	SERA [120]
NURBS	Continuous (control-point defined)	Low (manual/semi-automated control points)	Overlap risk, poor internal-structure fidelity	BNCT_Rtpe [120]
MultiCell / multi-resolution	Locally variable	Moderate–high	Implementation complexity	Argentine MultiCell TPS [121]

3.2.3 Tissue Composition and Boron Concentration Assignment

Unlike photon or electron dose calculation, which depends essentially only on the electron density derived from the CT, BNCT dose calculation depends on the full elemental composition of every tissue, since the boron, nitrogen, hydrogen, and photon dose components each depend on a different set of nuclear cross sections (Section 1.2.1). ICRU Report 44, integrated with ICRU Report 46, provides the reference elemental compositions and densities most commonly adopted in BNCT treatment planning [52]. Because tissue composition uncertainty propagates directly and non-trivially into the boron, nitrogen, and fast-neutron dose components, the choice of composition data set for a given anatomical structure is a documented source of systematic dosimetric uncertainty in BNCT, distinct from the interaction-coefficient uncertainty relevant to conventional dosimetry [52].

Beyond elemental composition, the ^{10}B concentration must be assigned separately for every tissue class, since the boron-dose KERMA factors used in Monte Carlo transport are conventionally tabulated per unit (1 ppm) ^{10}B concentration in any material and must be explicitly scaled by the tissue-specific concentration before being combined into a physical dose [123]. As discussed in Section 1.2.2, these concentrations are not measured directly for every tissue and patient, but are instead derived from blood boron measurements through assumed or pharmacokinetically modeled tumor-to-blood and mucosa-to-blood ratios; using a placeholder rather than a tissue-appropriate concentration value can suppress the boron-dose component by an order of magnitude and is therefore one of the most consequential configuration choices in a BNCT dose calculation.

Table 3.3 reports the MCNP material definitions used for the five tissue classes (bone, brain, skin, air, and GTV) in the OpenPINT validation benchmark described in Section 3.3.7. Brain, skin, and GTV are treated as hydrogenous soft-tissue materials and assigned the MCNP light-water thermal-scattering treatment, following the water thermal-treatment convention established in the NCTPlan/MacNCTPlan BNCT-planning lineage [12, 124]; bone and air are not assigned a thermal-scattering treatment in this benchmark material library. Note that the GTV composition differs from the surrounding brain only in its assigned ^{10}B mass fraction, isolating the effect of boron uptake from any assumed change in the underlying soft-tissue elemental composition.

Table 3.3: MCNP material definitions used in the cylindrical phantom benchmark.

Class	MCNP material definition
BONE	$\rho = 1.610\ \mathrm{g\,cm^{-3}}$; H 0.034, C 0.155, N 0.042, O 0.435, Na 0.001, Mg 0.002, P 0.103, S 0.002985, Ca 0.225, ^{10}B 0.000015; no thermal treatment.
BRAIN	$\rho = 1.040\ \mathrm{g\,cm^{-3}}$; H 0.107, C 0.145, N 0.022, O 0.712, Na 0.002, P 0.004, S 0.002, Cl 0.003, K 0.002985, ^{10}B 0.000015; `lwtr.01`.
SKIN	$\rho = 1.090\ \mathrm{g\,cm^{-3}}$; H 0.100, C 0.143, N 0.034, O 0.708, Na 0.002, P 0.003, S 0.005, Cl 0.002, K 0.002985, ^{10}B 0.000015; `lwtr.01`.
AIR	$\rho = 0.001205\ \mathrm{g\,cm^{-3}}$; C 0.000124, N 0.755268, O 0.231781, Ar 0.012827; no thermal treatment.
GTV	$\rho = 1.030\ \mathrm{g\,cm^{-3}}$; H 0.107, C 0.145, N 0.022, O 0.712, Na 0.002, P 0.004, S 0.002, Cl 0.003, K 0.002960, ^{10}B 0.000040; `lwtr.01`.

3.2.4 Neutron Source Definition

Constructing an accurate representation of the neutron source is among the most delicate steps of BNCT treatment planning, since it requires reproducing a five-dimensional probability distribution over three spatial coordinates, energy, and direction [54, 125]. Two complementary source-definition strategies are used in current practice, both ultimately relying on a preliminary, independent Monte Carlo characterization of the BSA.

Surface Source Files

A detailed simulation of the neutron beam is run once, and the phase-space (position, direction, energy, statistical weight) of every particle crossing a reference plane is recorded to a binary file. Subsequent patient-specific transport simulations resample particle histories directly from this file rather than re-simulating the full BSA, which both removes any approximation of the source characteristics and reduces computation time, at the cost of a finite achievable statistics set by the number of recorded source points (which can be partially mitigated by resampling the same points with different random seeds) [54].

Probability Density Functions (PDFs)

Alternatively, the source can be described analytically by sampling energy, position, and direction from probability density functions fitted to a prior beam characterization. These PDFs can be defined at different levels of the beam line: at the beam-port exit, upstream inside the collimator (requiring explicit modelling of part of the collimator

geometry), or, most demandingly, at the level of the full moderator/collimator/reflector assembly. Defining the source at the beam exit is computationally cheaper but generally requires a richer distribution to preserve accuracy, whereas defining it further upstream trades computational cost for a more complete, first-principles representation of the beam [54].

Both strategies presuppose a fixed source with respect to the treatment room; as detailed in Section 3.2.6, the patient, rather than the beam, is repositioned to achieve the desired irradiation geometry.

3.2.5 Monte Carlo Dose Calculation and the KERMA Approximation

Once patient geometry, material composition, boron concentration, and neutron source have been defined, the dose calculation itself proceeds by Monte Carlo transport of neutrons and photons through the patient model. MCNP, in the successive versions used across the BNCT TPS literature, remains the most widely used general-purpose transport engine for this task, owing to its validated cross-section libraries and its flexibility in describing arbitrary source and tally geometries [126]. Table 3.4 summarizes the standard MCNP tally types most relevant to BNCT dosimetry.

Table 3.4: Standard MCNP tally types relevant to BNCT dosimetric calculations [126].

Tally	**Description**
F1	Surface current
F2	Surface flux
F4	Track-length estimate of cell flux
F5	Flux at a point or ring detector
F6	Track-length estimate of energy deposition
F7	Track-length estimate of fission energy deposition
F8	Pulse-height tally
FMESH	Superimposed mesh tally (independent of problem geometry)

In practice, BNCT dose calculations almost universally rely on the superimposed-mesh F4-type tally (FMESH card), which allows particle fluence to be scored on a regular voxel mesh decoupled from the transport geometry itself, coupled with energy-dependent dose-function (DFn) and dose-energy (DEn) cards that convert the tallied fluence into KERMA through tabulated, energy-dependent KERMA factors [123, 126], such as those shown for soft tissue in Figure 3.6.

This KERMA approximation is justified because charged-particle equilibrium (CPE) is satisfied throughout most of the patient volume in BNCT owing to the short range of the secondary charged particles involved (Section 1.2.1). In this situation KERMA and absorbed dose coincide. A formal derivation of this equivalence, of the specific conditions under which CPE holds, and of the resulting dosimetric bias when it does not, is given in Appendix C.1. CPE, in fact, breaks down near tissue boundaries where secondary charged particles can escape before depositing their energy, most importantly at the skin surface; because skin is frequently the dose-limiting OAR, this is a documented limitation of the standard KERMA-based approach that some TPS implementations address through a dedicated, finer-grained dose-deposition calculation restricted to the near-surface region [114]. Where such a dedicated correction is not implemented, the resulting overestimate of the near-surface dose through the KERMA approximation is at least a conservative bias with respect to the OAR dose limit that ultimately determines treatment time.

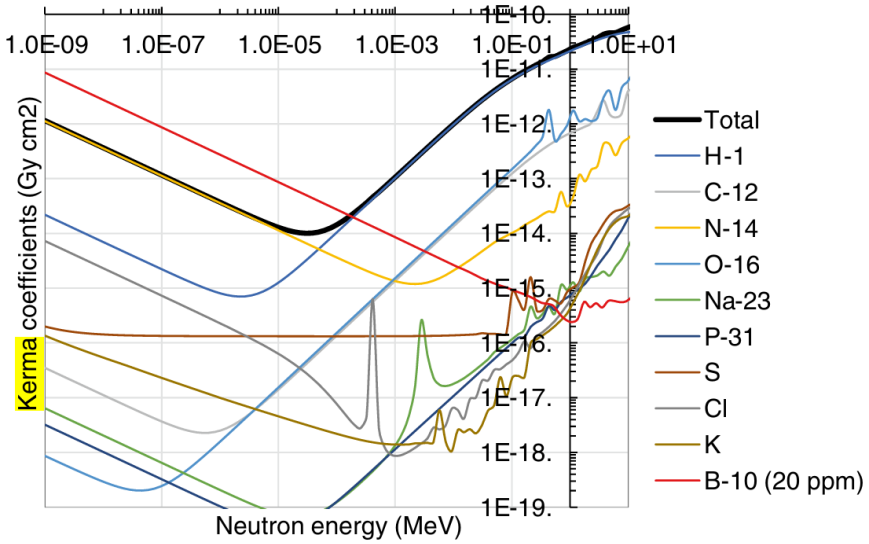


Figure 3.6: *The KERMA coefficients of soft tissue for neutrons calculated by the JENDL 4.0 dataset* [127]. *The KERMA coefficient of each element is standardized by the ratio of the elements that make up the average soft tissue. The KERMA coefficient of boron-10 is also shown for comparison* [117].

For a voxel x receiving contributions from the four dose components introduced in Section 1.3.1, the total absorbed dose accumulated over an irradiation time t can be written, following the OpenPINT dose-processing convention adopted for this thesis (Section 3.3), as

$$D_{\text{phys}}(x) = t\left[C_B(x)\,\dot{D}_{B10}(x) + \dot{D}_{N14}(x) + \dot{D}_n(x) + \dot{D}_\gamma(x)\right], \tag{3.1}$$

where $\dot{D}_{B10}$, $\dot{D}_{N14}$, $\dot{D}_n$, and $\dot{D}_\gamma$ are the four component dose rates obtained directly from the KERMA-weighted MCNP mesh tallies, each already normalized per unit source strength, and $C_B(x)$ is the local ^{10}B concentration used to scale the boron-dose channel, which is tabulated per unit (1 ppm) concentration and must therefore be supplied explicitly for every tissue class, as discussed in Section 3.2.3 [12, 123].

3.2.6 Patient Positioning for Fixed-Beam Facilities

Because both reactor-based and, until recently, most accelerator-based BNCT facilities deliver a spatially fixed, non-rotating epithermal neutron beam, angular beam optimization around the patient, routine in conventional external-beam radiotherapy, is not available in BNCT; instead, the patient must be positioned and, where multiple fields are used, repositioned so that the desired entry surface is placed in contact with, or close to, the beam-port collimator plane [128]. This constraint is compounded by the impossibility of observing the patient from a beam's-eye view, since the treatment position requires placing the patient in close proximity to the collimator, and by the comparatively long treatment times (tens of minutes to just over an hour) over which the position must be maintained without interruption [128].

The reference protocol for reactor-based BNCT of glioblastoma, developed at the Brookhaven Medical Research Reactor, addressed this problem through a dedicated mock-up room replicating the treatment-room geometry, including a transparent opening simulating the beam port and allowing a beam's-eye view of the patient during positioning rehearsal, together with a portable stereotactic frame used to place reproducible fiducial marks on the patient's head [128]. These marks, transferred to the planning CT/MRI images, allow the treatment planning system to compute the coordinates of the beam central axis' entry and exit points relative to the fiducial marks, and hence the distances from those marks to the collimator face required to reproduce the planned position in the treatment room [128]. Position verification during irradiation itself relied on triangulating laser lines converging on a virtual isocenter defined relative to the collimator, together with continuous video monitoring and audio-guided corrective instructions to the patient [128]. Retrospective dosimetric analysis of point doses computed at fixed offsets from the beam axis in this protocol showed variations of up to approximately 36% under small, uncorrected positioning errors, underscoring that, notwithstanding the diffusive, less sharply collimated nature of epithermal

neutron beams relative to photon beams, positioning accuracy remains dosimetrically consequential in BNCT [128].

The geometric logic underlying this reactor-era protocol, namely selecting a patient orientation that places a chosen entry point on the beam axis and as close as possible to the beam port, remains the operative principle in current accelerator-based practice, but modern pipelines increasingly compute the required rigid transform directly and reproducibly from the segmented patient geometry rather than from manually placed fiducial marks, also through the use of AI [10].

3.2.7 Plan Evaluation and Optimization

Once a dose distribution has been computed, a BNCT plan is evaluated, as in conventional radiotherapy (Section 3.1.2), through isodose curves superimposed on the anatomical images and through DVHs computed separately for the tumor and for each OAR, using tissue-specific boron concentrations and biological weighting factors. Because, as discussed in Section 3.2.1, the constraint that sets treatment time is normal-tissue tolerance rather than target dose, plan evaluation in BNCT is centered on verifying that the dose-volume constraints for the relevant OARs (historically expressed as a mean or maximum RBE-weighted dose, e.g., $D_{\text{mean}} \leq 2.5$ Gy-w and $D_{\text{max}} \leq 13$ Gy-w to the brain in early glioblastoma protocols [49]) are satisfied at the shortest possible irradiation time, with the resulting minimum and mean target dose reported for outcome interpretation.

For single-field plans, optimization reduces to comparing a small number of candidate beam orientations. For multi-field plans, selecting and weighting a non-overlapping subset of beams from a larger candidate pool by hand becomes impractical; this has motivated the application of linear-programming (simplex) optimization to BNCT planning, maximizing the minimum averaged boron dose to the target subject to the OAR dose-volume constraints, an approach demonstrated using dose data generated by NCT-Plan and re-imported into a finer MCNP voxel model at the Petten BNCT facility for melanoma treatment planning [116].

3.2.8 Treatment Planning Systems Developed for BNCT

Building on the geometry-reconstruction and dose-calculation strategies introduced above, a number of dedicated TPS have been developed at different BNCT facilities worldwide, each reflecting the specific requirements of its host institution. Table 3.5 summarizes the principal systems; a narrative overview follows.

Table 3.5: Overview of treatment planning systems developed for BNCT.

System	Institution	Geometry	MC engine
MacNCTPlan/ NCTPlan[a]	Harvard–MIT	Voxel	MCNP
SERA[b]	Idaho National Eng. and Env. Laboratory / Montana State University	Univel	SeraMC (MORSE-derived)
JCDS/ Tsukuba Plan[c]	University of Tsukuba	Voxel	MCNP
THORplan[d]	Tsing Hua Open-Pool Reactor	Voxel	MCNP
NeuMANTA[e]	Neuboron Therapy System	Custom voxel	COMPASS (dedicated engine)
NeuCure Dose Engine[f]	Neutron Therapeutics	Voxel/patient-specific	Geant4
MultiCell TPS[g]	CNEA, Argentina	MultiCell	MCNP
CARONTE	DIMNP, Pisa	Macroscopic boron distribution model	MCNP-4A
RayStation for BNCT[i]	RaySearch Laboratories; SHI, NT, TAE Life Sciences	Voxel/patient-specific	PHITS (SHI), Geant4 (NT), or TAE engine
OpenPINT[h]	INFN Pavia	Voxel (NIfTI-derived)	MCNP (PHITS-ready)

[a] Reference cranial NCT TPS; NCTPlan is the PC-based successor of the Macintosh-based MacNCTPlan [118, 124].
[b] Replaced the NURBS-based BNCT_Rtpe; used in early Brookhaven trials; discontinued and not maintained as a certified medical device [18, 120].
[c] JCDS developed for JRR-4 clinical trials; Tsukuba Plan its linac-based successor, reusing JCDS know-how [129, 130].
[d] Developed and verified for the Taiwanese reactor-based program [131].
[e] Accelerator-oriented; benchmarked against MCNP6 with comparable accuracy at roughly half the computation time [132].
[f] Dose engine of the first clinically approved accelerator-based BNCT system (JHN002 trial, Japan) [20, 54].
[g] Basis of the photon-isoeffective dose formalism used throughout this thesis; applied beyond cranial BNCT (e.g., thoracic geometries) [13, 121].
[h] Open-source, reproducible simulation-preparation and dose-analysis platform; adopted for this thesis (Section 3.3) [12].
[i] Integrates BNCT-specific dose engines into RayStation, a TPS platform originally developed for conventional radiotherapy; installed at Southern TOHOKU BNCT Research Center, Osaka Medical College, and Helsinki University Hospital [117]. SHI: Sumitomo Heavy Industries; NT: Neutron Therapeutics.

The chronologically earliest of these systems, MacNCTPlan and its PC-based successor NCTPlan, established the voxel-reconstruction, MCNP-based paradigm that most later BNCT TPS have followed in one form or another, and remain a reference point against which newer voxel-reconstruction strategies are benchmarked [118, 124]. SERA, developed jointly by the Idaho National Engineering and Environmental Laboratory and Montana State University, represented at the time a substantial improvement in geometric fidelity and execution speed over its NURBS-based predecessor BNCT_Rtpe, and supported early clinical trials at Brookhaven; however, its development was not continued, and it was not subsequently classified as Software as a Medical Device (SaMD) under the regulatory framework later adopted in Japan, which required the development of a new, certifiable TPS for clinical accelerator-based BNCT [18]. In Japan, the JAERI Computational Dosimetry System (JCDS), developed for intraoperative BNCT trials at the JRR-4 research reactor and verified against phantom and retrospective clinical measurements to within approximately 10% [129], was subsequently extended into the linac-oriented Tsukuba Plan, reusing JCDS methodology and nuclear data verification procedures for the transition to accelerator-based treatment [130]. THORplan performed an analogous role for the Taiwanese reactor-based program at the Tsing Hua Open-Pool Reactor [131]. More recently, accelerator-specific systems such as NeuMANTA have been developed with dedicated, purpose-built Monte Carlo engines optimized for computational speed while remaining benchmarked against MCNP6 [132], and the NeuCure Dose Engine underlies the first BNCT system to receive full, non-investigational regulatory approval, on the basis of the JHN002 phase II trial discussed in Section 1.4.2 [20].

More recently, RaySearch Laboratories has collaborated with Sumitomo Heavy Industries, Neutron Therapeutics, and TAE Life Sciences to integrate BNCT-specific dose engines, respectively PHITS-based (the NeuCure Dose Engine, marketed jointly as "SACRA planning"), Geant4-based, and independently developed, into RayStation, a TPS platform originally developed for conventional external-beam radiotherapy [117]. This extends BNCT treatment planning into an established, multi-vendor commercial TPS ecosystem, rather than relying on facility-specific software, and the resulting platform is reported to be in clinical use at the Southern TOHOKU BNCT Research Center, Osaka Medical College, and Helsinki University Hospital [117].

This last case is illustrative of a broader, non-purely-technical dimension of BNCT

treatment planning software: because a TPS translates directly into the delivered patient dose, it is classified, wherever BNCT has moved from research to routine clinical use, as a regulated medical device. In Japan, this classification (Software as a Medical Device, SaMD, under the Pharmaceutical and Medical Device Agency, PMDA) was a contributing factor in SERA's discontinued clinical use, since it was not maintained by a commercial supplier able to support certification, and the dose-calculation software developed for the JHN002 trial was reviewed and expedited under the SAKIGAKE designation system for innovative medical products, alongside the accelerator hardware and the borofalan drug itself [18]. This regulatory dimension is a further respect in which BNCT TPS development departs from a purely research-software context, and is directly relevant to positioning an open-source, non-clinically-certified platform such as OpenPINT (Section 3.3) explicitly as a research tool rather than as a clinical deliverable.

3.3 OpenPINT

OpenPINT (Open-source Planning for Isoeffective Nuclear Treatments) is the treatment planning system adopted for the Monte Carlo dosimetric pipeline of this thesis. It is developed at INFN Pavia in collaboration with the CNEA/CONICET group in Argentina that originated the photon-isoeffective dose formalism of Section 1.3.3 [12].

3.3.1 Motivation and Scope

As noted throughout Section 3.2.8, much of the historical BNCT TPS ecosystem has been facility-specific, closed-source, or tied to a particular clinical device and local beam model. This fragmentation is analogous, and directly comparable, to a broader methodological difficulty already discussed in Chapter 1: without a common, openly available computational reference, it is difficult to establish whether differences observed between independent BNCT dosimetric studies arise from genuine physical or biological differences, or merely from undocumented differences in geometry preparation, transport-output parsing, or dose-processing conventions [12]. OpenPINT is explicitly conceived as such a shared reference layer, drawing an analogy with open research platforms already established in conventional and particle radiotherapy, such as matRad, SlicerRT, and OpenTPS [12].

3.3.2 Software Architecture

OpenPINT is organized into three functional layers [12]:

1. an **input preparation layer**, which converts segmented NIfTI masks, material definitions, and simulation settings into MCNP-ready textual inputs;

2. a **transport-output decoding layer**, which parses MCNP mesh tallies and maps them onto structured, component-wise dose arrays;

3. an **analysis layer**, which exports the resulting component maps to CT-aligned NIfTI volumes, combines them across irradiation fractions where relevant, and reduces them to planning-oriented dosimetric metrics.

This layered separation allows each stage to be validated independently and keeps interfaces (file schema, coordinate mapping, component naming) explicit. While the transport engine used for this thesis is MCNP6.3.0, the architecture is not intrinsically tied to it: the output-decoding layer is designed to accommodate other codes exposing an equivalent component-wise tally structure, such as PHITS, by replacing the input-generation syntax and the output-reading layer [12].

3.3.3 From Segmented Geometry to MCNP Input

The input-preparation workflow starts from the segmented NIfTI masks produced by the pipeline of Chapter 2 (brain, skull/bone, skin, and GTV, in the current release), together with a user-defined ordering of material classes. Voxels not covered by any explicitly supplied mask are assigned to an AIR class by taking the geometric complement of the union of the supplied masks, which prevents unlabeled voxels from silently entering the transport geometry [12]. The individual masks are collapsed into a single simulation mask in which integer labels follow the chosen material order; this label volume is itself exported as a NIfTI object for independent inspection before being converted into an MCNP lattice.

During this conversion, the affine transform of the source NIfTI image is used to preserve anatomical orientation: voxel traversal is reversed along any axis with a negative affine direction cosine, so that the resulting MCNP fill order remains consistent with the medical-image coordinate convention [12]. This makes the resulting lattice

a deterministic, auditable translation from segmented image space to MCNP voxel geometry, rather than an unchecked array dump, directly addressing the coordinate-mismatch failure mode identified as a common source of silent error in fragmented BNCT computational chains [12]. Figure 3.7 shows an example of a synthetic phantom produced by this workflow.

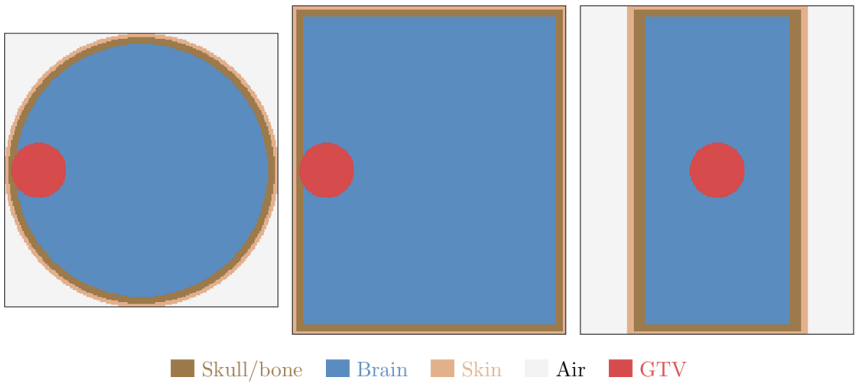


Figure 3.7: *Synthetic cylindrical phantom generated by the OpenPINT input-preparation workflow. The panels show orthogonal sections through the tumour centroid and preserve physical voxel aspect ratio; the axial panel spans 200mm by 200mm, whereas the coronal and sagittal panels span 200mm by 240mm* [12].

3.3.4 Automated Patient Positioning

Building on the geometric logic of Section 3.2.6, OpenPINT implements a geometric patient-positioning strategy that computes, directly from the segmented anatomy, a reproducible rigid transform aligning a chosen patient-surface entry point with the beam port and orienting the entry-to-target direction along the beam axis [12]. Unlike the manual, fiducial-mark-based reactor-era protocol described in Section 3.2.6, this workflow requires no physical mock-up room: the GTV centroid c_{GTV} is computed as the arithmetic mean, under uniform voxel weighting, of the RAS coordinates of the segmented tumor voxels, and a candidate skin-entry point is extracted either from an explicit skin mask or from a Hounsfield-unit threshold of the CT followed by morphological clean-up and largest-connected-component selection [12]. Among the candidate surface points $\mathcal{S}$, the selected entry point $\mathbf{s}_{\mathrm{entry}}$ minimizes the Euclidean distance to the GTV centroid,

$$\mathbf{s}_{\mathrm{entry}} = \arg\min_{\mathbf{s}\in\mathcal{S}} \|\mathbf{s} - \mathbf{c}_{\mathrm{GTV}}\|_2, \tag{3.2}$$

providing an automatic geometric approximation of the shortest skin-to-target path and defining the entry-to-target direction $\mathbf{v} = \mathbf{c}_{\mathrm{GTV}} - \mathbf{s}_{\mathrm{entry}}$, as shown in Figure 3.8 [12].

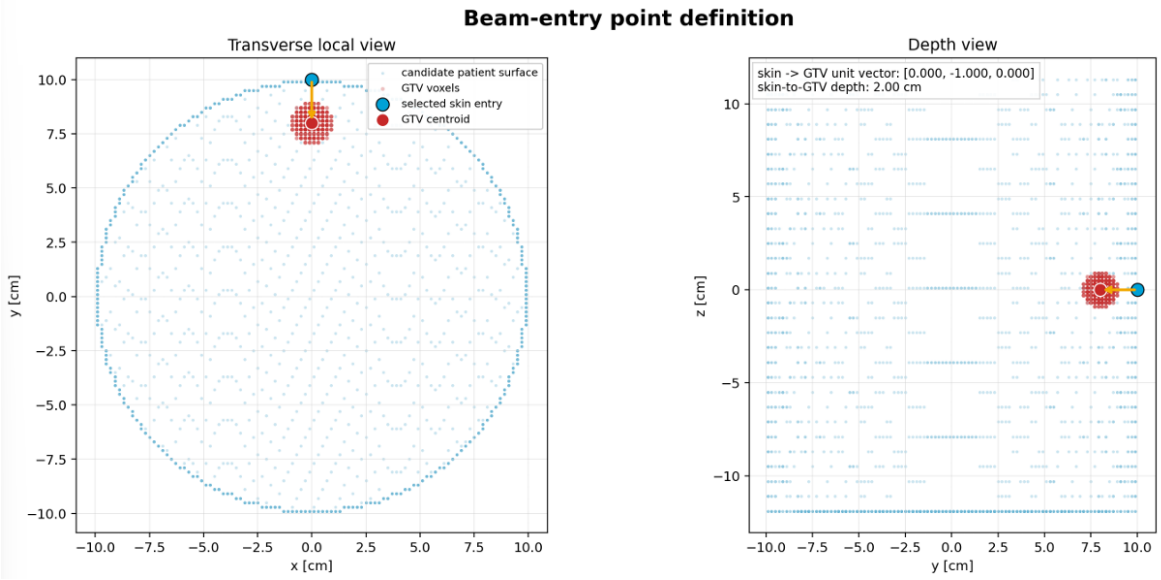


Figure 3.8: *Definition of the geometric points used for patient positioning. The generated beam-alignment image shows the candidate patient surface, GTV voxels, selected skin-entry point, GTV centroid, skin-to-GTV unit vector, and skin-to-GTV depth. The selected skin-entry point is the patient-surface point that minimizes the Euclidean distance to the GTV centroid* [12].

A rigid rotation is then computed so that this direction is aligned with the requested beam axis, as shown in Figure 3.9. This is optionally preceded by user-specified Euler pre-rotations that allow a preferred patient roll or orientation convention to be imposed before the corrective alignment; the translation is finally chosen so that the transformed entry point coincides with the beam-port coordinate, and the resulting rotation and translation are written directly into the MCNP input as a coordinate-transform card [12].

Every quantity involved (centroid, entry point, minimum distance, rotation matrix, translation vector, and the resulting transform card) is recorded in the case summary, making the positioning step fully auditable and allowing multiple candidate orientations to be generated and compared through the same downstream transport and dosimetric-analysis pipeline [12]. This is directly the computational counterpart to the manually derived positioning coordinates of Section 3.2.6, and, as will be discussed in Chapter 4, does not by itself resolve every degree of freedom of the positioning problem: a roll rotation about the entry-to-target axis remains under-constrained by Equation (3.2) alone and must be fixed through an additional convention or optimiza-

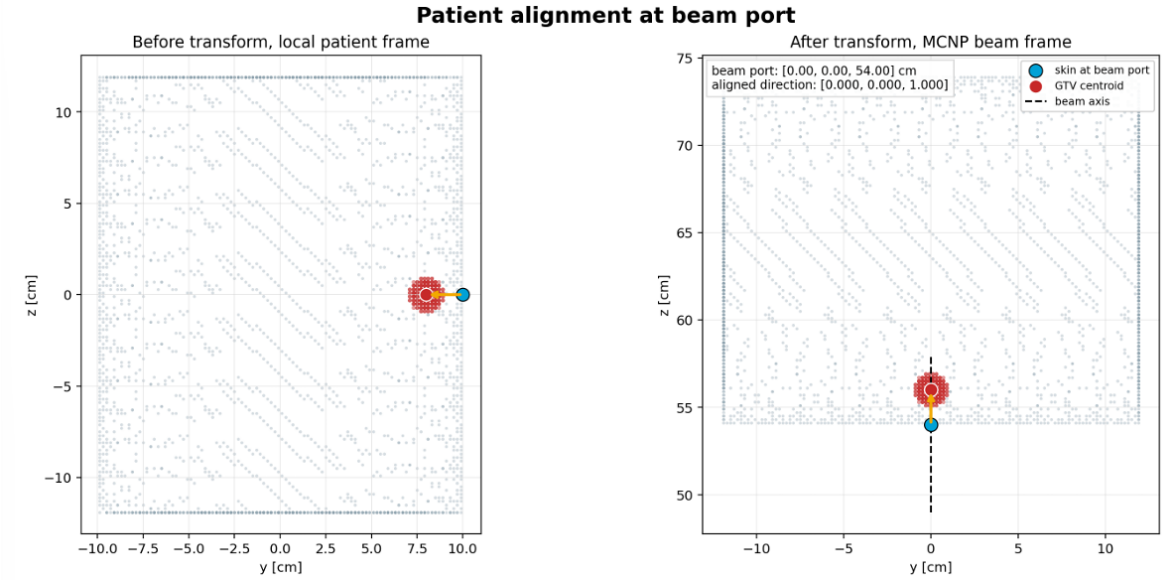


Figure 3.9: *Beam-port alignment after applying the computed rigid patient transform. This separate positioning example uses a spherical GTV of 10mm radius, compared with the 20mm radius GTV in Figure 3.7.*
The left panel shows the selected skin-entry point and GTV centroid before transformation in the local patient frame. The right panel shows the transformed geometry in the MCNP beam frame, where the skin-entry point is placed at the beam port and the entry-to-target direction is aligned with the requested +z beam axis [12].

tion criterion, a point of direct relevance to the patient-positioning pipeline developed for this thesis.

3.3.5 Dose-Component Extraction and Biological Weighting

For MCNP outputs, OpenPINT reads four component-specific flux mesh tallies, each already converted into a dose-rate contribution through the MCNP tally-multiplier formalism coupled with energy-dependent KERMA factors, following exactly the KERMA-approximation logic introduced in Section 3.2.5 [12]. The four channels follow a fixed semantic convention: `B10` (boron-dose contribution from $^{10}\mathrm{B}(n,\alpha)^{7}\mathrm{Li}$, tabulated per 1 ppm ^{10}B, from ENDF/B-VI cross sections and reaction Q-values as reported by Goorley et al. [123]); `N14` (nitrogen-capture contribution from $^{14}\mathrm{N}(n,p)^{14}\mathrm{C}$, computed for ICRU 44 brain material from ICRU Report 63 KERMA data with JENDL-3.2 cross sections [12]); `n` (non-capture, elastic-scattering neutron KERMA); and `g` (photon KERMA, computed from NIST photon mass-attenuation and mass-energy-absorption coefficients for ICRU 44 brain material [12]). Crucially, the parser preserves these four physical components *before* any boron-concentration scaling or biological weighting is

applied, so that alternative weighting factors or tissue boron concentrations can be explored by reprocessing the same transport output, without regenerating the Monte Carlo simulation itself [12]. The energy-dependent KERMA factors underlying these four channels are shown in Figure 3.10.

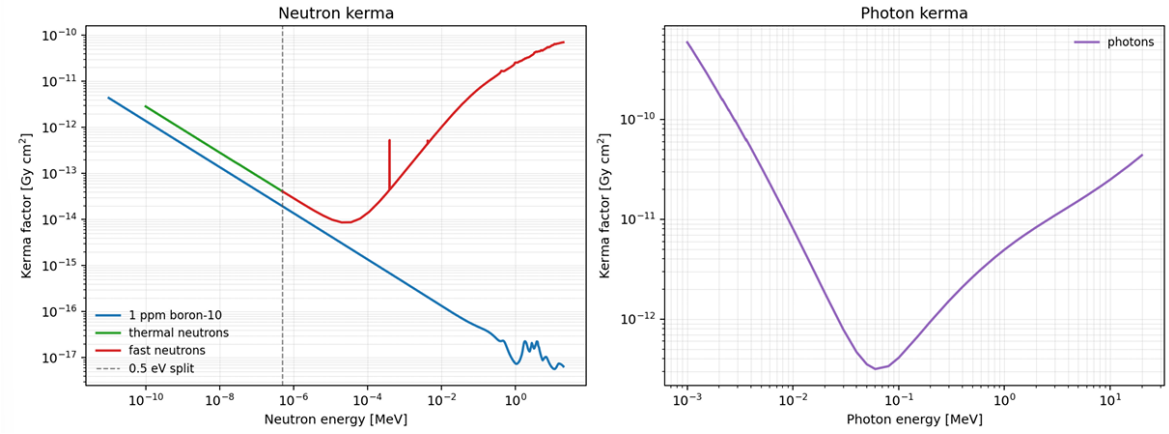


Figure 3.10: *Energy-dependent KERMA factors used by the MCNP component tallies in the simulation template. The left panel shows the neutron KERMA factors for 1ppm boron-10, thermal neutrons, and fast neutrons; the right panel shows the photon KERMA factors. The dashed line marks the 0.5eV thermal/fast neutron split used to separate the thermal and fast neutron KERMA channels* [12].

The absorbed dose of Equation (3.1) is complemented, for biologically weighted reporting, by a component-specific weighted-dose expression,

$$D_w(x) = t\left[r_B\, C_B(x)\, \dot{D}_{B10}(x) + r_{N14}\, \dot{D}_{N14}(x) + r_n\, \dot{D}_n(x) + r_\gamma\, \dot{D}_\gamma(x)\right], \tag{3.3}$$

where r_B is the CBE factor of the boron component and r_{N14}, r_n, r_γ are the RBE factors of the nitrogen-capture, non-capture-neutron, and photon components, respectively (Section 1.3.2) [12]. This clean separation between physical component maps and biologically weighted summaries is, again, an auditability design choice: the transport-derived information required for independent review is retained regardless of which biological weighting convention is subsequently applied for planning interpretation [12].

3.3.6 Radiobiological Model Database

Beyond the fixed RBE/CBE weighting of Equation (3.3), OpenPINT implements a broader database of dosimetric and radiobiological formalisms, exposed through a dedicated model layer for research use: physical absorbed-dose summation, RBE/CBE-weighted dose, the photon-isoeffective dose (IsoE) formalism of Section 1.3.3, Tumor

Control Probability (TCP) and Normal Tissue Complication Probability (NTCP) models (Section 1.3.4), Uncomplicated Tumor Control Probability (UTCP), Biologically Effective Dose (BED), and a simplified, dose-independent weighting approximation for normal-tissue photon-IsoE reporting [12]. The IsoE parameter database specifically includes *in vivo* H&N squamous cell carcinoma parameterizations, in addition to *in vitro* glioblastoma, gliosarcoma, and lip-mucosa models, making it directly applicable to the H&N dosimetric pipeline developed in this thesis [12]. This modular design is what allows the same validated transport-output processing chain to be reused for beam-quality evaluation studies and for cross-modality comparisons, such as the combination of BNCT and carbon-ion radiotherapy through the shared photon-isoeffective-dose formalism reported for a H&N case [12].

3.3.7 Validation: Analytic and Voxelized Phantom Benchmarks

The OpenPINT validation benchmark isolates the effect of the image-to-MCNP conversion pipeline itself, independent of any patient-specific anatomical variability, by comparing two representations of the same cylindrical phantom: an analytic MCNP geometry, in which the external boundary, internal shells, and a spherical GTV are described by continuous geometric surfaces, and a set of voxelized phantoms generated from segmented NIfTI masks through the same input-preparation workflow used for patient geometries [12]. Because the material boundaries of the analytic phantom are unaffected by image discretization, it provides a controlled reference against which the effects of voxel size, tally-mesh resolution, and interpolation-based remapping of MCNP mesh tallies back onto a CT-aligned NIfTI grid can be isolated [12].

The benchmark comprised 28 independent MCNP transport simulations (six analytic and 22 voxelized configurations, each run to the equivalent of 10^{12} source histories from a surface source file), evaluated against three complementary endpoints: the percent difference in the brain $D_{2\%}$ dose rate and in the resulting brain-limited irradiation time; structure-wise gamma-index pass rates; and weighted DVH metric differences for the GTV, brain, skull, and skin [12]. The irradiation time itself was defined, in direct analogy with the OAR-driven planning logic of Section 3.2.7, as the earlier of two candidate times set by a healthy-brain $D_{2\%} \leq 13$ Gy-w constraint and a $D_{50\%} \leq 2.5$ Gy-w

constraint,

$$t_{D_{2\%}} = \frac{13 \text{ Gy-w}}{\dot{D}_{2\%,\text{brain}}}, \qquad t_{D_{50\%}} = \frac{2.5 \text{ Gy-w}}{\dot{D}_{50\%,\text{brain}}}, \qquad t_{\text{brain}} = \min\left(t_{D_{2\%}}, t_{D_{50\%}}\right), \tag{3.4}$$

with the $D_{2\%}$ constraint found to be the more restrictive of the two in every evaluated configuration [12]. A general derivation of this dose-rate-to-time conversion from the underlying dose-volume constraint, extended to an arbitrary number of simultaneous constraints, is given in Appendix C.2.

Table 3.6 reports a representative subset of the reported voxelized-phantom results, compared against the 1 mm analytic reference; Figure 3.11 shows the resulting brain-limited irradiation time across the full sweep of tested configurations. Across the full benchmark, fine-resolution voxelized configurations (1–2 mm voxel and tally-mesh size) reproduced the analytic brain-limited-time endpoint within 0.13%, whereas the coarsest tested configurations (8–10 mm) showed deviations of up to 4.42% [12].

Table 3.6: Analytic cylindrical mesh sweep against the analytic 1 mm reference. Percent differences are relative to the 1 mm analytic case processed with the corresponding evaluation workflow. Gamma criteria were 3%/3 mm with a 10% low-dose threshold. Adapted from [12].

Mesh	Remapping	Time diff. (%)	Brain D_2 rate diff. (%)	GTV D_{98} diff. (%)	Skin γ pass (%)	Skull γ pass (%)	Brain γ pass (%)	GTV γ pass (%)
2	nearest	0.063	−0.063	0.01	99.42	99.31	99.92	100.00
4	nearest	0.092	−0.092	3.31	97.54	98.13	99.77	99.99
5	nearest	0.106	−0.106	−3.88	97.09	97.85	99.75	99.98
8	nearest	−0.005	0.005	7.92	96.32	97.51	99.70	98.55
10	nearest	−0.404	0.406	−17.30	95.48	97.28	99.65	89.48
2	linear	0.102	−0.101	0.09	98.95	98.63	99.83	100.00
4	linear	0.220	−0.220	1.20	97.82	97.72	99.72	99.98
5	linear	0.284	−0.283	0.94	97.54	97.56	99.71	99.96
8	linear	0.618	−0.615	1.50	93.15	97.05	99.68	99.59
10	linear	0.941	−0.933	0.60	88.23	96.75	99.69	98.69
2	cubic	0.070	−0.070	−0.03	99.41	99.15	99.90	100.00
4	cubic	0.117	−0.116	0.09	98.75	98.13	99.73	99.99
5	cubic	0.211	−0.211	−0.16	98.37	97.65	99.69	99.99
8	cubic	0.350	−0.348	6.13	95.13	95.80	99.57	99.93
10	cubic	0.176	−0.175	6.24	90.05	95.90	99.32	99.82

Patient-wide gamma-index pass rates remained at or above 99.60% across all evaluated mesh and interpolation combinations, with the lowest structure-specific pass rates occurring, as expected from the KERMA/CPE discussion of Section 3.2.5 and Appendix C.1, in boundary-dominated normal-tissue regions (skull and skin), while GTV pass rates remained above 99.99% in every case [12], as summarized across the full configuration sweep in Figure 3.12. Weighted DVH metrics were most robust for high-dose and mean-dose quantities (median absolute percent difference of 0.116% across all structures and metrics), whereas low-dose, boundary-sensitive quantities such as $D_{98\%}$ in thin normal-tissue structures showed markedly larger relative differences, reaching over 40% in skull in the coarsest configurations [12], as illustrated by the GTV DVH

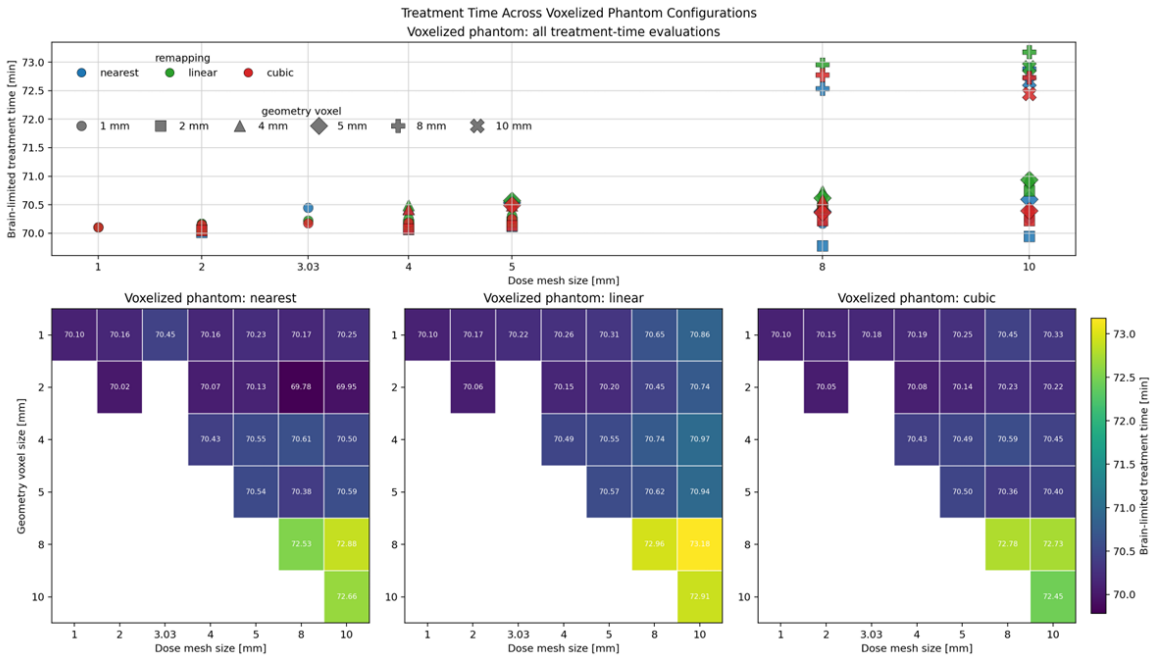


Figure 3.11: *Brain-limited irradiation time across the voxelized cylindrical-phantom sweep. The upper panel reports every voxelized configuration as a scatter plot, with point color indicating the remapping strategy and marker shape/size indicating the geometry voxel size. The lower panels report the same results as heatmaps, with geometry voxel size on the vertical axis, MCNP tally-mesh size on the horizontal axis, and one panel for each remapping strategy. Empty heatmap cells correspond to combinations that were not simulated* [12].

curves and residuals of Figure 3.13.

This pattern, large relative but small absolute deviations concentrated in the low-dose DVH tail of boundary-dominated structures, is consistent with the discretization limitations already discussed for voxel-based geometry reconstruction in Section 3.2.2, and is interpreted in the original manuscript as a discretization-sensitivity finding rather than a failure of the high-dose planning-relevant quantities [12].

At the 1 mm reference configuration, the resulting brain-limited weighted tumor-dose metrics were $D_{98\%} = 57.25$ Gy-w, $D_{50\%} = 85.55$ Gy-w, $D_{2\%} = 91.46$ Gy-w, and $D_{\text{mean}} = 82.85$ Gy-w, with the brain $D_{2\%}$ constrained to 13.0 Gy-w by construction and a brain-limited irradiation time of approximately 70–73 min across the tested mesh and interpolation choices [12]. These figures illustrate concretely the intended output of the workflow: component-wise Monte Carlo maps are reduced not only to image-level outputs for visual inspection, but also to the traceable, table-level planning summaries used throughout the results reported in Chapter 4.

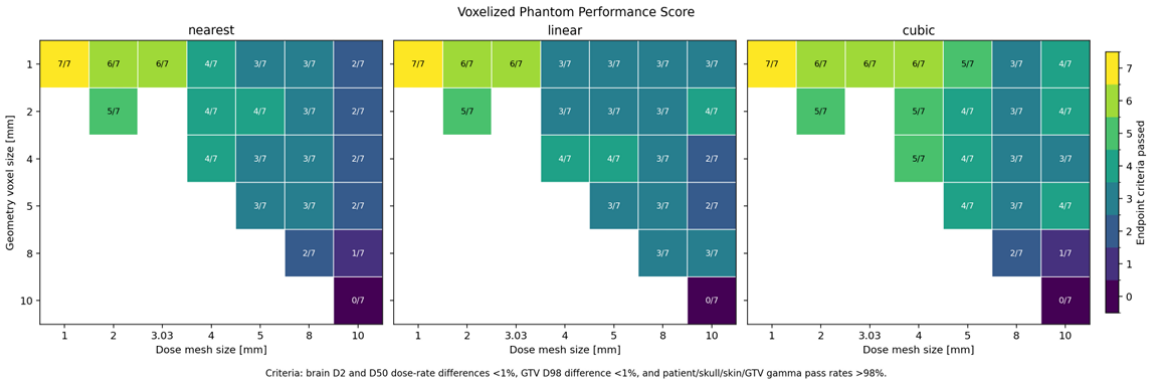


Figure 3.12: *Non-time performance score for the voxelized cylindrical-phantom sweep. Each cell reports the number of endpoint criteria satisfied out of seven. Criteria were brain D2 and D50 dose-rate differences below 1%, GTV D98 difference below 1%, and patient, skull, skin, and GTV gamma pass rates above 98%. Empty cells correspond to combinations that were not simulated* [12].

3.3.8 Limitations and Development

The validation described above concerns the computational core of OpenPINT rather than a complete clinical TPS, which brings with it some limitations.

First, the quantitative agreement benchmark is based on controlled cylindrical phantom geometries; the patient-positioning example demonstrates geometric workflow integration but does not, by itself, constitute patient-cohort dosimetric validation, a gap that the systematic multi-patient pipeline developed in this thesis (Chapter 4) is intended to help close.

Second, the current release does not calibrate boron pharmacokinetics, optimize beam arrangements, or estimate patient-specific biological response; these tasks require additional modeling layers and clinical validation beyond the present computational scope.

Third, KERMA-factor handling is presently based on reference BNCT KERMA tables (Section 3.3.5) rather than on composition-specific KERMA factors generated per material definition, which limits material specificity in heterogeneous patient models and is explicitly identified as a direction for future development, particularly in combination with automated material contouring [12].

Finally, and consistent with the low-dose DVH-tail sensitivity reported in Section 3.3.7, treatment-decision endpoints that depend on low-dose tails require inspection of structure geometry, mesh resolution, interpolation method, and propagated Monte Carlo uncertainty, rather than reliance on percent-difference summaries alone [12].

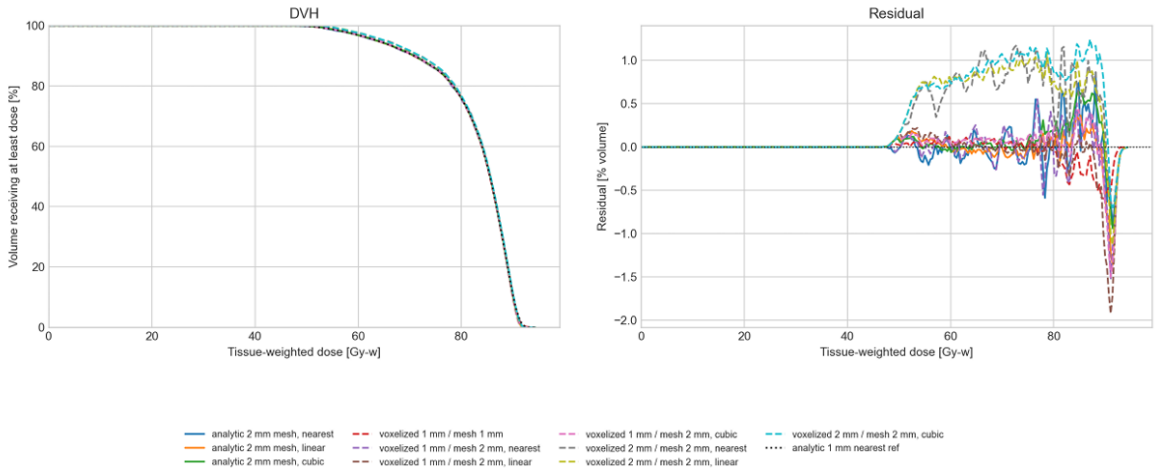


Figure 3.13: *Weighted DVH curves and DVH residuals for GTV. Analytic and voxelized comparisons are shown together in each panel. The dashed black curve is the analytic 1 mm nearest-neighbour reference, and residuals are computed as candidate DVH minus reference DVH in percent volume. The analytic comparison includes the 2 mm nearest-neighbour, linear, and cubic remappings. The voxelized comparison includes the matched 1 mm voxel/1 mm mesh case once because all remappings are equivalent on the matched evaluation grid, plus the 1 mm voxel/2 mm mesh and 2 mm voxel/2 mm mesh cases with nearest-neighbour, linear, and cubic remappings* [12].

The roadmap toward a full BNCT TPS built on this validated computational core includes optimization strategies, expanded biological-weighting models, plan-comparison interfaces, and an extended, patient-specific validation set, together with tighter integration between the `OpenPINT.mcgenerator` package and a complementary MCNP geometry-construction framework under parallel development by the same group [12]. The present thesis, through the systematic geometry-generation and dosimetric-analysis pipeline described in Chapter 4, is itself a direct contribution to this patient-specific validation effort, extending the OpenPINT computational core, previously validated on phantom geometries, to a cohort of H&N patients segmented as described in Chapter 2.

Chapter 4

AI-Based Treatment Planning and Dosimetric Analyses for Head and Neck Patients

Previous Chapters establish the physical and radiobiological framework of BNCT, the deep-learning segmentation pipeline evaluated against a manual reference, and the treatment-planning system, OpenPINT, that turns a patient CT and a set of segmented structures into an MCNP Monte Carlo dosimetric simulation.

This Chapter applies that full pipeline to a 24-patient H&N cohort, with the specific aim of quantifying how much the choice of GTV segmentation propagates into clinically relevant dosimetric quantities.

Every simulation in this Chapter is a full, independent MCNP Monte Carlo dose calculation, using the same beam model, material compositions, and radiobiological parameters that would be used for a real BNCT treatment plan, including the most updated radiobiological models and dosimetric strategies. What is not being claimed, however, is that any of the 24 plans presented here constitutes a validated, clinically deliverable treatment: patient selection was driven entirely by the need for a wide, representative range of segmentation agreement, not by clinical eligibility criteria; no plan was reviewed or approved by a radiation oncologist; and several methodological choices are appropriate for a comparative segmentation study but not necessarily for real treatment planning. This Chapter should therefore be read as a controlled dosimetric comparison between two segmentation strategies, using the same infrastructure

that a real treatment plan would use, rather than as a report on 24 real treatments.

Section 4.1 describes the patient cohort and the geometric indices used to quantify segmentation agreement. Section 4.2 documents how every anatomical structure needed for the Monte Carlo geometry, beyond the GTV itself, was obtained. Section 4.3 introduces the four-arm simulation design (manual, AI, swap and anti-swap) that disentangles segmentation-driven dose differences from the automated-positioning effect that couples to it. Sections 4.4 and 4.5 present the dosimetric analysis first for a single pilot patient and then pooled over the full cohort, covering TCP, the RBE-weighted versus photon-isoeffective dose models, mucosal complication risk, and gamma-index agreement between configurations. Appendix D tabulates the complete per-patient results underlying every pooled statistic in this Chapter.

4.1 Patient Cohort

4.1.1 Data Provenance

The H&N imaging data used throughout this thesis originate from *The Cancer Imaging Archive* (TCIA), a public repository of anonymized medical imaging organized into disease and site specific collections. The H&N test set assembled within the AI_MIGHT project (Section 2.5) and used for the nnU-Net evaluation of Chapter 2 draws on six TCIA collections, reflecting the greater anatomical and institutional heterogeneity of H&N cancer imaging. Every case included a pre-treatment planning CT together with a RTSTRUCT file containing, at minimum, an expert manual delineation of the gross tumor volume, the requirement that determined eligibility for inclusion in the test set.

The list below gives the six source collections, and the number of patients each source contributed to the resulting 329-patient nnU-Net test set:

- **HNSCC** (104 patients): a multi-institutional collection of H&N squamous cell carcinoma cases from the M.D. Anderson Cancer Center Head and Neck Quantitative Imaging Working Group [134].
- **OPC-Radiomics** (115 patients): oropharyngeal squamous cell carcinoma cases from the Princess Margaret Cancer Centre, originally curated for CT-based radiomic prognostic modeling [135].

- **Head-Neck-PET-CT** (56 patients): FDG-PET/CT and radiotherapy planning imaging from four institutions in Québec, Canada [136].
- **Head-Neck-Radiomics-HN1** (30 patients): CT imaging from the MAASTRO Clinic (Maastricht), originally assembled for a foundational radiomics phenotyping study [137].
- **Head-Neck Cetuximab** (22 patients): imaging from patients enrolled in RTOG-0522, a phase III trial of accelerated radiotherapy with concurrent cisplatin, with or without cetuximab, for locally advanced H&N carcinoma [138].
- **HNSCC-3DCT-RT** (2 patients): longitudinal pre-, mid-, and post-treatment CT imaging of H&N squamous cell carcinoma patients undergoing radiotherapy [139].

All DICOM series and their associated RTSTRUCT contours were converted to NIfTI format, the file format required by nnU-Net, following the same conversion procedure and directory convention described in Section 2.5; the resulting GTV segmentation is a binary mask volume with unit-valued voxels where tumor is present. Acquisition parameters (in-plane resolution, slice thickness, number of slices) vary appreciably across the six collections, consistent with their independent, multi-institutional origin; this heterogeneity is part of the motivation for the fixed-resolution voxel geometry reconstruction strategy (Section 3.2.2) that resamples every case onto a common simulation grid regardless of native acquisition resolution.

4.1.2 Cohort Selection

Of the cases in the nnU-Net test set, a subset of 24 patients was selected for the full Monte Carlo dosimetric analysis reported in this Chapter, through a two-stage selection process. First, cases were excluded from the eligible pool if they presented technical issues that would compromise their use as a Monte Carlo dosimetric geometry – most commonly an incomplete or artifact-affected planning CT series, or an ambiguous, fragmented, or otherwise unreliable RTSTRUCT contour. Second, among the remaining technically usable cases, patients were preferentially selected for a clearly delineated, well-defined manual GTV contour, identified by visual inspection rather

than by a quantitative image-quality metric, so that the manual segmentation used as the reference standard throughout this Chapter was itself a reliable one.

Within this pool of technically usable, clearly delineated cases, selection was further guided to provide representative coverage of the full range of nnU-Net segmentation agreement, rather than concentrating on the well-segmented, high-agreement cases that dominate the distribution. Figure 4.1 shows the distribution of Dice Coefficients, further explained in the section below, across the nnU-Net segmentation-quality evaluation set; the 24-patient dosimetric cohort was drawn to span this entire range, from a complete segmentation failure (DC = 0, patient HN_1019) to near-perfect agreement (DC = 0.90, patient HN_0717).

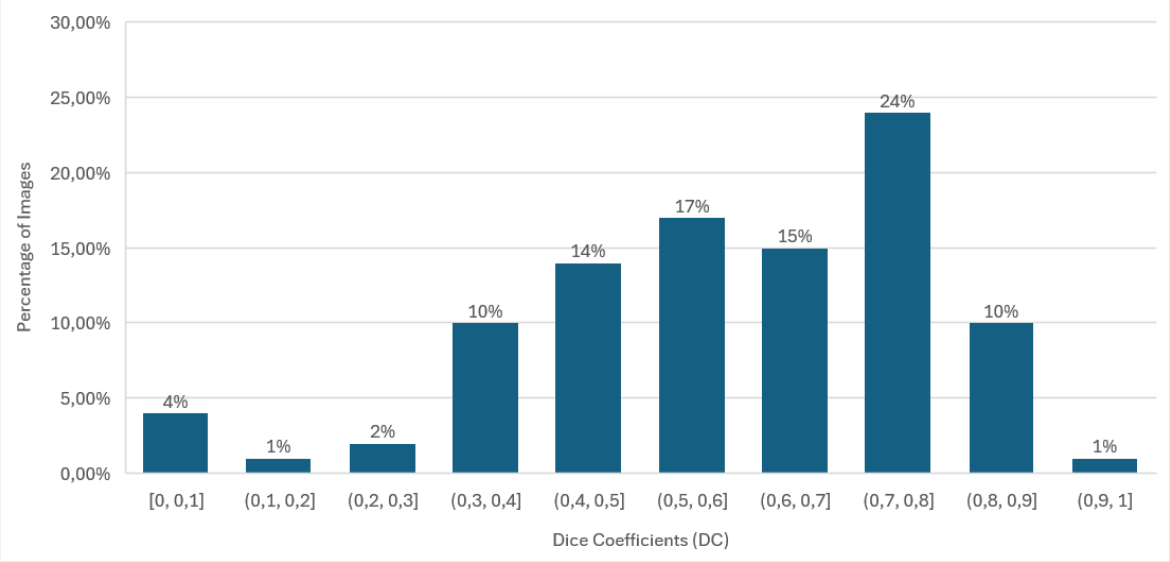


Figure 4.1: *Distribution of Dice Coefficients across the nnU-Net segmentation-quality evaluation set, used to guide the stratified selection of the 24-patient dosimetric cohort.*

4.1.3 Geometric Evaluation Metrics

First, segmentation agreement between the manual and AI GTV masks was quantified geometrically, directly on the binary masks used by the dosimetric pipeline. Three indices were used: the Dice Coefficient in its original formulation [140], and the Geometrical Miss Index and Discordance Index, defined as in [133].

Let G denote the manual GTV mask and S the AI GTV mask, both expressed as sets of voxels on a shared patient grid. The Dice Coefficient is

$$\mathrm{DC} = \frac{2\,|G \cap S|}{|G| + |S|}, \tag{4.1}$$

the Geometrical Miss Index quantifies the fraction of the manual (reference) volume missed by the AI segmentation,

$$\mathrm{GMI} = \frac{|G \setminus S|}{|G|}, \tag{4.2}$$

and the Discordance Index quantifies the fraction of the AI volume that exceeds the manual volume,

$$\mathrm{DI} = \frac{|S \setminus G|}{|S|}. \tag{4.3}$$

Dice Coefficient is thus an evaluation of the overlapping between the two masks[1], while the Geometrical Miss Index and the Discordance Index measure the dissimilarity between volumes.

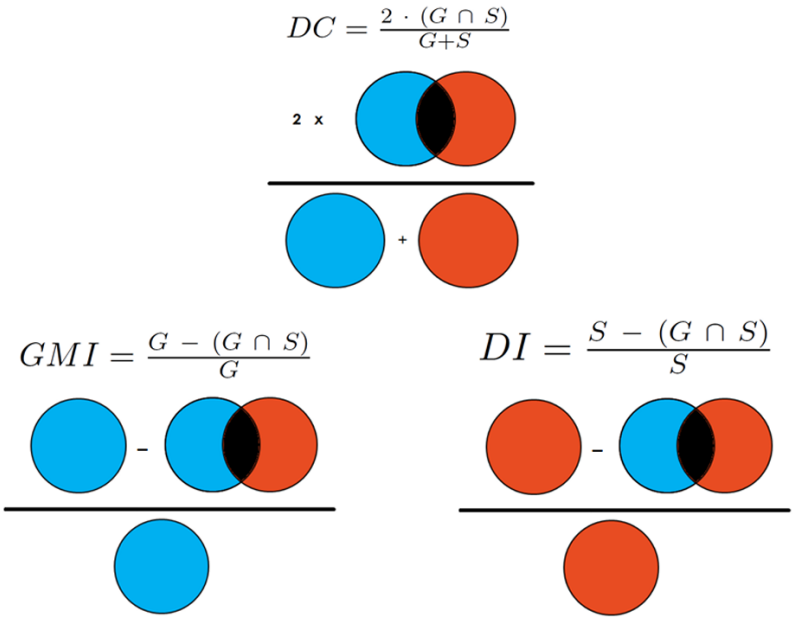


Figure 4.2: *Schematic illustration of the Dice Coefficient, the Geometrical Miss Index and the Discordance Index, respectively. The blue circle represents the ground-truth volume G, the red one is the simulated S, and the black part is the intersection between the two volumes* $G \cap S$ [133].

Table 4.1 reports, for every patient in the 24-patient dosimetric cohort, the three geometric indices of Equations (4.1)–(4.3) together with the manual and AI GTV volumes (computed as voxel count $\times 0.008\,\mathrm{cm}^3$, the volume of a single 2 mm voxel), sorted by increasing DC. The full per-patient voxel counts, including the intersection

[1] $G \cap S$ – the intersection of the manual and AI masks – reappears throughout this Chapter as the domain of comparison for gamma-index analyses (Section 4.4.7), for a reason developed there: it is the only region in which both segmentations agree the tissue is tumor.

volume $|G \cap S|$ used for gamma-index normalization, are reported in Table D.1 of Appendix D.

Table 4.1: Segmentation-agreement metrics and GTV volumes for the 24-patient dosimetric cohort, sorted by DC.

Patient	DC	GMI	DI	V_{manual} [cm^3]	V_{AI} [cm^3]
HN_1019	0.000	1.000	1.000	1.05	6.10
HN_0149	0.197	0.889	0.140	43.46	5.62
HN_0591	0.374	0.606	0.645	23.55	26.10
HN_0580	0.384	0.744	0.232	43.49	14.50
HN_0806	0.401	0.749	0.005	80.73	20.38
HN_0569	0.457	0.664	0.288	47.52	22.44
HN_0689	0.473	0.075	0.683	3.22	9.38
HN_0109	0.497	0.655	0.111	109.73	42.63
HN_0150	0.499	0.661	0.057	144.39	51.94
HN_1640	0.560	0.610	0.002	131.33	51.26
HN_1615	0.573	0.424	0.429	27.17	27.38
HN_1583	0.579	0.471	0.362	49.58	41.13
HN_1555	0.670	0.469	0.092	75.86	44.32
HN_0792	0.691	0.380	0.221	75.49	60.10
HN_1581	0.691	0.135	0.425	21.39	32.15
HN_0621	0.702	0.457	0.010	131.12	71.94
HN_1588	0.742	0.118	0.360	60.03	82.74
HN_1631	0.754	0.353	0.096	89.00	63.70
HN_0847	0.762	0.291	0.177	62.50	53.87
HN_0601	0.780	0.110	0.306	53.03	67.97
HN_0687	0.838	0.233	0.076	75.63	62.84
HN_0666	0.851	0.211	0.076	70.93	60.54
HN_1551	0.868	0.162	0.101	24.63	22.95
HN_0717	0.901	0.055	0.139	77.06	84.54

Two features of Table 4.1 deserve to be commented. First, DC and volume are not simply correlated: patient HN_0806 has a large manual GTV (80.7 cm^3) but a low DC (0.401), while HN_1551 has a small manual GTV (24.6 cm^3) but a high DC (0.868). Second, GMI and DI vary largely independently of one another, and of DC itself: HN_0689 combines a very low GMI (0.075, almost the entire manual volume is captured) with a very high DI (0.683, the AI mask substantially over-extends beyond the manual volume), the opposite pattern to the complete miss of HN_1019. This is the geometric basis for treating GMI and DI as distinct failure modes rather than collapsing them into DC alone, a distinction that is shown to impact on dosimetry in Section 4.5.5.

4.2 Image Segmentation

Before any Monte Carlo geometry can be built, every patient's planning CT must be converted into a complete, labeled anatomical model: not only the GTV, but every OAR and tissue compartment that influences the dose calculation or the patient-positioning transform. Section 3.2.2 described, in general terms, the fixed-resolution voxel geometry-reconstruction strategy adopted by OpenPINT: segmented organ masks are combined into a single labelled NIfTI volume and rasterized into an MCNP-compatible lattice. This Section documents how that strategy was instantiated for the present H&N cohort.

4.2.1 Organ-at-Risk Segmentation

Beyond the GTV mask, every OAR and anatomical structure required for the Monte Carlo geometry was segmented automatically from the CT using two complementary tools: TotalSegmentator, applied directly to the CT, and CERR own automated segmentation models, applied where a structure of interest is not covered by TotalSegmentator label set or where an independent estimate was desired.

Left/right pairs (e.g., carotid arteries, parotid and submandibular glands, masticatory muscles) are merged into a single bilateral mask immediately after loading, since the final MCNP geometry does not distinguish laterality for any structure. Structures sharing the same underlying tissue composition but with independent clinical relevance as separate dose-reporting volumes (e.g., pharyngeal constrictor muscles, salivary glands, thyroid) are kept as distinct masks and MCNP tally volumes even though they are later assigned to the same physical material.

The air-filled cavities of the upper aerodigestive tract – nasopharynx, oropharynx, hypopharynx, oral cavity, nasal cavities, laryngeal airway, and trachea – are combined into a single mask and assigned the internal-air material in the final MCNP geometry, as shown in Table 4.2. This is fundamental for the creation of the OAR mask, since no segmentation tool provides a direct mucosal-surface label.

Because acute mucositis is the clinically established, dose-limiting toxicity endpoint for H&N BNCT, a dedicated geometric procedure was developed to reconstruct an oropharyngolaryngeal mucosa mask from the lumen and soft-tissue segmentations already available, to use as the dose-limiting OAR in the irradiation-time calculation.

The mucosa is reconstructed as a peri-luminal shell around the airway lumen, in the following sequence:

1. **Reference lumen.** A reference lumen volume is built from the union of the oral cavity, oropharynx, hypopharynx, and laryngeal airway. The nasal cavities and nasopharynx are deliberately *excluded*: including them in early iterations of this procedure was found to systematically overestimate the mucosal volume.

2. **Preliminary expansion.** The reference lumen is morphologically closed and dilated by 2 mm, producing a continuous reference volume prior to shell construction.

3. **Raw shell.** The expanded lumen is dilated by a further 4 mm and the (already expanded) lumen itself is subtracted, yielding the ring of tissue immediately adjacent to the airway.

4. **Uniformization.** The raw shell is morphologically closed, with an explicit 1.5 mm respect margin excluded around the GTV mask, to prevent any overlap between the mucosa OAR and the tumor volume. This exclusion step is precisely what produces a different mucosa voxel count in every arm that uses a different GTV mask – illustrated concretely for patient HN_0717 in Section 4.4.6.

5. **Anatomical exclusion.** All non-mucosal structures that the purely geometric shell can erroneously include are subtracted explicitly: bone; cartilage; teeth; hard and soft palate; salivary glands; external auditory canal; masticatory, suprahyoid and pharyngeal constrictor neck muscles; tongue and laryngeal soft tissue. Neck structures with a per-patient-variable presence in the field of view (major neck vessels, thyroid gland, esophagus) are included in this exclusion set only if their segmented volume exceeds a minimum-voxel threshold for that patient, since not every CT field of view extends far enough caudally to include them.

6. **Hounsfield-unit safety filter.** Any residual voxel with HU < -200 is removed regardless of its origin: true mucosal tissue does not present pure-air density, so this filter is anatomically justified independently of which upstream step produced the residual voxel.

7. **Fragment removal.** Connected components smaller than 80 voxels are discarded as residual segmentation noise at tissue interfaces.

8. **Volume sanity check.** The final mucosal volume is checked against an expected physiological range of 40–150 cm^3, and flagged for manual review if outside this range.

The criterion selected for mucosa delineation has been chosen from the clinical experience conducted in Finland, adopted after personal communication with the medical physicists of the Helsinki hospital.

4.2.2 MCNP Geometry Generation

The overall patient body contour, `PATIENT` (`PAT`), is constructed as the union of every named anatomical mask produced above, together with a whole-body contour obtained directly from TotalSegmentator, used specifically to guarantee that `PAT` reflects the true external body shape even in regions not covered by any individual organ segmentation.

This union is refined by a sequence of morphological operations: a binary opening (erosion followed by dilation, common radius) to detach thin, non-anatomical protrusions left over from the mask union without systematically eroding the true external contour; a binary closing (dilation followed by erosion, an independent radius) to bridge external gaps and channels that the opening step does not close, such as a corridor left uncovered where a metal marker or artifact prevented any TotalSegmentator label from being assigned; two-dimensional, per-slice hole filling; and, after each morphological step that can alter voxel connectivity, retention of the largest connected component only. Both the erosion and dilation operations are implemented with explicit edge-padding of the cropped array before the operation (and cropping back afterward): without this, the array boundary is otherwise treated as background by the underlying morphology routines, spuriously eroding exactly the axial extremes of the volume – in this geometry, the cranial vertex and the base of the neck – an effect verified numerically by comparing the active voxel count on those extreme slices before and after each step. `AIR` is defined as the logical complement of the resulting `PAT` mask. The opening and closing radii are patient-specific parameters, calibrated by visual inspection for each case rather than fixed a priori. Figure 4.3 shows the resulting `PATIENT` mask for the pilot patient HN_0717.

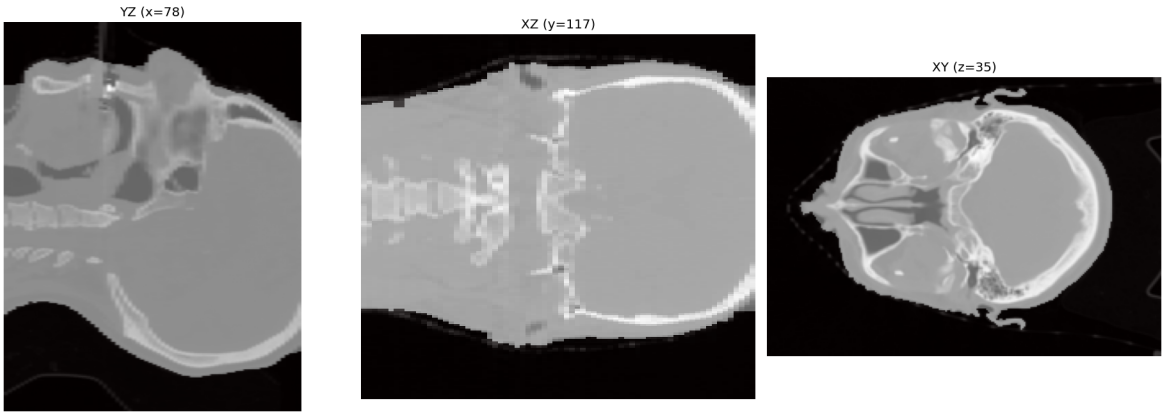


Figure 4.3: *PATIENT mask for patient HN_0717 (manual GTV mask), shown in orthogonal section.*

Table 4.2 lists the complete set of named structures, grouped by the MCNP material to which each is ultimately assigned. Coarse, clinically non-actionable skeletal or cartilaginous sub-structures (e.g., individual vertebrae, the zygomatic arch, the styloid process) are merged into their parent structure (`BONE_TOTAL`, `CARTILAGE_TOTAL`); structures with independent clinical relevance as a dose-reporting volume are kept as separate masks and MCNP tally regions even where they share the same underlying material composition as another structure.

Table 4.2: Anatomical structures grouped by assigned MCNP material.

MCNP material	Structures included
External air	`AIR`
Internal air	`UPPER_AERODIGESTIVE_LUMEN`
Generic soft tissue (fallback)	`HEAD`
Skin	`SKIN`
Mucosa	`MUCOSA_UNIFORM`
Bone	`BONE_TOTAL`
Cartilage	`CARTILAGE_TOTAL`
Teeth	`TEETH_TOTAL`
Paranasal sinus	`SINUS_TOTAL`
Muscle	`MASTICATORY_MUSCLES_TOTAL`, `SUPRAHYOID_TOTAL`, `NECK_MUSCLES_TOTAL`, `PHARYNGEAL_CONSTRICTORS_TOTAL`, `TONGUE`, `LARYNX_TISSUE`
Vascular	`NECK_MAJOR_VESSELS_TOTAL`
Glandular	`SALIVARY_GLANDS_TOTAL`, `THYROID_GLAND`
Palate	`PALATE`
Esophagus	`ESOPHAGUS`
Ocular	`OCULAR_COMPLEX_TOTAL`, `AUDITORY_CANAL`
Critical CNS	`CNS_CRITICAL`, `BRAIN`
Tumor	`GTV`

Figure 4.4 shows the resulting final MCNP geometry for the same patient, with its material-assignment legend. Once every structure is finalized, the complete set of masks

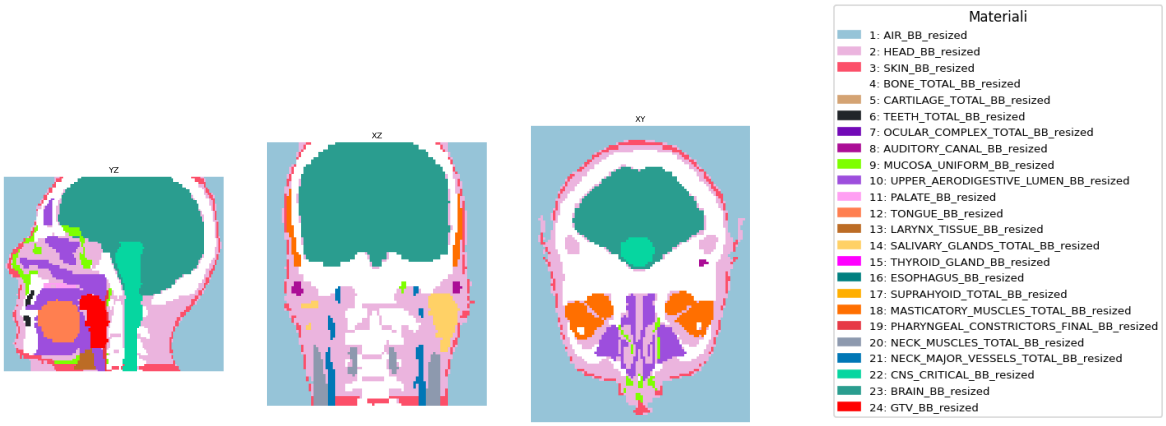


Figure 4.4: *Final MCNP geometry for patient HN_0717 (manual GTV mask), with the corresponding material-assignment legend (Table 4.2).*

is combined into a single multi-class label volume following an explicit priority order (generic-to-specific, with the GTV and the mucosa OAR assigned the highest priority so that no other structure can overwrite them in case of residual overlap); cropped to a bounding box around the patient body contour; resampled to the production voxel size; padded with a synthetic air margin on every side; and written out as an MCNP lattice input, following the input-preparation workflow of Section 3.3.3. An automated consistency check compares the declared voxel size and the affine-derived spacing of the resampled volume, and a second check verifies, structure by structure, that no voxel originally assigned to a given class was silently overwritten by a higher-priority structure during the final combination step without being explicitly logged.

4.3 Treatment Simulation

4.3.1 Four-Arm Simulation Design

For every patient in the 24-patient cohort, four dosimetric configurations were produced:

- **Manual**: a complete, independent Monte Carlo simulation built from the CT and the manually contoured GTV, following the OpenPINT workflow, including automated patient positioning computed from the manual GTV centroid.
- **AI**: a complete, independent Monte Carlo simulation built in the same way,

but from the nnU-Net-derived GTV mask, including its own automated patient positioning computed from the AI GTV centroid.

- **Swap**: the absorbed dose grid from the *manual* simulation, re-processed with the *AI* GTV mask applied only at the dose-extraction stage.
- **Anti-swap**: the absorbed dose grid from the *AI* simulation, re-processed with the *manual* GTV mask applied only at the dose-extraction stage.

This four-arm design disentangles two effects that are hidden in a naive manual-versus-AI comparison. Because the automated positioning in OpenPINT computes the beam-alignment transform from the GTV centroid, two segmentations of the same patient that disagree geometrically will, in general, also produce two different patient positions relative to the beam port.

The manual-versus-AI comparison therefore measures the *combined* effect of (i) a different target volume and (ii) a different treatment geometry. The manual-versus-swap comparison, sharing an identical dose grid and beam geometry with manual by construction, isolates effect (i) alone, evaluated at manual positioning. The manual-versus-anti-swap comparison, sharing an identical dose grid and beam geometry with AI by construction but evaluated on the manual mask, isolates effect (ii) alone, evaluated on the true tumor volume – the configuration of direct clinical relevance, since it is the one that would actually be delivered if treatment were planned on the AI segmentation. Swap and anti-swap are thus symmetric counterparts, isolating the two confounded effects from opposite ends of the same four-arm design. This logic, and how it plays out concretely, is illustrated end-to-end for patient HN_0717 in Section 4.4.6, and analyzed systematically over the full cohort in Section 4.5.

All four configurations were computed at the production voxel size of 2 mm, using MCNP6.3.0 with $N_{\text{particles}} = 1 \times 10^9$ per simulation. Boron concentrations were fixed at 50 ppm for the GTV and 30 ppm for healthy mucosa.

4.3.2 Patient Positioning

OpenPINT positions the patient automatically inside the BSA, shown in Figures 4.5 and 4.6, using the rigid-transform procedure of Equation (3.2): the beam-entry point and alignment direction are computed from the GTV centroid c_{GTV}, and the resulting

translation and rotation are written into the MCNP coordinate-transform card that places the patient geometry relative to the fixed BSA geometry. Because this transform is computed *directly* from the segmented GTV, it is the mechanism through which a geometric segmentation disagreement propagates into a difference in beam alignment, independently of any difference in the shape or volume of the GTV mask.

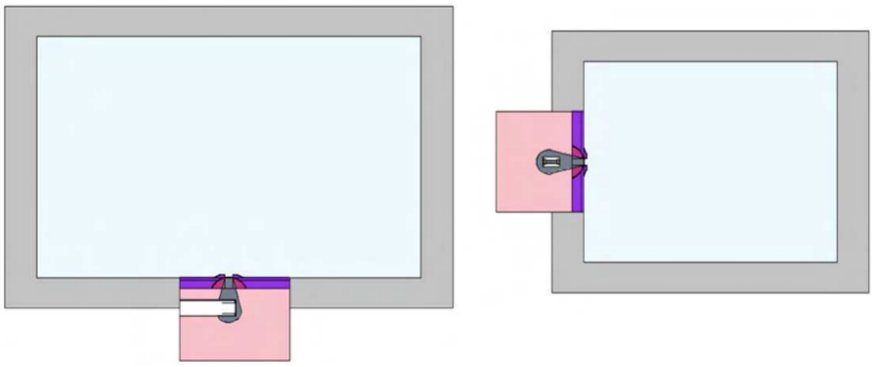


Figure 4.5: *Horizontal (left) and vertical (right) sections of the MCNP6 model of the treatment room. The simulated treatment room has dimensions of 3.30 × 4× 6 m3 with walls of 50 cm in borated concrete* [141].

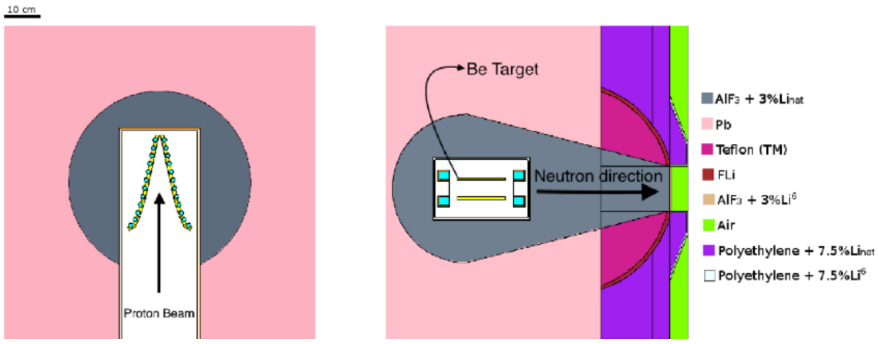
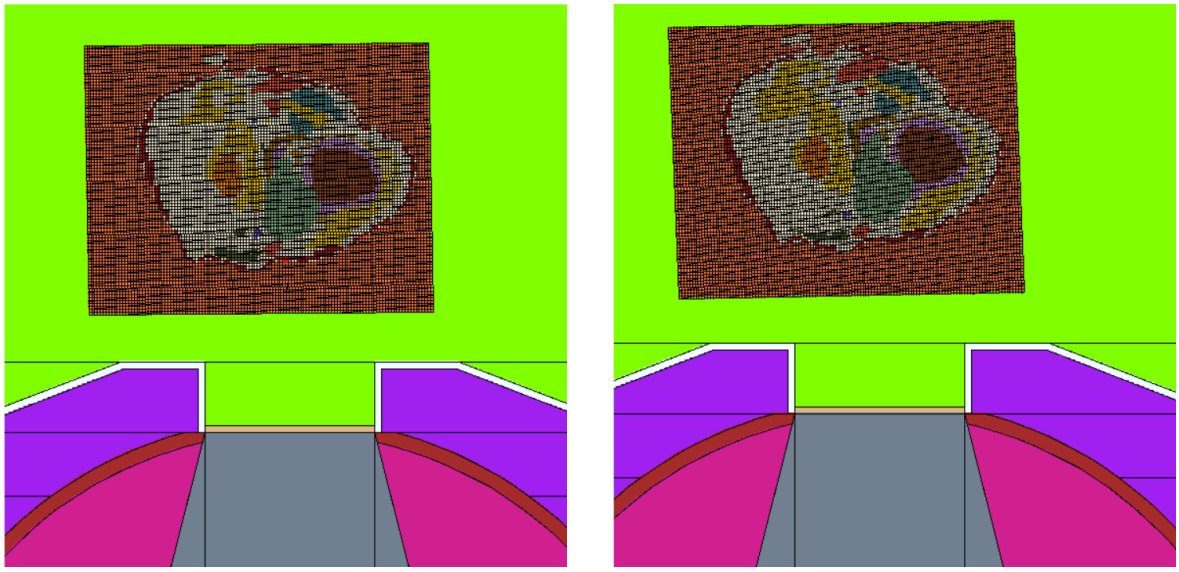


Figure 4.6: *Representation of the simulated BSA. Neutrons are produced via nuclear reaction of protons in the beryllium target and are moderated along the lithiated aluminum fluoride structure (grey)* [141].

For the pilot patient HN_0717, the GTV centroid c_{GTV} computed from the manual and AI masks differs by 1.9 mm, and the resulting rigid-transform translation vector written into the MCNP coordinate-transform card differs by 3.4 mm between the two configurations, while the beam-alignment rotation angle is unchanged. On a 2 mm voxel grid, a 3.4 mm shift of the entire simulated geometry relative to the beam is sufficient to displace essentially every voxel of the GTV by a non-negligible fraction of a voxel width relative to the beam-dependent dose gradient. This is not a peculiarity of HN_0717: it is the generic mechanism by which segmentation discordance and beam-



(a) *Manual configuration.* **(b)** *AI configuration.*

Figure 4.7: *Automated patient positioning for patient HN_0717. The two beam-alignment transforms differ by a 1.9 mm GTV-centroid shift and a 3.4 mm translation-vector difference (Section 4.3.2).*

geometry discordance become coupled in this pipeline, and its cohort-wide dosimetric consequence is documented directly in the gamma-index analysis of Section 4.5.4.

The positioning transform of Equation (3.2) places the patient so that the beam axis passes through the GTV centroid, but applying this transform would, for most H&N patients, bring the shoulders into direct conflict with the BSA aperture: the GTV centroid lies close to the head/neck region, while the shoulders extend laterally well beyond the collimator opening designed for a head-only treatment. To avoid this collision, an additional 7 cm longitudinal shift was applied on top of the centroid-derived transform, translating the patient further from the BSA along the beam axis so that the shoulders clear the aperture while the beam-entry point and alignment direction computed from the GTV centroid are preserved. This makes the treatment simulation more realistic compared to real BNCT, accounting for a certain degree of beam aperture and out-of-target dose, which may limit the irradiation time for healthy tissues irradiation.

4.3.3 Voxel-Size Sensitivity Study

A fixed-resolution voxel geometry requires a choice of voxel size that balances geometric fidelity, statistical precision, and computational cost. This choice was investigated on the pilot patient HN_0717 (manual segmentation only, chosen for its high segmentation

agreement so that the study would not be confounded by contour-quality effects) at five voxel sizes: 1, 2, 3, 4, and 5 mm, each simulated with a target of 1×10^9 particle histories. Five quantities were tracked as a function of voxel size: the TCP; the near-minimum GTV dose under the photon-isoeffective model, $D_{\min}$; the discretized GTV volume; the mucosa-limiting irradiation time; and, separately, the real (wall-clock) computation time and the median relative statistical error in the mucosa.

Dosimetric Convergence

Table 4.3 reports the percentage deviation of each dosimetric quantity from its value at the finest tested resolution (1 mm). TCP does not converge monotonically over the tested range: it decreases by 6.0% from 1 mm to 2 mm, then increases monotonically and substantially from 2 mm to 5 mm, reaching a deviation of +42.7% relative to 1 mm at the coarsest resolution tested. $D_{\min}$ shows a cleaner, strictly monotonic increase with voxel size (from 0% at 1 mm to +72.1% at 5 mm), while the discretized GTV volume decreases overall (from 0% to −7.6% at 5 mm), though not monotonically.

Table 4.3: Percentage deviation of dosimetric and geometric quantities from their value at 1 mm voxel size, patient HN_0717 (manual dataset).

Voxel size [mm]	TCP dev. [%]	$D_{\min}$ dev. [%]	Volume dev. [%]	t_{irr} dev. [%]	Mucosa rel. err. dev. [%]
1	0.0	0.0	0.0	0.0	0.0
2	−6.0	+9.6	−2.4	+22.9	−47.4
3	+23.6	+39.5	−4.8	+50.4	−63.9
4	+29.4	+59.1	−4.1	+59.7	−73.4
5	+42.7	+72.1	−7.6	+66.4	−78.6

This pattern is consistent with a partial-volume mechanism rather than genuine Monte Carlo convergence: as the voxel size increases, the low-dose peripheral voxels of the GTV mask are progressively eroded from the discretized volume, leaving a smaller, more centrally located residual volume with systematically higher minimum and near-minimum dose. Because the retrospective clinical literature reviewed in Section 1.4.3 identifies minimum GTV dose as the strongest independent predictor of H&N BNCT outcome, the non-convergent, strictly monotonic behavior of $D_{\min}$ over the tested range is the most direct evidence that the chosen production voxel size (2 mm) is a pragmatic compromise between a good statistics and geometrical precision.

Computational Cost and Statistical Precision

Table 4.4 reports the real (wall-clock) computation time and the mucosa median relative statistical error at each voxel size.

At fixed $N_{\text{particles}} = 1 \times 10^9$, the real computation time is essentially constant between 2 mm and 5 mm (within 3% of one another), with a distinct increase only at 1 mm (+14% relative to 2 mm), since the dominant computational cost is the transport of a fixed number of source particles rather than the tally-mesh bookkeeping. Over the same range, the mucosa median relative statistical error decreases continuously and substantially, from 5.8% at 5 mm to 27.1% at 1 mm – a factor of 4.7.

Table 4.4: Computational cost and statistical precision versus voxel size, patient HN_0717 (manual dataset, $N_{\text{particles}} = 1 \times 10^9$ target).

Voxel size [mm]	Wall-clock time [h]	nps completion	Mucosa median rel. error
1	11.03	100%	0.271
2	9.68	100%	0.142
3	9.79	100%	0.098
4	9.75	100%	0.072
5	9.59	100%	0.058

Taken together, Tables 4.3 and 4.4 support the following characterization of the 2 mm choice: it is the coarsest resolution at which the real computational cost is already at its plateau value, while remaining fine enough that a further refinement to 1 mm would not be justified, since 1 mm has *worse* statistical precision than 2 mm at fixed particle count. The choice of 2 mm is therefore defended as a resource-efficiency argument, not a convergence argument.

4.4 Dosimetric Analysis: Patient HN_0717

The same analysis pipeline is applied uniformly to all 24 patients; before the results of the entire study are presented in Section 4.5, this Section describes the dosimetric quantities computed on each patient. Patient HN_0717 is used throughout, both because it is the pilot patient of the voxel-size study and because its very high segmentation agreement (DC = 0.901, Table 4.1) makes it a clean illustration of the method.

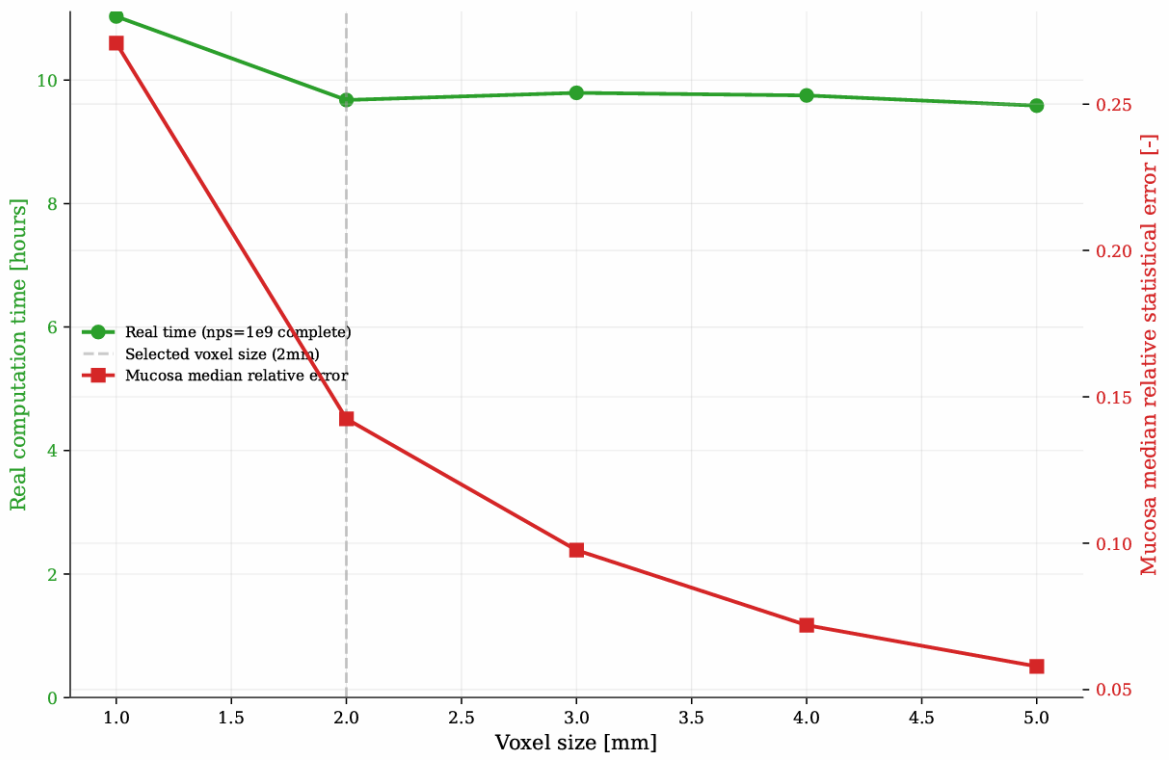


Figure 4.8: *Real (wall-clock) computation time and mucosa median relative statistical error as a function of voxel size, patient HN_0717 (manual dataset), at fixed* $N_{\text{particles}} = 1 \times 10^9$.

4.4.1 Two Radiobiological Dose Models

Every dosimetric quantity in the remainder of this Chapter can be expressed under either of two radiobiological dose models: the classical, fixed RBE/CBE-weighted dose (Section 1.3.2) and the photon-isoeffective (IsoE) dose (Section 1.3.3). Both were computed for every structure, patient, and configuration in the cohort.

The fixed RBE/CBE-weighted dose remains the quantity most commonly reported in the BNCT dosimetric literature, and is retained here mainly for continuity and comparability with that literature and with the glioblastoma benchmark of [10, 133] discussed throughout this Chapter. A fixed RBE or CBE factor does not correctly capture how biological effect scales with the dose level itself, as extensively explained by the authors of the model [14], so the first is not an appropriate input to the TCP model of Equation (1.28). TCP throughout this Chapter is therefore computed exclusively from the isoeffective GTV dose, following the approach demonstrated for BNCT tumor control by González et al. [14].

The same reasoning governs the treatment of mucosal complication risk (NTCP), with one further consideration specific to that endpoint. The dose-response parame-

terization used for NTCP is a model fitted to clinical outcome data from conventional photon-based external-beam radiotherapy; its TD_{50} threshold is therefore only physically meaningful when applied to a photon-equivalent dose. This makes the isoeffective mucosa dose the theoretically correct input to this model. In practice, however, applying the literature TD_{50} to the mucosa isoeffective dose of every patient in this cohort saturates the resulting probability numerically to $\sim 10^{-8}$–10^{-9}: the model is being evaluated far into the low-dose tail of its sigmoid, where it no longer discriminates inter-patient or inter-configuration differences. Rather than report a saturated, uninformative probability, the mean isoeffective mucosa dose itself is used throughout this Chapter as the primary comparative index of mucosal risk across patients and configurations. For reference, and for comparability with the absorbed-dose NTCP conventionally reported in earlier BNCT dosimetric work, the same dose-response model is also evaluated directly on the absorbed physical mucosa dose and reported alongside the isoeffective dose; this absorbed-dose NTCP is *not* claimed to be the theoretically preferred formalism for this endpoint, and is reported for comparison only, exactly as the fixed RBE/CBE-weighted GTV dose is retained for comparison alongside the isoeffective one.

4.4.2 Dose Components

The total absorbed dose at every voxel is the sum of the four components: the ^{10}B(n,α)^{7}Li boron neutron-capture reaction (B10), the ^{14}N(n,p)^{14}C nitrogen-capture reaction (N14), fast-neutron elastic scattering (n), and incident/secondary photons (g). These are combined into the RBE-weighted dose using the fixed compound-specific and tissue-specific weighting factors of Section 1.3.2, and, separately, into the photon-isoeffective dose using the formalism of Section 1.3.3.

Table 4.5 reports the mean, median, P98[2], and maximum of each component for the GTV and the mucosa under the absorbed and RBE-weighted representations, together with the corresponding isoeffective dose value.

Two features of Table 4.5 are worth flagging as generic to BNCT dosimetry rather

[2]P50 and P98 here are statistical percentiles of the per-voxel dose distribution within the structure (P98 is therefore a *near-maximum* value, close to Max). This is a different convention from the D_x notation used in the dose-volume-histogram figures of merit elsewhere in this chapter (e.g. Table 4.6), where D_{98} denotes the dose received by at least 98% of the structure's volume and is consequently a *near-minimum* value, close to $D_{\min}$.

Table 4.5: Dose-component breakdown, GTV and mucosa, patient HN_0717, manual configuration.

Structure	Component	Mean [Gy]	P50 [Gy]	P98 [Gy]	Max [Gy]	Mean fraction
GTV, absorbed	B10	3.749	3.655	6.021	7.332	64.3%
	N14	0.152	0.149	0.244	0.296	2.6%
	n	0.269	0.228	0.809	1.849	4.6%
	g	1.661	1.645	2.462	3.411	28.5%
	Total	5.831	5.687	8.855	10.411	100%
GTV, RBE-weighted	B10	14.247	13.891	22.881	27.861	82.6%
	N14	0.487	0.475	0.780	0.948	2.8%
	n	0.861	0.730	2.589	5.917	5.0%
	g	1.661	1.645	2.462	3.411	9.6%
	Total	17.256	16.800	27.294	32.130	100%
GTV, isoeffective	Total	18.383	—	—	27.610	—
Mucosa, absorbed	B10	0.848	0.762	2.191	2.934	39.0%
	N14	0.057	0.052	0.150	0.200	2.6%
	n	0.161	0.111	0.663	1.489	7.4%
	g	1.108	1.068	1.932	2.901	50.9%
	Total	2.174	2.027	4.298	6.000	100%

than specific to this patient. First, the boron component rises from 64.3% of the absorbed dose to 82.6% of the RBE-weighted dose, since the CBE factor applied to the boron component is far larger than the RBE factors applied to the other components. This reflects the higher radiobiological effectiveness of the high-LET component due to capture in boron. Second, the photon (g) component is, by definition, identical in the absorbed and RBE-weighted representations, as the RBE of photons is 1; RBE-weighting therefore only rescales the neutron-capture and fast-neutron components.

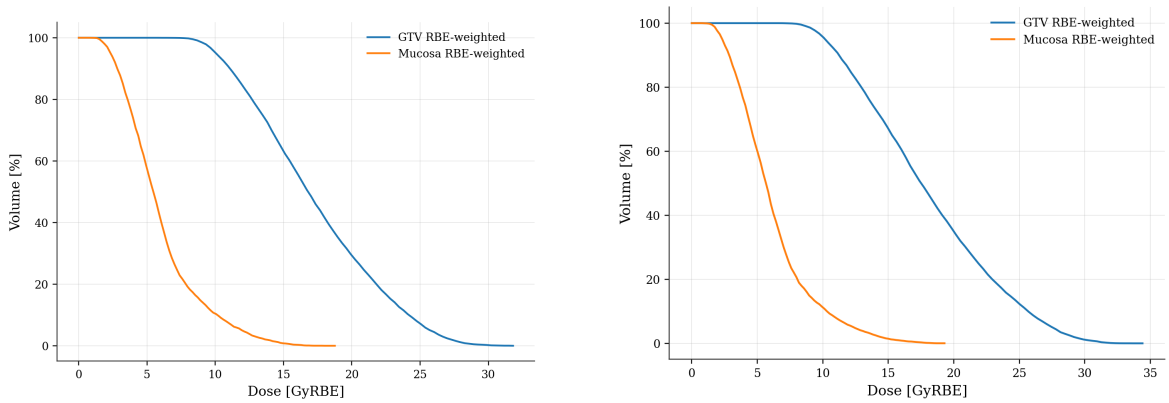
It is also worth noting that the GTV isoeffective mean dose (18.38 Gy) already exceeds the RBE-weighted mean (17.26 Gy) for this patient; this comparison, and why the two models diverge, is developed quantitatively for the full cohort in Section 4.5.2.

4.4.3 Dose-Volume Histograms

A dose-volume histogram (DVH) summarizes the full spatial dose distribution within a structure into a single curve, plotting the fraction of the structure's volume that receives at least a given dose against that dose; it is the standard tool for comparing target coverage and OAR sparing without having to inspect the dose field voxel by voxel. For HN_0717, DVHs were extracted for both the GTV and the mucosa under the RBE-weighted model, and, separately, for the GTV under the photon-isoeffective model; the two structures are always compared under the *same* radiobiological weighting.

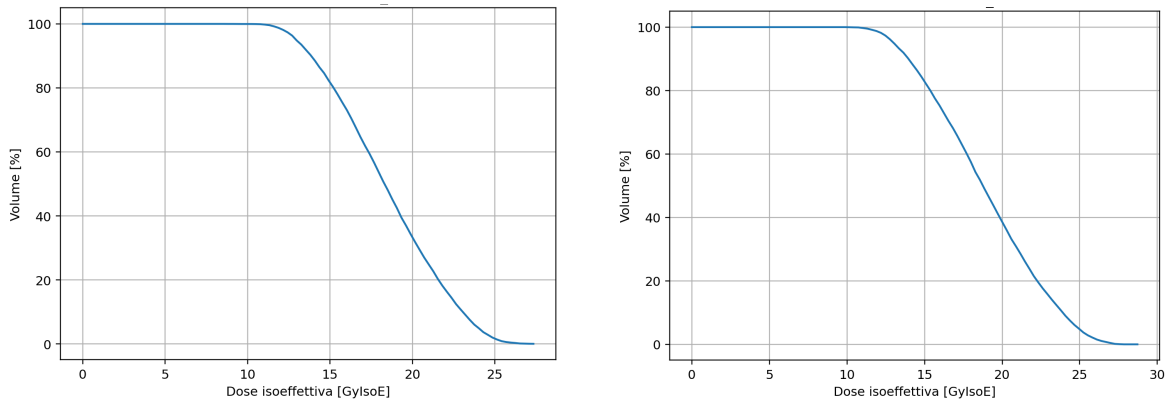
Figure 4.9 shows the GTV and mucosa RBE-weighted DVHs for patient HN_0717, manual versus AI configuration; Figure 4.10 shows the corresponding GTV photon-isoeffective (IsoE) DVH. Table 4.6 reports the dosimetric figures of merit for both configurations. Swap and anti-swap are not shown as separate DVHs here, since their GTV dose statistics are simply the manual and AI dose fields respectively, re-extracted under the other mask.

It bears emphasizing that the mucosa RBE-weighted DVH shown here is used only to compare target and OAR dose distributions on a common radiobiological scale: the mucosa-limiting irradiation time itself is computed from the *absorbed* dose rate, as in clinical experience in Finland.



(a) *Manual configuration.* **(b)** *AI configuration.*

Figure 4.9: *GTV and mucosa dose-volume histograms, both RBE-weighted, patient HN_0717, manual versus AI configuration.*



(a) *Manual configuration.* **(b)** *AI configuration.*

Figure 4.10: *GTV photon-isoeffective (IsoE) dose-volume histogram, patient HN_0717, manual versus AI configuration.*

Three aspects are worth noting for the interpretation of DVHs throughout this thesis.

Table 4.6: Dosimetric figures of merit, GTV and mucosa, patient HN_0717, manual versus AI configuration.

Structure	Configuration	$D_{\min}$ [Gy]	$D_{\max}$ [Gy]	D_{mean} [Gy]	D_{98} [Gy]	D_2 [Gy]
GTV, RBE-weighted	Manual	5.98	32.13	17.26	9.26	27.29
	AI	6.76	34.75	18.02	9.29	29.16
GTV, IsoE	Manual	8.78	27.61	18.38	12.18	24.84
	AI	9.94	29.01	18.85	12.26	25.90
Mucosa, absorbed	Manual	0.50	6.00	2.17	0.95	4.30
	AI	0.52	6.00	2.27	0.98	4.58

First, the mucosa *absorbed*-dose figures of merit reported in Table 4.6 are truncated at exactly 6.0 Gy in both configurations by construction.

Second, the GTV IsoE D_{mean} exceeds the RBE-weighted D_{mean} in both configurations for this patient. This may look surprising: at high absolute dose, the quadratic term of the isoeffective model typically pulls its estimate *below* the fixed-weighting one, so the more common pattern in this cohort is $D_{\text{IsoE}} > D_w$ only at comparatively low-to-moderate dose. HN_0717 sits precisely in that lower-dose regime: its manual mean IsoE dose (18.38 Gy) falls close to the 18.19 Gy average of the cohort majority for which $D_{\text{IsoE}} > D_w$ (Table 4.14), rather than near the 21.7 Gy average of the minority for which the opposite holds. HN_0717's direction is therefore consistent with the cohort-wide dose-dependent pattern documented in Section 4.5.2.

Third, for this particular patient, every figure of merit is systematically higher for the AI configuration than for the manual one, for both the GTV and the mucosa. This is the dosimetric consequence of the AI mask being larger than the manual mask for HN_0717 specifically (84.54 cm^3 vs. 77.06 cm^3, Table 4.1), which shifts the automated positioning transform and, in turn, the dose delivered to both structures. It is a patient-specific illustration of a general mechanism (segmentation size and shape feed into automated positioning), not a claim that AI values are systematically higher across the cohort: Section 4.5 shows the pooled GTV dose metrics are not, in fact, significantly different between manual and AI once averaged over all 24 patients.

4.4.4 Mucosa-Limiting Irradiation Time

The irradiation time is set by the first voxel of the mucosa OAR to reach its absorbed dose limit (6.0 Gy, Table 4.6), as explained above. For patient HN_0717 manual configuration, this limiting voxel is located at $(i, j, k) = (56, 58, 41)$, corresponding to physical

coordinates $(x, y, z) = (-112.0, -116.0, 82.0)$ mm, with a dose rate of 0.1240 Gy/min, giving an irradiation time of 48.40 min. For the AI configuration, the limiting voxel shifts to $(i, j, k) = (56, 57, 40)$, so $(x, y, z) = (-112.0, -114.0, 80.0)$ mm, leading to a 2 mm shift along y and z relative to the manual hotspot – with a lower dose rate of 0.1167 Gy/min, thus giving a longer irradiation time of 51.42 min (Table 4.7). Because anti-swap reuses the AI dose grid and mucosa hotspot search unchanged, its irradiation time for this patient is identical to AI's.

Table 4.7: Mucosa-limiting voxel and irradiation time, patient HN_0717.

	Manual	AI
Limiting voxel (i, j, k)	$(56, 58, 41)$	$(56, 57, 40)$
Position (x, y, z) [mm]	$(-112, -116, 82)$	$(-112, -114, 80)$
Dose rate [Gy/min]	0.1240	0.1167
Irradiation time [min]	48.40	51.42
Combined relative error [%]	9.60	9.04
Reference threshold [%]	10.0	10.0

The longer AI irradiation time is a direct consequence of the lower dose rate at the AI-configuration hotspot; since both configurations share the same 6.0 Gy mucosa limit, a lower limiting dose rate means a longer irradiation, which in turn raises the delivered dose everywhere else in the geometry, including the GTV – consistent with the systematically higher GTV figures of merit already noted for the AI configuration in Table 4.6.

Because this single voxel determines the delivered dose to the entire treatment, its statistical uncertainty is tracked and decomposed explicitly. Table 4.8 reports this decomposition, by dose component, for both configurations.

Table 4.8: Statistical-error decomposition at the mucosa-limiting voxel, patient HN_0717, manual versus AI configuration.

	Manual		AI	
Component	Dose-rate [Gy/min]	Rel. err. [%]	Dose-rate [Gy/min]	Rel. err. [%]
B10	0.0606	6.86	0.0563	6.93
N14	0.0041	6.97	0.0038	7.03
n	0.0100	57.68	0.0148	39.76
g	0.0492	19.39	0.0418	18.77
Combined	0.1240	9.60	0.1167	9.04

In both cases, the fast-neutron (n) component carries by far the largest relative error, consistent with its small absolute contribution to the dose rate at each hotspot: a

low-fluence tally is expected to converge more slowly in relative terms even where its contribution to the combined dose is modest. This is precisely the situation that made nps $= 1 \times 10^8$ statistically insufficient in early pipeline development: the components that matter least for the mean dose are frequently the ones that converge slowest, and it is the *combined* relative error that must be checked against threshold before an irradiation time is trusted.

Both hotspots for patient HN_0717 pass the common 10% reference-error threshold, so irradiation time is not compromised by insufficient Monte Carlo statistics.

4.4.5 Tumor Control Probability and Normal Tissue Complication Probability

TCP was computed with the OpenPINT radiobiological model layer using the general Poissonian form of Equation (1.28), extended to a voxel-wise integration over the full GTV dose-volume histogram under the photon-isoeffective model. Radiobiological parameters were taken from the in-vivo H&N squamous-cell-carcinoma dataset published by González et al. [14].

For patient HN_0717 manual configuration: $\mathrm{TCP}_{\mathrm{GTV}} = 0.1586$.

NTCP for the mucosa was computed with two distinct dose formalisms, presented together here and used consistently for every patient and configuration in the rest of this Chapter.

Isoeffective-dose. The same voxel-wise mucosa dose-rate components (B10, N14, n, g) are combined through the photon-isoeffective model, using the mucosa-specific parameterization, following González et al. [14]. For HN_0717 manual, the mean mucosa isoeffective dose is = 7.60 Gy. As explained in Section 4.4.1, this quantity is reported in Gy rather than converted into a complication probability, and is used throughout this Chapter as a *comparative index* across patients and configurations, following the same logic already established for TCP at low absolute values.

Absorbed-dose NTCP. The same dose-response formalism of Section 1.3.4, applied instead to the absorbed mucosa dose. For HN_0717 manual: mean NTCP $= 3.4 \times 10^{-5}$, $\mathrm{NTCP}_{P98} = 1.4 \times 10^{-4}$, $\mathrm{NTCP}_{\max} = 0.039$.

4.4.6 Configuration Comparison

The four-arm design is easiest to understand concretely, on a single patient, before it is applied to the full cohort. Table 4.9 reports the full dosimetric summary for HN_0717 in all four configurations.

Table 4.9: Full dosimetric summary, patient HN_0717, all four configurations.

	Manual	AI	Swap	Anti-swap
GTV voxel count (volume, cm^3)	9632 (77.06)	10568 (84.54)	10568 (84.54)	9632 (77.06)
Irradiation time [min]	48.40	51.42	48.40	51.42
Hotspot dose rate [Gy/min]	0.1240	0.1167	0.1240	0.1167
GTV RBE D_{mean} [Gy-w]	17.26	18.02	16.63	18.65
GTV IsoE D_{mean} [Gy]	18.38	18.85	17.94	19.28
TCP_{GTV}	0.1586	0.1577	0.1447	0.1708
Mean mucosal NTCP	3.43×10^{-5}	1.23×10^{-4}	3.14×10^{-5}	1.23×10^{-4}
Mucosa absorbed D_{mean} [Gy]	2.174	2.267	2.164	2.267
Mucosa isoE dose [Gy]	7.60	7.76	7.58	7.76

Two mechanical regularities are visible immediately, and both follow directly from how each configuration is constructed rather than being new physical findings. First, GTV voxel count/volume, GTV IsoE/RBE dose, and TCP take independent values in all four configurations, since these are extracted using the mask specific to each row (manual mask for manual and anti-swap; AI mask for AI and swap). Second, irradiation time, hotspot dose rate, and every mucosa-related quantity (absorbed-dose NTCP, absorbed mean dose, isoeffective comparative dose) are *identical* between AI and anti-swap, just as they are identical between manual and swap: these quantities depend only on the dose grid and the mucosa hotspot search, both of which anti-swap inherits unchanged from AI, exactly as swap inherits them unchanged from manual.

For this patient specifically, anti-swap TCP (0.1708) is the *highest* of the four configurations, exceeding even the manual baseline (0.1586) – a reminder that the cohort-wide tendency for anti-swap TCP to sit below manual is a pooled effect: HN_0717's near-perfect segmentation agreement (DC = 0.901) means the AI-based positioning used by anti-swap here is well aligned with the true tumor, so the higher dose delivered under AI slightly longer irradiation time benefits the true GTV.

It is also worth noting that the TCP deviation between manual and AI (0.1586 $\rightarrow$ 0.1577, a -0.6% relative change) is small for this particular patient, while the deviation between manual and swap (0.1586 $\rightarrow$ 0.1447, -8.8%) is larger, even though swap shares the manual dose grid exactly. This is not a contradiction – it confirms

that, for HN_0717, the mask-only effect (manual vs. swap) and the repositioning effect (which pulls AI back toward manual, since HN_0717's centroid shift is small), act with different signs and comparable magnitude, so the two effects partially cancel in the raw manual-vs-AI comparison. Section 4.5.4 shows that this cancellation is not guaranteed elsewhere in the cohort, and that isolating the mask effect from the repositioning effect requires the swap and anti-swap arms rather than the manual-vs-AI comparison alone.

4.4.7 Gamma-Index Analysis

TCP and the dosimetric figures of merit summarize the GTV dose distribution into single values; they cannot, by themselves, say whether a given TCP deviation stems from a spatially uniform dose shift or from a localized cold/hot spot. The gamma index [142] addresses this by comparing two dose fields *voxel by voxel*, combining a dose-difference criterion and a spatial distance-to-agreement (DTA) criterion into a single number per voxel:

$$\gamma(\mathbf{r}) = \min_{\mathbf{r}'} \sqrt{\left(\frac{D_{\mathrm{ref}}(\mathbf{r}) - D_{\mathrm{eval}}(\mathbf{r}')}{\Delta D}\right)^2 + \left(\frac{|\mathbf{r} - \mathbf{r}'|}{\Delta d}\right)^2}, \tag{4.4}$$

where ΔD is the dose-difference tolerance[3], Δd is the distance-to-agreement tolerance[4], and the minimum is taken over every candidate point $\mathbf{r}'$ within Δd of $\mathbf{r}$.

A voxel *passes* if $\gamma(\mathbf{r}) \leq 1$: its evaluated dose is within tolerance of *some* nearby reference-dose point, either because the doses themselves agree closely or because a small spatial shift would bring them into agreement.

Because the gamma index compares two dose *maps* voxel by voxel, rather than embodying a radiobiological effect model in its own right, the choice between the RBE-weighted and isoeffective representations does not carry the same interpretive weight here as it does for TCP or NTCP: gamma-index analysis asks only whether one dose distribution reproduces another one taken as reference, within a stated tolerance, and any consistent, physically meaningful dose map can be used for that comparison. The photon-isoeffective full-patient dose map is used here for consistency with the TCP calculation, on the shared 2 mm grid, without sub-voxel interpolation: the search over candidate points is restricted to the exact 3 mm radius, which is sufficient to classify

[3] 3% of the maximum reference dose within the region compared, $D_{\mathrm{max,ref}}$.
[4] 3 mm.

pass/fail correctly even though it does not recover the true numerical gamma minimum for voxels that fail.

Mask Overlap

Every gamma-index result in this Chapter is reported for a single region: the intersection of the two GTV masks involved in the comparison, $G \cap S$. Two independent reasons motivate this choice, one general to gamma-index methodology and one specific to BNCT.

First, the general one: the 3%/3 mm criterion itself was designed for photon-beam dose, where the tolerance reflects photon dosimetry precision and beam-penumbra characteristics [142]. Applying the same criterion to a BNCT dose field is a non-trivial extrapolation, since the dominant spatial dose gradient in BNCT is governed by boron microdistribution rather than photon beam penumbra.

Second, the BNCT-specific one, which is the primary reason for restricting to the overlap rather than a secondary caveat: the photon-isoeffective full-patient dose map used for this comparison is computed with a single, uniform boron concentration, 50 ppm applied to the *entire* GTV. The mask intersection $G \cap S$ is the one region both segmentations agree is genuine tumor, and therefore the one region in which the uniform 50 ppm assumption underlying *both* compared dose fields is physically justified, independently of which segmentation is trusted.

How the overlap is obtained differs mechanically across the three comparisons made in this Chapter, reflecting whether the two masks involved are actually different:

- **Manual vs. AI**: the manual and AI masks genuinely differ, so the union $G \cup S$ must first be computed and classified into its three sub-regions (intersection, reference-only, evaluated-only) before the intersection can be isolated – the procedure illustrated in Figure 4.11 and Table 4.10 below.

- **Manual vs. swap**: swap re-extracts the manual dose grid under the AI mask, so this comparison likewise requires the manual/AI union-and-intersection procedure, even though the underlying dose field is identical to manual's.

- **Manual vs. anti-swap**: both fields compared are evaluated on the *same, single* manual mask by construction. No union of two different masks is needed here at all: the comparison region is simply the manual GTV mask itself.

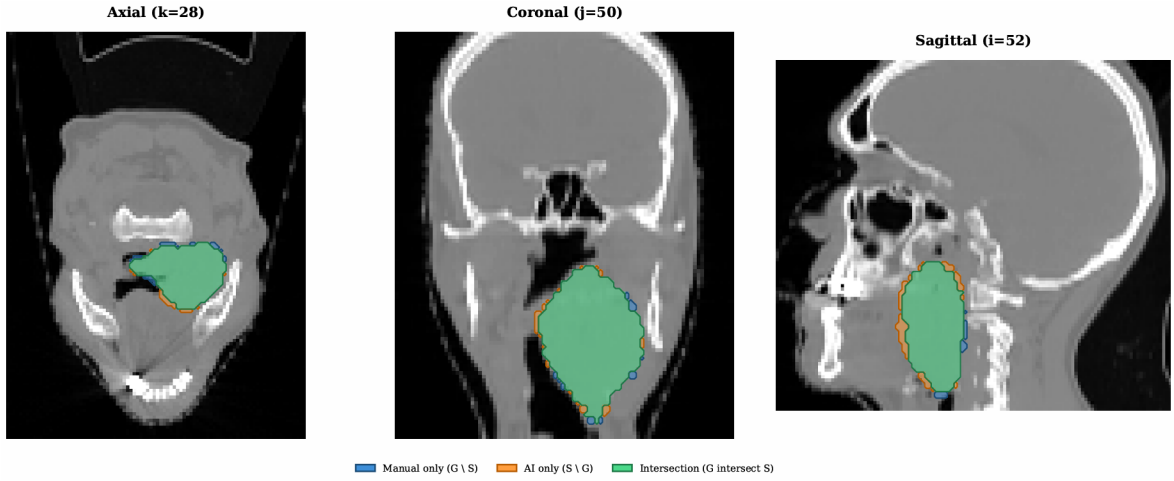


Figure 4.11: *GTV mask union overlaid on the CT, patient HN_0717: manual-only (blue), AI-only (orange), intersection (green).*

Manual vs. Swap as a Control

Before looking at the manual-vs-AI numbers, it is worth explaining why manual-vs-swap is reported alongside them, since on its own a gamma index that is exactly 100% everywhere doesn't add information on the comparison.

The swap configuration does not correspond to an independent Monte Carlo simulation: it reuses the *physical dose grid of the manual simulation, voxel for voxel, unchanged*, and only substitutes the AI GTV mask at the point where summary statistics are extracted. Consequently, when the gamma index is computed between the manual dose field and the swap dose field, it is – by construction – comparing an array against itself within the intersection region. The result is therefore not a measurement at all: it is guaranteed, for any patient, to be exactly $\gamma = 0$ everywhere the two fields are compared, hence a 100% pass-rate.

Two distinct roles follow directly from this:

- **Positive control on the calculation.** Because the manual-vs-swap result is known *a priori* to be exactly 100%, it functions as a check on the gamma-index code. Any deviation from 100% here would indicate a data or implementation bug rather than a genuine physical effect.

- **Logical control on the interpretation.** Swap changes *only* the GTV mask used to define the intersection region; it changes nothing about the underlying dose field. If a change of mask *alone* were capable of producing gamma disagreement within that region, the manual-vs-swap pass-rate would not be 100%. Since

mask choice alone is ruled out, the manual-vs-AI disagreement (Table 4.10) must originate in a genuine difference between the two *independently computed* dose fields.

Results for HN_0717

As shown in Table 4.10, the gap between the two criteria for manual-vs-AI is large: the pass-rate more than doubles, from 38.2% (dose-difference alone) to 86.6% (with the 3 mm DTA term), within the intersection region alone.

Table 4.10: Gamma-index pass-rate, overlap region only, patient HN_0717. For manual vs. anti-swap the overlap is the entire manual GTV mask ($n = 9632$ voxels); for the other two comparisons it is the mask intersection $G \cap S$ ($n = 9102$ voxels).

Comparison	Dose-difference only (3%)	True gamma (3%/3 mm)
Manual vs. AI	38.2%	86.6%
Manual vs. swap	100.0%	100.0%
Manual vs. anti-swap	—	86.8%

Because this region contains only voxels present in *both* GTV masks, this gap cannot be explained by mask-membership disagreement or by the boron-uniformity caveat above – it reflects the automated-positioning shift: a large fraction of the raw dose disagreement at a given voxel is recovered once the DTA term is allowed to search a 3 mm neighborhood, consistent with the ∼3.4 mm translation-vector shift measured directly for this patient. The manual-vs-swap control confirms this is a genuine spatial effect of the independently-computed AI dose field, not an artifact of the gamma calculation itself.

The manual-vs-anti-swap true-gamma pass-rate (86.8%) is close to the manual-vs-AI value (86.6%) for this particular patient, consistent with HN_0717's near-perfect segmentation agreement leaving little room for the two comparisons to diverge; this is not true for every patient in the cohort, as shown in Section 4.5.4.

Figure 4.12 shows the resulting voxel-wise gamma map (true-gamma criterion, overlap region) in three views; Figure 4.13 shows the corresponding distribution.
It is worth noting that high-statistical-uncertainty voxels are not excluded from the gamma calculation, so a fraction of the voxels reported as failing $\gamma \leq 1$ may reflect statistical noise in the dose field rather than a genuine dosimetric discrepancy between the manual and AI configurations.

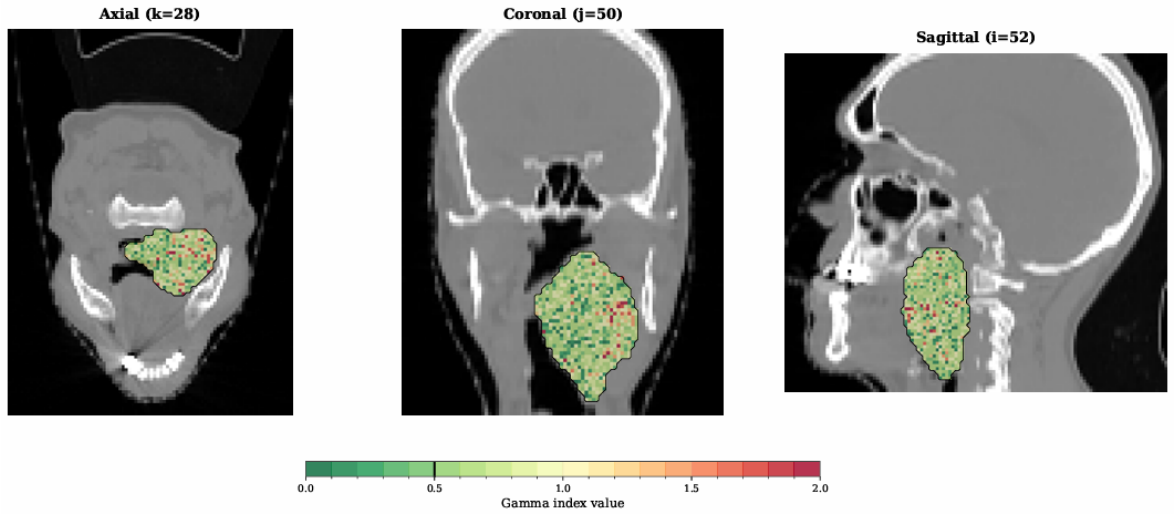


Figure 4.12: *Voxel-wise gamma index map (true gamma, 3%/3 mm), manual vs. AI, patient HN_0717, overlap region only. Green: $\gamma < 1$ (pass); red: $\gamma > 1$ (fail). Pass-rate: 86.6%.*

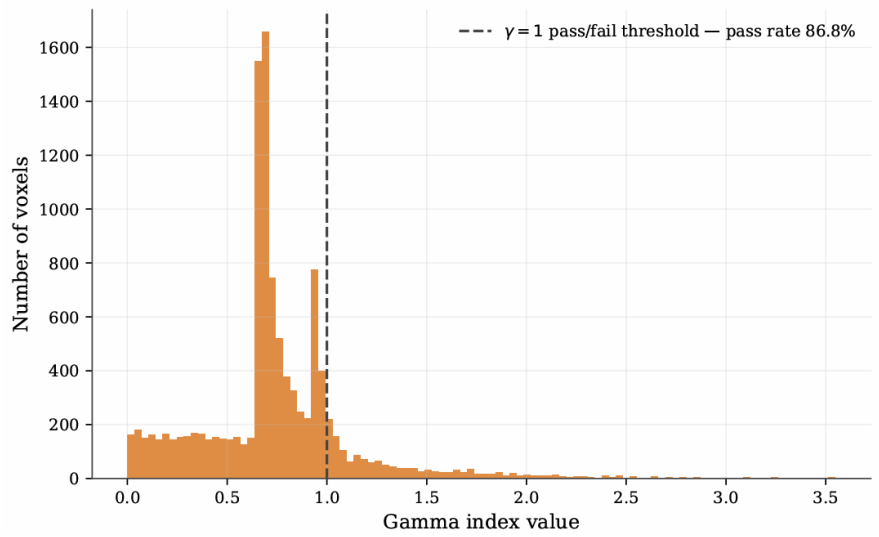


Figure 4.13: *Gamma-index value distribution over the mask overlap, manual vs. AI, patient HN_0717.*

4.5 Overall Results

Having established, on a single patient, exactly what each quantity means and how the four arms relate to one another, this Section reports the same quantities pooled over the full 24-patient cohort, always across all four configurations together. The complete per-patient values underlying every pooled statistic below are tabulated in Appendix D.

4.5.1 Tumor Control Probability

Figure 4.14 shows the per-patient TCP for all four configurations, sorted by Dice Coefficient; Figure 4.15 shows the corresponding pooled distribution. Table 4.11 reports the pooled descriptive statistics and the paired Wilcoxon signed-rank test between every pair of configurations; the complete per-patient TCP values and normalized deviations are tabulated in Table D.2 of Appendix D.

Table 4.11: Pooled TCP statistics (24 patients, 2 mm) and paired Wilcoxon signed-rank tests between every pair of configurations.

	Manual	AI	Swap	A-swap
Mean	0.198	0.206	0.203	0.185
Median	0.172	0.204	0.200	0.173
Std. dev.	0.086	0.074	0.077	0.075
Min	0.076	0.085	0.083	0.078
Max	0.399	0.365	0.434	0.376

Pair	Wilcoxon p
Manual vs. AI	0.375
Manual vs. swap	0.252
AI vs. swap	0.252
Manual vs. anti-swap	0.089
AI vs. anti-swap	**0.043**
Swap vs. anti-swap	0.074

Only one of the six pairwise comparisons reaches statistical significance at the pooled-group level ($\alpha = 0.05$): AI vs. anti-swap ($p = 0.043$). Anti-swap has the lowest pooled mean and median TCP of the four configurations, and the direction is consistent across all three comparisons involving it (manual, AI, swap vs. anti-swap all show anti-swap lower), even where individual tests do not reach significance – consistent with evaluating true tumor coverage under a treatment geometry (AI-based positioning) that was not optimized for that specific volume. Reporting only the pooled result would still be materially incomplete for manual vs. AI and manual vs. swap: 10 of the 24 patients (42%) show a normalized TCP deviation between the AI and manual configurations exceeding 30%, with the five largest reaching +80.5% (HN_0150), +65.2% (HN_0149), +63.8% (HN_1640), −52.9% (HN_0591), and +48.3% (HN_1615). The absence of a significant pooled difference in these two comparisons reflects cancellation

between positive and negative per-patient deviations – the same mechanism illustrated for HN_0717 – rather than genuine case-by-case agreement.

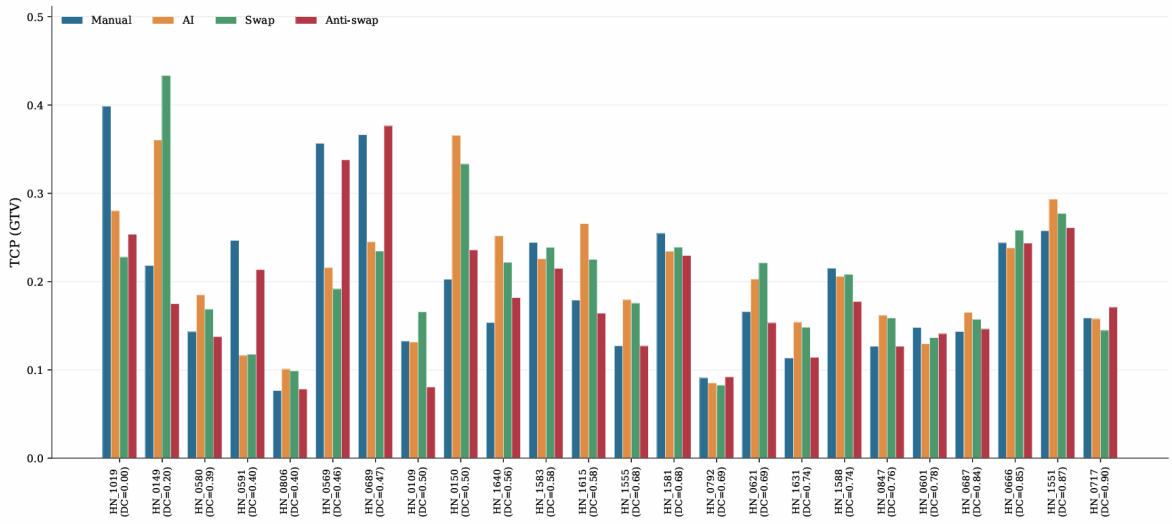


Figure 4.14: *Per-patient TCP for the manual, AI, swap, and anti-swap configurations, sorted by Dice coefficient (voxel size 2 mm).*

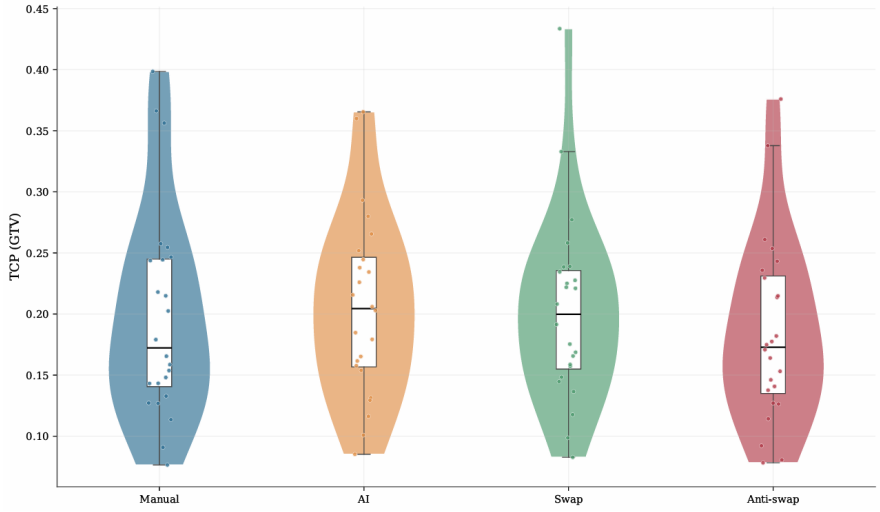


Figure 4.15: *Pooled TCP distribution across the 24-patient cohort for all four configurations (violin plot with overlaid boxplot and individual patient values).*

Mask-Only Risk versus Real Clinical Risk

Swap and anti-swap isolate the mask-only effect from two complementary perspectives, and it is worth stating precisely what each one asks.

Swap poses a counterfactual question about the manual treatment: had the AI segmentation been substituted for the manual one at the point of dose extraction, with

the treatment geometry left unchanged – i.e., with positioning still derived from the manual, ground-truth GTV – how would TCP have differed? Anti-swap poses the converse, clinically realistic question: given that a treatment was actually planned and delivered on the AI segmentation, with positioning derived from it, how would TCP have differed had the true tumor volume been used to evaluate that same delivered dose distribution? Comparing the two, both expressed as the normalized deviation $|\Delta\text{TCP}\,\%|$ against the same manual baseline, tests whether swap is a reliable proxy for anti-swap, the configuration of genuine clinical relevance.

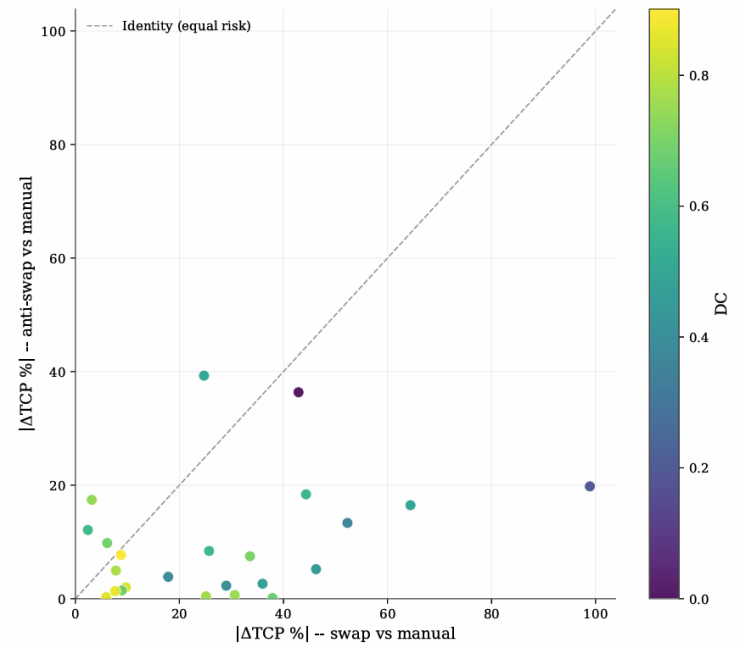


Figure 4.16: *Theoretical mask-only risk (swap vs. manual) versus real clinical risk (anti-swap vs. manual), both normalized, 24 patients, colored by DC. The identity line marks equal risk.*

As shown in Figure 4.16, it is not a reliable proxy. Only 4 of the 24 patients (17%) show the real risk (anti-swap) exceeding the theoretical mask-only risk (swap); in the remaining 20 patients, swap *overestimates* the real-world consequence of segmentation discordance, often substantially – for instance HN_0149 shows a swap deviation of 98.9% against an anti-swap deviation of only 19.8%. The two deviations are, moreover, not significantly correlated with one another (Spearman $\rho = 0.255$, $p = 0.230$, $n = 24$): knowing the theoretical mask-only risk for a given patient does not reliably predict their real clinical risk.

A plausible mechanism is that AI-based automated positioning, computed from the AI GTV centroid, is optimized – by construction – for the AI-delineated volume;

when re-evaluated on the true manual volume, this positioning is often still reasonably well-aligned, partially compensating for the mask disagreement that swap, evaluated at manual positioning, cannot benefit from. This finding argues against using the swap configuration alone as a conservative upper bound on the clinical risk of AI segmentation errors: it is frequently pessimistic, and does not rank patients consistently with the configuration that is actually clinically relevant.

TCP as a Comparative Index

The pooled TCP values above (mean $\approx$ 0.20, range 0.076–0.434) are far below the range in which TCP is conventionally interpreted as an estimate of actual tumor-control likelihood (typically 0.5–0.95 in clinical BNCT) – the same regime already seen for HN_0717 individually. At such low absolute values, the Poissonian TCP model of Equation (1.28) operates on the steepest part of its dose-response curve, where a small change in the underlying dose distribution is amplified into a proportionally larger change in TCP. This section tests whether the correlation reported in Section 4.5.5 survives when computed directly on the underlying dose metrics, i.e., whether it is a genuine dosimetric effect that TCP happens to make more visible, or an artifact of TCP nonlinearity.

For each patient, the normalized percentage deviation between the AI and manual configurations was computed both for TCP and for four dose metrics from the same GTV dose-volume histogram under the photon-isoeffective model: D_{min}, D_2 (near-minimum, statistically more robust than the single-voxel D_{min}), D_{mean}, and D_{98} (near-maximum). Table 4.12 reports the resulting distributions.

Table 4.12: Distribution of $|\Delta X|$ (%, AI vs. manual, normalized, 24 patients) for TCP and four GTV dose metrics under the photon-isoeffective model.

	TCP	D_{min}	D_{mean}	D_2	D_{98}
Mean	28.0	22.4	7.6	13.2	5.7
Median	28.2	20.9	6.3	8.3	5.0
Std. dev.	22.3	17.6	6.3	11.9	4.3
Max	80.5	59.3	25.8	50.9	19.9

The median TCP deviation (28.2%) is 1.35 times the median D_{min} deviation, and this amplification factor is markedly patient-dependent, ranging up to 6.2$\times$ for HN_0687 and 4.7$\times$ for HN_1581.

Table 4.13 reports the Spearman correlation between the *signed* TCP deviation and the signed deviation of each dose metric: TCP tracks the near-minimum dose D_2 most strongly ($\rho = 0.923$), more strongly than D_{min} ($\rho = 0.806$) or D_{mean} ($\rho = 0.750$), and only weakly D_{98} ($\rho = 0.413$). Table 4.13 re-computes the DC correlation of Section 4.5.5 using each dose metric directly, in place of TCP.

Table 4.13: Left: Spearman correlation between the signed TCP deviation and the signed deviation of each dose metric (AI vs. manual, 24 patients). Right: Spearman correlation between $|\Delta X|$ (normalized) and DC, computed on TCP versus each raw dose metric.

Dose metric	ρ	p
D_2 (near-min.)	0.923	< 0.0001
D_{min}	0.806	< 0.0001
D_{mean}	0.750	< 0.0001
D_{98} (near-max.)	0.413	0.0448

Metric	ρ vs. DC	p
TCP	-0.582	0.0029
D_{min}	-0.617	0.0013
D_{mean}	-0.522	0.0089
D_2	-0.503	0.0121
D_{98}	$+0.235$	0.270

The correlation with DC is of comparable strength, and if anything marginally stronger, when computed directly on D_{min} than on TCP. The association between segmentation accuracy and dosimetric discrepancy is therefore *not* an artifact of TCP amplification at low absolute values – it is present, with comparable statistical strength, in the underlying near-minimum GTV dose itself. TCP percentages elsewhere in this chapter should be read as a sensitive *comparative index*, appropriate for ranking patients or flagging cases of concern, not as an estimate of clinical viability for any individual plan in this cohort.

4.5.2 RBE-Weighted and Photon-Isoeffective Dose Models

The two dosimetric models introduced in Sections 1.3.2 and 1.3.3 were compared on the mean GTV dose for both the manual and AI configurations, extending the single-patient observation of to the full cohort.

Table 4.14: Left: percentage discrepancy between the RBE-weighted and isoeffective mean GTV dose, $100 \times (D_{w,\text{mean}} - D_{\text{IsoE,mean}})/D_{\text{IsoE,mean}}$, 24 patients. Right: mean isoeffective GTV dose, split by the sign of that discrepancy (manual configuration).

	Manual	AI
Mean discrepancy	-3.34%	-2.26%
Std. dev.	6.37%	6.64%
Min	-13.87%	-15.19%
Max	$+9.74\%$	$+12.03\%$
$D_w > D_{\text{IsoE}}$	8/24 (33%)	7/24 (29%)

Patient group	Mean IsoE [Gy]
$D_w > D_{\text{IsoE}}$ (8 pts.)	21.71
$D_w < D_{\text{IsoE}}$ (16 pts.)	18.19

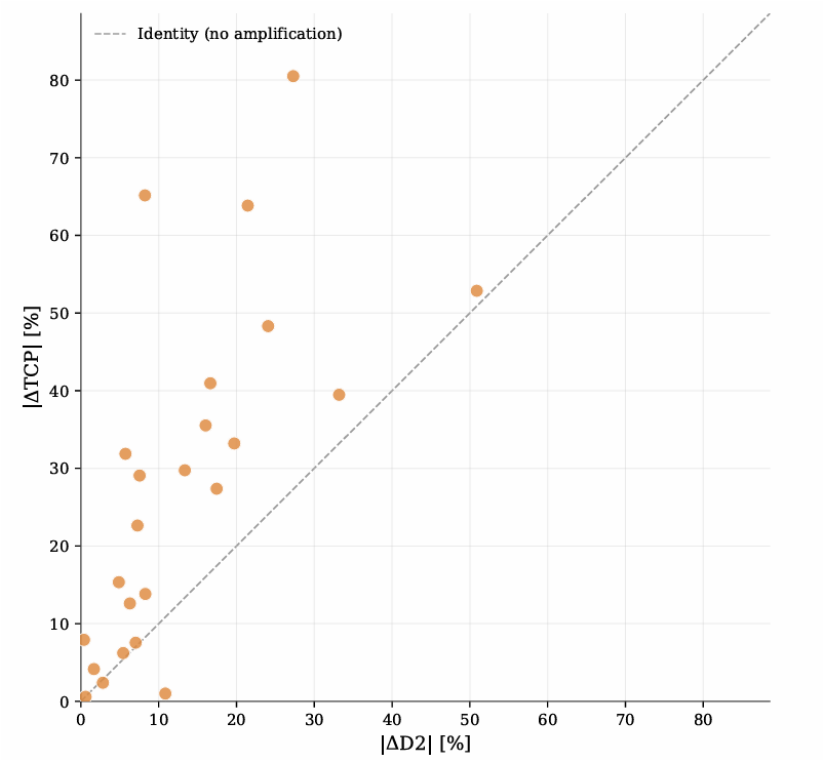


Figure 4.17: $|\Delta\text{TCP}|$ *versus* $|\Delta D_2|$ *(near-minimum GTV dose deviation), 24 patients, with the identity line indicating no amplification.*

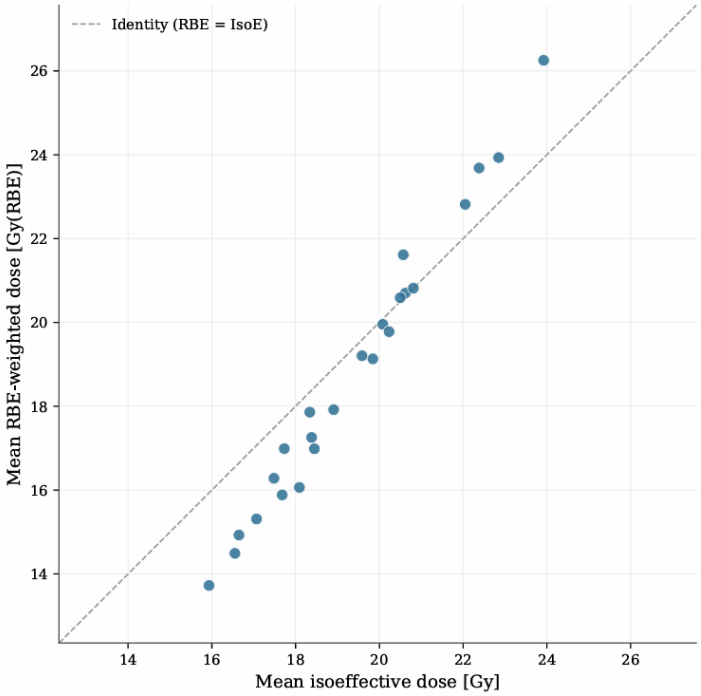


Figure 4.18: *Mean GTV dose, RBE-weighted versus photon-isoeffective model, manual configuration, 24 patients, with the identity line indicating no discrepancy between models.*

Only 8 of 24 patients (manual configuration) show the RBE-weighted dose exceeding the isoeffective dose; the majority (16/24, 67%) show the opposite pattern, consistent with what was observed for HN_0717 individually. A Spearman correlation between this discrepancy and DC shows no significant relationship ($p = 0.69$), consistent with the discrepancy being a structural property of the two radiobiological models rather than a consequence of segmentation quality.

Discrepancy vs. Absolute Dose Magnitude

If the discrepancy is not explained by segmentation quality, a physically motivated alternative is that it is explained by dose magnitude itself: the LQ isoeffective model quadratic dose term (Equation (1.21)) contributes proportionally more at higher absolute dose than at lower dose, so the two models are expected to diverge in a dose-dependent way regardless of which GTV mask is used.

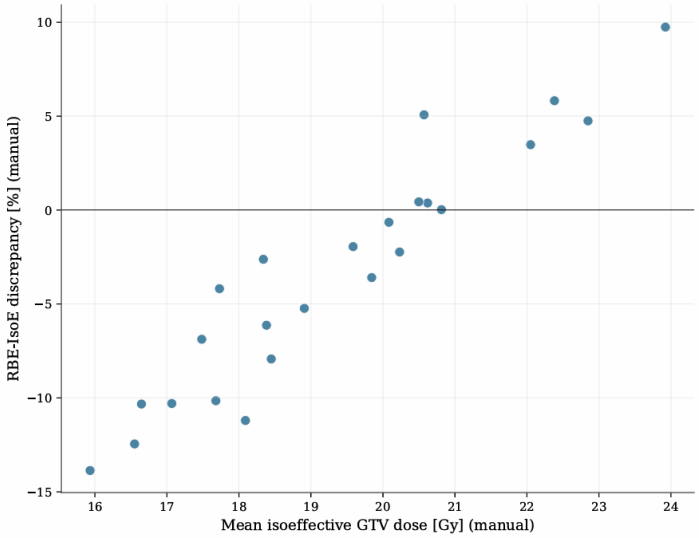


Figure 4.19: *RBE-IsoE discrepancy versus mean isoeffective GTV dose, manual configuration, 24 patients.*

The correlation is strong and highly significant: Spearman $\rho = 0.940$, $p < 0.0001$ ($n = 24$), by a wide margin the strongest correlation reported anywhere in this chapter. Table 4.14 above confirms the direction: the mean isoeffective GTV dose is 21.7 Gy among the 8 patients where the RBE-weighted dose exceeds the isoeffective dose, against 18.2 Gy among the 16 patients showing the opposite pattern. This confirms the hypothesis: at higher absolute dose, the quadratic term of the isoeffective model contributes proportionally less to the total relative to the linear term under this pa-

rameterization, pulling the isoeffective estimate below the fixed-weighting one; at lower dose the opposite holds – exactly the mechanism invoked for HN_0717, whose own mean IsoE dose (18.38 Gy manual) sits close to the 18.2 Gy low-dose regime rather than the 21.7 Gy high-dose one. The RBE-IsoE gap in this cohort is therefore best understood as a property of the two radiobiological models at the dose levels actually delivered, not as an artifact of manual versus AI segmentation. Per-patient absolute GTV isoeffective dose statistics (minimum, mean, maximum, manual and AI) are tabulated in Table D.4 of Appendix D, and can be used to verify directly, patient by patient, which side of this dose-dependent split each case falls on.

4.5.3 Mucosal Isoeffective Dose

Section 4.4.5 introduced the mucosa isoeffective dose as the primary comparative index of mucosal risk, reported alongside the absorbed-dose NTCP retained for comparability. This section presents the pooled, cohort-wide behavior of both mucosal quantities across all four configurations; the complete per-patient values are tabulated in Table D.6 of Appendix D.

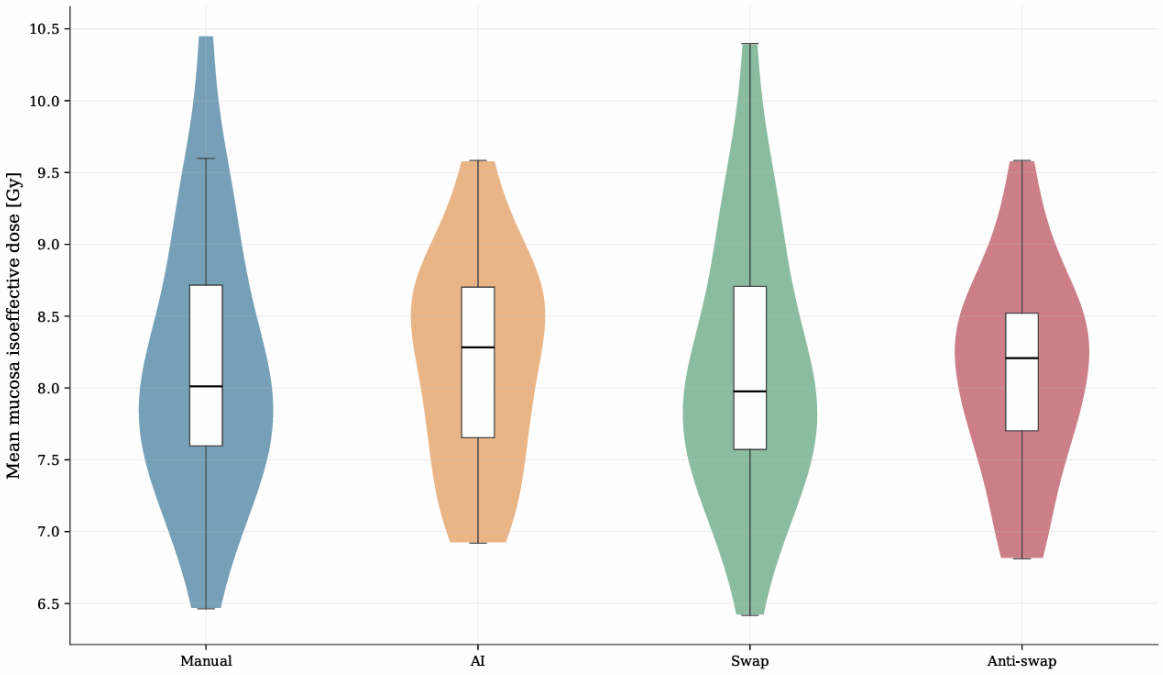


Figure 4.20: *Pooled mean mucosa isoeffective dose (comparative index, not a probability) across the four configurations, 24 patients.*

Unlike TCP, the pooled mucosa isoeffective dose is remarkably stable across all four configurations (8.11–8.19 Gy mean, a 1% range), and only two of the four pairs listed

reach significance: manual vs. swap and AI vs. anti-swap – in both cases comparing configurations that share the same dose grid but differ in mask, i.e. the mask-only effect is detectable in the mucosa dose even though its pooled magnitude is small. Manual vs. AI and manual vs. anti-swap show no detectable pooled difference. This stability is consistent with the mucosa dose depending primarily on the shared irradiation-time constraint rather than on which GTV mask defines the exclusion margin around it. The absorbed-dose NTCP shows the same qualitative pattern of pairwise significance, reinforcing that both mucosal metrics are primarily sensitive to the dose-grid identity (manual-grid pair vs. AI-grid pair) rather than to which mask defines the exclusion margin.

Table 4.15: Pooled mean mucosa isoeffective dose [Gy] and absorbed-dose NTCP, four configurations, 24 patients, with paired Wilcoxon signed-rank tests for the four pairs where a p-value is available.

	Manual	AI	Swap	A-swap
IsoE dose, mean [Gy]	8.151	8.187	8.134	8.113
IsoE dose, std. [Gy]	0.932	0.732	0.930	0.710
NTCP absorbed, mean	1.13×10^{-4}	1.16×10^{-4}	1.07×10^{-4}	8.9×10^{-5}

Pair	Wilcoxon p
Manual vs. AI	0.60
Manual vs. swap	**0.0003**
Manual vs. A-swap	1.00
AI vs. A-swap	**0.0022**

4.5.4 Gamma-Index Analysis

The gamma-index methodology, including the boron-concentration and photon criterion caveats motivating the overlap-only convention, was introduced in full for patient HN_0717. This section applies the identical procedure to the full 24-patient cohort, for three comparisons: manual-vs-AI, manual-vs-swap (positive control), and manual-vs-anti-swap.

Results

One patient, HN_1019 (DC = 0), has an empty manual-AI mask intersection by definition and is therefore excluded from the manual-vs-AI and manual-vs-swap statistics ($n = 23$); it remains included in the manual-vs-anti-swap comparison ($n = 24$), since anti-swap requires no union of two different masks and is defined for any patient with a non-empty manual mask. The complete per-patient pass-rates are tabulated in Table D.5 of Appendix D.

Table 4.16: Left: gamma-index pass-rate, overlap region only (true gamma, 3%/3 mm), three comparisons, 24 patients (23 for manual-vs-AI and manual-vs-swap). Right: Spearman correlation between DC and the overlap pass-rate for the two comparisons where it is defined (manual-vs-swap is excluded: every value is exactly 100%, so the correlation has zero variance and is undefined, not merely non-significant).

	M. vs. AI	M. vs. swap	M. vs. A-swap
Mean [%]	79.5	100.0	74.6
Std. dev. [%]	13.3	0.0	22.4
Min [%]	50.0	100.0	2.3
Max [%]	91.6	100.0	100.0

Comparison	ρ	p	n
M. vs. AI	0.338	0.115	23
M. vs. A-swap	0.503	0.012	24

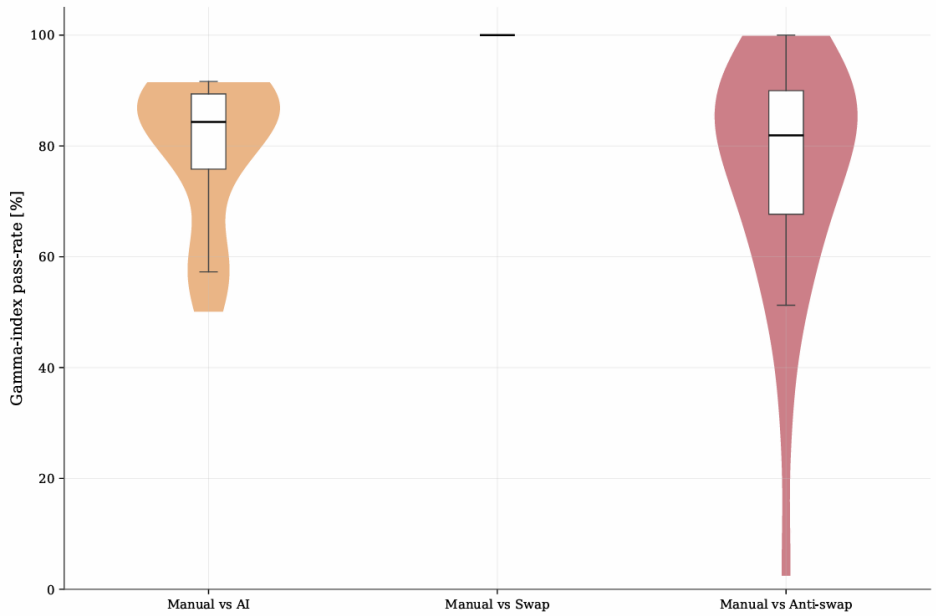


Figure 4.21: *Pooled gamma-index pass-rate (overlap region only), manual-vs-AI, manual-vs-swap (positive control), and manual-vs-anti-swap.*

The manual-vs-swap overlap pass-rate is exactly 100% for every one of the 23 patients, confirming at cohort scale – not only for HN_0717 – that the positive control holds throughout. The manual-vs-AI overlap pass-rate (mean 79.5%, range 50.0–91.6%) is lower and patient-dependent. The manual-vs-anti-swap pass-rate (mean 74.6%) is close to, but somewhat lower and substantially more variable than, manual-vs-AI (std. dev. 22.4 pp vs. 13.3 pp); its minimum (2.3%) is driven by HN_1019, whose manual GTV lies almost entirely outside the region where the AI-based positioning places any meaningful dose.

Correlation with Segmentation Accuracy

The manual-vs-AI correlation (Table 4.16) does not reach significance at this cohort size, while manual-vs-anti-swap does. This is a weaker relationship, for manual-vs-AI, than the one found between DC and the near-minimum GTV dose deviation, consistent with the gamma index summarizing a broader, spatially distributed pattern of agreement not fully captured by a single global geometric overlap measure. That the anti-swap correlation is both stronger and significant is consistent with Section 4.4.7: anti-swap removes the mask-membership variability that manual-vs-AI still carries even after restricting to the overlap (the overlap region itself is defined differently for every patient, in a way that depends on DC), isolating the positioning-driven component of dose disagreement more cleanly.

The Role of Patient Repositioning

The pattern above extends, at cohort scale, the mechanism already identified for HN_0717: because automated positioning is computed from the GTV centroid, a geometric segmentation disagreement generically produces a beam-geometry disagreement, and this affects the dose field even within the region both masks agree is tumor – precisely why the manual-vs-swap positive control is essential for ruling out a purely mask-driven explanation of the manual-vs-AI disagreement in Table 4.16.

This has a direct methodological consequence: a manual-versus-AI comparison in this OpenPINT-based pipeline measures the *combined* effect of segmentation discordance and the consequent change in beam geometry, and the gamma-index analysis alone cannot isolate the dosimetric impact of segmentation quality from that of repositioning using the manual-vs-AI comparison in isolation. The four-arm design isolates

each effect from a different comparison: swap isolates the mask effect at manual positioning; anti-swap isolates the positioning effect at the true tumor volume.

4.5.5 Segmentation Accuracy and Dosimetric Discrepancy

Table 4.17 reports the Spearman rank correlation between each geometric index and the normalized TCP deviation

$$|\Delta\mathrm{TCP}\,\%| = 100 \times |\mathrm{TCP}_{\mathrm{config}} - \mathrm{TCP}_{\mathrm{manual}}|/\mathrm{TCP}_{\mathrm{manual}}, \tag{4.5}$$

for all three non-manual configurations, laid out as a single compact matrix (index × configuration) for ease of comparison; each cell gives ρ with the corresponding p-value in parentheses.

Table 4.17: Spearman correlation ρ (p-value) between each geometric index and the normalized $|\Delta\mathrm{TCP}\,\%|$, for the three non-manual configurations, 24 patients.

Index	AI vs. manual	Swap vs. manual	Anti-swap vs. manual
DC	−0.582 (0.0029)	−0.661 (0.0004)	−0.518 (0.0095)
GMI	+0.535 (0.0071)	+0.617 (0.0013)	+0.390 (0.060)
DI	−0.070 (0.747)	−0.115 (0.593)	+0.279 (0.187)

A single further correlation, on the raw (unnormalized), signed manual-vs-AI TCP deviation rather than on any of the three normalized columns above, is worth reporting separately: DI correlates significantly with its *sign* ($\rho = -0.625$, $p = 0.0011$), even though it does not correlate with the *magnitude* of any of the three normalized deviations above. Table 4.18 checks the robustness of the DC/GMI/DI–AI column with and without patient HN_1019 (DC = 0), whose complete segmentation failure could in principle dominate the correlation on its own.

Table 4.18: Robustness of the correlation between the normalized $|\Delta\mathrm{TCP}\,\%|$ (AI vs. manual) and geometric indices, with and without patient HN_1019 (DC = 0).

Index	ρ ($n = 24$)	p ($n = 24$)	ρ ($n = 23$)	p ($n = 23$)
DC	−0.582	0.0029	−0.598	0.0026
GMI	+0.535	0.0071	+0.555	0.0059
DI	−0.070	0.747	−0.093	0.673

DC correlates most strongly with the *swap* deviation among the three configurations, consistent with swap being the configuration that isolates the mask effect most directly (manual dose grid, only the mask changes); the anti-swap correlation is weaker

(DC: −0.518 vs. −0.661), consistent with the partial positioning compensation discussed in Section 4.5.1 – the anti-swap deviation still correlates significantly with DC ($p = 0.0095$), confirming that segmentation quality remains a relevant predictor of dosimetric risk even under the clinically realistic configuration, just less tightly than for the mask-only effect isolated by swap. GMI follows the same pattern across all three configurations; DI does not reach significance for any of the three normalized comparisons, though it does correlate significantly with the sign of the raw manual-vs-AI deviation, as noted above: a high DI (over-segmentation) is associated with a systematically *lower* AI TCP relative to manual. A plausible interpretation is that under-segmentation (GMI) removes clonogenic tumor voxels from the dose calculation entirely, directly reducing TCP, whereas over-segmentation (DI) adds peripheral, comparatively lower-dose tissue without removing correctly identified high-dose voxels, diluting the mean without eliminating the contribution of well-irradiated tumor – consistent with a smaller, more variable effect on the magnitude of the deviation but a directionally consistent downward pull. Figure 4.22 shows all three normalized deviations plotted directly against DC, GMI, and DI.

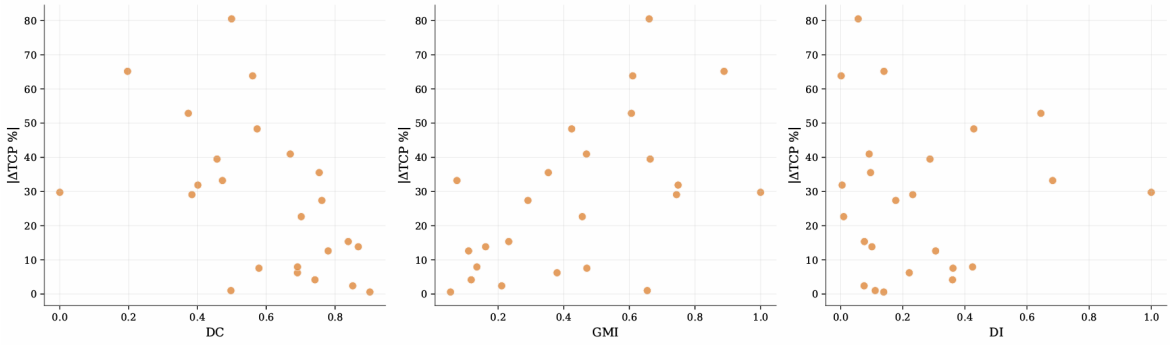


Figure 4.22: *Normalized* $|\Delta\mathrm{TCP}\,\%|$ *(AI vs. manual) versus DC, GMI, and DI, 24 patients.*

4.5.6 Further Geometric and Dosimetric Analyses

Further analyses complement the targeted correlation test above mentioned with a broader exploratory view, and provide the point of comparison with the glioblastoma cohort of [10] referenced in the Introduction.

GTV Volume and Segmentation Accuracy

The correlation between GTV volume and DC is weak and not statistically significant ($\rho = 0.280$, $p = 0.185$, $n = 24$; mean GTV volume 63.4 cm^3, std. dev. 38.8 cm^3, consistent with the range visible in Table 4.1). This suggests that, for the H&N nnU-Net model evaluated here, volume alone is not a strong predictor of segmentation accuracy – plausibly because of the heterogeneous range of anatomical sub-sites (oral cavity, oropharynx, larynx, salivary glands) represented in the training and evaluation set, which weakens any simple volume-based trend relative to a more anatomically homogeneous target.

Pooled Dose-Metric Distributions and DC Reliability Threshold

None of the three pooled dose metrics (Table 4.19, left) differs significantly between configurations: the absence of a pooled difference coexists with, and should not be read as contradicting, the substantial per-patient variability documented throughout this Chapter. The reduction in dispersion above the empirical DC = 0.7 threshold indicates that higher segmentation agreement is associated with more reliable dose reproduction, but the H&N cohort does not show a sharp, near-deterministic stabilization at this threshold: even above DC = 0.7, more than half of the patients fall outside the ±10% band. This is consistent with a continuous, rather than threshold-like, relationship between geometric and dosimetric agreement.

Table 4.19: Left: pooled GTV dose (photon-isoeffective model), 24 patients. Right: minimum-dose ratio (AI/manual) reliability above and below the empirical DC = 0.7 threshold.

Metric	Manual [Gy]	AI [Gy]	Diff. [Gy]	p
$D_{\min}$	9.76	9.85	+0.55	0.406
D_{mean}	19.36	19.64	+0.10	0.509
$D_{\max}$	27.43	27.66	+0.24	0.603

	n	Ratio (std.)	±10%
DC ≥ 0.7	9	1.109 (0.119)	44%
DC < 0.7	15	1.019 (0.348)	13%

Conclusions

This thesis set out to answer a question that is easy to state and hard to answer: when a deep-learning model is used to segment the gross tumor volume for BNCT treatment planning, how much does that choice actually change the dose the patient would receive, compared to a manual, expert-drawn contour? Chapter 4 answered it not with a single number, but with a four-arm simulation design – manual, AI, swap, and anti-swap – built precisely so that the answer could not be collapsed into a single, misleadingly simple statement. The conclusions below draw together what that design revealed, what it could not resolve, and what a natural next phase of this work would need to address.

Segmentation Accuracy and Dosimetric Risk

The most direct question this thesis could ask was whether patients with a better geometric match between the manual and AI segmentations also show a smaller dosimetric discrepancy. The answer is yes, but only once the analysis is set up to ask that question cleanly. The Dice Coefficient and the Geometrical Miss Index both correlate with the magnitude of the dosimetric deviation with a consistency that held up under every robustness check attempted – including after removing the one patient whose segmentation failed outright, and including when the correlation was recomputed directly on the underlying near-minimum GTV dose rather than on tumor control probability, precisely to rule out the possibility that the correlation was merely an artifact of how sharply TCP responds to small dose changes at the low absolute values seen throughout this cohort. This confirms that segmentation quality is not dosimetrically irrelevant, and a geometric metric does carry real information about the dose consequences of trusting an automated contour.

At the same time the Discordance Index showed a different behavior. It failed to

correlate with the *magnitude* of the dosimetric deviation in any of the three configurations tested, yet it correlated clearly with its *sign*: patients whose AI segmentation substantially over-extended beyond the manual volume tended to show a lower AI-derived tumor control probability, not a higher one. A possible interpretation is that under-segmentation and over-segmentation are not mirror-image failure modes. Missing genuine tumor voxels removes clonogenic cells from the calculation outright and depresses TCP directly, while adding peripheral, lower-dose tissue around the true tumor dilutes the mean dose without discarding the contribution of the voxels that were correctly identified. The two geometric indices were measuring two genuinely different ways in which a segmentation can go wrong, with two genuinely different dosimetric signatures. Collapsing them into the Dice Coefficient alone, as is common practice in segmentation-evaluation literature, would have hidden exactly this asymmetry.

Patient Positioning

This thesis work shows that a discrepancy in the segmentation translates directly to a positioning disagreement, since the beam-alignment transform is computed automatically from the segmented tumor's centroid. Segmentation and positioning are coupled in a way that a standard manual-versus-AI geometrical comparison index cannot see nor separate.

This connection was investigated thoroughly by this thesis work. The gamma-index positive control was constructed specifically to prove that a change of mask alone, with the underlying dose field held fixed, could not produce the dosimetric disagreement observed between manual and AI. Once that possibility was closed off, the substantial disagreement that remained between the two independently computed dose fields was found in the position of the patient relatively to the beam. A geometric shift in the tumor centroid of only a couple of millimetres was enough, on the voxel grid used, to displace the entire simulated geometry by a fraction of a voxel relative to a dose gradient steep enough that the consequence was clearly visible in the gamma-index pass-rate.

This is also, on reflection, the reason the swap configuration, which reuses the manual dose field and only substitutes the AI mask at extraction, and the anti-swap configuration, which does the opposite, were not redundant with each other, and why building both was meaningful.

Swap answers a theoretical, counterfactual question: what would the AI mask have cost us, had positioning not moved at all? Anti-swap answers the question that actually matters clinically: given that positioning *did* move, because that is what would really happen if a treatment were planned on the AI segmentation, what does the true tumor actually receive? The comparison between the two showed unambiguously that the theoretical, mask-only estimate of risk is not a trustworthy stand-in for the real one: for the great majority of patients, treating the mask-only deviation as a proxy for clinical risk would have overestimated the dose to normal tissue, sometimes considerably, precisely because the automated positioning derived from the AI segmentation is optimized for a different volume.

The practical conclusion is that any future evaluation of an AI segmentation model intended for a treatment-planning pipeline with automated, contour-derived positioning cannot stop at reporting Dice Coefficients, or even at reporting a naive dose comparison between the two independently simulated plans. Both segmentation quality *and* its downstream effect on beam placement need to be assessed, and, as this thesis has shown, they need to be assessed with a design capable of telling them apart.

Tumor Control Probabilities

The absolute values of TCP for this cohort sit well below the range in which TCP is conventionally read as a genuine estimate of clinical efficacy. This is because TCP was used as a comparative index rather than as a clinical prediction. In fact, TCP is a radiabiological figure of merit which encompasses the whole tridimensional dose distribution into a single value, useful for an immediate comparison of different irradiation set ups.

H&N tumors targeted by BNCT are not, anatomically, a single compact mass the way a glioblastoma tumor often is. They tend to be scattered across the mucosal surfaces of the oropharyngeal and laryngeal region, frequently presenting as several discontiguous deposits embedded within, or immediately adjacent to, the mucosa itself, rather than as one contiguous, well-circumscribed volume. A single BNCT irradiation, delivered from a single beam entry and a single patient position, is usually not enough to cover a target of this shape uniformly. Some part of the tumor may receive a high maximum dose, however geometrically distant portions could receive a substantially lower dose due to poor beam penetration. The TCP computed over the whole

GTV inevitably reflects this: it is dominated by the coldest, most under-dosed subvolume, exactly as the strong correlation found between TCP and the near-minimum GTV dose would predict. This is almost certainly why real clinical BNCT protocols for H&N disease routinely deliver more than one irradiation, often from different beam angles, specifically to bring each fragment of a spatially scattered target within adequate range of the beam at some point during treatment. The single-position, single-field simulation used throughout this thesis, chosen deliberately to isolate the segmentation and positioning effects under study, is therefore expected to understate what a real multi-field H&N treatment would achieve. The low TCP values found here should be read in this light.

The RBE-Weighted and Isoeffective Dose Models

A secondary, but methodologically important point is the relationship between the two radiobiological dose models: the fixed-weighting RBE/CBE model, retained mainly for comparability with earlier BNCT dosimetric work, and the photon-isoeffective dose model that this thesis used as the correct choice to calculate TCP and NTCP.

The two models were found to diverge from one another in a systematic dose-dependent way: at higher absolute dose isoeffective dose values are below fixed-weighted dose, as expected, while at lower absolute dose the isoeffective estimate exceeds the fixed-weighted one. This depends on the fact that the irradiation set-up produces low doses in the GTV (as explained above) thus the behavior of the isoeffective dose resembles more what normally happens for the healthy tissues, i.e. a higher value compared to RBE-weighted dose values.

Gamma Index

The gamma-index analysis reinforced the conclusions about the positioning from a different perspective, but it also showed its own limits. Once restricted to the region of true geometric agreement between the two segmentations, the gamma-index pass-rate between manual and AI showed only a weak, non-significant correlation with the Dice Coefficient at the cohort size studied here. That correlation strengthened, and became statistically significant, once the comparison was repeated for anti-swap instead, where no mask-membership ambiguity remains to blur the signal. The pattern is consistent with everything else this thesis found about positioning: a global geometric overlap

measure like the Dice Coefficient captures only part of what determines whether two dose fields agree voxel by voxel, and the remaining part is largely the automated-positioning shift that a purely geometric index has no way of seeing directly.

Future Work

Every dose calculation performed in this thesis, for every patient and every configuration, rests on a simplification: a single, uniform boron-10 concentration assigned to the entire tumor volume, and a second, different but still uniform value assigned to the healthy mucosa. Real boron biodistribution is neither uniform nor identical across patients; it depends on delivery route, timing, tumor vascularity and cellular uptake, and it can vary substantially within a single tumor volume, particularly one as anatomically heterogeneous as the scattered H&N targets discussed above. This thesis could not incorporate that variability, because no patient-specific, spatially resolved boron map was available for this retrospective cohort. It is, however, the single simplification whose removal would most change the clinical realism of everything presented here, and it defines the most natural direction for the future work.

The most consequential extension of this work would replace the uniform-concentration assumption with a patient-specific, spatially resolved boron map derived from boron-PET imaging, registered to the same treatment-planning geometry used throughout this thesis. This is not a purely incremental improvement: it would change the isoeffective and TCP calculations from an idealized best case into an estimate that reflects what a given patient's tumor actually received, voxel by voxel.

A PET-derived boron map would also make possible a comparative tool that does not yet exist for BNCT and that this thesis argues is urgent: an index built on the same logic as the gamma index used throughout Chapter 4, but extended so that two distributions are judged to *agree* only if they agree simultaneously in physical dose, in spatial location, and in local boron concentration. The gamma index used in this thesis inherits a comparison tool designed for photon-beam dose, where the beam penumbra is what actually shapes the dose gradient. A specific BNCT comparative index, incorporating boron concentration as a third dimension of agreement alongside dose and distance, would remove that imperfect fit at its root. Developing and validating such an index, using the PET-derived boron maps described above, is the natural continuation of the methodological groundwork laid in this thesis.

Given the anatomical argument made above for why the single-field TCP values found in this cohort are low, a direct and valuable next step would be to re-simulate this same 24-patient cohort under a more clinically realistic scheme: multiple irradiation ports, with the patient repositioned between fields, in order to maximize the coverage of the whole tumor volume. Because the manual and AI segmentations, and the full four-arm infrastructure built in this thesis, already exist for every patient in this cohort, extending the simulation to a multi-field regime would not require rebuilding the pipeline from scratch; it would instead test directly whether the low, single-field TCP values documented here rise toward clinically meaningful levels once the delivery geometry is allowed to do what real BNCT protocols for this disease site already do in practice.

Finally, because this thesis identified automated, centroid-derived positioning as the central mechanism coupling segmentation error to dosimetric error, a natural and important validation step would compare the automated positioning used throughout this work against positioning performed directly by an experienced medical physicist, on the same patients, using their clinical judgment rather than a purely geometric rule. Such a comparison would clarify how much of the positioning-driven disagreement documented in this thesis is a true, unavoidable consequence of segmentation disagreement propagating through *any* reasonable positioning rule, geometric or expert-driven, and how much of it is instead specific to the particular centroid-based algorithm. This thesis could only explore one positioning strategy, which was useful to standardize the comparison. However, working in close connection with clinical decision makers and using positioning strategies that also take into consideration the protection of radiation sensitive structures in the beam, will add to the relevance of the obtained results.

Altogether, a new index explicitly defined for BNCT, and a realistic irradiation set-up for the whole cohort of patients, will produce a robust comparison of the expected BNCT effectiveness based on radiobiological figures of merit, adding to the prediction power of the dose simulated in patients and on the capacity to improve the speed and the accuracy of the treatment planning.

Appendix A

A.1 The Factor G_{ij}

In its general form, $G_{ij}(\theta)$ accounts for the repair of pairs of sublesions produced during the irradiation time θ, one sublesion originating from radiation i and the other from radiation j. If the production of one sublesion does not influence the production of the other, $G_{ij}(\theta)$ can be shown to decompose into a weighted sum of two independent, single-radiation G factors,

$$G_{ij}(\theta) = a_i\, G_i(\theta) + a_j\, G_j(\theta), \tag{A.1}$$

where a_i and a_j denote the relative proportion of each component, with $a_i + a_j = 1$.

Assuming further that the repair kinetics of each component follows a bi-exponential decay characterized by a fast and a slow repair time, t_{0f} and t_{0s}, independent of LET or, equivalently, independent of which component of the BNCT mixed field is being considered, the single-radiation factor for component k can itself be split into its fast and slow contributions:

$$G_k(\theta, t_{0f}, t_{0s}) = a_{kf}\, G(\theta, t_{0f}) + a_{ks}\, G(\theta, t_{0s}), \tag{A.2}$$

where a_{kf} and a_{ks} are the fractions of sublesions of component k repaired through the fast and slow kinetics, respectively, with $a_{kf} + a_{ks} = 1$. Here $G(\theta, t_0)$ is the generalized Lea–Catcheside time factor governing the simultaneous build-up and repair of radiation damage under constant-dose-rate irradiation,

$$G(\theta, t_0) = \frac{2t_0}{\theta} - 2\left(\frac{t_0}{\theta}\right)^2 \left(1 - e^{-\theta/t_0}\right). \tag{A.3}$$

Combining Equations (A.1)–(A.3) and simplifying, one obtains

$$G_{ij}(\theta) = G(\theta, t_{0s}) - \big(a_i a_{if} + a_j a_{jf}\big)\Big(G(\theta, t_{0s}) - G(\theta, t_{0f})\Big). \tag{A.4}$$

For the irradiation times typical of cell-survival experiments performed in BNCT (of the order of 10–30 min), together with the fast and slow repair times generally reported in the literature (about 30 min and 14 h, respectively), Equation (A.4) is well approximated by

$$G_{ij}(\theta) \simeq G(\theta, t_0 = 1\,\mathrm{h}), \qquad \forall\, i, j. \tag{A.5}$$

In other words, choosing a single convenient value $t_0 = 1$ h allows all sixteen individual G_{ij} functions to be well approximated by one common function G. As an illustration, for a 15-minute irradiation combining low- and high-LET radiation in relative proportions of 0.2 and 0.8, with 53% and 20% of the respective sublesions repaired through the fast kinetics, the exact expression gives $G_{ij}(\theta = 15\,\mathrm{min}) = 0.94$, compared with $G(\theta = 15\,\mathrm{min}, t_0 = 1\,\mathrm{h}) = 0.92$ from the single-factor approximation – a relative difference below 3% [13].

Appendix B

B.1 Loss Functions and Optimization Algorithms

This appendix expands on two concepts introduced only briefly in Chapter 2 — the loss function used to quantify a network's prediction error (Section 2.2.3) and the gradient descent algorithm used to minimize it — by giving a more complete, self-contained account of the loss functions and optimization algorithms actually in use in modern deep learning, including the ones adopted by the segmentation networks discussed in Chapter 2 (U-Net, V-Net, nnU-Net).

A brief note on notation: Chapter 2 denotes the network's output as g, matching the label used in Figure 2.4. In this appendix and in Appendix B.2, the same quantity is instead written $\hat{y}$, the standard notation in the sources cited here (Ruder, Kingma & Ba, Zeiler, Elgendy, and others); the symbol g (and g_t, with a time subscript) is reserved throughout both appendices for the *gradient* of the loss, $g_t \equiv \nabla_\theta J(\theta)$, to keep the two uses distinct.

B.1.1 Loss Functions

Section 2.2.3 introduces θ as the parameters — weights and biases — that a network's training procedure adjusts. Throughout this appendix, θ denotes these parameters collected into a single vector, simply the collection of every weight $w_{ij}^{(l)}$ and bias $b_i^{(l)}$ across all layers $l = 1, \ldots, L$. Training then means finding the θ that minimizes a scalar *loss function* $J(\theta)$, quantifying how far the network's prediction $\hat{y}$ is from the true, known target y.

Elgendy, cited throughout this section for the specific loss functions below, keeps weights W and biases b as separate arguments rather than combining them into a single vector, writing the loss as $E(W, b)$ — also noting the equivalent notation $J(W, b)$ [83].

This appendix instead adopts the single-vector convention θ and $J(\theta)$ throughout, consistent with the optimization literature of Section B.1.2 (Ruder, Kingma & Ba, Duchi, Zeiler), so that the same symbols can be used across both loss functions and optimizers without switching notation partway through.

Regression Losses

For tasks in which the target is a continuous, real-valued quantity, the most common choice is the *mean squared error* (MSE),

$$J_{\text{MSE}}(\theta) = \frac{1}{N}\sum_{i=1}^{N}(\hat{y}_i - y_i)^2, \tag{B.1}$$

where N is the number of training examples [83]. Squaring the residual $(\hat{y}_i - y_i)$ serves two purposes: it guarantees that the error contributed by every example is non-negative, so that errors of opposite sign cannot cancel out when averaged (an error of $+10$ and one of -10 would otherwise average to a misleading 0); and it penalizes large errors disproportionately more than small ones [83]. This second property, however, also makes MSE highly sensitive to outliers, since a single badly mispredicted example can dominate the loss. When this sensitivity is undesirable, the *mean absolute error* (MAE),

$$J_{\text{MAE}}(\theta) = \frac{1}{N}\sum_{i=1}^{N}\left|\hat{y}_i - y_i\right|, \tag{B.2}$$

provides a more robust alternative, since it does not amplify the contribution of large residuals [83].

Classification Losses: Cross-Entropy

Cross-entropy is best introduced independently of any specific network notation, as a way of comparing two probability distributions directly. Given two discrete probability distributions over the same m outcomes — a predicted distribution $p = (p_1, \ldots, p_m)$ and a target (reference) distribution $q = (q_1, \ldots, q_m)$, both satisfying $p_i, q_i \in [0, 1]$ and $\sum_{i=1}^{m} p_i = \sum_{i=1}^{m} q_i = 1$ — the cross-entropy loss between them, weighted by the target distribution, is

$$J_{\text{CE}} = -\sum_{i=1}^{m} q_i \log(p_i). \tag{B.3}$$

For a fixed target q, this quantity is minimized exactly when $p = q$, and grows the more the prediction p diverges from the target — which is what makes it a sensible loss for comparing a predicted distribution against a target one.[5]

In the classification setting of this thesis, the predicted distribution $p = \hat{y} = (\hat{y}_1, \ldots, \hat{y}_m)$ is the direct output of the softmax activation function introduced in Section 2.2.1, with $\hat{y}_i \in [0, 1]$ the predicted probability that the input belongs to class i; the target distribution $q = y = (y_1, \ldots, y_m)$ is typically a one-hot vector, with $y_i = 1$ for the correct class and $y_i = 0$ otherwise. Substituting into Equation (B.3) gives the classification loss actually used in training,

$$J_{\mathrm{CE}}(\theta) = -\sum_{i=1}^{m} y_i \log(\hat{y}_i). \tag{B.4}$$

Because y_i is zero for every incorrect class, only the predicted probability $\hat{y}_i$ assigned to the *correct* class contributes to the sum: the loss is small when the network assigns high probability to the correct class, and grows (in principle without bound, as $\hat{y}_i \to 0$) as the network becomes more confidently wrong. For example, if the correct class is "dog" (true distribution $[0, 1, 0]$ over cat/dog/fish) and the network predicts the probability distribution $[0.2, 0.3, 0.5]$, the cross-entropy loss is $-\log(0.3) \approx 1.2$; a better prediction that assigns higher probability to "dog" would yield a correspondingly smaller loss [83].

For a two-class problem, Equation (B.4) reduces to the more familiar binary cross-entropy form,

$$J_{\mathrm{BCE}}(\theta) = -\left[y \log(\hat{y}) + (1 - y) \log(1 - \hat{y})\right], \tag{B.5}$$

where $y \in \{0, 1\}$ is the true label and $\hat{y} \in [0, 1]$ is the predicted probability of the positive class.

Segmentation Losses: Dice Loss and Composite Losses

Neither MSE nor cross-entropy alone is ideal for the dense, per-voxel classification task of image segmentation (Section 2.3), because a purely voxel-wise loss treats every voxel independently and can therefore be dominated by the background class whenever the

[5] The standard information-theoretic notation $H(p, q) = -\sum_i p_i \log(q_i)$ conventionally takes the *first* argument as the true/reference distribution and the second as the model's approximation. To keep p for "prediction" and q for "target" throughout this thesis, the roles are reversed relative to that convention; Equation (B.3) is therefore deliberately *not* labeled $H(p, q)$, to avoid it being misread against the standard definition.

structure of interest — a GTV, for instance — occupies only a small fraction of the volume. The Dice similarity coefficient addresses this by measuring the overlap between two binary volumes directly, independently of the size of the background. Given p the predicted segmentation and q for the ground-truth segmentation — let $p_i, q_i \in \{0, 1\}$ indicate whether voxel i is, respectively, predicted and actually part of the structure of interest, out of N total voxels. The Dice coefficient is then

$$D = \frac{2 \sum_{i=1}^{N} p_i q_i}{\sum_{i=1}^{N} p_i^2 + \sum_{i=1}^{N} q_i^2}, \tag{B.6}$$

ranging from $D = 0$ (no overlap) to $D = 1$ (perfect overlap), and the corresponding Dice loss $\mathcal{L}_{\text{Dice}} = 1 - D$ is differentiable and can therefore be minimized by gradient descent [87].

B.1.2 Gradient-Based Optimization Algorithms

Section 2.2.3 introduced the general gradient descent update rule,

$$\theta \leftarrow \theta - \eta \, \nabla_\theta J(\theta),$$

and its three standard variants — batch, stochastic, and mini-batch gradient descent — which differ in how much data is used to compute each gradient estimate [88]. In practice, however, plain ("vanilla") gradient descent, even in its mini-batch form, converges slowly and unreliably on the loss surfaces encountered in deep learning, for two main reasons:

- **Ravines.** Around many local optima, the loss surface curves much more steeply in one dimension than in others. In such a ravine, vanilla gradient descent oscillates back and forth across the steep direction while making only slow progress along the shallow direction toward the minimum.
- **Saddle points.** At a saddle point, the surface has positive curvature along one dimension and negative curvature along another. Vanilla gradient descent, and even some of its refinements, can struggle to break the symmetry of such points and may stall in their vicinity.

The following algorithms, in widespread use throughout modern deep learning were developed specifically to address these problems.

Momentum and Nesterov Accelerated Gradient

Momentum accelerates gradient descent in the relevant direction and damps the oscillations characteristic of ravines by accumulating a fraction of the previous update into the current one,

$$v_t = \gamma v_{t-1} + \eta \nabla_\theta J(\theta), \qquad \theta \leftarrow \theta - v_t, \tag{B.7}$$

where the momentum coefficient γ is typically set to around 0.9 [88]. The intuition is mechanical: as a ball rolling downhill accumulates speed, the update accumulates contributions from directions along which the gradient has consistently pointed in recent steps, while updates along directions that keep changing sign are damped.

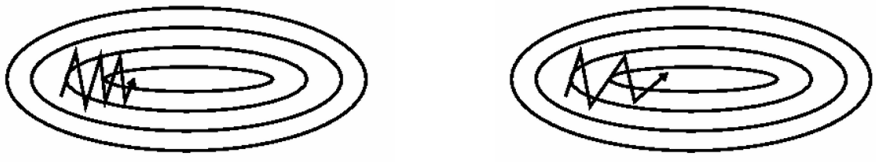


Figure B.1: *SGD without momentum (left) and SGD with momentum (right). For most points on the surface, the gradient does not point towards the minimum, and successive steps of gradient descent can oscillate from one side to the other, progressing only very slowly to the minimum (left). The addition of momentum helps to speed up convergence to the minimum by damping these oscillations (right)* [88].

A refinement, *Nesterov Accelerated Gradient* (NAG), gives the momentum term a degree of foresight: rather than computing the gradient at the current parameters θ, NAG first takes an approximate step in the direction of the accumulated momentum, and then evaluates the gradient at that anticipated, near-future position,

$$v_t = \gamma v_{t-1} + \eta \nabla_\theta J(\theta - \gamma v_{t-1}), \qquad \theta \leftarrow \theta - v_t. \tag{B.8}$$

This look-ahead correction allows NAG to slow down before the loss surface starts rising again, rather than blindly overshooting the minimum as plain momentum tends to do [88].

Adaptive Learning-Rate Methods: Adagrad, Adadelta, and RMSprop

Momentum and NAG adapt the *direction* of the update, but still apply the same learning rate η to every parameter. The algorithms here listed instead adapt the learning rate itself; to describe them it is convenient to abbreviate the gradient at time step t as $g_t \equiv \nabla_\theta J(\theta)$, evaluated at the current θ, with $g_{t,i}$ denoting its i-th component (i.e., the partial derivative with respect to a single parameter θ_i) — the same quantity computed by the backpropagation algorithm of Appendix B.2.

Adagrad adapts the learning rate on a per-parameter basis, performing larger updates for parameters that are updated infrequently and smaller updates for parameters that are updated frequently — a property that makes it particularly well suited to sparse data [88, 101]. For each parameter θ_i, Adagrad divides the (fixed) global learning rate η by the square root of the sum of the squares of all past gradients with respect to θ_i,

$$\theta_{t+1,i} = \theta_{t,i} - \frac{\eta}{\sqrt{G_{t,ii} + \epsilon}}\, g_{t,i}, \tag{B.9}$$

where $G_{t,ii} = \sum_{\tau=1}^{t} g_{\tau,i}^2$ is that accumulated sum up to time step t, and ϵ is a small smoothing term that prevents division by zero [101]. Adagrad's main practical advantage is that it removes the need to hand-tune the learning rate; its main weakness is that the accumulated sum in the denominator grows monotonically throughout training, so the effective learning rate shrinks continuously and can eventually become so small that learning effectively stops [88, 101].

Adadelta and *RMSprop* were developed independently, at around the same time, specifically to correct this weakness [102]. Instead of accumulating *all* past squared gradients, both methods replace the sum with an exponentially decaying average of squared gradients,

$$E[g^2]_t = \gamma\, E[g^2]_{t-1} + (1-\gamma)\, g_t^2, \tag{B.10}$$

with γ typically set to 0.9, so that old gradients are progressively forgotten rather than accumulated indefinitely. RMSprop uses this decaying average directly in place of Adagrad's growing sum,

$$\theta_{t+1} = \theta_t - \frac{\eta}{\sqrt{E[g^2]_t + \epsilon}}\, g_t, \tag{B.11}$$

with a recommended learning rate of $\eta = 0.001$ [88].

Adadelta additionally eliminates the need to set a learning rate at all, by replacing η with a second decaying average, this time of the past *parameter updates* themselves rather than of the gradients — a substitution motivated by the observation that, in SGD, Momentum, and Adagrad alike, the units of the parameter update do not match the units of the parameter itself, whereas a second-order method such as Newton's method does have consistent units; Adadelta's update is constructed to restore this dimensional consistency without actually computing second-order (Hessian) information [102].

It is convenient to introduce the shorthand $\mathrm{RMS}[X]_t \equiv \sqrt{E[X^2]_t + \epsilon}$ for the *root mean square* of any decaying-average quantity $E[X^2]_t$ of the type in Equation (B.10). This is the square root of the running average of X's squared values, with ϵ added for numerical stability, and is also the origin of RMSprop's name: Equation (B.11) is exactly

$$\theta_{t+1} = \theta_t - \frac{\eta}{\mathrm{RMS}[g]_t} g_t, \tag{B.12}$$

dividing the learning rate by the root mean square of recent gradients.

Adadelta uses the same $\mathrm{RMS}[g]_t$ for the gradient, and additionally defines an analogous decaying average of the squared parameter updates $E[\Delta\theta^2]_t$ with $\mathrm{RMS}[\Delta\theta]_t \equiv \sqrt{E[\Delta\theta^2]_t + \epsilon}$. The Adadelta update is then

$$\Delta\theta_t = -\frac{\mathrm{RMS}[\Delta\theta]_{t-1}}{\mathrm{RMS}[g]_t}\, g_t, \qquad \theta_{t+1} = \theta_t + \Delta\theta_t, \tag{B.13}$$

where $\mathrm{RMS}[\Delta\theta]_{t-1}$ — the numerator, lagging one step behind — plays the role that the fixed learning rate η plays in every other method above, and is instead estimated directly from the recent history of the optimization itself [102].

Adam and Its Variants

Adaptive Moment Estimation (Adam) combines the two ideas introduced above: like Momentum, it keeps an exponentially decaying average of past (non-squared) gradients, m_t; like Adadelta and RMSprop, it also keeps an exponentially decaying average of past squared gradients, v_t,

$$m_t = \beta_1 m_{t-1} + (1-\beta_1) g_t, \qquad v_t = \beta_2 v_{t-1} + (1-\beta_2) g_t^2, \tag{B.14}$$

which are estimates of, respectively, the first moment (mean) and the (uncentered) second moment of the gradient — hence the name of the method [103].

Because m_t and v_t are initialized at zero, they are biased toward zero especially in the first few iterations; Adam corrects for this with bias-corrected estimates $\hat{m}_t = m_t/(1-\beta_1^t)$ and $\hat{v}_t = v_t/(1-\beta_2^t)$, which are then used in the final update rule,

$$\theta_{t+1} = \theta_t - \frac{\eta}{\sqrt{\hat{v}_t} + \epsilon}\, \hat{m}_t, \tag{B.15}$$

with recommended default values $\eta = 0.001$, $\beta_1 = 0.9$, $\beta_2 = 0.999$, and $\epsilon = 10^{-8}$ [103].

Kingma and Ba further show that, under bounded-gradient assumptions, Adam achieves an $O(\sqrt{T})$ regret bound in the online convex optimization setting, comparable to the best known bounds for this class of problem, and report empirically that Adam converges as fast as or faster than SGD with Nesterov momentum, Adagrad, and RMSprop on logistic regression, multilayer, and convolutional network benchmarks [103].

Two further variants extend Adam: *AdaMax* generalizes the second-moment term from the ℓ_2 norm to the (numerically more stable) ℓ_∞ norm of past gradients [103], and *Nadam* incorporates the Nesterov look-ahead correction of Equation (B.8) directly into Adam's momentum term, which Dozat shows speeds up convergence and improves both training and validation loss relative to plain Adam on a convolutional autoencoder benchmark [104].

Empirical Comparison on Convolutional Networks

The theoretical trade-offs summarized above are corroborated by empirical benchmarks. Dogo et al. trained the same convolutional network architecture with eight different optimizers (vanilla SGD, SGD with momentum, SGD with momentum and Nesterov correction, Adagrad, Adadelta, RMSprop, Adam, AdaMax, and Nadam) across three independent image classification datasets, and compared convergence time, final classification accuracy, and final loss [105]. Table B.2 reports their results on one of the three benchmarks (a natural-images dataset with eight object classes); the relative ranking of optimizers was broadly, though not perfectly, consistent across all three datasets tested [105].

Table B.1: Summary of the gradient descent optimization algorithms discussed in this appendix.

Algorithm	Core idea	Typical use / limitation
SGD (mini-batch) [88]	Update using the gradient of a small, randomly sampled subset of the data	Simple, widely used baseline; can be slow in ravines and near saddle points
Momentum [88]	Accumulates a decaying average of past gradients to accelerate descent and damp oscillation	Faster convergence in ravines; can overshoot the minimum
NAG [88]	Momentum with a look-ahead gradient evaluation	Corrects momentum's overshooting tendency
Adagrad [101]	Per-parameter learning rate, scaled by the accumulated sum of squared past gradients	Well suited to sparse data; learning rate shrinks to zero over long training
Adadelta [102]	Restricts Adagrad's accumulation to a decaying window; eliminates the need for a learning rate	Addresses Adagrad's diminishing learning rate
RMSprop [102]	Divides the learning rate by a decaying average of squared gradients	Similar to Adadelta; robust default choice
Adam [103]	Combines RMSprop's adaptive learning rate with Momentum's decaying gradient average, plus bias correction	Often the best overall default choice [88]
AdaMax [103] / Nadam [104]	ℓ_∞-norm variant of Adam / Nesterov-corrected Adam	Used in specific settings requiring extra numerical stability or responsiveness

Consistent with Ruder's theoretical discussion, the plain SGD and Adagrad configurations are the weakest performers on this benchmark, while momentum-based and adaptive methods (SGD with momentum, Adam, AdaMax, Nadam) achieve both higher accuracy and lower final loss within a comparable convergence time [105]. It should be noted that optimizer performance is architecture- and dataset-dependent, and that the ranking in Table B.2 should not be read as a universal recommendation, but as one empirical data point supporting the qualitative guidance summarized in the previous section.

Table B.2: Empirical comparison of gradient descent-based optimizers trained on the same convolutional network architecture and natural-image dataset (8 classes, 6899 images). Convergence time, final classification accuracy, and final loss as reported by [105].

Optimizer	Convergence time (h)	Accuracy	Loss
SGD (vanilla)	5.06	0.775	1.228
SGD + momentum	4.86	0.891	0.501
SGD + momentum + Nesterov	5.02	0.851	0.584
Adagrad	5.02	0.699	1.589
RMSprop	5.17	0.797	0.619
Adam	5.06	0.892	0.920
Adadelta	5.48	0.739	1.669
AdaMax	4.90	0.909	0.440
Nadam	5.44	0.913	0.225

B.2 The Backpropagation Algorithm

Section 2.2.3 introduced the general gradient descent update, $\theta \leftarrow \theta - \eta \nabla_\theta J(\theta)$, and Appendix B.1 detailed the family of algorithms used to turn that gradient into an effective parameter update. Neither addresses a more basic question: for a network with millions of weights organized into dozens of layers, how is the gradient $\nabla_\theta J(\theta)$ itself computed? This appendix answers that question by deriving the *backpropagation* algorithm, first popularized by Rumelhart, Hinton and Williams [106], which remains the computational backbone of essentially every deep learning framework in use today — including the ones underlying nnU-Net [92].

B.2.1 Why Backpropagation Is Needed

In principle, the partial derivative of the loss J with respect to any individual weight could be estimated numerically, by perturbing that weight slightly and observing the change in J. In practice this is computationally hopeless: a modern segmentation network has millions of weights, and a separate forward pass through the entire network would be needed for every single one of them just to estimate one gradient.

Backpropagation avoids this by exploiting the layered, compositional structure of a neural network directly: because feedforward propagation (Section 2.2.2) is nothing more than a composition of simple functions, the calculus *chain rule* makes it possible to compute the gradient with respect to *every* weight in the network using only two passes, (i) one forward pass to compute the network's output, and (ii) one backward pass, of comparable computational cost, that reuses intermediate quantities from the forward pass rather than recomputing them from scratch for each weight [83].

B.2.2 The Chain Rule

The chain rule states that the derivative of a composition of functions is the product of the derivatives of each function in the chain,

$$\frac{d}{dx} f(h(x)) = f'(h(x))\, h'(x). \tag{B.16}$$

A feedforward network is precisely such a composition: the output of layer 1 feeds into layer 2, whose output feeds into layer 3, and so on, until the final loss J is computed

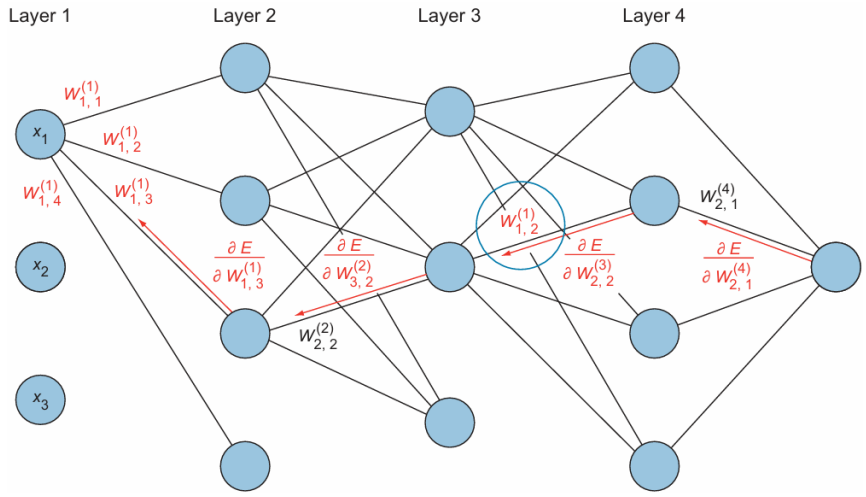


Figure B.2: *Backpropagation for a multilayer perceptron with many weight variables* [83].

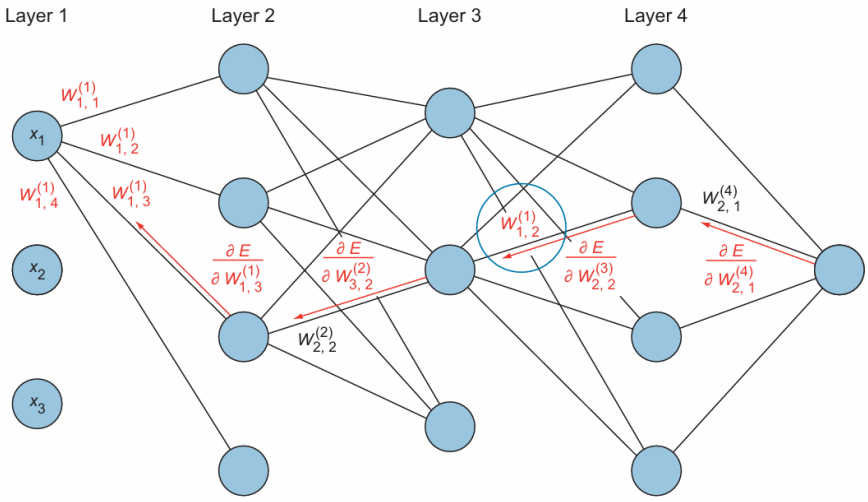


Figure B.3: *Backpropagation uses the chain rule to flow gradients back through the network* [83].

from the network's output. The effect of a deep weight on J is therefore the *product* of the local effects of every intermediate computation lying on the path between that weight and J: the effect of the weight on its own layer's pre-activation, the effect of that pre-activation on the layer's output, the effect of that output on the next layer's pre-activation, and so on, all the way to the loss [83]. Backpropagation is simply a systematic, efficient bookkeeping scheme for multiplying out these chains of local derivatives for every weight in the network at once, working backward from the loss toward the input — hence the name, illustrated in Figure B.3.

B.2.3 Formal Derivation

We adopt the same layer-index notation introduced in Section 2.2.2 for feedforward propagation. Recall that the pre-activation and activation of unit i in layer l are

$$z_i^{(l)} = \sum_j w_{ij}^{(l)} a_j^{(l-1)} + b_i^{(l)}, \qquad a_i^{(l)} = \sigma\Big(z_i^{(l)}\Big), \tag{B.17}$$

consistent with Equations (2.3)–(2.4), with $a^{(0)} \equiv x$ (the input features) and $a^{(L)} \equiv \hat{y}$ (the network's output, layer L being the output layer). Our goal is to compute $\partial J/\partial w_{ij}^{(l)}$ and $\partial J/\partial b_i^{(l)}$ for every weight and bias in the network, where J is any of the loss functions of Appendix B.1.1.

The Error Term δ

The key device that makes backpropagation efficient is to introduce, for every unit i in every layer l, an intermediate quantity

$$\delta_i^{(l)} \equiv \frac{\partial J}{\partial z_i^{(l)}}, \tag{B.18}$$

informally interpretable as "how much unit i's pre-activation is responsible for the final loss." As shown below, once $\delta_i^{(l)}$ is known for every unit, the gradient with respect to every weight and bias follows immediately and cheaply; the bulk of the derivation therefore consists of finding an efficient, recursive way to compute $\delta_i^{(l)}$ for every layer, working backward from the output.

Error at the Output Layer

At the output layer L, $\delta_i^{(L)}$ follows directly from the chain rule applied to Equation (B.17): the loss depends on $z_i^{(L)}$ only through $a_i^{(L)} = \sigma(z_i^{(L)})$, so

$$\delta_i^{(L)} = \frac{\partial J}{\partial a_i^{(L)}} \cdot \sigma'\Big(z_i^{(L)}\Big). \tag{B.19}$$

The first factor, $\partial J/\partial a_i^{(L)}$, is simply the derivative of whichever loss function is in use (Appendix B.1.1) with respect to the network's raw output: for instance, for the mean squared error of Equation (B.1) this factor is $2(\hat{y}_i - y_i)/N$.
The second factor, $\sigma'(z_i^{(L)})$, is the derivative of the output layer's activation function,

evaluated at the current pre-activation — one further reason, alongside those discussed in Section 2.2.3, why the choice of activation function (and, in particular, avoiding functions whose derivative vanishes over large regions) directly affects how well a network trains.

Error at Hidden Layers: The Backward Recursion

For a hidden layer $l < L$, unit i does not connect directly to the loss; it influences J only indirectly, through every unit k in the next layer that it feeds into. The chain rule must therefore sum over all such downstream paths,

$$\delta_i^{(l)} = \sigma'\left(z_i^{(l)}\right) \sum_k w_{ki}^{(l+1)}\, \delta_k^{(l+1)}, \tag{B.20}$$

where the sum runs over every unit k in layer $l+1$ that receives a connection from unit i, and $w_{ki}^{(l+1)}$ is the weight of exactly that connection [107]. Equation (B.20) is the computational heart of backpropagation: it expresses the error term of layer l entirely in terms of quantities already computed for layer $l+1$ (the $\delta_k^{(l+1)}$ terms) and quantities already computed during the forward pass (the pre-activation $z_i^{(l)}$). Starting from $\delta^{(L)}$ (Equation (B.19)) and applying Equation (B.20) repeatedly for $l = L-1, L-2, \ldots, 1$, the error term of every unit in the network can be obtained with exactly one pass through the network, moving backward from the output to the input — the "backward pass" that gives the algorithm its name.

From Error Terms to Weight Gradients

Once every $\delta_i^{(l)}$ is known, the gradient with respect to any weight or bias in the network follows from one further, local application of the chain rule to Equation (B.17): since $z_i^{(l)}$ depends on $w_{ij}^{(l)}$ only through the term $w_{ij}^{(l)} a_j^{(l-1)}$, and on $b_i^{(l)}$ directly,

$$\frac{\partial J}{\partial w_{ij}^{(l)}} = \delta_i^{(l)}\, a_j^{(l-1)}, \qquad \frac{\partial J}{\partial b_i^{(l)}} = \delta_i^{(l)}. \tag{B.21}$$

These are precisely the quantities g_t (or $g_{t,i}$, in the per-parameter notation of Appendix B.1.2) that feed into every gradient descent variant discussed there: backpropagation supplies the gradient, and the optimization algorithm (vanilla SGD, Momentum, Adam, and so on) decides how to use that gradient to update the weights [88].

B.2.4 The Algorithm

Collecting the derivation above, one full backpropagation step for a network with L layers, given a training example (x, y), consists of:

1. **Forward pass.** Propagate x forward through the network using Equation (B.17), storing every pre-activation $z_i^{(l)}$ and activation $a_i^{(l)}$ for every layer $l = 1, \ldots, L$; the final activation $a^{(L)} = \hat{y}$ is the network's prediction, and the loss J is evaluated by comparing $\hat{y}$ to the true label y (Appendix B.1.1).

2. **Output error.** Compute $\delta_i^{(L)}$ for every output unit using Equation (B.19).

3. **Backward pass.** For $l = L-1, L-2, \ldots, 1$, compute $\delta_i^{(l)}$ for every unit in layer l using the recursion of Equation (B.20).

4. **Gradient extraction.** Compute $\partial J/\partial w_{ij}^{(l)}$ and $\partial J/\partial b_i^{(l)}$ for every weight and bias using Equation (B.21).

5. **Parameter update.** Hand the resulting gradient to the chosen optimization algorithm of Appendix B.1.2 to update every weight and bias.

Steps 1–4 are repeated for every mini-batch of training data (Section 2.2.3), with the gradients of Equation (B.21) averaged (or summed) over the examples in the batch before step 5 is applied. Because both the forward and backward pass require, for each layer, an amount of computation proportional to the number of connections into that layer, one full backpropagation step costs only a small constant multiple of a single forward pass regardless of how many weights the network has, which is precisely what makes training networks with millions or billions of parameters computationally feasible in the first place.

Appendix C

C.1 Derivation of the KERMA Approximation and the Charged-Particle-Equilibrium Condition

Section 3.2.5 introduces the KERMA approximation used throughout the Monte Carlo dose-calculation pipeline of this thesis, and states, without derivation, that KERMA and absorbed dose coincide under conditions of charged-particle equilibrium (CPE). This appendix derives that equivalence and makes explicit the conditions under which it holds, together with the direction of the resulting bias when those conditions are violated, as is the case near the skin surface.

C.1.1 KERMA and Absorbed Dose

Neutrons and photons are indirectly ionizing radiations: they do not deposit energy in matter directly, but instead transfer energy to charged secondary particles (protons, α-particles, recoil nuclei, or electrons), which in turn deposit that energy locally through Coulomb interactions with the surrounding medium. Two distinct dosimetric quantities follow from this two-step process [109].

KERMA (Kinetic Energy Released per unit MAss) is defined, for a mass element $\mathrm{d}m$ at a point P, as the initial kinetic energy transferred to *all* charged particles liberated by interactions of uncharged particles within $\mathrm{d}m$, irrespective of where that kinetic energy is subsequently deposited [109]:

$$K = \frac{\mathrm{d}E_{\mathrm{tr}}}{\mathrm{d}m}, \tag{C.1}$$

where $\mathrm{d}E_{\mathrm{tr}}$ is the sum of the initial kinetic energies of all charged particles produced in $\mathrm{d}m$. For a photon or neutron energy fluence Ψ, KERMA is related to the medium's

mass energy-transfer coefficient μ_{tr}/ρ (for photons) or to the corresponding KERMA factor $k(E)$ (for neutrons, as introduced for each of the four BNCT dose components in Section 1.3.1) by

$$K = \Psi\left(\frac{\mu_{\mathrm{tr}}}{\rho}\right) \qquad \text{(photons)}, \qquad\qquad K = \int_E \phi(E)\, k(E)\, \mathrm{d}E \qquad \text{(neutrons)}, \quad \text{(C.2)}$$

with $\phi(E)$ the energy-differential neutron fluence. A fraction g of the transferred energy may subsequently be re-radiated by the charged particles as bremsstrahlung rather than deposited locally; the *collision KERMA* $K_{\mathrm{c}} = K(1-g)$ excludes this radiative fraction and is the quantity of direct dosimetric relevance [109]. At the photon and secondary-electron energies relevant to BNCT (well below the energies at which bremsstrahlung becomes significant in low-Z tissue), $g \ll 1$ and $K_{\mathrm{c}} \approx K$; this approximation is adopted throughout Section 3.2.5 and is not restated explicitly there.

Absorbed dose D, by contrast, is defined directly in terms of the energy actually imparted to $\mathrm{d}m$, regardless of where that energy was originally transferred to a charged particle:

$$D = \frac{\mathrm{d}E_{\mathrm{dep}}}{\mathrm{d}m}, \tag{C.3}$$

where $\mathrm{d}E_{\mathrm{dep}}$ is the net energy actually deposited within $\mathrm{d}m$ by all charged particles passing through it, whether they originated inside $\mathrm{d}m$ or entered it from elsewhere. Formally, $\mathrm{d}E_{\mathrm{dep}}$ is defined as the sum of all energy entering the volume of interest minus all energy leaving it, accounting for any mass-energy conversion within the volume [109] – a balance made explicit in Section C.1.2 below.

K and D are therefore conceptually distinct: K accounts for energy at the moment and point of *transfer* to a charged particle, while D accounts for energy at the point of *deposition*. Because charged particles travel a finite distance (their range) before losing all their kinetic energy, a particle produced at one point deposits its energy over an extended track that may cross into neighboring mass elements, or even leave the medium altogether [109]. The two quantities coincide only under the specific balance condition derived next.

C.1.2 The CPE condition

Consider a small volume element $\mathrm{d}V$ of mass $\mathrm{d}m = \rho\,\mathrm{d}V$ centered on a point P. Neglecting nuclear reaction Q-values (treated separately for each BNCT dose component

in Section 1.3.1), energy conservation for dV can be written, following directly from the definition of absorbed dose as energy imparted [109], as

$$D\,\mathrm{d}m = T_{\mathrm{in}} + T_{\mathrm{local}}, \tag{C.4}$$

where T_{local} is the kinetic energy deposited within dV by charged particles *produced* within dV, and T_{in} is the kinetic energy deposited within dV by charged particles produced *outside* dV that enter it and deposit some or all of their remaining energy there. By the definition of KERMA,

$$K\,\mathrm{d}m = T_{\mathrm{local}} + T_{\mathrm{out}}, \tag{C.5}$$

where T_{out} is the kinetic energy carried *out* of dV by charged particles that were produced within dV but escape it while still carrying kinetic energy. Subtracting Equation (C.5) from Equation (C.4) gives

$$\big(D - K\big)\,\mathrm{d}m = T_{\mathrm{in}} - T_{\mathrm{out}}. \tag{C.6}$$

Equation (C.6) is the central result: $D = K$ if and only if the energy carried into dV by charged particles born elsewhere exactly balances, in expectation, the energy carried out of dV by charged particles born within it. This balance, $T_{\mathrm{in}} = T_{\mathrm{out}}$, is precisely the condition of *charged-particle equilibrium*, consistent with the buildup-region picture of CPE given by Podgorsak for the photon-beam case [109].

The three conditions below, under which CPE holds at P, are a standard restatement, adapted to the present neutron/photon mixed-field context, of the compositional-homogeneity and fluence-uniformity requirements for CPE discussed generally in the radiological-physics literature; they are not quoted verbatim from a single source and should be cross-checked against a dedicated dosimetry reference (e.g., Attix, *Introduction to Radiological Physics and Radiation Dosimetry*) if used beyond this thesis. CPE holds at P when the following three conditions are simultaneously satisfied over a region of radius at least $R_{\max}$ around P, where $R_{\max}$ is the maximum range of the secondary charged particles produced by the local radiation field:

1. **Compositional homogeneity.** The atomic composition and density of the medium are uniform over the region, so that the rate and energy spectrum of

charged-particle production is the same at every point within range of P; a material interface within $R_{\max}$ of P breaks this condition, since the production rate on the two sides of the interface generally differs.

2. **Fluence uniformity.** The uncharged-particle (neutron or photon) fluence is approximately constant over the region, i.e., attenuation of the primary field over a distance $R_{\max}$ is negligible, so that charged particles entering $\mathrm{d}V$ from a distance $R_{\max}$ away were produced under essentially the same fluence as those produced locally.

3. **Absence of a nearby boundary in the propagation geometry.** No external boundary of the irradiated medium (e.g., the patient's skin surface) lies within $R_{\max}$ of P in the up-field direction, since charged particles cannot be produced in vacuum or air at the rate required to balance the outflow of particles produced in tissue near such a boundary.

When all three conditions hold, $T_{\mathrm{in}} = T_{\mathrm{out}}$ on average over the ensemble of particle histories, and Equation (C.6) gives $D = K$ (or, retaining the radiative correction of Section C.1.1, $D = K_{\mathrm{c}}$).

C.1.3 Breakdown of CPE

Section 3.2.5 notes that CPE breaks down specifically near tissue boundaries, most importantly at the skin surface, and that the resulting bias is conservative with respect to the dose-volume constraint that determines treatment time. This can now be derived directly from Equation (C.6).

Consider a point P in tissue, at a depth less than $R_{\max}$ below the skin surface, with air occupying the region above the surface. Condition 1 of Section C.1.2 is violated: the air above the surface has a much lower density, and hence a much lower rate of charged-particle production per unit volume, than the tissue below it. Charged particles produced within tissue near P that travel upward, toward the surface and into the air, still carry away kinetic energy ($T_{\mathrm{out}} > 0$, as in the homogeneous case), but the corresponding inward contribution T_{in} is reduced, since the low-density air produces far fewer charged particles capable of traveling back down into tissue to deposit energy

at P. Equation (C.6) then gives

$$T_{\text{in}} < T_{\text{out}} \quad \Longrightarrow \quad D < K \tag{C.7}$$

at points sufficiently close to the tissue-air interface. The KERMA approximation therefore *overestimates* the true absorbed dose near the skin surface. Because skin is frequently the dose-limiting organ at risk in BNCT (Section 3.2.5), and because the irradiation time is set by the requirement that the dose to the dose-limiting OAR not exceed its tolerance (Section C.2 below), an overestimate of the near-surface dose translates directly into an *underestimate* of the maximum permissible irradiation time: the resulting treatment plan is dosimetrically conservative for the OAR in question, at the cost of a correspondingly shorter, and therefore potentially suboptimal, treatment time for the target. This is the precise mechanism underlying the qualitative statement made in Section 3.2.5.

C.2 Derivation of the Dose-Rate-to-Time Conversion Under a Dose-Volume Constraint

Section 3.3.7 introduces, in Equation (3.4), the expression used to convert a Monte Carlo dose-*rate* map into a maximum permissible irradiation *time* under a set of dose-volume constraints on the dose-limiting organ at risk. This appendix derives that expression from the underlying dose-volume constraint and makes explicit the approximation on which it rests.

C.2.1 Dose Accumulation

Let $\dot{D}_w(x)$ denote the biologically weighted dose *rate* at voxel x, obtained directly from the Monte Carlo mesh tallies of Equation (3.3), normalized per unit source strength and per unit time. If the neutron source strength and the tissue boron concentration $C_B(x)$ are both treated as constant over the course of the irradiation, an approximation to the true time-dependent pharmacokinetics of the boron carrier discussed in Section 1.2.2, then the weighted dose accumulated at x after an irradiation time t is simply

$$D_w(x,t) = t \cdot \dot{D}_w(x), \tag{C.8}$$

i.e., dose accumulates linearly with time at every voxel, with a spatially varying but time-independent proportionality constant $\dot{D}_w(x)$. This is the same constant-dose-rate assumption already used in Equation (3.1) of Section 3.2.5 to write the absorbed dose as $D_{\text{phys}}(x) = t\,[\cdots]$, and carries the same limitation: it neglects any variation of $C_B(x)$, and hence of the boron-dose component specifically, over the duration of a single irradiation fraction. The approximation is standard in BNCT treatment planning and is reasonable provided the irradiation time is short compared with the characteristic timescale of boron washout from blood and tissue (Section 1.2.2), but it is nonetheless an approximation, not an exact statement, and should be borne in mind when interpreting the resulting irradiation times.

C.2.2 From Dose-Volume Constraint to Maximum Irradiation Time

A dose-volume constraint on an OAR specifies that the dose received by a given percentile of the structure's volume must not exceed a prescribed limit. Denote by $D_{q\%}(t)$ the dose received by $q\%$ of the OAR's volume after an irradiation time t, i.e., the value read off the cumulative DVH of the structure (Section 3.1.2) at the $q\%$-volume point, and let $D^{\text{lim}}_{q\%}$ be the corresponding prescribed dose limit (e.g., $D^{\text{lim}}_{2\%} = 13$ Gy-w or $D^{\text{lim}}_{50\%} = 2.5$ Gy-w for the brain, as used in Section 3.3.7). The constraint requires

$$D_{q\%}(t) \leq D^{\text{lim}}_{q\%}. \tag{C.9}$$

Because Equation (C.8) states that dose accumulates linearly in time at *every* voxel with the same proportionality constant $\dot{D}_w(x)$ at that voxel, the entire spatial dose distribution at time t is simply the dose-*rate* distribution scaled by t. The cumulative DVH is a monotonic transform of the spatial dose distribution (it counts, for each dose value d, the fractional structure volume receiving at least d); scaling every voxel's dose by the same time-independent factor t therefore scales the entire DVH curve, and in particular its $q\%$-volume point, by that same factor:

$$D_{q\%}(t) = t \cdot \dot{D}_{q\%}, \tag{C.10}$$

where $\dot{D}_{q\%}$ is the $q\%$-volume point of the *dose-rate* DVH, computed once directly from the static Monte Carlo dose-rate map, independently of t. Substituting Equation (C.10) into the constraint of Equation (C.9) and solving for the largest t that still satisfies the constraint with equality gives the maximum irradiation time *permitted by that single constraint*:

$$t_{q\%} = \frac{D_{q\%}^{\lim}}{\dot{D}_{q\%}}. \tag{C.11}$$

C.2.3 Multiple Simultaneous Constraints

In general, more than one dose-volume constraint may be prescribed simultaneously for the same OAR (for instance, both a $D_{2\%}$ and a $D_{50\%}$ limit on the brain, as in Section 3.3.7), or constraints may be prescribed on more than one OAR at once. Each such constraint i, with limit $D_i^{\lim}$ evaluated at its own dose-rate percentile $\dot{D}_i$, defines its own candidate maximum time $t_i = D_i^{\lim}/\dot{D}_i$ via Equation (C.11). Because every constraint must be satisfied simultaneously, and because dose increases monotonically with t at every point (Equation (C.8)), the actual maximum permissible irradiation time is set by whichever constraint is reached *first* as t increases from zero, i.e., the smallest of the candidate times:

$$t = \min_i (t_i) = \min_i \left(\frac{D_i^{\lim}}{\dot{D}_i} \right). \tag{C.12}$$

For the specific two-constraint case of Section 3.3.7, with $i \in \{D_{2\%}, D_{50\%}\}$ evaluated on the brain, Equation (C.12) reduces exactly to Equation (3.4):

$$t_{D_{2\%}} = \frac{13 \text{ Gy-w}}{\dot{D}_{2\%,\text{brain}}}, \qquad t_{D_{50\%}} = \frac{2.5 \text{ Gy-w}}{\dot{D}_{50\%,\text{brain}}}, \qquad t_{\text{brain}} = \min\left(t_{D_{2\%}}, t_{D_{50\%}}\right). \tag{C.13}$$

The empirical finding reported in Section 3.3.7, namely that the $D_{2\%}$ constraint was found to be the more restrictive of the two in every evaluated configuration, is equivalent to observing that $t_{D_{2\%}} < t_{D_{50\%}}$ held throughout the benchmark; Equation (C.12) makes explicit that this is an empirical property of the specific dose-rate distribution obtained for that beam and geometry, not a general consequence of the constraint formalism itself, which would in principle allow either constraint to be the more restrictive one for a sufficiently different dose distribution.

Appendix D

D.1 Per-Patient Data

Every pooled statistic in this chapter is a summary of 24 individual patients whose behavior, as shown throughout, can differ substantially from the pooled mean. This appendix tabulates the complete per-patient results underlying those statistics, organized by topic, so that any individual case can be inspected directly rather than only through cohort-level aggregates. All tables are sorted by Dice coefficient (DC), consistent with the ordering used throughout the main chapter.

Table D.1: Geometric indices and GTV voxel counts for all four configurations, 24 patients, sorted by DC.

Patient	DC	GMI	DI	$\|G\|$ (manual, vox.)	$\|S\|$ (AI, vox.)	$\|G \cap S\|$ (vox.)
HN_1019	0.000	1.000	1.000	131	762	0
HN_0149	0.197	0.889	0.140	5432	702	604
HN_0591	0.374	0.606	0.645	2944	3262	1159
HN_0580	0.384	0.744	0.232	5436	1813	1392
HN_0806	0.401	0.749	0.005	10091	2548	2536
HN_0569	0.457	0.664	0.288	5940	2805	1998
HN_0689	0.473	0.075	0.683	402	1172	372
HN_0109	0.497	0.655	0.111	13716	5329	4736
HN_0150	0.499	0.661	0.057	18049	6493	6126
HN_1640	0.560	0.610	0.002	16416	6407	6396
HN_1615	0.573	0.424	0.429	3396	3422	1955
HN_1583	0.579	0.471	0.362	6197	5141	3281
HN_1555	0.670	0.469	0.092	9483	5540	5031
HN_0792	0.691	0.380	0.221	9436	7513	5853
HN_1581	0.691	0.135	0.425	2674	4019	2312
HN_0621	0.702	0.457	0.010	16390	8992	8905
HN_1588	0.742	0.118	0.360	7504	10343	6617
HN_1631	0.754	0.353	0.096	11125	7963	7196
HN_0847	0.762	0.291	0.177	7813	6734	5539
HN_0601	0.780	0.110	0.306	6629	8496	5898
HN_0687	0.838	0.233	0.076	9454	7855	7255
HN_0666	0.851	0.211	0.076	8866	7567	6995
HN_1551	0.868	0.162	0.101	3079	2869	2580
HN_0717	0.901	0.055	0.139	9632	10568	9102

Table D.2: TCP, four configurations, and normalized |ΔTCP %| vs. manual for AI, swap and anti-swap, 24 patients, sorted by DC.

Patient	DC	Manual	AI	Swap	Anti-swap	|Δ%| AI	|Δ%| swap	|Δ%| anti-swap
HN_1019	0.000	0.3986	0.2800	0.2276	0.2536	29.8	42.9	36.4
HN_0149	0.197	0.2180	0.3600	0.4336	0.1748	65.2	98.9	19.8
HN_0591	0.374	0.2466	0.1162	0.1176	0.2137	52.9	52.3	13.3
HN_0580	0.384	0.1431	0.1848	0.1687	0.1376	29.1	17.8	3.9
HN_0806	0.401	0.0765	0.1008	0.0987	0.0782	31.9	29.0	2.3
HN_0569	0.457	0.3563	0.2157	0.1915	0.3378	39.5	46.3	5.2
HN_0689	0.473	0.3663	0.2447	0.2345	0.3759	33.2	36.0	2.6
HN_0109	0.497	0.1328	0.1315	0.1656	0.0806	1.0	24.7	39.3
HN_0150	0.499	0.2025	0.3655	0.3329	0.2358	80.5	64.4	16.5
HN_1640	0.560	0.1537	0.2519	0.2219	0.1820	63.8	44.3	18.4
HN_1615	0.573	0.1790	0.2655	0.2250	0.1640	48.3	25.7	8.4
HN_1583	0.579	0.2444	0.2260	0.2386	0.2149	7.5	2.4	12.1
HN_1555	0.670	0.1272	0.1793	0.1754	0.1270	41.0	37.9	0.1
HN_0792	0.691	0.0908	0.0852	0.0827	0.0922	6.2	8.9	1.4
HN_1581	0.691	0.2545	0.2344	0.2390	0.2296	7.9	6.1	9.8
HN_0621	0.702	0.1655	0.2029	0.2210	0.1531	22.6	33.6	7.5
HN_1588	0.742	0.2149	0.2060	0.2081	0.1775	4.2	3.1	17.4
HN_1631	0.754	0.1135	0.1539	0.1483	0.1143	35.5	30.6	0.6
HN_0847	0.762	0.1268	0.1615	0.1587	0.1263	27.4	25.1	0.4
HN_0601	0.780	0.1481	0.1294	0.1366	0.1407	12.6	7.8	5.0
HN_0687	0.838	0.1433	0.1652	0.1572	0.1461	15.3	9.7	2.0
HN_0666	0.851	0.2438	0.2379	0.2582	0.2431	2.4	5.9	0.3
HN_1551	0.868	0.2575	0.2931	0.2772	0.2610	13.8	7.6	1.3
HN_0717	0.901	0.1586	0.1577	0.1447	0.1708	0.6	8.8	7.7

Table D.3: Mean GTV dose, RBE-weighted versus photon-isoeffective model, manual and AI configurations, 24 patients, sorted by DC.

		Manual [Gy]			AI [Gy]		
Patient	DC	RBE	IsoE	%Δ	RBE	IsoE	%Δ
HN_1019	0.000	16.06	18.09	−11.21	15.29	17.21	−11.14
HN_0149	0.197	20.70	20.62	+0.37	24.15	22.95	+5.24
HN_0591	0.374	20.59	20.50	+0.44	19.25	19.29	−0.23
HN_0580	0.384	17.86	18.34	−2.62	22.84	21.51	+6.20
HN_0806	0.401	16.99	17.73	−4.19	15.85	16.55	−4.18
HN_0569	0.457	23.93	22.85	+4.75	20.17	20.33	−0.81
HN_0689	0.473	19.13	19.84	−3.60	16.59	18.09	−8.27
HN_0109	0.497	17.92	18.91	−5.23	13.06	15.40	−15.19
HN_0150	0.499	22.82	22.05	+3.48	27.88	24.88	+12.03
HN_1640	0.560	15.88	17.68	−10.15	18.13	19.27	−5.92
HN_1615	0.573	14.49	16.55	−12.45	20.72	20.81	−0.47
HN_1583	0.579	20.82	20.81	+0.02	19.81	20.20	−1.95
HN_1555	0.670	13.72	15.93	−13.87	15.35	17.18	−10.69
HN_0792	0.691	16.28	17.48	−6.88	16.45	17.49	−5.91
HN_1581	0.691	16.99	18.45	−7.93	17.35	18.65	−6.97
HN_0621	0.702	19.95	20.08	−0.65	19.30	19.77	−2.34
HN_1588	0.742	23.68	22.38	+5.82	24.88	23.04	+8.01
HN_1631	0.754	14.93	16.65	−10.33	16.90	18.09	−6.56
HN_0847	0.762	19.20	19.59	−1.95	18.84	19.55	−3.61
HN_0601	0.780	21.61	20.57	+5.07	20.14	19.71	+2.19
HN_0687	0.838	15.31	17.07	−10.30	16.72	18.04	−7.33
HN_0666	0.851	26.25	23.92	+9.74	24.23	22.80	+6.26
HN_1551	0.868	19.78	20.23	−2.24	22.01	21.62	+1.82
HN_0717	0.901	17.26	18.38	−6.13	18.02	18.85	−4.37

Table D.4: Absolute GTV dose under the photon-isoeffective model, minimum/mean/maximum, manual and AI configurations, 24 patients, sorted by DC.

Patient	DC	Manual			AI		
		$D_{\min}$ [Gy]	D_{mean} [Gy]	$D_{\max}$ [Gy]	$D_{\min}$ [Gy]	D_{mean} [Gy]	$D_{\max}$ [Gy]
HN_1019	0.000	16.12	18.09	21.33	12.69	17.21	24.12
HN_0149	0.197	10.72	20.62	30.26	13.29	22.95	28.11
HN_0591	0.374	12.64	20.50	28.11	5.14	19.29	26.97
HN_0580	0.384	8.48	18.34	29.82	4.50	21.51	30.80
HN_0806	0.401	4.25	17.73	28.59	5.11	16.55	28.76
HN_0569	0.457	15.84	22.85	28.12	10.99	20.33	26.68
HN_0689	0.473	13.63	19.84	25.29	9.67	18.09	25.75
HN_0109	0.497	5.81	18.91	26.43	7.28	15.40	23.45
HN_0150	0.499	10.03	22.05	29.16	15.70	24.88	31.67
HN_1640	0.560	8.90	17.68	24.30	14.01	19.27	24.15
HN_1615	0.573	9.84	16.55	23.58	11.26	20.81	27.87
HN_1583	0.579	12.50	20.81	28.87	10.96	20.20	27.04
HN_1555	0.670	8.27	15.93	24.25	10.12	17.18	24.99
HN_0792	0.691	3.18	17.48	27.20	3.37	17.49	28.12
HN_1581	0.691	11.55	18.45	24.85	11.75	18.65	25.17
HN_0621	0.702	10.07	20.08	30.02	12.01	19.77	29.12
HN_1588	0.742	10.73	22.38	29.86	10.93	23.04	31.73
HN_1631	0.754	6.89	16.65	26.54	8.83	18.09	26.47
HN_0847	0.762	5.90	19.59	28.00	7.44	19.55	26.26
HN_0601	0.780	7.84	20.57	33.39	7.34	19.71	31.90
HN_0687	0.838	9.15	17.07	24.86	9.38	18.04	27.28
HN_0666	0.851	11.05	23.92	31.32	11.24	22.80	29.95
HN_1551	0.868	12.00	20.23	26.60	13.38	21.62	28.54
HN_0717	0.901	8.78	18.38	27.61	9.94	18.85	29.01

Table D.5: Left: gamma-index overlap pass-rate [%] (true gamma, 3%/3 mm), three comparisons, 24 patients. Right: minimum-dose ratio (AI/manual).

Patient	M. vs. AI	M. vs. swap	M. vs. A-swap
HN_1019	–	–	2.3
HN_0149	57.3	100.0	51.3
HN_0591	60.8	100.0	69.1
HN_0580	84.3	100.0	89.5
HN_0806	90.7	100.0	90.5
HN_0569	81.5	100.0	74.3
HN_0689	91.1	100.0	91.3
HN_0109	54.6	100.0	32.2
HN_0150	60.3	100.0	54.3
HN_1640	76.5	100.0	75.1
HN_1615	50.0	100.0	63.4
HN_1583	83.4	100.0	74.5
HN_1555	89.2	100.0	90.0
HN_0792	82.4	100.0	73.5
HN_1581	89.8	100.0	85.8
HN_0621	88.8	100.0	88.5
HN_1588	91.3	100.0	57.6
HN_1631	89.6	100.0	90.0
HN_0847	75.2	100.0	78.0
HN_0601	91.6	100.0	91.8
HN_0687	88.7	100.0	92.3
HN_0666	85.9	100.0	100.0
HN_1551	79.8	100.0	88.5
HN_0717	86.6	100.0	86.8

Patient	$D_{\min}$ ratio
HN_1019	0.787
HN_0149	1.240
HN_0591	0.407
HN_0580	0.531
HN_0806	1.205
HN_0569	0.694
HN_0689	0.709
HN_0109	1.254
HN_0150	1.565
HN_1640	1.573
HN_1615	1.144
HN_1583	0.877
HN_1555	1.224
HN_0792	1.058
HN_1581	1.017
HN_0621	1.192
HN_1588	1.019
HN_1631	1.281
HN_0847	1.261
HN_0601	0.937
HN_0687	1.025
HN_0666	1.017
HN_1551	1.115
HN_0717	1.132

Table D.6: Mucosa isoeffective dose and absorbed dose NTCP, four configurations, 24 patients, sorted by DC.

Patient	Mucosa isoE dose [Gy]				NTCP absorbed			
	Manual	AI	Swap	Anti-swap	Manual	AI	Swap	Anti-swap
HN_1019	10.453	9.057	10.398	9.057	5.6×10^{-4}	2.9×10^{-4}	5.1×10^{-4}	2.9×10^{-4}
HN_0149	9.100	8.698	9.100	8.355	6.9×10^{-5}	1.5×10^{-4}	6.9×10^{-5}	1.5×10^{-4}
HN_0591	8.208	8.656	8.199	8.357	1.0×10^{-4}	1.2×10^{-4}	1.0×10^{-4}	1.2×10^{-4}
HN_0580	6.463	7.110	6.416	6.811	2.9×10^{-5}	3.5×10^{-5}	2.7×10^{-5}	3.4×10^{-5}
HN_0806	8.281	9.346	8.281	9.220	2.0×10^{-5}	5.4×10^{-5}	2.0×10^{-5}	5.3×10^{-5}
HN_0569	7.243	7.918	7.221	7.918	3.5×10^{-5}	5.1×10^{-5}	3.4×10^{-5}	5.1×10^{-5}
HN_0689	9.471	9.584	9.434	9.584	1.2×10^{-4}	2.9×10^{-4}	1.2×10^{-4}	2.9×10^{-4}
HN_0109	7.277	8.748	7.276	8.670	1.3×10^{-5}	4.0×10^{-5}	1.3×10^{-5}	4.0×10^{-5}
HN_0150	7.575	7.658	7.562	7.652	3.0×10^{-5}	3.5×10^{-5}	2.9×10^{-5}	3.5×10^{-5}
HN_1640	9.598	8.780	9.598	8.477	2.9×10^{-5}	1.0×10^{-4}	2.9×10^{-5}	9.6×10^{-5}
HN_1615	8.906	8.469	8.872	8.427	1.6×10^{-4}	9.5×10^{-5}	1.6×10^{-4}	9.3×10^{-5}
HN_1583	8.379	8.305	8.379	8.181	5.5×10^{-5}	7.7×10^{-5}	5.5×10^{-5}	7.6×10^{-5}
HN_1555	8.004	8.456	7.985	8.386	4.0×10^{-5}	6.4×10^{-5}	3.9×10^{-5}	6.6×10^{-5}
HN_0792	7.622	7.639	7.620	7.523	1.1×10^{-4}	6.6×10^{-5}	1.1×10^{-4}	6.9×10^{-5}
HN_1581	7.859	7.980	7.817	7.980	6.6×10^{-5}	7.7×10^{-5}	6.4×10^{-5}	7.7×10^{-5}
HN_0621	7.773	8.054	7.773	7.834	6.1×10^{-5}	9.3×10^{-5}	6.1×10^{-5}	9.0×10^{-5}
HN_1588	8.019	7.200	7.966	7.199	9.3×10^{-5}	6.4×10^{-5}	9.2×10^{-5}	6.4×10^{-5}
HN_1631	7.003	7.151	6.996	7.121	4.0×10^{-5}	6.6×10^{-5}	4.0×10^{-5}	6.4×10^{-5}
HN_0847	9.099	8.714	9.089	8.654	1.1×10^{-4}	7.7×10^{-5}	1.1×10^{-4}	7.6×10^{-5}
HN_0601	7.142	6.920	7.127	6.911	1.6×10^{-4}	1.2×10^{-4}	1.6×10^{-4}	1.2×10^{-4}
HN_0687	8.179	8.261	8.152	8.235	1.2×10^{-4}	8.5×10^{-5}	1.2×10^{-4}	8.6×10^{-5}
HN_0666	7.722	7.344	7.722	7.718	1.5×10^{-4}	1.1×10^{-4}	1.5×10^{-4}	5.5×10^{-5}
HN_1551	8.653	8.682	8.653	8.682	5.9×10^{-5}	5.5×10^{-5}	5.9×10^{-5}	5.5×10^{-5}
HN_0717	7.604	7.755	7.575	7.755	3.4×10^{-5}	1.2×10^{-4}	3.1×10^{-5}	1.2×10^{-4}

Bibliography

[1] WHO Classification of Tumours Editorial Board, *Head and Neck Tumours*, WHO Classification of Tumours, 5th ed., vol. 9, International Agency for Research on Cancer (IARC), Lyon, France, 2024. ISBN 978-92-832-4514-8.

[2] F. Bray, J. Ferlay, I. Soerjomataram, R. L. Siegel, L. A. Torre, and A. Jemal, Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries, *CA: A Cancer Journal for Clinicians*, vol. 68, no. 6, pp. 394–424, 2018.

[3] S. Takeno, Y. Yoshino, T. Aihara, M. Higashino, Y. Kanai, N. Hu, R. Kakino, R. Kawata, K. Nihei, and K. Ono, Preliminary outcomes of boron neutron capture therapy for head and neck cancers as a treatment covered by public health insurance system in Japan: real-world experiences over a 2-year period, *Cancer Medicine*, vol. 13, no. 11, p. e7250, 2024.

[4] R. F. Barth, M. G. H. Vicente, O. K. Harling, W. S. Kiger, K. J. Riley, P. J. Binns, F. M. Wagner, M. Suzuki, T. Aihara, I. Kato, and S. Kawabata, Current status of boron neutron capture therapy of high grade gliomas and recurrent head and neck cancer, *Radiation Oncology*, vol. 7, p. 146, 2012.

[5] T. D. Malouff, D. S. Seneviratne, D. K. Ebner, W. C. Stross, M. R. Waddle, D. M. Trifiletti, and S. Krishnan, Boron neutron capture therapy: a review of clinical applications, *Frontiers in Oncology*, vol. 11, p. 601820, 2021.

[6] C. L. Brouwer, R. J. Steenbakkers, J. Bourhis, W. Budach, C. Grau, V. Grégoire, et al., CT-based delineation of organs at risk in the head and neck region: DAHANCA, EORTC, GORTEC, HKNPCSG, NCIC CTG, NCRI, NRG Oncology and TROG consensus guidelines, *Radiotherapy and Oncology*, vol. 117, no. 1, pp. 83–90, 2015.

[7] O. Ronneberger, P. Fischer, and T. Brox, U-Net: Convolutional Networks for Biomedical Image Segmentation, in *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015*, LNCS vol. 9351, pp. 234–241, 2015.

[8] F. Isensee, P. F. Jaeger, S. A. A. Kohl, J. Petersen, and K. H. Maier-Hein, nnU-Net: a self-configuring method for deep learning-based biomedical image segmentation, *Nature Methods*, vol. 18, pp. 203–211, 2021.

[9] S. Nikolov, S. Blackwell, A. Zverovitch, R. Mendes, M. Livne, J. De Fauw, et al., Clinically applicable segmentation of head and neck anatomy for radiotherapy: deep learning algorithm development and validation study, *Journal of Medical Internet Research*, vol. 23, no. 7, p. e26151, 2021.

[10] C. Pezzi, S. Bortolussi, I. Postuma, B. Marcaccio, R. L. Ramos, V. Vercesi, F. Morosato, L. Bagnale, and S. Fatemi. "Artificial intelligence in BNCT: a dosimetric assessment of treatment planning using human- or AI-segmented medical images". In: *Physica Medica* (2026).

[11] K. Clark, B. Vendt, K. Smith, J. Freymann, J. Kirby, P. Koppel, S. Moore, S. Phillips, D. Maffitt, M. Pringle, L. Tarbox, and F. Prior, The Cancer Imaging Archive (TCIA): maintaining and operating a public information repository, *Journal of Digital Imaging*, vol. 26, no. 6, pp. 1045–1057, 2013.

[12] I. Postuma, S. J. González, S. Fatemi, C. Pezzi, C. Ruzzon, O. Nicrosini, V. Vercesi, and S. Bortolussi, OpenPINT: Open-source Planning for Isoeffective Nuclear Treatments in BNCT research, *arXiv preprint* arXiv:2606.21476, 2026. Preprint manuscript, not yet peer-reviewed at the time of writing.

[13] S. J. González and G. A. Santa Cruz, The photon-isoeffective dose in boron neutron capture therapy, *Radiation Research*, vol. 178, no. 6, pp. 609–621, 2012.

[14] S. J. González, E. C. C. Pozzi, A. Monti Hughes, L. Provenzano, H. Koivunoro, D. G. Carando, S. I. Thorp, M. Casal, S. Bortolussi, V. A. Trivillin, 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, *Physics in Medicine & Biology*, vol. 62, no. 20, p. 7938, 2017.

[15] G. L. Locher, Biological effects and therapeutic possibilities of neutrons, *American Journal of Roentgenology and Radium Therapy*, vol. 36, pp. 1–13, 1936.

[16] W. H. Sweet, Early history of development of boron neutron capture therapy of tumors, *Journal of Neuro-Oncology*, vol. 33, pp. 19–26, 1997.

[17] W. H. Sweet, The uses of nuclear disintegration in the diagnosis and treatment of brain tumor, *New England Journal of Medicine*, vol. 245, no. 23, pp. 875–878, 1951.

[18] International Atomic Energy Agency, *Advances in Boron Neutron Capture Therapy*, IAEA CRCP/BOR/002 (IAEAL 23-01601), IAEA, Vienna, 2023. ISBN 978-92-0-132623-2.

[19] J. A. Coderre, J. C. Turcotte, K. J. Riley, P. J. Binns, O. K. Harling, and W. S. Kiger III, Boron neutron capture therapy: Cellular targeting of high linear energy transfer radiation, *Technology in Cancer Research & Treatment*, vol. 2, no. 5, pp. 355–375, 2003.

[20] K. Hirose, A. Konno, J. Hiratsuka, S. Yoshimoto, T. Kato, K. Ono, N. Otsuki, J. Hatazawa, H. Tanaka, K. Takayama, H. Wada, M. Suzuki, M. Sato, H. Yamaguchi, I. Seto, Y. Ueki, S. Iketani, S. Imai, T. Nakamura, T. Ono, H. Endo, Y. Azami, Y. Kikuchi, M. Murakami, and Y. Takai, Boron neutron capture therapy using cyclotron-based epithermal neutron source and borofalan (^{10}B) for recurrent or locally advanced head and neck cancer (JHN002): an open-label phase II trial, *Radiotherapy and Oncology*, vol. 155, pp. 182–187, 2021.

[21] H. Kanno, H. Nagata, A. Ishiguro, et al., Designation products: boron neutron capture therapy for head and neck carcinoma, *The Oncologist*, vol. 26, no. 7, pp. e1250–e1255, 2021.

[22] W. H. Jin, C. Seldon, M. Butkus, W. Sauerwein, and H. B. Giap, A Review of Boron Neutron Capture Therapy: Its History and Current Challenges, *International Journal of Particle Therapy*, vol. 9, no. 1, pp. 71–82, 2022.

[23] L. Kankaanranta, T. Seppälä, H. Koivunoro, K. Saarilahti, T. Atula, J. Collan, E. Salli, M. Kortesniemi, J. Uusi-Simola, P. Välimäki, A. Mäkitie,

M. Seppänen, H. Minn, H. Revitzer, M. Kouri, P. Kotiluoto, T. Serén, I. Auterinen, S. Savolainen, and H. Joensuu, Boron neutron capture therapy in the treatment of locally recurred head-and-neck cancer: final analysis of a phase I/II trial, *International Journal of Radiation Oncology, Biology, Physics*, vol. 82, no. 1, pp. e67–e75, 2012.

[24] Y. Sun, Boron neutron capture therapy: moving towards targeted therapy for locally recurrent head and neck squamous cell carcinoma, *Military Medical Research*, vol. 6, p. 32, 2019.

[25] X. Sun, Y. Liu, Z. Xue, B. Cheng, R. Zhang, L. Li, W. Huang, Q. Song, and Y. Cheng, Efficacy and safety of boron neutron capture therapy for locally recurrent head and neck cancer, *Critical Reviews in Oncology/Hematology*, vol. 218, p. 105101, 2026.

[26] J. A. Coderre and G. M. Morris, The radiation biology of boron neutron capture therapy, *Radiation Research*, vol. 151, no. 1, pp. 1–18, 1999.

[27] H. Kumada and K. Takada, Treatment planning system and patient positioning for boron neutron capture therapy, *Therapeutic Radiology and Oncology*, vol. 2, p. 50, 2018.

[28] International Atomic Energy Agency, *Current Status of Neutron Capture Therapy*, IAEA-TECDOC-1223, IAEA, Vienna, 2001.

[29] L. Provenzano, H. Koivunoro, I. Postuma, J. M. Longhino, E. F. Boggio, R. O. Farías, S. Bortolussi, and S. J. González, The essential role of radiobiological figures of merit for the assessment and comparison of beam performances in boron neutron capture therapy, *Physica Medica*, vol. 67, pp. 9–19, 2019.

[30] I. Postuma, S. González, M. S. Herrera, L. Provenzano, M. Ferrarini, C. Magni, N. Protti, S. Fatemi, V. Vercesi, G. Battistoni, U. Anselmi Tamburini, Y. H. Liu, L. Kankaanranta, H. Koivunoro, S. Altieri, and S. Bortolussi, A Novel Approach to Design and Evaluate BNCT Neutron Beams Combining Physical, Radiobiological, and Dosimetric Figures of Merit, *Biology*, vol. 10, no. 3, p. 174, 2021.

[31] F. Ballarini, J. G. Bakeine, S. Bortolussi, P. Bruschi, A. M. Clerici, A. De Bari, P. Dionigi, C. Ferrari, M. A. Gadan, N. Protti, S. Stella, C. Zonta, A. Zonta, and

S. Altieri, Nuclear Physics meets Medicine and Biology: Boron Neutron Capture Therapy, *Journal of Physics: Conference Series*, vol. 41, pp. 561–571, 2009.

[32] M. A. Dymova, S. Y. Taskaev, V. A. Richter, and E. V. Kuligina, Boron neutron capture therapy: Current status and future perspectives, *Cancer Communications*, vol. 40, pp. 406–421, 2020.

[33] R. F. Barth, J. A. Coderre, M. G. H. Vicente, and T. E. Blue, Boron neutron capture therapy of cancer: Current status and future prospects, *Clinical Cancer Research*, vol. 11, no. 11, pp. 3987–4002, 2005.

[34] C. Hu, J. Chen, W. Zhong, J.-J. Pan, Y.-H. Liu, Y.-C. Teng, and X.-B. Xia, Application of a proven treatment planning system for accelerator-based boron neutron capture therapy, *Applied Radiation and Isotopes*, vol. 226, 2025.

[35] H. Kumada, T. Sakae, and H. Sakurai, Current development status of accelerator-based neutron source for boron neutron capture therapy, *EPJ Techniques and Instrumentation*, vol. 10, p. 18, 2023.

[36] J. Kopecky, *Atlas of Neutron Capture Cross Sections*, International Atomic Energy Agency, Vienna, Austria, 1997.

[37] Japan Atomic Energy Agency, Nuclear Data Center, ENDF/JENDL Graph, `https://wwwndc.jaea.go.jp/ENDF_Graph/`, 2020.

[38] R. F. Barth, P. Mi, and W. Yang, Boron delivery agents for neutron capture therapy of cancer, *Cancer Communications*, vol. 38, p. 35, 2018.

[39] L. W. Wang, Y. W. Chen, C. Y. Ho, Y. W. Hsueh Liu, F. I. Chou, Y. H. Liu, H. M. Liu, J. J. Peir, S. H. Jiang, C. W. Chang, C. S. Liu, K. H. Lin, S. J. Wang, P. Y. Chu, W. L. Lo, S. Y. Kao, and S. H. Yen, Fractionated boron neutron capture therapy in locally recurrent head and neck cancer: a prospective phase I/II trial, *International Journal of Radiation Oncology, Biology, Physics*, vol. 95, pp. 396–403, 2016.

[40] R. F. Barth, N. Gupta, and S. Kawabata, Evaluation of sodium borocaptate (BSH) and boronophenylalanine (BPA) as boron delivery agents for neutron capture therapy (NCT) of cancer: an update and a guide for the future clinical eval-

uation of new boron delivery agents for NCT, *Cancer Communications*, vol. 44, pp. 893–909, 2024.

[41] R. O. Farias, et al., Boron carriers for boron neutron capture therapy: an updated review, *Cells*, vol. 13, p. 2065, 2024.

[42] T. Zhou, et al., The current status and novel advances of boron neutron capture therapy, *International Journal of Medical Sciences*, vol. 21, no. 5, pp. 794–808, 2024.

[43] H. Koivunoro, L. Kankaanranta, T. Seppälä, A. Haapaniemi, A. Mäkitie, and H. Joensuu, Boron neutron capture therapy for locally recurrent head and neck squamous cell carcinoma: an analysis of dose response and survival, *Radiotherapy and Oncology*, vol. 137, pp. 153–158, 2019.

[44] W. S. Kiger, M. R. Palmer, K. J. Riley, R. G. Zamenhof, and P. M. Busse, A pharmacokinetic model for the concentration of boron-10 in blood after boronophenylalanine-fructose administration in humans, *Radiation Research*, vol. 155, no. 4, pp. 611–618, 2001.

[45] Y.-C. Teng, Y. Cao, X. Zhu, L. Huo, J. Hong, X. Lin, Y. Lai, P. Zhou, Y. Bian, J. Chen, W. Zhong, J. Ping, Q. Lin, J. Pan, and Y.-H. Liu, PET-Guided Modeling of Mucosal Boron Heterogeneity Improves Dose-Toxicity Prediction in BNCT, *International Journal of Radiation Oncology, Biology, Physics*, 2026, in press.

[46] T. Morita, et al., Dynamic changes in ^{18}F-borono-L-phenylalanine uptake in unresectable, advanced, or recurrent squamous cell carcinoma of the head and neck and malignant melanoma during boron neutron capture therapy patient selection, *Radiation Oncology*, vol. 13, p. 4, 2018.

[47] J. A. Coderre, A. D. Chanana, D. D. Joel, et al., Biodistribution of boronophenylalanine in patients with glioblastoma multiforme: boron concentration correlates with tumor cellularity, *Radiation Research*, vol. 149, no. 2, pp. 163–170, 1998.

[48] O. K. Harling and K. J. Riley, Fission reactor neutron sources for neutron capture therapy – a critical review, *Journal of Neuro-Oncology*, vol. 62, pp. 7–17, 2003.

[49] Y. Bae, D.-S. Kim, H. J. Seo, J.-U. Han, H. J. Yoon, J. J. Hwang, et al., Advances of LINAC-based boron neutron capture therapy in Korea, *AAPPS Bulletin*, vol. 32, p. 34, 2022.

[50] M. Suzuki, Boron neutron capture therapy (BNCT): a unique role in radiotherapy with a view to entering the accelerator-based BNCT era, *International Journal of Clinical Oncology*, vol. 25, pp. 43–50, 2020.

[51] P. J. Binns, K. J. Riley, Y. Ostrovsky, W. Gao, J. R. Albritton, W. S. Kiger, and O. K. Harling, Improved dose targeting for a clinical epithermal neutron capture beam using optional lithium-6 filtration, *International Journal of Radiation Oncology, Biology, Physics*, vol. 67, no. 5, pp. 1484–1491, 2007.

[52] International Commission on Radiation Units and Measurements, *Tissue Substitutes in Radiation Dosimetry and Measurement*, ICRU Report 44, *Journal of the ICRU*, ICRU, Bethesda, MD, 1989.

[53] D. A. Brown, M. B. Chadwick, R. Capote, A. C. Kahler, A. Trkov, M. W. Herman, A. A. Sonzogni, Y. Danon, et al., ENDF/B-VIII.0: The 8th Major Release of the Nuclear Reaction Data Library with CIELO-project Cross Sections, New Standards and Thermal Scattering Data, *Nuclear Data Sheets*, vol. 148, pp. 1–142, 2018.

[54] W. A. G. Sauerwein, A. Wittig, R. Moss, and Y. Nakagawa (Eds.), *Neutron Capture Therapy: Principles and Applications*, Springer Science & Business Media, 2012. Chap. 16.

[55] A. M. Kellerer and H. H. Rossi, The theory of dual radiation action, *Current Topics in Radiation Research Quarterly*, vol. 8, pp. 85–158, 1972.

[56] A. M. Kellerer and H. H. Rossi, A generalized formulation of dual radiation action, *Radiation Research*, vol. 75, pp. 471–488, 1978.

[57] E. L. Alpen, *Radiation Biophysics*, 2nd ed., Academic Press, San Diego, CA, 1998.

[58] E. J. Hall and A. J. Giaccia, *Radiobiology for the Radiologist*, 8th ed., Wolters Kluwer, Philadelphia, PA, 2019.

[59] D. E. Lea and D. G. Catcheside, The mechanism of the induction by radiation of chromosome aberrations in *Tradescantia*, *Journal of Genetics*, vol. 44, pp. 216–245, 1942.

[60] D. J. Brenner, L. R. Hlatky, P. J. Hahnfeldt, Y. Huang, and R. K. Sachs, The linear-quadratic model and most other common radiobiological models result in similar predictions of time-dose relationships, *Radiation Research*, vol. 150, pp. 83–91, 1998.

[61] M. Zaider and H. H. Rossi, The synergistic effects of different radiations, *Radiation Research*, vol. 83, pp. 732–739, 1980.

[62] R. Railton, R. C. Lawson, and D. Porter, Interaction of gamma-ray and neutron effects on the proliferative capacity of Chinese hamster cells, *International Journal of Radiation Biology and Related Studies in Physics, Chemistry, and Medicine*, pp. 75–82, 1975.

[63] R. E. Durand and P. L. Olive, Irradiation of multi-cell spheroids with fast neutrons versus X-rays: a qualitative difference in sub-lethal damage repair capacity or kinetics, *International Journal of Radiation Biology and Related Studies in Physics, Chemistry, and Medicine*, vol. 30, no. 6, pp. 589–592, 1976.

[64] S. Hornsey, U. Andreozzi, and P. R. Warren, Sublethal damage in cells of the mouse gut after mixed treatment with X rays and fast neutrons, *The British Journal of Radiology*, pp. 513–517, 1977.

[65] S. Suzuki, A theoretical model for simultaneous mixed irradiation with multiple types of radiation, *Journal of Radiation Research*, vol. 39, no. 3, pp. 215–221, 1998.

[66] J. A. Coderre, et al., Derivations of relative biological effectiveness for the high-LET radiations produced during boron neutron capture irradiations of the 9L rat gliosarcoma in vitro and in vivo, *International Journal of Radiation Oncology, Biology, Physics*, pp. 1121–1129, 1993.

[67] S. J. González, M. Casal, M. D. Pereira, G. A. Santa Cruz, D. G. Carando, H. Blaumann, M. Bonomi, O. Calzetta Larrieu, D. Feld, C. Fernández, S. Gossio, R. Jiménez Rebagliati, J. Kessler, J. Longhino, P. Menéndez, S. Nievas,

B. M. C. Roth, and S. J. Liberman, Tumor control and normal tissue complications in BNCT treatment of nodular melanoma: a search for predictive quantities, *Applied Radiation and Isotopes*, vol. 67, pp. S153–S156, 2009.

[68] S. J. González, D. G. Carando, et al., Normal tissue analysis and dose-area histograms for BNCT skin dosimetry, *Applied Radiation and Isotopes*, vol. 67, pp. S60–S63, 2009.

[69] I. Kato, K. Ono, Y. Sakurai, M. Ohmae, A. Maruhashi, Y. Imahori, M. Kirihata, M. Nakazawa, and Y. Yura, Effectiveness of BNCT for recurrent head and neck malignancies, *Applied Radiation and Isotopes*, vol. 61, pp. 1069–1073, 2004.

[70] T. Aihara, J. Hiratsuka, N. Morita, M. Uno, Y. Sakurai, A. Maruhashi, K. Ono, and T. Harada, BNCT for advanced or recurrent head and neck cancer, *Applied Radiation and Isotopes*, vol. 88, pp. 12–15, 2014.

[71] T. Aihara, N. Morita, N. Kamitani, H. Kumada, K. Ono, J. Hiratsuka, and T. Harada, Boron neutron capture therapy for advanced salivary gland carcinoma in head and neck, *International Journal of Clinical Oncology*, vol. 19, no. 3, pp. 437–444, 2014.

[72] L. Kankaanranta, T. Seppälä, H. Koivunoro, K. Saarilahti, T. Atula, J. Collan, E. Salli, M. Kortesniemi, J. Uusi-Simola, A. Mäkitie, M. Seppänen, H. Minn, H. Revitzer, M. Kouri, P. Kotiluoto, I. Auterinen, and S. Savolainen, Boron neutron capture therapy in the treatment of locally recurred head and neck cancer, *International Journal of Radiation Oncology, Biology, Physics*, vol. 69, no. 2, pp. 475–482, 2007.

[73] A. Haapaniemi, L. Kankaanranta, R. Saat, H. Koivunoro, K. Saarilahti, A. Mäkitie, T. Atula, and H. Joensuu, Boron neutron capture therapy in the treatment of recurrent laryngeal cancer, *International Journal of Radiation Oncology, Biology, Physics*, vol. 95, no. 1, pp. 404–410, 2016.

[74] L. W. Wang, S. J. Wang, P. Y. Chu, C. Y. Ho, S. H. Jiang, Y. W. Liu, Y. H. Liu, H. M. Liu, J. J. Peir, F. I. Chou, S. H. Yen, and C. S. Liu, BNCT for locally recurrent head and neck cancer: preliminary clinical experience from a phase I/II

trial at Tsing Hua Open-Pool Reactor, *Applied Radiation and Isotopes*, vol. 69, pp. 1803–1806, 2011.

[75] L. W. Wang, Y. H. Liu, F. I. Chou, and S. H. Jiang, Clinical trials for treating recurrent head and neck cancer with boron neutron capture therapy using the Tsing-Hua Open Pool Reactor, *Cancer Communications*, vol. 38, p. 37, 2018.

[76] K. Hirose and M. Sato, Clinical results and prognostic factors in boron neutron capture therapy for recurrent squamous cell carcinoma of the head and neck under the Japan National Health Insurance System: a retrospective study of the initial 47 patients, *International Journal of Radiation Oncology, Biology, Physics*, vol. 120, no. 3, pp. 796–804, 2024.

[77] David Teguh et al. "Clinical Validation of Atlas-Based Auto-Segmentation of Multiple Target Volumes and Normal Tissue (Swallowing/Mastication) Structures in the Head and Neck". In: *International Journal of Radiation Oncology, Biology, Physics* 81 (2010), pp. 950–957.

[78] Jordan Wong et al. "Comparing deep learning-based auto-segmentation of organs at risk and clinical target volumes to expert inter-observer variability in radiotherapy planning". In: *Radiotherapy and Oncology* 144 (2020), pp. 152–158.

[79] Youheng Zhang. "A Historical Review and Philosophical Examination of the two Paradigms in Artificial Intelligence Research". In: *European Journal of Artificial Intelligence and Machine Learning* 2.2 (2023), pp. 24–33. doi: 10.24018/ejai.2023.2.2.23.

[80] F. Rosenblatt. "The Perceptron: A Probabilistic Model for Information Storage and Organization in the Brain". In: *Psychological Review* 65.6 (1958), pp. 386–408.

[81] Kunihiko Fukushima and Sei Miyake. "Neocognitron: A new algorithm for pattern recognition tolerant of deformations and shifts in position". In: *Pattern Recognition* 15.6 (1982), pp. 455–469.

[82] 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 (2019), pp. 15–25.

[83] M. Elgendy. *Deep Learning for Vision Systems*. Manning Publications, 2020. isbn: 9781617296192.

[84] Andrew L. Maas, Awni Y. Hannun, and Andrew Y. Ng. "Rectifier Nonlinearities Improve Neural Network Acoustic Models". In: *Proceedings of the 30th International Conference on Machine Learning (ICML)* (2013).

[85] Vinod Nair and Geoffrey E. Hinton. "Rectified Linear Units Improve Restricted Boltzmann Machines". In: *Proceedings of the 27th International Conference on Machine Learning (ICML)* (2010), pp. 807–814.

[86] Christoph von der Malsburg. "Frank Rosenblatt: Principles of Neurodynamics: Perceptrons and the Theory of Brain Mechanisms". In: *Brain Theory* (Jan. 1986), pp. 245–248. doi: 10.1007/978-3-642-70911-1_20.

[87] Fausto Milletari, Nassir Navab, and Seyed-Ahmad Ahmadi. "V-Net: Fully Convolutional Neural Networks for Volumetric Medical Image Segmentation". In: *Proceedings of the Fourth International Conference on 3D Vision (3DV)* (2016), pp. 565–571.

[88] Sebastian Ruder. "An overview of gradient descent optimization algorithms". In: *arXiv preprint arXiv:1609.04747* (2016).

[89] K. Harrison, H. Pullen, C. Welsh, O. Oktay, J. Alvarez-Valle, and R. Jena. "Machine Learning for Auto-Segmentation in Radiotherapy Planning". In: *Clinical Oncology* 34 (2022), pp. 74–88.

[90] Jonathan Long, Evan Shelhamer, and Trevor Darrell. "Fully Convolutional Networks for Semantic Segmentation". In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (2015), pp. 3431–3440.

[91] Özgün Çiçek et al. "3D U-Net: Learning Dense Volumetric Segmentation from Sparse Annotation". In: *Medical Image Computing and Computer-Assisted Intervention (MICCAI)* (2016), pp. 424–432.

[92] Fabian Isensee et al. "nnU-Net: Self-adapting Framework for U-Net-Based Medical Image Segmentation". In: *arXiv preprint arXiv:1809.10486* (2018).

[93] Wenya Linda Bi et al. “Artificial intelligence in cancer imaging: Clinical challenges and applications”. In: *CA: A Cancer Journal for Clinicians* 69.2 (2019), pp. 127–157.

[94] R. A. Castellino. “Computer aided detection (CAD): an overview”. In: *Cancer Imaging* 5 (2005), pp. 17–19.

[95] Michael Sherer et al. “Metrics to evaluate the performance of auto-segmentation for radiation treatment planning: A critical review”. In: *Radiotherapy and Oncology* 160 (2021), pp. 185–191.

[96] Aditi Iyer et al. “Prospectively-validated deep learning model for segmenting swallowing and chewing structures in CT”. In: *Physics in Medicine & Biology* 67 (2022), 024001.

[97] Jakob Wasserthal et al. “TotalSegmentator: robust segmentation of 104 anatomic structures in CT images”. In: *Radiology: Artificial Intelligence* 5.5 (2023), e230024.

[98] Tugba Akinci D’Antonoli et al. “TotalSegmentator MRI: Robust Sequence-independent Segmentation of Multiple Anatomic Structures in MRI”. In: *Radiology* 314.2 (2025).

[99] Joseph O. Deasy, Angel I. Blanco, and Vanessa H. Clark. “CERR: a computational environment for radiotherapy research”. In: *Medical Physics* 30 (2003), pp. 979–985.

[100] Aditya P. Apte et al. “Library of deep-learning image segmentation and outcomes model implementations”. In: *Physica Medica* 73 (2020), pp. 190–196.

[101] John Duchi, Elad Hazan, and Yoram Singer. “Adaptive Subgradient Methods for Online Learning and Stochastic Optimization”. In: *Journal of Machine Learning Research* 12 (2011), pp. 2121–2159.

[102] Matthew D. Zeiler. “ADADELTA: An Adaptive Learning Rate Method”. In: *arXiv preprint arXiv:1212.5701* (2012).

[103] Diederik P. Kingma and Jimmy Lei Ba. "Adam: A Method for Stochastic Optimization". In: *Proceedings of the 3rd International Conference on Learning Representations (ICLR)* (2015).

[104] Timothy Dozat. "Incorporating Nesterov Momentum into Adam". In: *ICLR 2016 Workshop Track* (2016).

[105] Eustace M. Dogo et al. "A Comparative Analysis of Gradient Descent-Based Optimization Algorithms on Convolutional Neural Networks". In: *Proceedings of the 2018 International Conference on Computational Techniques, Electronics and Mechanical Systems (CTEMS)* (2018), pp. 92–99.

[106] David E. Rumelhart, Geoffrey E. Hinton, and Ronald J. Williams. "Learning representations by back-propagating errors". In: *Nature* 323 (1986), pp. 533–536.

[107] Jeremy Watt, Reza Borhani, and Aggelos Katsaggelos. *Machine Learning Refined: Foundations, Algorithms, and Applications*. Cambridge University Press, 2020. isbn: 9781108480727.

[108] Q. Duhoux, V. Marchesi, F. Lorchel, D. Azria, U. Schick, and I. Barillot, External irradiation treatment process: 2025 update, *Cancer/Radiothérapie*, 2025. doi: 10.1016/j.canrad.2025.104773.

[109] E. B. Podgorsak, *Radiation Oncology Physics: A Handbook for Teachers and Students*, IAEA, Vienna, 2005. Chap. 11.

[110] D. Jones, ICRU Report 50 – Prescribing, Recording, and Reporting Photon Beam Therapy, *Journal of the ICRU*, vol. 21, 1994.

[111] A. Wambersie, ICRU Report 62, Prescribing, Recording and Reporting Photon Beam Therapy (Supplement to ICRU 50), *Journal of the ICRU*, vol. 32, 1999.

[112] B. Emami, J. Lyman, A. Brown, L. Coia, M. Goitein, J. E. Munzenrider, B. Shank, L. J. Solin, and M. Wesson, Tolerance of normal tissue to therapeutic irradiation, *International Journal of Radiation Oncology, Biology, Physics*, vol. 21, pp. 109–122, 1991.

[113] W. A. G. Sauerwein, Prescribing and Reporting in BNCT, in *Neutron Capture Therapy: Principles and Applications*, 2nd ed., W. A. G. Sauerwein, A. Wittig,

R. L. Moss, Y. Nakagawa, K. Ono (Eds.), Springer Nature Switzerland, 2025, pp. 481–493.

[114] D. W. Nigg, Computational dosimetry and treatment planning considerations for neutron capture therapy, *Journal of Neuro-Oncology*, vol. 62, pp. 75–86, 2003.

[115] D. W. Nigg, F. J. Wheeler, D. E. Wessol, J. Capala, and M. Chadha, Computational dosimetry and treatment planning for Boron Neutron Capture Therapy, *Journal of Neuro-Oncology*, vol. 33, pp. 93–104, 1997.

[116] V. A. Nievaart, G. G. Dacquino, and R. L. Moss, Monte Carlo based treatment planning systems for Boron Neutron Capture Therapy in Petten, The Netherlands, *Journal of Physics: Conference Series*, vol. 74, 012012, 2007.

[117] H. Kumada, K. Takada, T. Mukawa, and H. Koivunoro, Treatment Planning, in *Neutron Capture Therapy: Principles and Applications*, 2nd ed., W. A. G. Sauerwein, A. Wittig, R. L. Moss, Y. Nakagawa, K. Ono (Eds.), Springer Nature Switzerland, 2025, pp. 449–480.

[118] S. J. González, D. G. Carando, G. A. Santa Cruz, and R. G. Zamenhof, Voxel model in BNCT treatment planning: performance analysis and improvements, *Physics in Medicine & Biology*, vol. 50, pp. 441–458, 2005.

[119] R. G. Zamenhof, E. Redmond II, G. Solares, D. Katz, K. J. Riley, S. Kiger III, and O. K. Harling, Monte Carlo-Based Treatment Planning for Boron Neutron Capture Therapy Using Custom-Designed Models Automatically Generated from CT Data, *International Journal of Radiation Oncology, Biology, Physics*, vol. 35, no. 2, pp. 383–397, 1996.

[120] F. J. Wheeler, D. E. Wessol, C. A. Wemple, D. W. Nigg, C. L. Albright, M. T. Cohen, M. W. Frandsen, G. J. Harkin, and M. B. Rossmeier, SERA – an advanced treatment planning system for neutron therapy, in *Advances in Neutron Capture Therapy*, International Atomic Energy Agency, Vienna, 1999–2001, pp. 223–232.

[121] R. O. Farias and S. J. 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, pp. 148–150.

[122] R. O. Farías, S. Bortolussi, P. R. Menéndez, and S. J. González, Exploring Boron Neutron Capture Therapy for non-small cell lung cancer, *Physica Medica*, vol. 30, no. 8, pp. 888–897, 2014.

[123] J. T. Goorley, W. S. Kiger III, and R. G. Zamenhof, Reference dosimetry calculations for neutron capture therapy with comparison of analytical and voxel models, *Medical Physics*, vol. 29, no. 2, pp. 145–156, 2002.

[124] S. J. González, G. A. Santa Cruz, W. S. Kiger III, J. T. Goorley, M. R. Palmer, P. M. Busse, and R. G. Zamenhof, NCTPlan, the new PC version of MacNCTPlan: Improvements and verification of a BNCT treatment planning system, in *Research and Development in Neutron Capture Therapy*, W. Sauerwein et al. (Eds.), Monduzzi Editore, Bologna, 2002, pp. 557–561.

[125] J. R. Albritton, Computational aspects of treatment planning for neutron capture therapy, Ph.D. thesis, Nuclear Science and Engineering Department, Massachusetts Institute of Technology, Cambridge, 2009.

[126] J. A. Kulesza, T. R. Adams, J. C. Armstrong, S. R. Bolding, F. B. Brown, J. S. Bull, T. P. Burke, A. R. Clark, R. A. Forster III, J. F. Giron, T. S. Grieve, C. J. Josey, et al., MCNP Code Version 6.3.0 Theory & User Manual, Technical Report LA-UR-22-30006, Los Alamos National Laboratory, 2022.

[127] K. Shibata, O. Iwamoto, T. Nakagawa, N. Iwamoto, A. Ichihara, S. Kunieda, S. Chiba, K. Furutaka, N. Otuka, T. Ohsawa, T. Murata, H. Matsunobu, A. Zukeran, S. Kamada, and J. Katakura, JENDL-4.0: A New Library for Nuclear Science and Engineering, *Journal of Nuclear Science and Technology*, vol. 48, no. 1, pp. 1–30, 2011.

[128] L. Wielopolski, J. Capala, N. E. Pendzick, and A. D. Chanana, Patient Positioning in Static Beams for Boron Neutron Capture Therapy of Malignant Glioma, *Radiation Medicine*, vol. 18, no. 6, pp. 381–387, 2000.

[129] H. Kumada, K. Yamamoto, A. Matsumura, T. Yamamoto, Y. Nakagawa, K. Nakai, and T. Kageji, Verification of the computational dosimetry system in JAERI (JCDS) for boron neutron capture therapy, *Physics in Medicine & Biology*, vol. 49, pp. 3353–3365, 2004.

[130] H. Kumada, K. Takada, K. Yamanashi, T. Sakae, A. Matsumura, and H. Sakurai, Verification of nuclear data for the Tsukuba plan, a newly developed treatment planning system for boron neutron capture therapy, *Applied Radiation and Isotopes*, vol. 106, pp. 111–115, 2015.

[131] T. Y. Lin and Y. W. H. Liu, Development and verification of THORplan – A BNCT treatment planning system for THOR, *Applied Radiation and Isotopes*, vol. 69, no. 12, pp. 1878–1881, 2011.

[132] J. Chen, Y.-C. Teng, W.-B. Zhong, H.-B. Yang, Q. Hong, and Y.-H. Liu, Development of Monte Carlo based treatment planning system for BNCT, *Journal of Physics: Conference Series*, vol. 2313, 012012, 2022.

[133] C. Pezzi, *Evaluation of BNCT treatment outcome using AI-based tumor segmentation for glioblastoma multiforme*. MSc thesis, Università degli Studi di Pavia, 2024. AI_MIGHT project.

[134] A. Grossberg, A. Mohamed, H. Elhalawani, W. Bennett, K. Smith, T. Nolan, S. Chamchod, M. Kantor, T. Browne, K. Hutcheson, G. B. Gunn, A. Garden, S. Frank, D. Rosenthal, J. Freymann, and C. Fuller, "Data from Head and Neck Cancer CT Atlas (Version 2) [Dataset]," *The Cancer Imaging Archive*, 2017. DOI: 10.7937/K9/TCIA.2017.umz8dv6s.

[135] J. Y. Y. Kwan, J. Su, S. H. Huang, *et al.*, "Data from radiomic biomarkers to refine risk models for distant metastasis in oropharyngeal carcinoma [Dataset]," *The Cancer Imaging Archive*, 2019. DOI: 10.7937/tcia.2019.8dho2gls.

[136] M. Vallières, E. Kay-Rivest, L. J. Perrin, X. Liem, C. Furstoss, N. Khaouam, P. F. Nguyen-Tan, C.-S. Wang, and K. Sultanem, "Data from Head-Neck-PET-CT [Dataset]," *The Cancer Imaging Archive*, 2017. DOI: 10.7937/K9/TCIA.2017.8oje5q00.

[137] H. J. W. L. Aerts, L. Wee, E. Rios Velazquez, *et al.*, "Data from Head-Neck-Radiomics-HN1 [Dataset]," *The Cancer Imaging Archive*, 2019. DOI: 10.7937/tcia.2019.8kap372n.

[138] W. R. Bosch, W. L. Straube, J. W. Matthews, and J. A. Purdy, "Data from Head-Neck Cetuximab [Dataset]," *The Cancer Imaging Archive*, 2015. DOI: 10.7937/K9/TCIA.2015.7AKGJUPZ.

[139] T. Bejarano, M. De Ornelas-Couto, and I. B. Mihaylov, "Head-and-neck squamous cell carcinoma patients with CT taken during pre-treatment, mid-treatment, and post-treatment (HNSCC-3DCT-RT) [Dataset]," *The Cancer Imaging Archive*, 2018. DOI: 10.7937/K9/TCIA.2018.13upr2xf.

[140] L. R. Dice, "Measures of the amount of ecologic association between species," *Ecology*, vol. 26, no. 3, pp. 297–302, 1945.

[141] Chiara Magni. "Experimental and computational studies for an Accelerator-Based Boron Neutron Capture Therapy clinical facility: a multidisciplinary approach". PhD thesis. University of Pavia, 2022.

[142] D. A. Low, W. B. Harms, S. Mutic, and J. A. Purdy, "A technique for the quantitative evaluation of dose distributions," *Medical Physics*, vol. 25, no. 5, pp. 656–661, 1998.