

Aus der Radiologischen Klinik der Universität Heidelberg
(Geschäftsführender Direktor: Prof. Dr.med. Dr.rer.nat. Jürgen Debus)

Abteilung für RadioOnkologie und Strahlentherapie
(Ärztlicher Direktor: Prof. Dr.med. Dr.rer.nat. Jürgen Debus)

**PROTON DOSE CALCULATION USING ARTIFICIAL
NEURAL NETWORKS**

Inauguraldissertation
zur Erlangung des Doctor scientiarum humanarum
an der

Medizinischen Fakultät Heidelberg
der

Ruprecht-Karls-Universität

vorgelegt von
Ahmad Neishabouri

aus
Damaskus, Syrien

2024

Dekan: Herr Prof. Dr. Michael Boutros

Doktorvater: Herr Prof. Dr. Dr. Jürgen Debus

“The important thing is not to stop questioning. Curiosity has its own reason for existing.”

— Albert Einstein

Dedicated to Mohsen Neishabouri

CONTENTS

| | |
|--|-----------|
| Contents | iv |
| List of Figures | vi |
| List of Tables | x |
| 1 INTRODUCTION | 1 |
| 1.1 Introduction to Cancer Therapy | 1 |
| 1.1.1 Overview of Cancer Treatment Modalities | 1 |
| 1.1.2 Radiotherapy Across the Globe | 2 |
| 1.1.3 Particle Therapy: The Frontier in Radiotherapy | 3 |
| 1.2 Introduction to Proton Therapy | 4 |
| 1.2.1 Characteristics of Proton Therapy | 4 |
| 1.2.2 Proton Range | 6 |
| 1.2.3 Proton Therapy in Clinics | 6 |
| 1.2.4 Patient Data Acquisition | 6 |
| 1.2.5 Beam Selection and Delivery | 7 |
| 1.2.6 Proton Dose Calculation | 8 |
| 1.3 Introduction to Artificial Intelligence | 12 |
| 1.3.1 Classification of AI Problems | 12 |
| 1.3.2 Problem dimensionality | 12 |
| 1.3.3 Sequence-to-sequence learning | 14 |
| 1.4 Previous Work | 15 |
| 1.5 Aims | 23 |
| 2 MATERIAL AND METHODS | 26 |
| 2.1 Feasibility of RNN-based Models for Full-Filed Dose Estimation | 27 |
| 2.1.1 Problem parameterization | 27 |
| 2.1.2 Model Architecture: Deterministic LSTM Model | 28 |
| 2.1.3 Model Training | 30 |
| 2.1.4 Incorporated Metrics | 32 |
| 2.2 Uncertainty Quantification: BayesDose | 32 |
| 2.2.1 Dataset | 33 |
| 2.2.2 BayesDose Implementation | 34 |
| 2.2.3 BayesDose Architecture | 36 |
| 2.2.4 BayesDose Training | 37 |
| 2.2.5 Incorporated Metrics and Experimental Design | 38 |
| 2.3 Toward Real-Time Adaptive Proton Therapy | 41 |
| 2.3.1 CC-LSTM Architecture | 41 |
| 2.3.2 LSTM-133 Architecture | 44 |
| 2.3.3 Data Preparation | 45 |
| 2.3.4 Model Evaluation Update | 50 |

| | | |
|-------|---|-----|
| 2.3.5 | Estimating Full-Field SOBP Dose Distributions | 51 |
| 3 | RESULTS | 55 |
| 3.1 | Feasibility of RNN-based Models for Full-Field Dose Estimation | 55 |
| 3.2 | BayesDose evaluations outcomes | 57 |
| 3.3 | CC-LSTM: A New Benchmark | 67 |
| 3.4 | Full-Field Dose Estimations | 74 |
| 3.4.1 | Estimating IES Dose Distributions | 74 |
| 3.4.2 | SOBP Dose Estimations | 76 |
| 3.5 | run time Analysis | 83 |
| 4 | DISCUSSION | 87 |
| 5 | SUMMARY | 99 |
| 6 | ZUSAMMENFASSUNG | 101 |
| | REFERENCES LIST | 104 |
| | PERSONAL CONTRIBUTION TO DATA ACQUISITION/ASSESSMENT AND PERSONAL PUBLICATIONS | 119 |
| | Curriculum Vitae | 122 |
| | ACKNOWLEDGEMENTS | 123 |
| | AFFIDAVIT | 125 |

LIST OF FIGURES

| | | |
|------------|--|----|
| Figure 1.1 | Depth dose profiles for three distinct proton energies | 5 |
| Figure 1.2 | active proton beam delivery system | 8 |
| Figure 1.3 | Step-by-step visualization of a proton’s path from vacuum to material, highlighting interaction locations, energy, and movement directions. | 10 |
| Figure 1.4 | Illustrations depicting (a) the internal configuration for a simple RNN, and (b) for an LSTM, adapted from Christopher Olah’s blog post “Understanding LSTM Networks.” . . | 16 |
| Figure 1.5 | Setup for phantom study to explore geometric changes in proton therapy, modifying slab positioning and densities. . . | 17 |
| Figure 1.6 | Demonstrates the setup for lung case simulations to enrich training datasets by adjusting gantry angles and isocenter positioning. | 18 |
| Figure 1.7 | Comparative analysis of the (b) RNN, (c) LSTM, and (d) PB techniques against the (a) baseline MC calculation for a selected sample. | 20 |
| Figure 1.8 | Dose estimation outcomes for a test sample (104.25 MeV) using the LSTM network and the PB algorithm. | 21 |
| Figure 1.9 | Distribution of voxels failing the γ -analysis criteria. | 21 |
| Figure 2.1 | Illustration of dose distribution from a 104.25 MeV proton beam, highlighting the targeted area in a CT scan. | 28 |
| Figure 2.2 | Describes the method for sequential and spatial-temporal proton dose calculations using RNN/LSTM models to produce accurate $m \times m$ dose outputs. | 29 |
| Figure 2.3 | Analysis of a proton beam’s dose profile alteration due to a heterogeneity with RSP of 2.5, creating a dual-peaked Bragg peak. | 30 |
| Figure 2.4 | Box plot illustrating the variance in RSP values across the tested patient geometries. Patient 0 represents the initial lung patient detailed in 2.1.3. | 34 |
| Figure 2.5 | Schematic of a BLSTM cell showing stochastic parameters distribution; includes input, hidden state, output, and cell state with activation functions. | 35 |
| Figure 2.6 | Flow diagram of the BayesDose model showing data processing through BLSTM cells into a Bayesian network for final output generation. | 37 |

| | | |
|-------------|---|----|
| Figure 2.7 | Graphical representation of the loss function evolution across 1000 epochs during patient training, displaying separate components of the loss | 39 |
| Figure 2.8 | Detailed diagram of the CC-LSTM architecture, showing integration of spatial input layers, temporal pathways, and spatial output layers for dose distribution mapping. | 42 |
| Figure 2.9 | Illustration of CC-LSTM architecture segmented into spatial feature extraction, output layers, and temporal pathway for enhanced dose prediction accuracy. | 43 |
| Figure 2.10 | Illustration of the LSTM-133 architecture. The model follows closely the parameterization and the information flow described in Figure 2.2. | 44 |
| Figure 2.11 | Statistical overview showing the number and energy of pencil beams used in the study. | 46 |
| Figure 2.12 | Histogram of incorporated gantry and couch angles in the cohort. Reds are the selected test patients. | 48 |
| Figure 2.13 | Histogram of unique patients, couch angles, and gantry angles per trained energy | 48 |
| Figure 2.14 | Comprehensive visual of dose distribution simulation for single pencil beams, showcasing different views including coronal, sagittal, and axial views, along with RSP cube and dose distribution post-extraction. | 49 |
| Figure 2.15 | The outlook of integration procedure in transition from PB domain to CT grid | 52 |
| Figure 2.16 | a comparison between 1 mm and 3 mm isotropic extraction schemes. | 52 |
| Figure 2.17 | Comparison between (a) 1 mm and (b) 3 mm resolution extraction and back-interpolation, assuming 100 % model accuracy. | 53 |
| Figure 2.18 | multiple GPU setup for AI-based dose calculation. | 54 |
| Figure 3.1 | Dose estimation results for representative samples from two energy ranges. | 56 |
| Figure 3.2 | Optimal prediction by the Bayesian LSTM network. | 59 |
| Figure 3.3 | Least effective prediction by the Bayesian LSTM network. | 60 |
| Figure 3.4 | Most significant quantification of uncertainty for the phantom case. | 61 |
| Figure 3.5 | Depiction of the most precise prediction by BayesDose on the lung patient dataset. | 62 |
| Figure 3.6 | Least accurate BayesDose prediction for the patient dataset. | 63 |
| Figure 3.7 | Most substantial quantification of uncertainty in the patient dataset. | 64 |
| Figure 3.8 | Dose estimation comparison for the worst-case scenario: (a) LSTM model versus (b) the newly developed CC-LSTM model. | 67 |

| | | |
|-------------|---|-----|
| Figure 3.9 | Dose estimation comparison for a representative sample: (a) LSTM model versus (b) the newly developed CC-LSTM model. | 69 |
| Figure 3.10 | Dose estimation results for a representative sample: (a) LSTM model versus (b) the newly designed CC-LSTM model. | 69 |
| Figure 3.11 | Dose estimation results for a representative sample: (a) LSTM model versus (b) the output of the newly designed CC-LSTM model. | 70 |
| Figure 3.12 | Distribution of the average number of dose voxels with dose values greater than zero. | 70 |
| Figure 3.13 | Lateral distribution of 'failed' voxels across the lateral axis. | 70 |
| Figure 3.14 | The average lateral distribution of 'failed' voxels across the lateral axis for both axes (a) and (b). | 71 |
| Figure 3.15 | Comparison of dose estimation for a sample case: (a) CC-LSTM model and (b) DoTA model. | 74 |
| Figure 3.16 | Comparison of IES dose distributions for low-range, mid-range, and high-range energies in the BEV. | 75 |
| Figure 3.17 | γ analysis ([1%, 2 mm], local) pass rates of all defined energies. | 76 |
| Figure 3.18 | γ analysis comparison of dose estimation for an unseen, real-patient (Ho3476) case. | 77 |
| Figure 3.19 | DVH comparison for patient Ho3476, illustrating the performance of MC, CC-LSTM, and LSTM-133 models for both the Planning Target Volume (PTV) and Optimization Ring. | 78 |
| Figure 3.20 | γ analysis comparison of dose estimation for an unseen, real-patient (Ho2520) case. | 80 |
| Figure 3.21 | γ analysis comparison of dose estimation for an unseen, real-patient (Ho2520) case. | 81 |
| Figure 3.22 | DVH comparison for patient Ho2520, illustrating the performance of the MC, CC-LSTM, and LSTM-133 models for the target PTV, as well as two OARs: the left inner ear and the brainstem. | 82 |
| Figure 3.23 | γ analysis comparison of dose estimation for an unseen, real-patient (Ho2348) case. | 82 |
| Figure 3.24 | DVH comparison for patient Ho2348, illustrating the performance of the CC-LSTM and LSTM-133 models for the OAR in proximity to highly heterogeneous geometry, specifically the left optical nerve, as shown in Figure 3.23. | 83 |
| Figure 6.1 | DVH comparison for patient Ho2794, illustrating the performance of MC, CC-LSTM, and LSTM-133 models. | 112 |
| Figure 6.2 | γ analysis comparison of dose estimation for an unseen, real-patient (Ho2794) case. | 113 |
| Figure 6.3 | γ analysis comparison of dose estimation for an unseen, real-patient (Ho3037) case. | 114 |

| | | |
|------------|--|-----|
| Figure 6.4 | DVH comparison for patient Ho3037, illustrating the performance of MC, CC-LSTM, and LSTM-133 models. | 115 |
| Figure 6.5 | DVH comparison for patient Ho2593, illustrating the performance of MC, CC-LSTM, and LSTM-133 models. | 115 |
| Figure 6.6 | γ analysis comparison of dose estimation for an unseen, real-patient (Ho2593) case. | 116 |
| Figure 6.7 | γ analysis comparison of dose estimation for an unseen, real-patient (Ho2651) case. | 117 |
| Figure 6.8 | DVH comparison for patient Ho2651, illustrating the performance of MC, CC-LSTM, and LSTM-133 models. | 118 |

LIST OF TABLES

| | | |
|------------|---|----|
| Table 1.1 | HU to RSP Conversion of an example look-up table used in this study. | 7 |
| Table 1.2 | Comparison of γ -index values (1 %, 3 mm) among two neural network models and a PB algorithm against the MC simulation in a phantom setup. | 19 |
| Table 1.3 | MAE and MSE between the network models and the MC simulation in the phantom case (Experiment 1 - 1). | 19 |
| Table 1.4 | γ -index analysis (1 %, 3 mm), MAE, and MSE of the LSTM model and PB algorithm compared to MC calculations for the lung patient case (Experiment 1 - 2). | 20 |
| Table 2.1 | Statistics of the patient cohort. | 47 |
| Table 3.1 | γ -index analysis for five different lung cancer patients (1 %, 1 mm). | 55 |
| Table 3.2 | γ -index analysis on datasets with three distinct energies of proton pencil beams (Experiment 1 - 2). | 56 |
| Table 3.3 | Comparative analysis of execution times between MC calculations, LSTM predictions, and the PB algorithm. | 57 |
| Table 3.4 | Comparison of γ -index results, MAE, and MSE for different BayesDose model configurations against MC simulations in a phantom study (Experiment 2 - 1). | 58 |
| Table 3.5 | Proportion of voxels deviating from MC simulation beyond specified confidence intervals in the phantom study. | 58 |
| Table 3.6 | γ -index analysis (1 %, 3 mm) and MAE and MSE comparisons among BayesDose and the deterministic LSTM. | 58 |
| Table 3.7 | Percentage of voxels where the average prediction deviates from the MC simulation outside specified confidence intervals. | 62 |
| Table 3.8 | γ -index analysis ([1 %, 3 mm]) for three proton beamlets with varying initial energies (Experiment 2 - 3). | 63 |
| Table 3.9 | γ -index evaluation (1 %, 3 mm) for five lung cancer patients. | 65 |
| Table 3.10 | γ -index analysis ([1 %, 3 mm]) for lung cancer patients before and after 10 epochs of transfer learning on Patient 5 (Experiment 2 - 5). | 66 |
| Table 3.11 | Evaluative comparison of an established network (LSTM) and an advanced model (CC-LSTM) against a MC simulation benchmark at an initial energy of 79.17 MeV. | 68 |
| Table 3.12 | Local γ -index analysis ([1 % , 2 mm]), MAE, and MSE of the CC-LSTM model compared to MC calculations for the three representative energies | 72 |

| | | |
|------------|--|----|
| Table 3.13 | Evaluation of the CC-LSTM model performance relative to ground truth data from MC simulations at a low-range initial energy of 79.17 MeV, employing three unique cube parametrizations along the lateral axis. | 73 |
| Table 3.14 | Comparison of CC-LSTM (left) against DoTA (right) using gamma index analysis for three distinct PB energies. | 73 |
| Table 3.15 | γ -index evaluation ([2 %, 2 mm]; global) of CC-LSTM and LSTM-133 models on all test patient cases, against the ground truth MC simulations. | 78 |
| Table 3.16 | CC-LSTM run times vs its predecessor and DoTA | 84 |
| Table 3.17 | CC-LSTM run times when introducing customized parameterization in the longitudinal axis | 84 |
| Table 3.18 | CC-LSTM run times to generate the dose distribution of an entire IES (*: showcased in sub-figures outlined in 3.16) | 85 |
| Table 3.19 | Comparison of CC-LSTM and LSTM-133 models on all test patient cases, highlighting the γ -index pass rates ([2 %, 2 mm]; global) and average run times (mean \pm std) based on 5 measurements. | 86 |

ABBREVIATIONS

AI: Artificial Intelligence
ANN: Artificial Neural Network
APT: Adaptive Particle Therapy
BEV: Beam's Eye View
BLSTM: Bayesian Long Short-Term Memory
CC-LSTM: Convolutional Convolutional Long Short-Term Memory
COVID: Coronavirus Disease
CPU: Central Processing Unit
CSDA: Continuous Slowing Down Approximation
CT: Computed Tomography
DD: Dose Difference
DEGRO: German Radiation Oncology Society
DNA: Deoxyribonucleic Acid
DVH: Dose Volume Histogram
ELBO: Evidence Lower Bound
FC: Fully Connected
GB: Gigabyte
GPU: Graphics Processing Unit
GTX: GeForce GTX
HIT: Heidelberg Ion-Beam Therapy
HU: Hounsfield Units
ID: Identification
IES: Iso Energy Surface
IGRT: Image-Guided Radiotherapy
IMPT: Intensity-Modulated Particle Therapy
IMRT: Intensity-Modulated Radiotherapy
KL: Kullback-Leibler
LGG: Low-Grade Glioma
LHC: Large Hadron Collider

LSTM: Long Short-Term Memory
MAE: Mean Absolute Error
MC: Monte Carlo
MCS: Multiple Coulomb Scatterings
MSE: Mean Squared Error
NLP: Natural Language Processing
NVIDIA: NVIDIA Corporation
PB: Pencil Beam
PBA: Pencil Beam Algorithms
PBS: Pencil Beam Scanning
PC: Personal Computer
PDP: Peak Deviation Percentage
PR: Pass Rate
QK: Query Key
RAM: Random Access Memory
RNN: Recurrent Neural Network
RSP: Relative Stopping Power
RT: Radiotherapy
RTX: Ray Tracing Texel eXtreme
SD: Standard Deviation
SI: International System of Units
SMSE: Scaled Mean Squared Error
SOBP: Spread-Out Bragg Peak
SOTA: State-Of-The-Art
SPIDER: Superfast Proton-Ion Dose Engine based on Recurrent Neural Networks
TL: Transfer Learning
TOPAS: Tool for Particle Simulations

INTRODUCTION

1.1 INTRODUCTION TO CANCER THERAPY

Global Incidence and Trends

Recent studies show an increase in cancer incidence worldwide, affecting people of all ages. Notably, there is an increase in cancer diagnoses among those under 50. This change is linked to a mix of genetic, lifestyle, and environmental factors. With a higher incidence in younger patient populations, improving our diagnostic, preventive, and treatment strategies is imperative to enhance patient outcomes (Koh et al., 2023).

The rise in cancer among younger individuals challenges current treatment methods and demands new approaches and innovative treatments. Particle therapy is especially promising as it targets cancer cells while sparing healthy tissue. This approach could improve outcomes for younger patients, emphasizing the need for therapies that match the changing landscape of cancer (Pakela et al., 2022).

1.1.1 *Overview of Cancer Treatment Modalities*

Cancer treatment is inherently multidisciplinary, combining various treatment modalities tailored to the tumor's type, stage, location, and overall health. The main treatment options are chemotherapy, surgery, and radiotherapy, administered either separately or in combination (Zhao et al., 2022).

Chemotherapy

Chemotherapy employs cytotoxic agents to target and destroy cancer cells and is fundamental in managing widespread and metastatic cancers. Its efficacy depends on the cancer's molecular and genetic characteristics, determining its responsiveness to treatment. Chemotherapy is often combined with

other treatments, either before surgery to reduce tumor size (neoadjuvant therapy (Pattyn et al., 2022)) or post-surgery to eradicate any residual cancer cells (adjuvant therapy (Garapati et al., 2017)).

Surgery

Surgery is often the treatment of choice for localized solid tumors (Holcomb, 1999). The surgical procedure aims to completely remove the tumor and achieve clear margins to ensure no malignant cells are left at the surgical borders. Surgical techniques vary from minimally invasive procedures to extensive resections, depending on the tumor's location and the extent of the disease.

Radiotherapy

Radiotherapy (RT) utilizes high-energy ionizing radiation to eliminate cancer cells (Wood et al., 2022). It is administered as either external beam radiation or internal radiation therapy (brachytherapy). Advances in RT, such as intensity-modulated radiotherapy (IMRT) and image-guided radiotherapy (IGRT), have significantly improved the precision of treatments, particularly for photon-based therapies. The development of Intensity-Modulated Proton Therapy (IMPT) further enables highly conformal treatments that optimize radiation to surrounding organs at risk and normal tissues. This capability is critical for reducing toxicities and improving patient quality of life during and after treatment. Radiotherapy is indispensable in clinical routine, from reducing tumor volume as a primary treatment to serving as a palliative measure in advanced cancer stages.

1.1.2 Radiotherapy Across the Globe

Nearly 50% of cancer patients globally receive RT at some point during their treatment course (Hack et al., 2019). Therefore, the demand for this treatment modality is expected to rise significantly, driven by demographic shifts toward an aging population and rising cancer incidence rates.

In Europe, Germany has the highest number of RT centers (n=289) (Rosenblatt et al., 2013). As a pioneer in introducing novel treatment modalities and techniques, Germany continues to set benchmarks, exemplified by the

recent commissioning and first patient treatment of raster-scanned helium ions beams at the Heidelberg Ion Therapy (HIT) facility (Tessonnier et al., 2017; Tessonnier et al., 2023).

In contrast, access to advanced RT options is severely restricted in many developing regions due to economic constraints, a lack of technological advancements, and a scarcity of medical facilities equipped with state-of-the-art technologies. Additionally, there is a critical need for specialized training programs for medical professionals to implement advanced radiotherapy techniques effectively.

In Iran, these challenges are aggravated by international sanctions that severely impact the healthcare sector. The sanctions restrict economic resources and hinder the import of advanced medical equipment and technologies. This creates significant barriers to integrating state-of-the-art therapy facilities, such as those required for particle therapy, further deepening disparities in cancer care.

Global health organizations are increasingly focusing on improving RT access in underserved areas. Such initiatives aim to bridge the treatment gap and promote equity in cancer care. Their efforts ensure effective treatments are available globally, irrespective of patients' location.

1.1.3 *Particle Therapy: The Frontier in Radiotherapy*

Particle therapy represents a major advancement in RT, providing targeted oncological treatment with potentially minimized damage to surrounding healthy tissues. This advanced technique is particularly characterized by its use of charged particles, such as protons and heavier ions like carbon, which possess unique physical properties. In contrast to conventional photon therapy, which employs X-rays, particle therapy can precisely deliver radiation doses with minimal out-of-target dose deposition due to the unique Bragg peak phenomenon of charged particles. This peak allows the maximum energy to be deposited at a specific depth, with a sharp drop-off beyond that point, therefore significantly limiting the radiation dose received by healthy tissue adjacent to the tumor. This precision is crucial when targeting tumors near critical organs or within the central nervous system, where conventional RT may pose significant risks. This capability of particle therapy makes it a favorable option for complex cases.

The benefits of particle therapy are particularly evident when treating younger patients, who are especially susceptible to long-term radiation-

induced side effects, such as secondary cancers or growth abnormalities. Therefore, minimizing radiation exposure is of utmost importance (Mizumoto et al., 2021).

The integration of advanced imaging and treatment planning systems, increasingly augmented by artificial intelligence (AI), enhances the capabilities of particle therapy. These advancements improve dose conformity, thus advancing clinical use and research.

1.2 INTRODUCTION TO PROTON THERAPY

The discovery of X-rays by Wilhelm Röntgen in 1895 revolutionized medical diagnostics and treatment, leading to extensive research into their therapeutic applications. X-rays interact with biological matter through photon absorption and scattering, resulting in ionization and excitation of cellular molecules, which can result in DNA damage and cell death (Kamada et al., 2015; Loeffler and Durante, 2013). The core objective of RT is to optimize dose delivery to the target while minimizing exposure to surrounding healthy tissues. This principle guides the development of advanced techniques that enhance treatment efficacy and patient safety. Early clinical observations demonstrated the effectiveness of X-rays in treating superficial tumors, highlighting their potential for broader therapeutic applications (Kamada et al., 2015).

Despite decades of advancements, photon RT remains the major modality. However, the unique properties of proton therapy have positioned it as an attractive alternative, gaining daily adoption and integration in clinical applications. Proton therapy, first proposed by physicist Robert Wilson in 1946, offers significant advantages, such as a lower exit dose due to the finite range of the proton beam, which minimizes radiation exposure to non-target tissues (Wilson, 1946). The foundations laid down in the early 1960s, following the first human treatment with proton beams at the University of California in 1954, have established routine clinical applications of proton therapy, transforming it into a precise and innovative treatment modality (Newhauser, 2009). The subsequent section will detail the unique characteristics of proton therapy.

1.2.1 *Characteristics of Proton Therapy*

Protons mainly interact with matter through Multiple Coulomb Scatterings (MCS), involving both inelastic interactions with atomic electrons and elastic

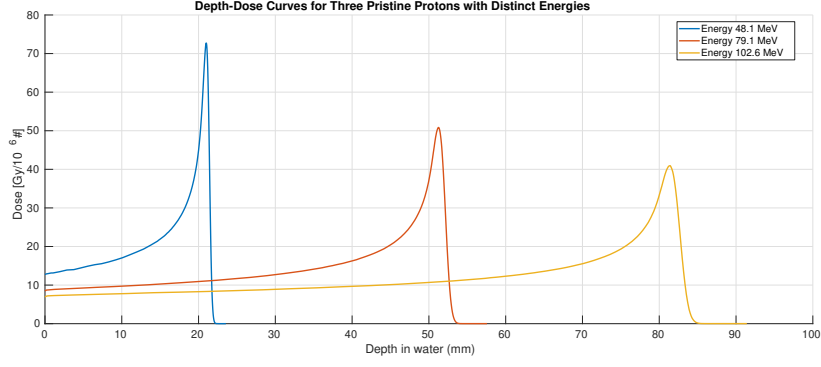


Figure 1.1: Depth dose profiles of protons with initial energy E_0 in water

interactions with nuclei. While interactions with electrons are more frequent and lead to a continuous energy loss, they rarely cause significant proton deflection due to the mass difference between protons and electrons. On the other hand, less frequent elastic nuclear interactions can significantly alter the proton's path, introducing uncertainty in its trajectory. This is crucial for the establishment of the Bragg peak. At higher energies, above the typical therapeutic range, protons become likely to induce nuclear reactions, though this is not a concern in standard proton therapy.

The behavior of protons in media is often described in terms of the mass stopping power, which is independent of the material's mass density:

$$\frac{S}{\rho} = -\frac{dE}{\rho dz}, \quad (1)$$

where ρ represents the mass density, E the energy, and z the distance traveled. The Bethe-Bloch formula is outlined below:

$$-\frac{dE}{dz} = \frac{4\pi n_e}{m_e c^2 \beta^2} \left(\frac{e^2}{4\pi\epsilon_0} \right)^2 \left[\ln \left(\frac{2m_e c^2 \beta^2}{I(1-\beta^2)} \right) - \beta^2 \right], \quad (2)$$

provides a detailed description of the mean energy loss per unit path length, emphasizing that the energy loss rate inversely correlates with the square of the proton's velocity (n_e is the material electron density, β is the incident proton speed relative to the speed of light m_e is the electron mass, I is the mean excitation potential of the target medium and finally ϵ_0 is the vacuum permittivity). This relationship is evident in the profile of the Bragg curve, highlighting the increased energy deposition as protons decelerate towards the end of their range (Fig. 1.1).

1.2.2 Proton Range

The range of a proton, or the total distance it travels through a medium, can be determined using the Continuous Slowing Down Approximation (CSDA):

$$R(E_0) = \int_0^{E_0} \frac{1}{-\frac{dE}{dz}} dE, \quad (3)$$

where E_0 represents the initial proton energy. This equation applies to a beam of protons rather than individual particles.

Due to the stochastic nature of proton interactions with matter, the Range Straggling phenomenon leads to a variance in proton energies within the beam, resulting in the widening of the beam. Figure 1.1 contrasts the 1D dose profiles of PB at low and high energies, demonstrating the range straggling effect.

1.2.3 Proton Therapy in Clinics

This section discusses the advantages of proton therapy's superior dose conformity and the critical aspects of its clinical implementation. Key factors facilitating the integration of particle therapy into treatment protocols are explored, and the prerequisites necessary for executing a successful treatment plan in clinical settings are discussed.

1.2.4 Patient Data Acquisition

In the context of RT, a computed tomography (CT) scan is performed before treatment planning. It uses X-ray beams to create cross-sectional images of the scanned body part by measuring the beams' attenuations. This attenuation is then converted into numerical values, known as Hounsfield Units (HU). The HUs are linearly correlated to the tissue's electron density, which is a key parameter for calculating the stopping power of protons as they traverse through the body. Accurate determination of the electron density is crucial for proton therapy treatment planning, as protons deposit most of their energy at the end of their range in the body. The location of the Bragg peak is highly dependent on the stopping power of the tissues (as outlined in Eq. 2), which is directly related to the electron density.

Table 1.1: HU to RSP Conversion of an example look-up table used in this study.

| Hounsfield Unit | Relative Stopping Power |
|-----------------|-------------------------|
| -1025 | 0.0001 |
| -1000 | 0.001 |
| -200 | 0.8 |
| -100 | 0.94 |
| -30 | 0.995 |
| 0 | 1 |
| 78 | 1.075 |
| 183 | 1.11 |
| 1385 | 1.7 |
| 1974 | 1.778 |
| 3096 | 2.077 |

Relative Stopping Power (RSP) to water quantifies the energy loss rate of a proton beam in a specific tissue compared to water. RSP calculations utilize look-up tables based on the electron density derived from Hounsfield Unit (HU) values. These tables enable physicists to transform CT images into detailed maps of RSP. Utilizing RSP maps, the properties of the proton beam can be precisely tailored to maximize tumor energy deposition, thereby optimizing therapeutic efficacy and minimizing damage to surrounding healthy tissues. This approach underscores the critical role of CT in both capturing anatomical detail and facilitating the complex physics calculations required for effective proton therapy planning.

1.2.5 Beam Selection and Delivery

The process of selecting appropriate proton beam energies and impinging angles is crucial for ensuring effective tumor coverage. The Spread-out Bragg Peak (SOBP) is usually implemented to create a uniform dose distribution across a tumor by superimposing multiple mono-energetic beams. These beams are accelerated to the selected energies using a cyclotron or synchrotron at the treatment facility. The beam can be delivered either through passive scattering or active scanning, with the latter providing more precise control by manipulating the beam's path using magnetic fields. These beams, also known as pencil beams (PB) due to their very narrow full width at half maximum, are the building blocks of a comprehensive treatment plan (Paganetti and Bortfeld, 2005).

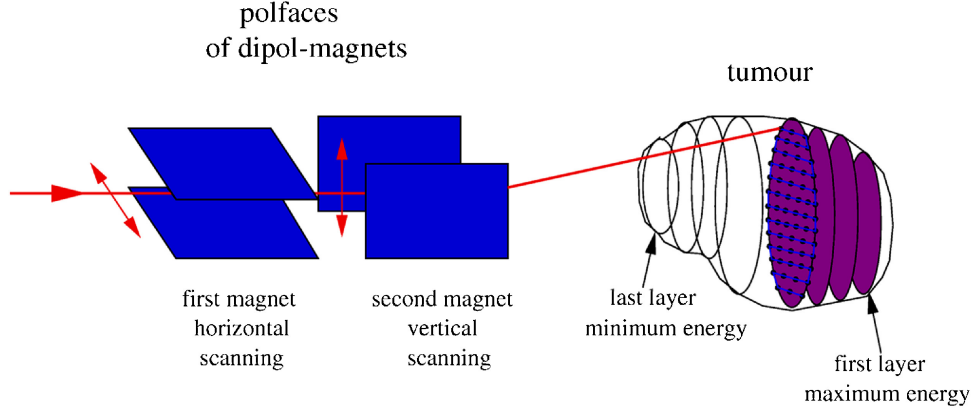


Figure 1.2: Principles of active scanning system (Reprint from (Jäkel, Kraft, and Karger, 2022))

Active scanning, also known as pencil beam scanning, directs the proton beam in a pre-defined pattern across the tumor, adjusting the energy dynamically to modify the beam's range (Fig. 1.2). This technique allows for a highly conformal dose distribution, which is especially beneficial for irregularly shaped or deep-seated tumors. Equation 4 demonstrates how the dose at point (x, y, z) can be evaluated:

$$D(x, y, z) = \sum_i w_i \cdot d(E_i, x, y, z), \quad (4)$$

where w_i denotes the fluence assigned to each beam, $d(E_i, x, y, z)$ represents the dose delivered by a beam with energy E_i at the coordinate (x, y, z) . To calculate the cumulative dose delivered by an active scanning system, it is essential to assess the dose contribution from each pencil beam. This process, integral to treatment planning, involves aggregating these contributions to determine the total therapeutic dose administered to the patient.

1.2.6 Proton Dose Calculation

The precise calculation of proton therapy dose distribution requires a thorough understanding of various factors beyond basic single-particle track interactions. Initial dose calculations focus on energy deposition in each voxel along the track using stopping power equations, but they are insufficient for real proton therapy planning. For example, the range in tissue can be dependent on uncertainties governed by the imaging modality, patient setup, beam delivery, and dose calculation, as listed by (Paganetti, 2012). Basic models do not encompass machine beam characteristics and how they affect scat-

tering and nuclear reaction distributions, significantly affecting the ultimate dose outcome. Therefore, typically, treatment facilities incorporate empirical beam data measurements into their beam model. The data are based on the treatment room delivery system and accelerator phase space configuration upstream to the patient.

With this background, two primary methods are predominantly used in the field: Monte Carlo (MC) simulations and analytical dose calculation methods. Each method offers distinct advantages and limitations in a clinical context, particularly concerning computation time and accuracy.

Monte Carlo Simulations

MC simulations are widely regarded for their high accuracy in modeling the transport and interaction of protons with matter. Named after the famous gambling city in southern France due to their stochastic nature, MC simulations are often considered the gold standard for simulating particle behavior in matter. One of the primary advantages of MC simulations is their ability to handle complex geometries and tissue heterogeneities with high accuracy. This method can accurately simulate dose distributions in regions with varying densities, such as bone-tissue interfaces, and can account for the effects of beam modifiers and patient-specific factors.

MC simulations operate by discretizing the path of a proton into small steps and modeling the process through a series of discrete subprocesses. Each step in a proton's trajectory involves a sequence of interactions that are modeled probabilistically, based on the probability distributions of potential interactions within the matter. These steps include:

1. **Distance Sampling:** The distance to the next step is determined by sampling from the total cross-section, which quantifies the probability of interactions per unit distance.
2. **Interaction Type Sampling:** The specific type of interaction (e.g., scattering event) is randomly determined, followed by the appropriate modeling of the event.
3. **Particle Transport:** The proton is then transported to the next step, or a secondary particle process is initiated, based on the results from the previous interaction sampling.

The accuracy and reliability of MC simulations are highly dependent on the 'history' of each particle, a term used to describe the record of all simulated

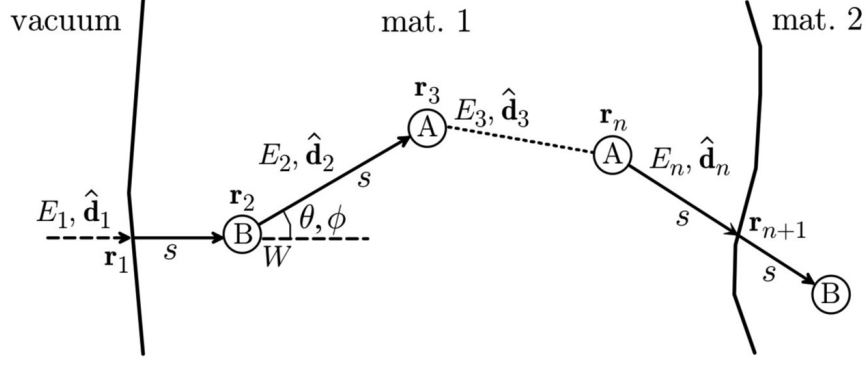


Figure 1.3: Detailed step-by-step discretization of a particle trajectory as it moves from a vacuum into material 1 and, after several interactions, proceeds to material 2. Here, \mathbf{r}_n denotes the location of the n -th scattering event, while E_n and $\hat{\mathbf{d}}_n$ represent the energy and the directional cosines of the movement path, respectively, characterized by the scatter polar angle θ and the azimuthal scattering angle ϕ . **A** and **B** indicate the type of interaction occurring at each step. (Salvat, Fernández-Varea, and Sempau, 2007)

subprocesses and interactions for a particle. Executing these simulations requires a long sequence of uniformly distributed pseudo-random numbers to perform the stochastic sampling necessary for the simulation. Given the large number of interactions that protons undergo, this process is time-consuming. Therefore, the high computational demand of MC simulations poses a significant challenge, particularly for real-time applications.

Efforts to accelerate MC simulations include parallelizing the algorithms using GPU technology (Schiavi et al., 2017), streamlining physical models (Lysakovski et al., 2021), and developing fast MC codes for clinical use. These advancements aim to reduce computation times while maintaining the accuracy of traditional MC methods. By leveraging these approaches, MC simulations can be made more practical for clinical applications, ensuring precise and efficient dose calculations in proton therapy.

Analytical Dose Calculation

Analytical dose calculation methods, such as the Pencil Beam Algorithm (PBA) (Hong et al., 1996), provide a computationally efficient alternative to MC simulations by using mathematical models to approximate dose distributions. These methods segment the proton beam into numerous narrow, focused beams, known as pencil beams, and calculate the dose for each beam

separately. The results are then summed to generate the overall dose distribution for the treatment field.

The PBA is preferred for its computational simplicity and satisfactory accuracy. In modeling proton beam transport, including energy loss and scatter. Developed initially for electron beams, the PBA was adapted in the 1990s for protons and heavier charged particles. The work by (Hong et al., 1996) provided the formalism widely adopted in current analytical dose computation engines.

The core of the PBA involves calculating the dose distribution for each pencil beam using a combination of central-axis and off-axis dose components. The central-axis dose quantifies the dose along the proton beam's path, while the off-axis component models the lateral spread due to MCS and nuclear interactions. These components are typically modeled using Gaussian functions to describe the dose distribution around the beam's central axis.

Despite the computational efficiency, analytical methods like the PBA exhibit limitations, particularly in heterogeneous tissues (Schaffner, Pedroni, and Lomax, 1999). The simplified assumptions may fail to capture the complex interactions and variations in tissue density accurately, leading to potential inaccuracies in dose calculations. Corrections, such as pencil beam splitting and convolution superposition algorithms, have been developed to enhance accuracy by better modelling the lateral spread and interactions in heterogeneous media (Soukup, Fippel, and Alber, 2005).

Recent advancements continue to refine analytical dose calculation methods, integrating improvements in computational speed and accuracy. Hybrid approaches that synergize elements of analytical models with Monte Carlo simulations are being explored to balance efficiency and precision in clinical applications.

Need for Speed

Proton dose calculation currently faces a critical challenge in balancing accuracy with computational speed. Specifically, for online adaptive radiotherapy, efficient on-table re-planning is essential due to patient anatomical changes or tumor movements during therapy. Currently, neither conventional MC simulations nor standard analytical methods can meet the necessary computational speeds in particle therapy. This challenge underscores the need for developing new computational approaches that can deliver both the precision of MC simulations and the speed necessary for real-time treatment adaptation.

1.3 INTRODUCTION TO ARTIFICIAL INTELLIGENCE

AI addresses a wide spectrum of computational challenges, each characterized by unique data structures, intended outcomes, and algorithms. A comprehensive understanding of these problems is essential to understanding the evolution and application of advanced deep learning models like LSTM (Long Short-Term Memory), ConvLSTM (Convolutional LSTM), and Transformers, especially in handling complex three-dimensional (3D) data in spatiotemporal contexts.

1.3.1 *Classification of AI Problems*

AI challenges can generally be divided into several groups: classification, regression, clustering, and sequence prediction. Each category is characterized by specific objectives and employs unique data types and modeling techniques.

- **Classification:** Classification tasks in AI involve assigning data to predefined categories. The output of these systems is discrete and often used in applications such as image recognition and segmentation.
- **Regression:** Regression problems deal with predicting a continuous quantity. This type of AI application is employed in estimating time series analysis.
- **Clustering:** Clustering techniques group a set of objects so that objects in the same group are more similar to each other than those in other groups.
- **Sequence Prediction:** This involves predicting the next sequence in a series of data points. Models like LSTM are particularly suited for this type of problem, where the order and context of historical data points play a crucial role in making accurate predictions.

1.3.2 *Problem dimensionality*

The dimensionality of both input and output data in AI-based models plays a critical role in defining the complexity and type of model used. For instance, 1D data, such as time series from stock prices or audio signals, is processed linearly, often using models like recurrent neural networks (RNN) that can

interpret temporal sequences. For two-dimensional (2D) data, including images or spatial grids, convolutional neural networks (CNNs) are employed to capture spatial hierarchies and patterns. These models excel in handling an image's width and height dimensions, extracting features layer by layer. In the context of 3D data, such as volumetric images from medical scans or video sequences, the complexity increases as the data incorporates depth alongside height and width, or time in the case of videos. Processing 3D data often requires more advanced AI techniques, such as 3D CNNs or LSTM networks combined with convolutional layers, to effectively capture the additional spatial or temporal relationships present. Each increase in dimensionality demands more computational power and sophisticated algorithms to accurately model and interpret the intricate structures inherent in the data.

Additionally, we could incorporate a *sequence* of the above-mentioned problems across various dimensions. For instance, stock prices are a sequence of numerical data, a sentence can be a sequence of 1D integers serving as a marker for the corresponding word in corpus (i.e., one-hot encoded vector), and a video can be presented as a sequence of 2D image frames. Such problems with a sequential nature are observed repeatedly in physics. These challenges frequently involve handling data sequences, including integer, biological, or event sequences. An important area where such sequences are pivotal is within 3D problem contexts, particularly in medical applications. Medical imaging utilizes 3D data to provide details about anatomical structures or diseases, which are critical for precise diagnosis and treatment planning.

Depending on the underlying physics of the task at hand, transforming 3D inputs into a sequence of 2D inputs can be beneficial for managing the task effectively. This approach allows for the use of techniques that are better adapted for processing sequential data. Following the successful performance of recently developed transformer models in processing sequences of data, this approach has been extensively adapted across various problem types.

Ultimately, the capability to manipulate 3D data, reformat it into a series of 2D inputs, and subsequently address sequence-to-sequence challenges presents new opportunities for employing advanced deep learning models like LSTM networks. These models are particularly efficient at recognizing patterns in sequential data and play a crucial role in addressing intricate AI challenges, especially in fields such as medical imaging and biology.

1.3.3 *Sequence-to-sequence learning*

In the field of AI, sequence-to-sequence (also known as seq2seq problems) learning represents an approach that addresses the challenge of mapping sequences of inputs to an output (many-to-one), an input to a sequence of outputs (one-to-many), and a sequence of input to sequence of output (many-to-many). This method is particularly beneficial in tasks where both the input and output are best expressed as sequences whose lengths are not predetermined, such as speech recognition, text translation, and more complex dynamic interactions.

The core mechanism of sequence-to-sequence learning involves two primary components: an encoder and a decoder, both often implemented using some variation of RNN-based networks such as LSTM networks. The encoder processes the input sequence into a fixed-dimensional vector representation, capturing the encoded representation of the input data in a compressed form. This vector, which encapsulates the entire input sequence, serves as the context for the decoder. The decoder then interprets this context to produce the output sequence, one element at a time, effectively translating the encoded data into a new sequence representing the desired output.

This model architecture allows for the handling of sequences with variable lengths and supports the learning of dependencies that may span the entire sequence length. The effectiveness of this approach is demonstrated in its ability to maintain performance even with long input sequences, which poses a significant challenge in many traditional models due to issues like the vanishing gradient problem (Pascanu, Mikolov, and Bengio, 2012). Section 2.1 provides examples of various neural network architectures that have been pivotal in advancing Seq2Seq problems.

1.4 PREVIOUS WORK

Feasibility of Recurrent Neural Networks for Seq2Seq problems

The groundwork for the present study was laid during earlier research, which explored the application of RNN-based networks for sequence-to-sequence parameterization of the task for single energy pencil beams. This preliminary investigation, conducted as part of a Master's thesis, has significantly informed the current research. The findings presented in section 3 build upon these initial outcomes.

A key insight extracted from the preliminary work was the identification of sequence-to-sequence parameterization as a particularly suitable approach for the learning task at hand. The investigation primarily centered on assessing the capabilities of RNN-based networks, encompassing both simple RNN structures and more complex configurations. Various preprocessing and post-processing schemes were examined for feature extraction and dimensionality reduction. This section provides a concise summary of the results from the preliminary study that are most relevant to the current research.

Basic RNNs

Recurrent Neural Networks (RNNs) are the foundation of most seq2seq models. They are designed to handle sequence data by maintaining a hidden state that updates as they process each element of an input sequence. The basic form of an RNN can be described by the equation:

$$h_t = \tanh(W_{hh}h_{t-1} + W_{hx}x_t + b_h)$$

where h_t is the hidden state at time t , x_t is the input at time t , W_{hh} and W_{hx} are weights, and b_h is a bias term.

Long Short-Term Memory (LSTM)

LSTMs are an enhancement over basic RNNs, designed to deal with the vanishing gradient problem by introducing a memory cell that can maintain information and propagate this memory across its recurrence. An LSTM unit includes three gates: an input gate, a forget gate, and an output gate, which

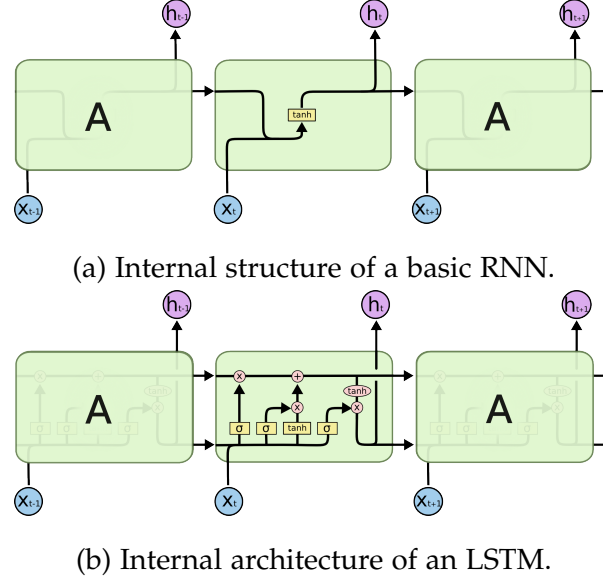


Figure 1.4: Illustrations depicting (a) the internal configuration for a simple RNN, and (b) for an LSTM, adapted from Christopher Olah’s blog post “Understanding LSTM Networks.”

control the flow of information into and out of the cell. The equations for these gates are:

$$\begin{aligned}
 i_t &= \sigma(W_{xi}x_t + W_{hi}h_{t-1} + b_i) \\
 f_t &= \sigma(W_{xf}x_t + W_{hf}h_{t-1} + b_f) \\
 o_t &= \sigma(W_{xo}x_t + W_{ho}h_{t-1} + b_o) \\
 g_t &= \tanh(W_{xg}x_t + W_{hg}h_{t-1} + b_g) \\
 c_t &= f_t \odot c_{t-1} + i_t \odot g_t \\
 h_t &= o_t \odot \tanh(c_t)
 \end{aligned}$$

where σ is the sigmoid function and \odot denotes element-wise multiplication.

Data Preparation: Phantom Cases

For an idealized evaluation of the neural network’s performance in dose calculation, simulations were initially performed on phantom models. These models were composed of cuboidal inhomogeneities with varying dimensions (2 mm to 14 mm on both z' and x' axes) and densities (0.1 RSP to 2.5 RSP), embedded within a water phantom (0.8 RSP to 1.2 RSP) as depicted in figure 1.5. 2500 such samples were generated, each paired with dose distributions computed via TOPAS MC simulations. To augment the dataset, considering the rotational symmetry about the beam axis, rotated (90°) CT and dose cubes were included, avoiding interpolation to preserve

the fidelity of MC simulations, thereby expanding the dataset to 10 000 training samples. This augmentation exploited the neural network’s sensitivity to input orientation to enhance training efficiency (Shorten and Khoshgoftaar, 2019). The samples were segregated into 2000 for testing (Experiment 1 - 1), 6000 for training, and 2000 for validation to optimize hyper-parameters. All simulations were conducted with an average of $\sim 1.1 \times 10^6$ histories, achieving statistical uncertainty below 1 %. Post-simulation, dose cubes were normalized by their fluence and then by the integral dose^a. Dose distributions from a standard pencil beam algorithm were used for baseline comparisons, normalized identically.

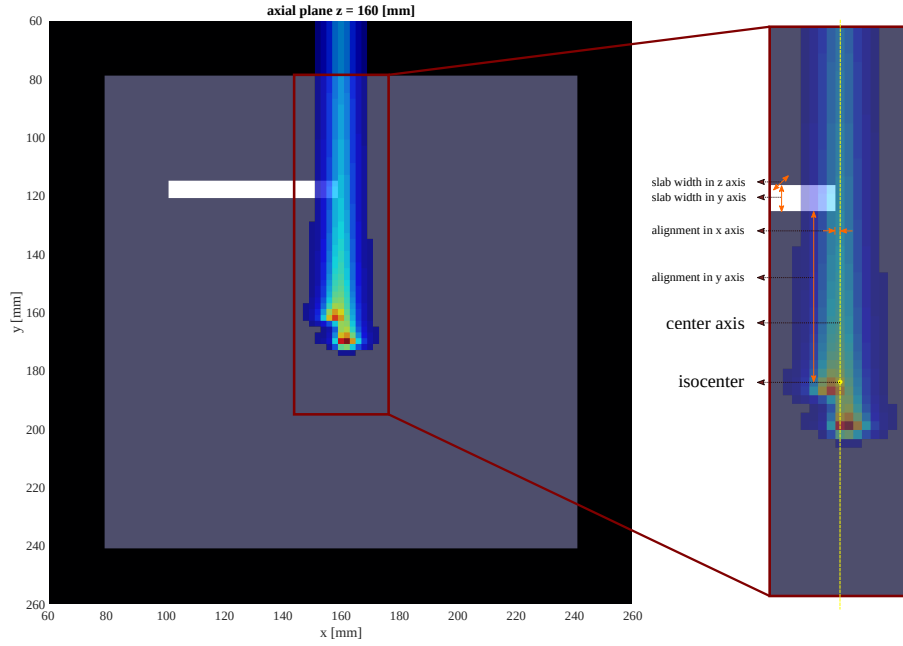


Figure 1.5: Setup for a phantom study; this configuration explores various geometric challenges by altering the dimensions of the slabs along the y and z axes, adjusting the slab’s positioning on the x and y axes, and modifying the densities of both the water and the slabs (Neishabouri et al., 2021b).

Data Preparation: Patient Case

To extend the study to more complex, real-world scenarios, lung patient cases characterized by significant anatomical heterogeneities, such as variations between normal and lung tissues and dense structures like the rib cage and spine, were included. For this purpose, 1000 lung case samples were generated, each derived from TOPAS MC simulations of the same patient, with variations in beam orientation in 5° increments from 0° to 355° and isocenter positions adjusted by 10 mm shifts along the z axis, as shown in figure 1.6.

These samples were simulated with 2.5×10^6 histories to ensure a statistical uncertainty of 1 % to 2 %. Augmenting these samples through 90° rotations, as previously described, increased the total to 4000 samples, which were distributed among 2400 for training, 800 for validation, and 800 for testing (Experiment 1 - 2). For uniformity, the original CT scans were downsampled to an isotropic 2 mm resolution

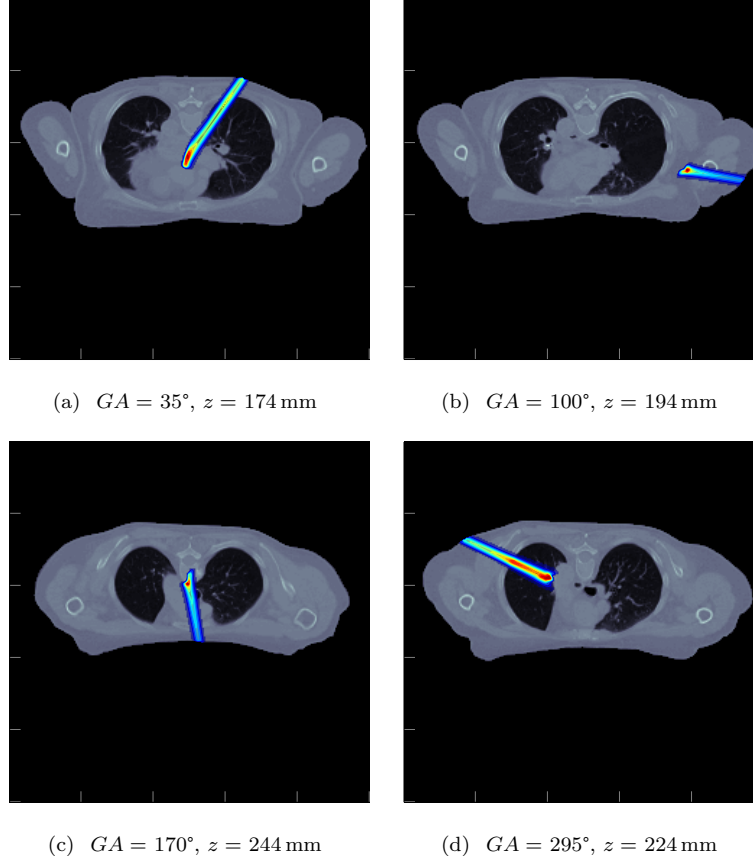


Figure 1.6: Setup for the lung case; this involves creating various geometric scenarios to enrich the training dataset by adjusting the gantry angles (GA) and altering the isocenter's position along the z axis (Neishabouri et al., 2021b).

The dataset prepared for the water box phantom facilitated the training of both a simple RNN and an LSTM network. The dosimetric performance of these networks was evaluated using the prepared test set. Table 1.2 showcases the results of the γ -analysis, which compares the dose estimations from both networks and PBA, against the ground truth MC calculations.

Both networks demonstrated high dosimetric accuracy in dose calculation, achieving mean pass rates exceeding 97.57 %. Notably, the LSTM network surpassed the RNN by a margin of 1.0 percentage point. This enhancement is particularly pronounced in scenarios with significant heterogeneities, where

Table 1.2: Comparison of γ -index values ([1 %, 3 mm]) among two neural network models and a PB algorithm against the MC simulation in a phantom setup (Experiment 1 - 1) (Neishabouri et al., 2021b).

| Model | Mean (%) | Std (%) | Min (%) | Max (%) |
|-------|----------|---------|---------|---------|
| RNN | 97.57 | 1.38 | 91.17 | 99.31 |
| LSTM | 98.57 | 0.84 | 92.37 | 99.81 |
| PB | 97.83 | 0.86 | 88.53 | 98.94 |

the LSTM’s superior performance becomes evident. An exemplary comparison, depicted in Figure 1.7, illustrates the LSTM’s capability to accurately better predict the bimodal Bragg peak behind a density interface, leading to an approximate 1.3 percentage point increase in the γ -index pass rate. This figure also compares the performance of the PB algorithm and includes a γ -analysis for a representative test sample. A notable observation is the distinct cut in the dose profiles behind interfaces (Schaffner, Pedroni, and Lomax, 1999), contrasting with the bimodal Bragg peak observed in both the ground truth MC and the estimated LSTM dose profile.

Given the comparable mean γ -index pass rates across all methods, the MAE and the MSE for the generated dose cubes relative to the ground truth MC simulations for the entire test set are also detailed in Table 1.3.

Table 1.3: MAE and MSE between the network models and the MC simulation in the phantom case (Experiment 1 - 1) (Neishabouri et al., 2021b).

| | LSTM | RNN | PB |
|------------------------|----------------------|----------------------|----------------------|
| MAE (Gy) | 3.3×10^{-3} | 6.1×10^{-3} | 3.8×10^{-3} |
| MSE (Gy ²) | 4.4×10^{-4} | 1.6×10^{-3} | 6.7×10^{-4} |

The LSTM network was further trained using the lung patient dataset. Table 1.4 provides a summary of the dosimetric comparison with the ground truth MC simulations, detailing the results from the γ -analysis, MAE, and MSE for the designated test set.

Figure 1.8 displays the performance of the trained LSTM network in comparison to the PBA on a selected test sample.

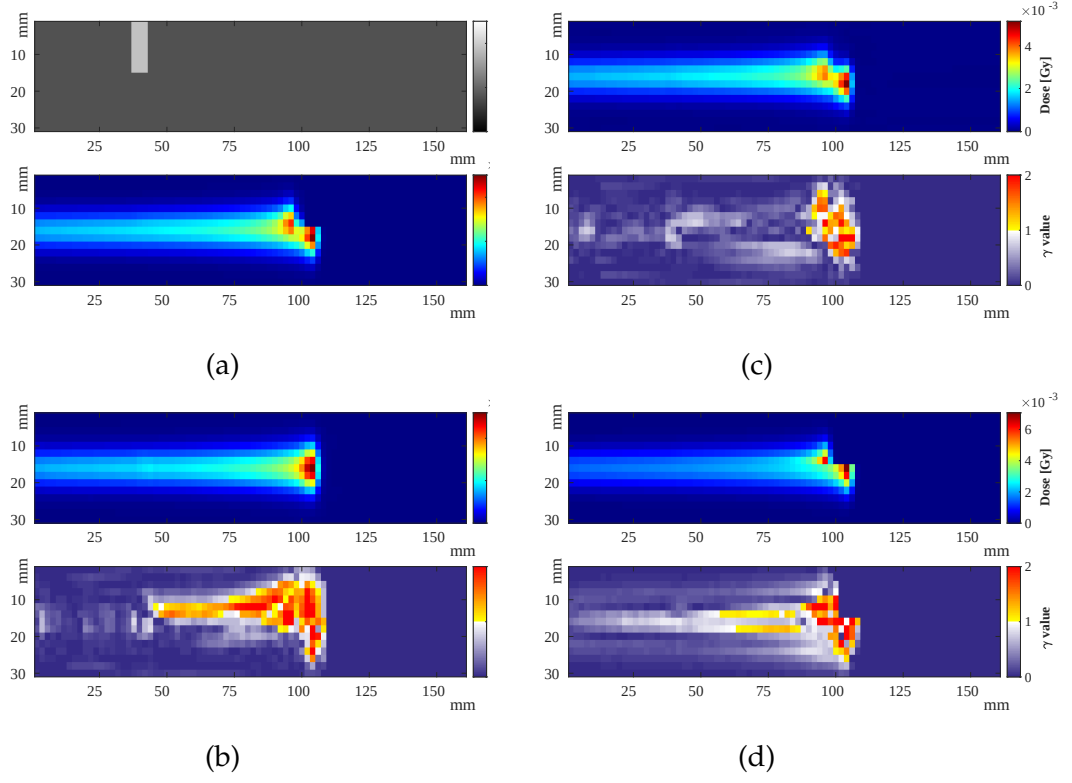


Figure 1.7: Comparative analysis of the (b) RNN, (c) LSTM, and (d) PB techniques against the (a) baseline MC calculation for a selected sample. The details include an energy of 104.25 MeV with a slab width of 6 mm and a 1.9 relative stopping power. The evaluation used a γ -analysis criterion of [1 %, 3 mm] (Neishabouri et al., 2021b).

Table 1.4: γ -index analysis [1 %, 3 mm]), MAE, and MSE of the LSTM model and PB algorithm compared to MC calculations for the lung patient case (Experiment 1 - 2) (Neishabouri et al., 2021b).

| | Mean (%) | SD (%) | Min (%) | Max (%) |
|------|----------------------|--------|------------------------|---------|
| LSTM | 98.50 | 1.00 | 93.93 | 99.82 |
| PB | 99.15 | 1.26 | 92.16 | 99.93 |
| | MAE (Gy) | | MSE (Gy ²) | |
| LSTM | 6.9×10^{-3} | | 6.8×10^{-4} | |
| PB | 4.7×10^{-3} | | 1.5×10^{-3} | |

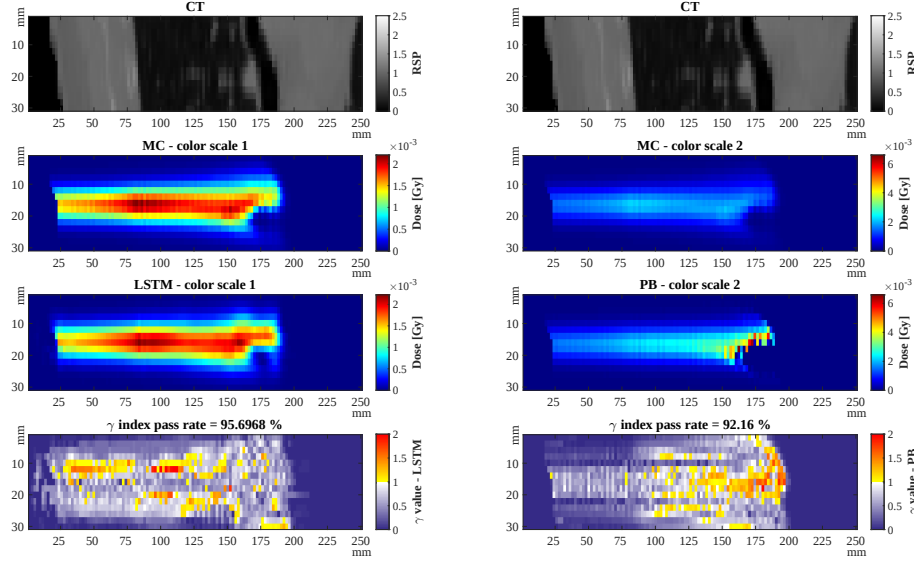


Figure 1.8: Dose estimation outcomes for a test sample (104.25 MeV) using the LSTM network (left) and the PB algorithm (right). The sequence starts with the patient's CT at the top, followed by the ground truth MC dose distribution, the LSTM's and PB's dose estimates, and their respective γ -index maps ([1 %, 3 mm]). Given the significant variation in estimated dose ranges, the reference MC displays in two distinct color scales tailored to the dose range of each comparison cube. To enhance visibility, the cubes are truncated along the longitudinal axis, showing 125 voxels (250 mm) rather than the full 150 voxels (300 mm) (Neishabouri et al., 2021b).

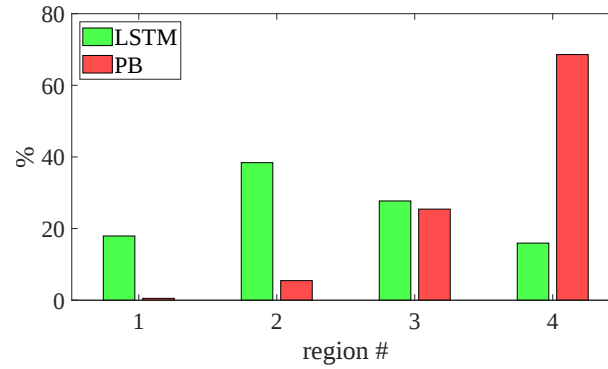


Figure 1.9: Distribution of voxels failing the γ -analysis criteria, segmented by their position along the longitudinal axis. These regions are evenly divided into quarters across the pencil beam's range (Neishabouri et al., 2021b).

To identify where the γ -index criteria violations occur, each pencil beam was divided into four equidistant quarter regions along their longitudinal range. The number of failed voxels in each quarter was counted and the distribution of these failing voxels along the longitudinal axis is presented as a percentage of the total for the entire test set in Figure 1.9.

1.5 AIMS

PB scanning in proton therapy is a key method for achieving precise tumor targeting while reducing radiation exposure to organs at risk. In this context, a PB is a focused stream of particles, accelerated to a distinct, predefined energy. A treatment plan typically comprises several thousand of PBs, each targeted to reach a certain depth within the tumor in the patient. Within each distinct energy layer, each PB is targeted to the tumor area—i.e. iso-energy surface (IES)—with fluence determined through inverse planning. Therefore, the effectiveness of these plans depends on accurately estimating the dose each constituent pencil beam delivers.

MC simulation, when based on accurate physical models (Tessonnier et al., 2016; Parodi et al., 2012) and sufficient particle histories, is the gold standard for dose calculation (Paganetti and Bortfeld, 2005). This high dosimetric precision results from the algorithm's capabilities to meticulously simulate each particle's trajectory, in a representative subset. However, the computational demand of MC often surpasses clinical time frames, particularly when multiple dose calculations are necessary, e.g., in robust optimization (Unkelbach et al., 2018), radiobiological studies (An et al., 2017), and arc therapy (Mein et al., 2021). To address this, streamlined analytical dose calculation methods such as PBA (Hong et al., 1996) have been developed. Despite being less accurate for complex geometries, the low computational costs and quantifiable uncertainties of these algorithms quickly led to their clinical adoption.

Concurrently, efforts to accelerate MC simulations are making significant progress. These include initiatives to parallelize the MC algorithm, utilizing the multi-thread capabilities of GPUs (Jia et al., 2012; Schiavi et al., 2017), and streamlining physical models (Lysakovski et al., 2021, 2024). However, implementing these advancements necessitates a certain level of simplification in the algorithm, differentiating them from standard MC simulation codes, e.g. FLUKA (Böhlen et al., 2014) and Geant4 (Jiang and Paganetti, 2004; Paganetti et al., 2008). The success of these efforts, however, is leading to the incorporation of fast-MC algorithms into routine clinical planning (Laboratories, 2017), particularly due to the proven reliability of MC methods compared to PBA.

Despite these advancements in accelerating MC simulations, and despite improvements in real-time imaging modalities, a significant gap remains in performing dose calculations for real-time adaptive proton therapy (APT) (Paganetti et al., 2021; Keall, Poulsen, and Booth, 2019). Current methods still struggle to provide the fast and accurate dose calculations necessary for effective real-time dose adaptation in therapeutic settings.

The inherently parallelizable nature of AI algorithms shows promise in this domain. Initial approaches can be categorized into methods that propose the *correction* of fast yet inaccurate algorithms (Wu et al., 2021). However, while these approaches are extremely fast, their performance is constrained by the underlying computational time of the primary algorithms they enhance. Moreover, as these methods estimate the entire proton field simultaneously, they fail to provide crucial data necessary for subsequent optimization, (re-)planning, and assessing radiobiological impacts in PT and personalized therapy.

This PhD thesis builds upon the previous work where various approaches were incorporated to investigate the feasibility of ANNs in estimating the single energy pencil beam dose distributions. To this end, a preliminary evaluation of RNN based networks were conducted. Moreover, the main finding of this previous work was the capability of LSTM networks in preserving information along the sequence length. The focus is on evaluating their suitability for dose estimation in the Beam's Eye View (BEV) domain or on the dose grid for full-field dose estimation. Specifically, this research aims to achieve the following objectives:

1. **Develop and Evaluate ANN Models:** The formulation of proton dose estimation as a supervised learning problem where CT input images are mapped to their respective dose distributions. This task will require creating different ANN architectures, especially Long Short-Term Memory (LSTM) networks, that can capture the spatial dependence and heterogeneity of patient anatomy. To train these models, a large dataset containing pre-calculated dose distributions produced through MC simulations will be used as ground truth. As part of this objective, the CC-LSTM model will be developed, taking into account the sequential nature in which doses are deposited while introducing the LSTM-133 model, which employs an innovative feature extraction technique aimed at enhancing output resolution. The performance evaluation should be done with thorough cross-validation methods to compare accuracy levels alongside computational efficiency between proposed ANN models and conventional dose calculation techniques, hence ensuring their technical validity as well as practicality within clinical settings.
2. **Error Quantification and Uncertainty Estimation:** To quantify the prediction uncertainties and model errors associated with dose estimations based on ANNs. This will be done by using Bayesian LSTM models that give probabilistic predictions, hence enabling the calculation of confidence intervals as well as identifying possible sources of mistakes. The research targets making the dose predictions more robust and dependable for clinical implementation through the inclusion of uncertainty

quantification. Techniques like Bayesian inference and statistical error metrics shall be applied to measure and analyze dose estimation uncertainties therefore enhancing explanation and comprehension of these estimations by ANNs.

3. **Real-time Adaptive Proton Therapy:** Investigate the ability of ANNs in meeting the primary goal of this study which is creating dose distributions with ultrafast execution times necessary for real-time adaptive proton therapy. This encompasses proposing models, workflows, and code libraries that will allow fast computation of dose distributions. The study aims to ensure that dose predictions can be made on a real-time basis by concentrating on improving the computational efficiency of the ANNs while maintaining a comparable accuracy performance relative to MC simulations.

MATERIAL AND METHODS

The objectives of this PhD thesis were conducted in three phases. The initial phase evaluated the feasibility of an AI-based dose calculation approach in particle therapy for full-field dose estimation. Building on the feasibility analysis presented in Section 1.4, the established framework was optimized, and additional data were generated to investigate the generalization of the proposed method to other therapeutic energies and unforeseen patient scenarios.

In the second phase, detailed in section 2.2, model robustness was evaluated by exploring methods to quantify the inherent uncertainties. This involved the integration of Bayesian neural networks into our deterministic LSTM-based framework, allowing for an assessment of prediction confidence levels across various scenarios. This Bayesian approach enhances the LSTM model by incorporating model uncertainty, which is essential for clinical decision-making and quality assurance. The Bayesian LSTM (B-LSTM) models provide dose predictions and uncertainty quantification, enabling a better informed evaluation of treatment plans.

The final phase of our study, outlined in section 2.3, focuses on the primary objective of this work: the development of a fully AI-based dose engine capable of performing comprehensive full-field dose estimations. Leveraging insights from the feasibility study and the uncertainty quantifications, this phase aimed to upscale the mono-energetic simulations' approach to full-field dose distribution estimations. Furthermore, the challenge of simulating real-world clinical scenarios involving diverse energies and complex patient geometries is addressed by leveraging a custom-designed AI model.

The model was meticulously developed to integrate and process a holistic dataset based on previously delivered treatment plans to meet the strict accuracy and run time requirements essential for real-time APT. This approach not only ensures the robustness of the simulations but also guarantees that the dose predictions are both precise and computationally efficient, addressing the critical demands of modern RT. The details of this AI-driven model and its operational capabilities are discussed in section 2.3.

2.1 FEASIBILITY OF RNN-BASED MODELS FOR FULL-FILED DOSE ESTIMATION

2.1.1 Problem parameterization

The foundational aspect of our research is the dose calculation for a single proton pencil beam within IMPT. This focus allows us to examine the inherent properties of LSTM-based dose prediction, avoiding the obscuring effects of averaging that occur in treatment plans utilizing thousands of beams (Agostinelli, 2003; Perl et al., 2012; Wieser et al., 2017).

MC simulations are regarded as the most accurate, serving as the ground truth for our model's training. These simulations were implemented using the Topas (TOol for PArticle Simulation (Perl et al., 2012)) wrapper for Geant4, alongside PBA dose calculations performed using the matRad software toolkit (Wieser et al., 2017)¹.

The 3D dose distribution \mathcal{D} of a single pencil beam is influenced by the particle's initial phase space \mathcal{P} , consisting of position and momentum, and the patient's 3D geometry \mathcal{G} :

$$\mathcal{D} = f(\mathcal{P}, \mathcal{G}) \quad (5)$$

To facilitate neural network training for accurate proton dose calculations, it is crucial to establish the mapping f from the initial particle phase space \mathcal{P} and the patient geometry \mathcal{G} to the resulting dose distribution \mathcal{D} . We simplified this task by limiting the learning to mono-energetic PBs, thereby reducing the complexity and increasing the data sampling density across potential geometrical and dosimetric scenarios. Training for additional energies is treated as a distinct learning task.

Adapting to a beam's eye view, the dose calculation becomes simpler as the deposition aligns along the z' -axis (as shown in figure 2.1). This setup allows for a finite range and lateral extent of the particles, facilitating the clipping of the region of interest to a manageable size for computational efficiency. The models use an isotropic resolution of 2 mm with $m = 15$ voxels laterally and $l = 150$ voxels longitudinally for each patient setup:

¹ <http://www.matrad.org>

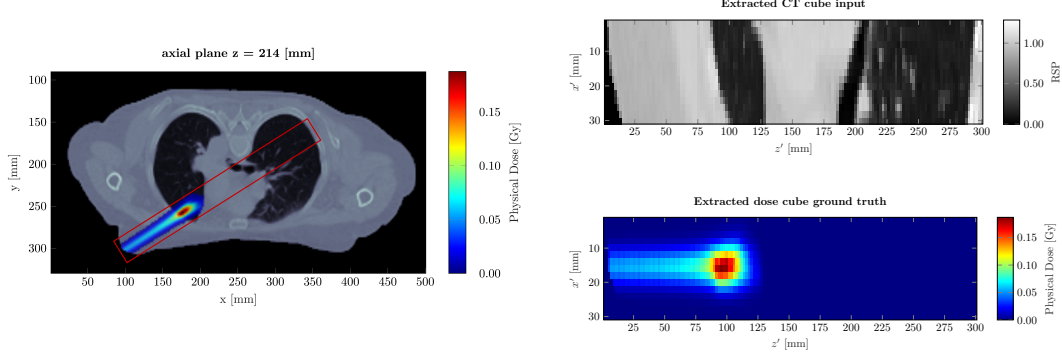


Figure 2.1: (a) Distribution of a dose from a single pencil beam with an initial energy of 104.25 MeV, entering from a gantry angle of 240° and superimposed on the patient's CT scan. The highlighted area is outlined in red. (b) The corresponding CT slice and (c) the dose distribution from the perspective of the beam's eye view coordinate system (Neishabouri et al., 2021b).

$$\mathcal{G}_i \in \mathbb{R}_+^{l \times m \times m} \rightarrow \mathcal{D}_i \in \mathbb{R}_+^{l \times m \times m} \quad (6)$$

Instead of relying solely on HU maps, our approach uses RSP maps, calculated based on the reduction of range in the patient relative to water, which are standard in conventional PB algorithms for density adjustments (Schaffner, Pedroni, and Lomax, 1999; Wieser et al., 2017). The RSP values, ranging from almost 0 (Air) to 2.5 (dense bone), are converted using HU look-up tables to the corresponding water densities utilized in MC simulations.

2.1.2 Model Architecture: Deterministic LSTM Model

The challenge of particle dose calculation is characterized by a unique geometrical aspect, motivating the adoption of a specialized neural network architecture. The deposition of dose in proton therapy occurs in a distinct *upstream-to-downstream* sequence. This means that energetic protons generally move in a forward direction with minimal lateral deviation until they reach their stopping point. Such directional behavior facilitates the representation of 3D data as a series of 2D slices, as depicted in figure 2.2.

This directional sequence in dose calculation shares notable similarities with the spatio-temporal analysis found in video processing, particularly in tasks like action recognition. In these tasks, models extract spatial details from individual frames and analyze temporal dynamics to understand object move-

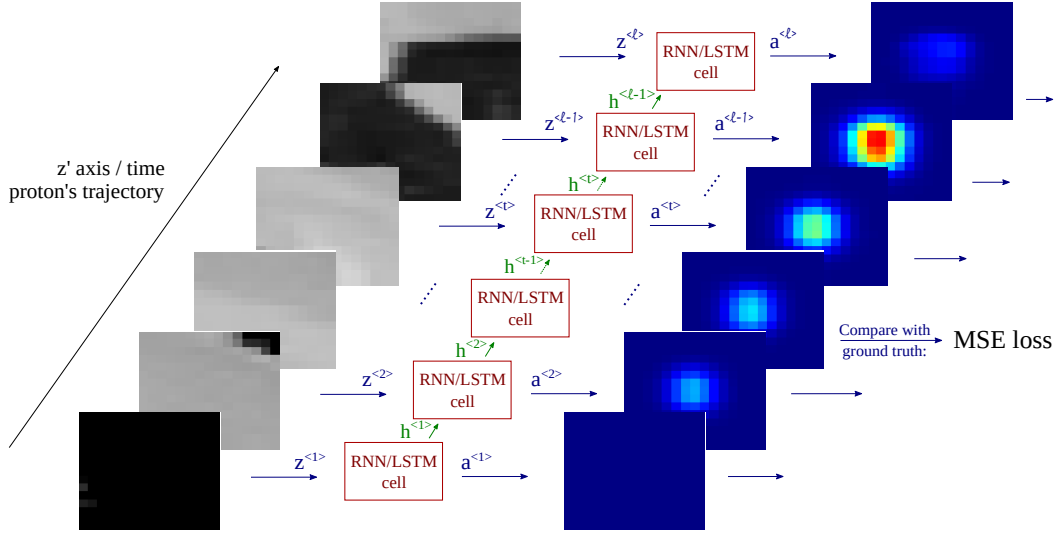


Figure 2.2: This illustrates the sequential and spatial-temporal approach to solving the proton dose calculation task. Each $m \times m$ segment of the input is transformed into a one-dimensional input array $z^{<t>}$. This array feeds into an RNN/LSTM cell, which produces both a hidden state $h^{<t>}$ and an output $a^{<t>}$. The hidden state informs the next slice processing for a total of l slices, and the output connects to a dense neural network that reconstructs an $m \times m$ output segment. This output is then evaluated against the actual ground truth using the mean squared error loss (Neishabouri et al., 2021b).

ments. Similarly, modeling the movement of protons through patient tissue involves tracking their trajectory through varying densities, particularly influenced by the initial material they traverse, thereby creating a causal relationship from upstream to downstream in the 3D volume. High-gradient regions in RSP values, such as transitions at bone interfaces or air cavities, are critical as they predominantly influence dose delivery at the proton's stopping point as shown in figure 2.3. Thus, any model designed for simulating particle dose deposition must extract these spatio-temporal features and accurately convey the influence of these heterogeneities along the proton path.

To handle long sequence dependencies effectively, a model capable of transmitting information throughout the sequence is required. RNNs, with their hidden states, link multiple conventional neural networks in a series, creating a framework suited for many-input-to-many-output configurations. LSTM networks are particularly capable of managing long information sequences due to their internal gating mechanism. Furthermore, unidirectional LSTM models are exceptionally well-suited for modeling the upstream-to-downstream proton propagation, thus avoiding unnecessary dependencies from downstream to upstream, significantly reducing the model's parameter count.

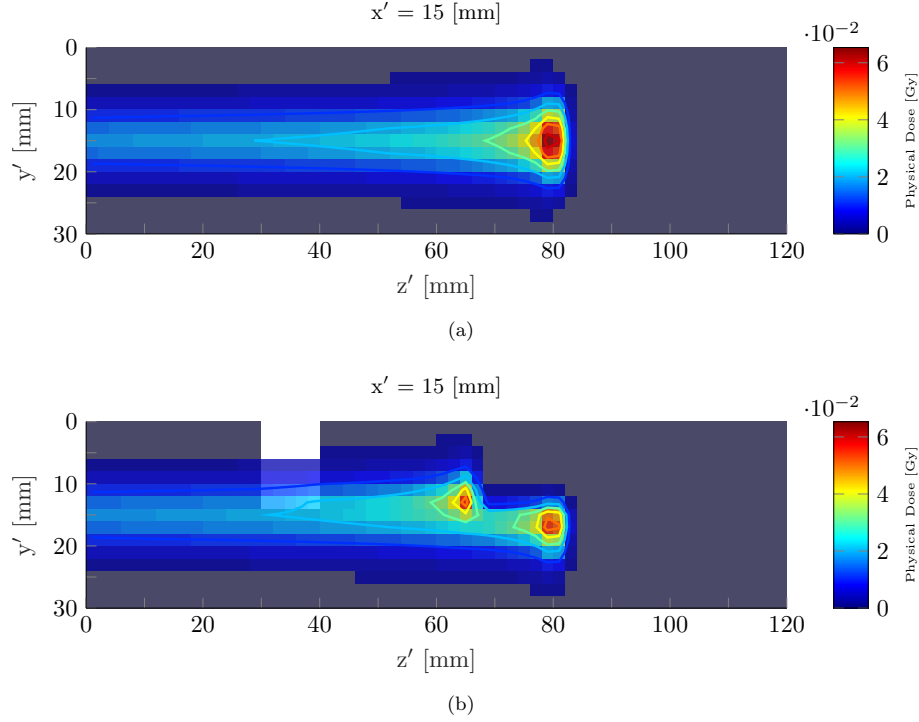


Figure 2.3: Exploring the impact of heterogeneity on the dose profile of a pencil beam. This beam comprises 10^6 protons, each with an initial energy of 104.25 MeV, traversing through (a) water with a RSP of 1.0. (b) The rectangular heterogeneity with an RSP of 2.5 measures 10 mm in width along the z' axis and 14 mm along the x' axis, positioned 1 mm from the center of the proton beam. This heterogeneity primarily affects the dose profile, notably producing a dual-peaked Bragg peak within the range of approximately 60 mm to 80 mm, or 20 mm past the heterogeneity (Neishabouri et al., 2021b).

2.1.3 Model Training

The development of LSTM networks has evolved significantly, driven by substantial advancements in neural network technologies (Gers, Schmidhuber, and Cummins, 2000; Gers and Schmidhuber, 2000; Graves and Schmidhuber, 2005). The LSTM model was implemented using the Pytorch framework². This framework facilitated the model's training with an Adam optimizer (Kingma and Ba, 2014), which is advantageous for its adaptive learning rate capabilities, crucial in managing the large dataset employed in this study.

The LSTM architecture designed for this study consists of a single layer with 1000 neurons, which feeds into a backend fully connected neural network. This network comprises a hidden layer of 100 neurons and an output layer

² <https://pytorch.org/docs/stable/generated/torch.nn.LSTM.html>

sized to produce m^2 neurons, matching the output dimensions required for our dose slices. Tanh activation functions were employed in the LSTM, with ReLU functions used in the backend to facilitate non-linear learning processes. This design was optimized to normalize the dose cubes between 0 to 1 and to handle RSP input cubes within the range of 0 to 2.5, thus maintaining the physical properties of the input data within expected thresholds.

The learning rate was set to 10^{-5} throughout the training phase to balance the training speed and model accuracy. Our observations indicated that the model reached a plateau in test loss after about 100 epochs, marking the start of overfitting beyond this point. The training sessions typically lasted between 3 to 4 hours on a Geforce GTX 970 GPU, emphasizing the computational demands and efficiency of our model given the complexity of the data processed. The model's efficacy was empirically validated through its performance across multiple test scenarios, where it demonstrated robustness without significant decreases in test loss, indicating stable learning and generalization capabilities. This operational efficiency is critical, particularly when considering the application of such models in clinical APT, where both accuracy and speed have to be optimized.

Patient Cases and Experiment Design

Based on the initial successful results outlined in section 1.4, the LSTM model's generalization to new patients was tested by evaluating its performance on five different lung cancer patients not included in the training set. For each patient, 200 pencil beams with varied gantry angles and isocenter positions were prepared, and their doses were calculated via MC simulations (Experiment 1 - 1).

Furthermore, to verify the model's adaptability to different proton energies within a clinical setting, we created datasets for three distinct energy levels: low-range (68.33 MeV), mid-range (104.25 MeV), and high-range (134.22 MeV) proton beams (Experiment 1 - 2). Each energy level consisted of 1000 randomly chosen samples, handled similarly to the patient cases described above. For the high-energy beams, the dose calculation cubes were extended to $l = 200$ voxels in the longitudinal direction, culminating in 12 000 samples post-augmentation, split following the same ratio as earlier experiments.

2.1.4 *Incorporated Metrics*

To assess the accuracy of 3D dose distributions, we conducted a gamma analysis (Low et al., 1998) applying a criterion of 1 % for dose difference (DD) and 3 mm for distance-to-agreement [1 %, 3 mm]). This was applied across both phantom and patient case studies. The 1 % dose difference threshold was specifically chosen in light of the MC simulations' statistical uncertainty, which is also around 1 %. This ensures a rigorous evaluation of the dose calculations, avoiding the analysis of noise characteristics inherent in MC simulations. The 3 millimeter distance-to-agreement criterion was selected to guarantee a search area extending at least one voxel in each direction, considering the 2 millimeter resolution of our dose cubes.

We visualize performance using gamma index distributions, highlighting how well the predicted dose distributions conform to those from MC simulations, particularly in response to variations in tissue density along the trajectory of the pencil beams. Additionally, we quantify the accuracy of our dose predictions using the gamma-index pass rate, the Mean Absolute Error (MAE), and the Mean Squared Error (MSE). These metrics help to condense the comparative analysis of the 3D dose distributions into single values, greatly simplifying the evaluation process when dealing with thousands of samples across training, validation, and testing phases.

2.2 UNCERTAINTY QUANTIFICATION: BAYESDOSE

Recent advancements in AI have introduced strategies that either enhance the accuracy of fast but approximate dose calculation methods or entirely replace traditional algorithms with AI-driven dose calculation models (Kontaxis et al., 2020; Martinot et al., 2021; Neishabouri et al., 2021b; Pastor-Serrano and Perkó, 2022; Pastor-Serrano et al., 2023). These approaches are generally trained using datasets generated from MC simulations to ensure high accuracy (as described in section 2.1.3). As it was reported in section 2.1.3, the training phase, while not overly time-restricted, allows for the development of tailored models for specific energies or patient categories. Subsequently, these models can perform dose calculations during clinical application, potentially outperforming conventional numerical methods in speed while maintaining accuracy comparable to MC calculations.

A critical aspect of adopting these AI models in clinical practice is their "explainability," particularly regarding the accuracy of their predictions (Barragán-Montero et al., 2022). Unlike systematic errors in pencil-beam calcu-

lations, which can be traced back to specific assumptions and simplifications, the interpretability of errors in neural network predictions is less straightforward.

Bayesian Neural Networks (BNN) offer a robust framework to address these concerns. In BNNs, the model parameters, such as weights and biases, are treated as stochastic variables defined by parametric probability distributions learned during training (Gal, 2016; MacKay, 1992). After training, sampling from these distributions can produce a variety of neural networks, each providing different dose predictions for the same input. This capability enables the derivation of statistical insights into the model's predictive accuracy, such as the expected dose and its standard deviation, for individual patient CT images.

Building on previous research that employed LSTM networks for individual proton beamlet dose calculations (described in section 2.1 and in (Neishabouri et al., 2021b)), this section introduces "BayesDose," a Bayesian LSTM (BLSTM)-based model. This model demonstrates how BNNs can produce statistical dose predictions, helping to alleviate concerns regarding the unclear nature of AI models in clinical settings and supporting quality assurance and decision-making processes in AI-based dose calculation systems.

2.2.1 Dataset

The datasets utilized for the BayesDose model were previously compiled and are detailed in sections 1.4 and 2.1.3. Utilizing these datasets allows for a consistent comparison between the BayesDose model and its deterministic counterpart under equivalent conditions.

The water phantom data is incorporated to study the uncertainty estimation under extreme, notable scenarios, serving as an intuitive and explainable feasibility check for the developed BayesDose. Furthermore, the patient's lung case was incorporated for a real patient scenario, exhibiting notable anatomical heterogeneity between normal tissue, lung tissue, and bony anatomy.

Since different geometric problems were created by altering the beam orientation, the training data resulted in cases where the gantry angle was oblique in relation to the CT axis. Consequently, these cubes experience strong wavering behavior due to the occurring interpolation errors (example can be seen in figure 3.6). However, it was decided to include these samples with high interpolation artifacts to validate the comparison to the deterministic LSTM

and to analyze how they interfere with the model uncertainty predicted by the BayesDose model.

Finally, the generalizability data, which focuses on the model’s ability to generalize across different energy levels, is incorporated. Additional generalization testing involved dose simulations for four more lung patients, enhancing our understanding of the model’s adaptability to diverse patient anatomies. Each scenario involved 200 randomly chosen proton beamlets, varying in RSP values, which provided a robust basis for evaluating the predictive consistency across different patient geometries as shown in Figure 2.4.

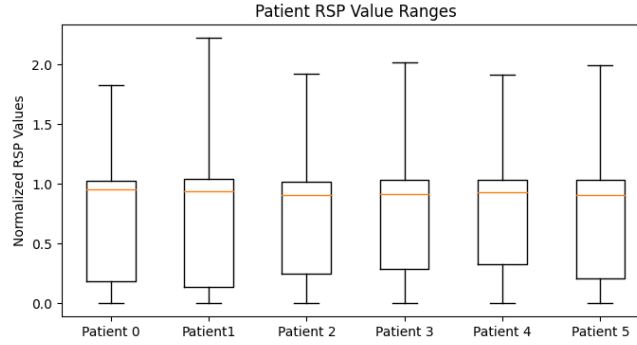


Figure 2.4: Box plot illustrating the variance in RSP values across the tested patient geometries. Patient 0 represents the initial lung patient detailed in 2.1.3.

2.2.2 BayesDose Implementation

The BayesDose model was developed using Python 3.10.5 and utilizes the PyTorch 1.12.0 framework, based on the structure from section 2.1.2. The Bayesian layers within the network are implemented through the Blitz framework 0.2.8 (Esposito, 2020). Enhancements have been introduced to optimize data handling and training, facilitating the use of larger batch sizes and improving training speed and convergence.

A pivotal aspect of our model is the BLSTM layer, known as BayesianLSTM in the Blitz framework. This layer transforms each weight and bias in the conventional LSTM layers into stochastic variables defined by probability distributions. This transformation introduces stochasticity to the cell in each iteration, impacting the dynamics of every iteration in the BLSTM layer. This structure is depicted in figure 2.5.

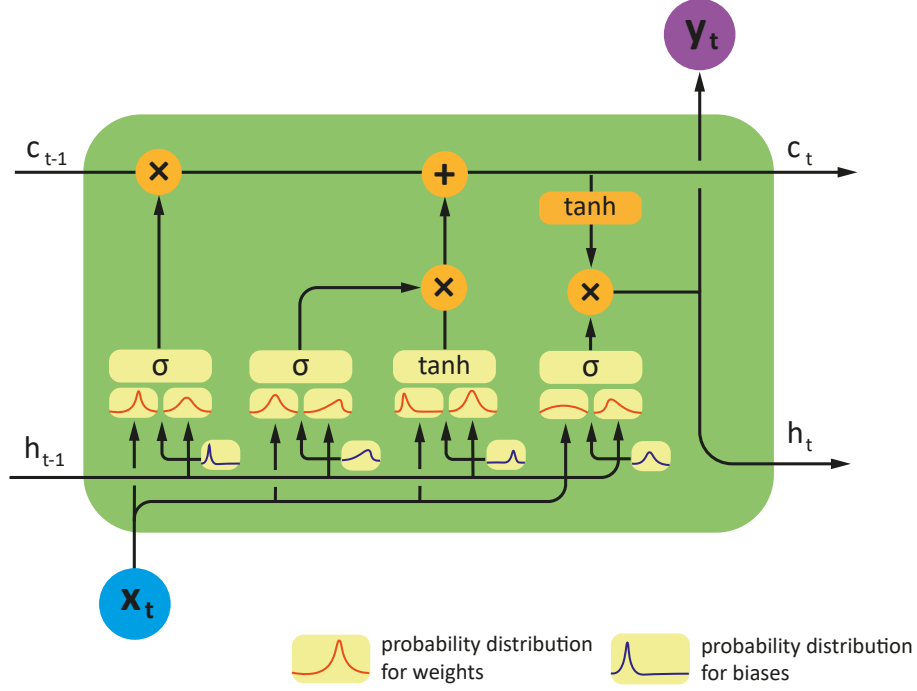


Figure 2.5: Schematic of a BLSTM cell illustrating the distribution-based stochastic parameters, including weights and biases. Shown are the input x_t , the hidden state h_t (also the output y_t), and the cell state c_t with activation functions sigmoid (σ) and tanh (\tanh).

The Bayesian inference process used in our BLSTM layers employs Bayes' theorem to evolve the prior distribution of the weights $p(w)$ toward a posterior distribution $p(w|D)$, where D represents the observed data:

$$p(w|D) = \frac{p(D|w)p(w)}{\int p(D|w)p(w)dw} . \quad (7)$$

The data likelihood $p(D|w)$ and the evidence $p(D)$ are crucial, yet directly computing $p(D)$ is often impractical.

To circumvent this computational challenge, the Blitz framework adopts variational inference to estimate the posterior distribution $p(w|D)$ as a variational posterior $q_\psi(w)$. This approach converts the complex integration task into an optimization problem, focusing on minimizing the negative Expected Lower Bound (ELBO):

$$\text{ELBO} = -\mathbb{E}_{q_\psi(w)}\{\log p(D|w)\} + \text{KL}\{q_\psi(w)||p(w)\} , \quad (8)$$

aiming to balance the model and data fit while managing the regularization via the Kullback-Leibler (KL) divergence. This dual objective helps balance the model's accuracy against the risk of overfitting.

Training involves updating the probability distributions of the network's weights and biases via the *Bayes by Backprop* method (Blundell et al., 2015), allowing the network to derive an ensemble of predictions from sampled parameters:

$$\mathbf{w}_i = \mathcal{N}(0, 1) \times \log(1 + \rho_i) + \mu_i, \quad (9)$$

$$\mathbf{b}_i = \mathcal{N}(0, 1) \times \log(1 + \rho_i) + \mu_i, \quad (10)$$

which reflects the conversion of stochastic elements into operational parameters for the network's predictive tasks.

To fine-tune the model parameters, including the stochastic components, a hyperparameter optimization was conducted using the Optuna Framework (Akiba et al., 2019), involving 200 trials that determined optimal settings for the network's performance characteristics. This optimization focused on the mixture components and the variance settings essential for the Gaussian prior scale mixture, resulting in specific values that enhance the predictive capability and efficiency of the model.

2.2.3 *BayesDose Architecture*

The architecture of our BayesDose model closely aligns with the deterministic model detailed in section 2.1.2. Each 15×15 slice from a sequence of lateral 2D dose slices is reshaped into a vector of 225 elements and fed into the BLSTM cell to process the input data. Both the cell and hidden states within each BLSTM cell consist of 1000 neurons. These neurons update their states with each slice processed, sequentially passing on updated information to the next slice in the sequence. This sequential processing ensures that each image slice is analyzed in the context of the information processed from previous slices.

The output from the BLSTM layer comprises a sequence of 1000-element vectors, where each vector corresponds to a processed image slice. Subsequent fully connected layers convert these vectors back to their original dimensionality of 225, which are then reformed into 15×15 slices.

Significant modifications are evident in the back-end network of the BayesDose compared to the deterministic model. Specifically, traditional linear layers are replaced with BayesianLinear layers, and the ReLU activation functions are substituted with SiLU activation functions (Elfwing, Uchibe, and Doya, 2018) to facilitate smoother posterior probability distributions for the weights and biases.

The redesigned back-end network consists of a Bayesian linear input layer with 1000 neurons, followed by a hidden layer of 100 neurons, and culminating in a Bayesian output layer of 225 neurons. SiLU activation functions are applied between these layers to enhance non-linear processing capabilities.

The BayesDose model samples weights and biases from the optimized variational posterior distribution to generate predictions and estimate uncertainty. This sampling produces a series of predictions for the same input sequence, enabling the computation of statistical measures such as the mean and standard deviation. These statistics represent the consolidated output of the network and its associated uncertainty.

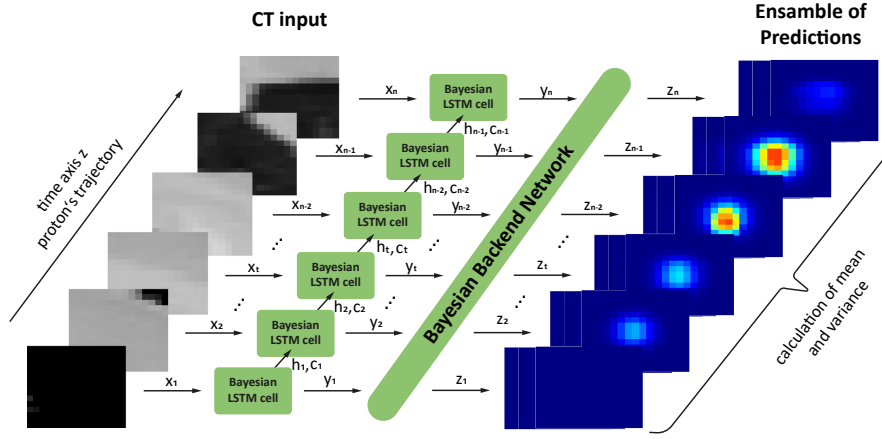


Figure 2.6: Illustration detailing the data flow within the BayesDose model. The depth dimension z is utilized as the time dimension t for the BLSTM. Each 15×15 slice is transformed into a vector x_t , which the BLSTM cell processes to produce cell states c_t , hidden states h_t , and the resulting outputs y_t . These outputs are then input into a fully connected Bayesian network, which generates the final outputs z_t .

2.2.4 BayesDose Training

The loss function utilized for training BayesDose comprises the ELBO loss function, which includes the MSE and the KL divergence components (refer to equation 8). To accommodate the Bayesian characteristics of the model, the average ELBO loss from three calculations is employed for backpropagation. Low sample numbers or even a single sample are typically adequate for Bayes by Backprop, balancing ELBO noise and training duration.

The MSE is approximately a factor of $1 \cdot 10^5$ smaller than the KL divergence loss, causing the latter to dominate the loss function. This disparity leads

to a slow convergence as the optimizer prioritizes minimizing the KL divergence. To ensure both components impact the optimization process equally, the MSE is scaled up by the aforementioned factor during training, referred to as scaled MSE (SMSE).

As proposed by (Smith, 2017), a learning rate range test is conducted on the SMSE and KL divergence loss separately to determine the optimal learning rates for each component during training. This involves progressively increasing the learning rate while monitoring the losses to identify the point just before each loss increases, indicating the optimal learning rates for each component. The OneCycleLR policy from PyTorch (Smith and Topin, 2019), which includes a warm-up phase with increasing learning rates followed by an annealing phase with decreasing rates, is implemented to manage the learning rates effectively.

For the phantom dataset, the starting value of the OneCycleLR scheduler is set at $1.3 \cdot 10^{-3}$, enabling rapid convergence of the SMSE loss and gradual convergence of the KL divergence loss. The learning rate increases to a maximum of $6.5 \cdot 10^{-3}$, the highest value where the SMSE loss remains stable, and then reduces to $4.13 \cdot 10^{-5}$ in the latter half of the training to facilitate total loss convergence. For the patient dataset, slightly higher learning rates are permissible, starting at $3.3 \cdot 10^{-4}$, reaching a maximum of $1 \cdot 10^{-3}$, and decreasing to a minimum of $3.3 \cdot 10^{-6}$.

The training regimen employs the Adam optimizer (Kingma and Ba, 2017) with the OneCycleLR scheduler and a batch size of 32, which is the most effective in balancing the learning rates between the loss components. The network undergoes training for 600 epochs for the phantom dataset and 1000 epochs for the patient dataset, after which no substantial improvements in performance are observed. The training duration for the phantom dataset approximates 10 hour and for the patient dataset around 11 hours on an NVIDIA RTX A6000 GPU. The progression of the total loss and its components during training is depicted in figure 2.7, illustrating the initial reduction in SMSE followed by a gradual decrease in KL divergence.

2.2.5 *Incorporated Metrics and Experimental Design*

Evaluating the BayesDose model's performance involved generating 100 prediction samples to accurately estimate the mean and standard deviation, ensuring efficient prediction times. The model's performance metrics were categorized into two types: accuracy metrics, based on the ensemble mean, and uncertainty metrics, derived from the ensemble standard deviation.

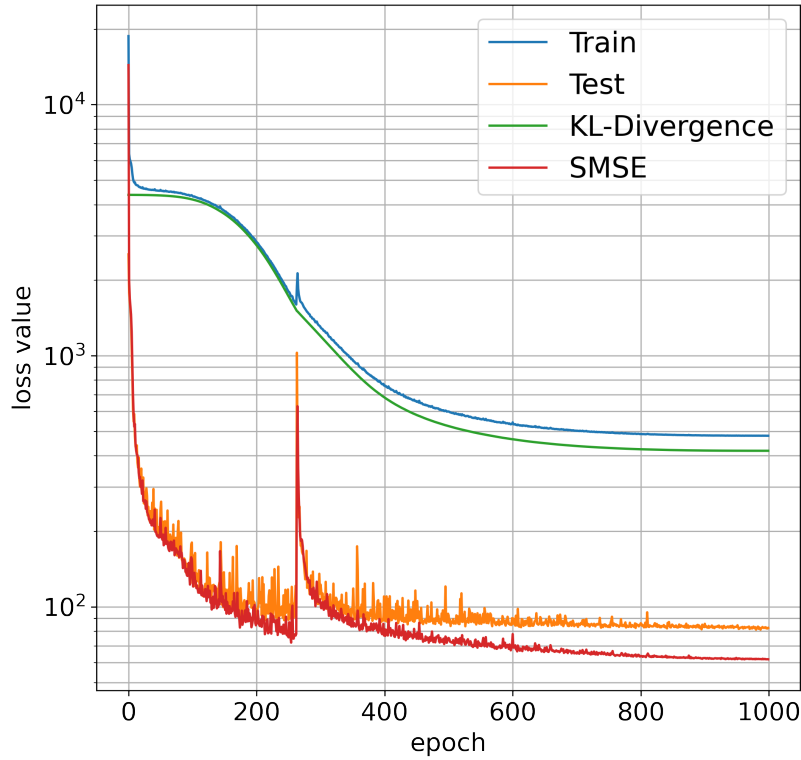


Figure 2.7: Evolution of the loss function across the 1000 epochs during patient training. It illustrates the training loss, comprising KL-divergence and squared mean squared error (SMSE) depicted separately. The test loss, evaluated every 10 epochs and based solely on SMSE, is derived from individual prediction samples.

For assessing the predictive accuracy, a comprehensive gamma analysis (Low et al., 1998) using a [1 %, 3 mm] criterion was conducted. To mitigate the issue of gamma pass-rates clustering at 100 %, doses under 0.1 % of the maximum were excluded from the analysis, establishing a stricter evaluation criterion than that used in section 2.1.4. This adjustment addresses the potential for near-zero values, which could artificially inflate pass rates due to the neural network’s numerical inaccuracies. In patient-specific cases, the gamma computation utilized 10 interpolation points, and the results for the original model were recalculated using this more rigorous standard. In addition to gamma pass-rates, the model’s performance was quantified by reporting the MSE and MAE of the DD.

To examine the quality of the uncertainty predictions, the proportion of voxels inaccurately predicted within their respective $n\sigma$ confidence intervals was

analyzed, with σ representing the standard deviation of the predicted dose ensemble within a voxel, for n values ranging from 1 to 5. Although the predicted dose does not conform to a Gaussian distribution, comparing the relative frequency of voxels deviating beyond $n\sigma$ with Gaussian expectations aids in assessing whether the model tends to significantly overestimate or underestimate its confidence. Similar data preparation methods were applied as in the gamma analysis, disregarding DDs less than 1 % and differences below 0.1 % of the maximum dose.

For comparative evaluation, the Bayesian LSTM network was benchmarked against the deterministic LSTM model from the previous section to discern any performance disparities stemming from architectural differences, such as using SiLU activation functions or the extended training protocols involved in transitioning to Bayesian Network layers.

The experimental setup for evaluating the BayesDose model involves five distinct experiments, where as described earlier, each designed to assess different aspects of model performance and robustness. These experiments utilize the datasets and models described previously.

EXPERIMENT 2 - 1: In this initial experiment, the model is trained, tested, and evaluated using the phantom dataset as detailed in section 1.4. The data is divided into a 60-20-20 distribution for training, validation, and testing, respectively, mirroring the approach in section 2.1.3.

EXPERIMENT 2 - 2: The model undergoes a similar evaluation process on the dataset for an individual lung patient, as described in section 2.1.3, with the same 60-20-20 data split applied for training, validation, and testing phases.

EXPERIMENT 2 - 3: This experiment focuses on assessing model performance across different initial energies by training, validating, and testing the model separately on the low-range and high-range energy datasets as outlined in section 2.2.1.

EXPERIMENT 2 - 4: The generalizability of the model is tested by applying the trained model from experiment 2 - 2 to four additional patients, assessing how well the model performs on new, unseen patient data.

EXPERIMENT 2 - 5: This experiment tests the model's adaptability to new information by initially training on data from patient 0 and then fine-tuning on data from patient 5, who presents a wider range of RSP values. The model is then re-tested on the remaining patients to compare results and assess any changes in performance.

Experiments 2 - 4 and 2 - 5 additionally explore how predicted uncertainty correlates with various gamma criteria, aiming to establish decision-making guidelines based on model output and associated uncertainty. Experiment 2 - 5 extends this analysis to examine the effects of re-training the model on previously unseen data.

2.3 TOWARD REAL-TIME ADAPTIVE PROTON THERAPY

This section thoroughly outlines and compares the development and benchmarking of the full field dose engine and the core AI model with state-of-the-art methodologies. Specifically, this section will highlight the advantages of RNN-based models, which, despite their earlier development, continue to demonstrate high efficiency in managing complex computational challenges such as dose calculation in particle therapy. Building on the strengths of RNN models, the steps that led to the design of the CC-LSTM model will be described. Additionally, by leveraging the run time efficiency of the pre-implemented LSTM cell module in PyTorch, LSTM-133 model is introduced, which formulates the problem efficiently to deliver sub-second run time performance.

2.3.1 CC-LSTM Architecture

As detailed in section 2.1.2, the rationale for representing input and output as a sequence of 2D slices in the BEV was explained. While physically equivalent, these representations are interpreted differently in the context of AI, particularly regarding feature extraction and subsequent architectural decisions. The former involves extracting spatial features across all axes, whereas the latter requires a spatio-temporal feature scheme—spatial in the lateral dimensions and temporal in the longitudinal dimension, as illustrated in figure 2.8. A critical distinction is the coherence along these axes; temporal features involve changes or patterns over time, necessitating an algorithm that can maintain this information. Spatial features related to object positions in space suggest the need for algorithms focusing on local receptive fields.

CNNs are utilized to capture physical characteristics such as size, shape, and orientation of heterogeneities. They can also capture spatial hierarchies. Advanced forms of RNNs, such as LSTM networks, attention-based Transformers (Vaswani et al., 2023), and Retnets (Sun et al., 2023), are considered for sequence handling and temporal information retention.

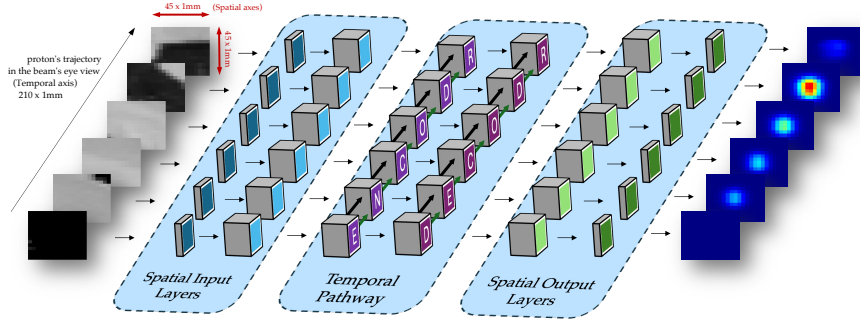


Figure 2.8: This diagram illustrates the CC-LSTM architecture, unfolded to show its three main components: the spatial input layers that compress each BEV slice from 3-to-1 before feature extraction via convolutional layers; the temporal pathway, which extends the features of each slice along the longitudinal axis to develop correlations up to time step t ; and the spatial output layer, where outputs from the temporal module for each time step are transformed into dose distributions for the corresponding time step. Importantly, the network maintains the spatial integrity of features by avoiding 'flattening' throughout the progression.

While Retnets have not been explored in this study, the decision to forego Transformer-based architectures was due to their extensive parameterization and high resource demands (Sun et al., 2023). Emphasizing the need for an architecture that balances robust performance with efficient inference speed is crucial. The dual nature of this task, the availability of data versus the need for super-fast, full-field dose inference, presents both opportunities and challenges. Each model modification must be meticulously considered, as their impacts are significantly magnified by the number of pencil beams in an IES or the entire field.

Scaling to the HIT required adjustments due to broader proton cross sections compared to earlier models. Previously, a 30-millimeter clipping was adequate, but HIT necessitates at least a 45-millimeter lateral clipping. This adjustment, coupled with the finer resolution scheme of isotropic 1mm, increases input neurons, raising run time and requiring more hidden neurons to preserve performance. However, the quadratic scaling of LSTM parameters with the increase in hidden neurons, as shown below, complicates maintaining real-time APT frame rates:

$$\theta_{\text{lstm}} = 4 \times ((x + h) \times h + h), \quad (11)$$

where x denotes the number of input neurons, and h is the number of hidden neurons. This increased parameter count significantly slows the algorithm.

Additionally, the original algorithm's capacity to maintain spatial coherence and effectively propagate temporal features diminishes with the substantial increase in input size.

The spatiotemporal, physics-informed, mono-energetic model (CC-LSTM) was introduced to address these challenges. It maintains the efficient sequential $O(1)$ inference cost characteristic of LSTM models while addressing the spatial feature coherence loss between BEV slices due to its 2D treatment of these slices.

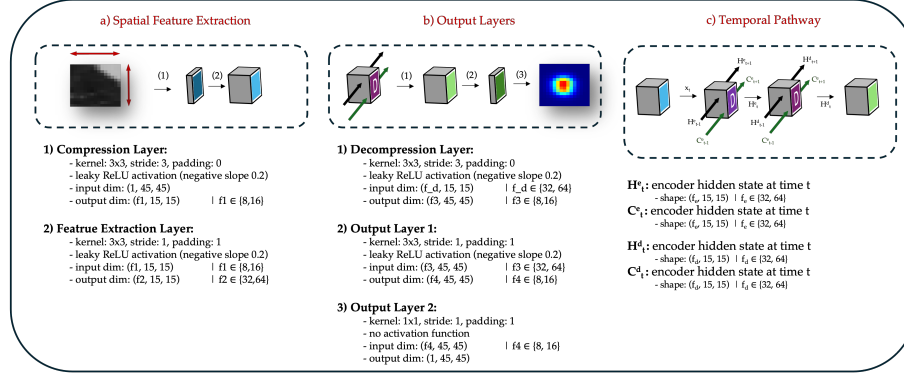


Figure 2.9: Illustration of the CC-LSTM architecture through time step t , segmented into critical components: a) Spatial Feature Extraction, where each BEV slice undergoes a 3:1 compression followed by convolutional layers for feature extraction; b) Output Layers, where the temporal module's outputs at each time step are converted into the forecasted dose distribution; and c) Temporal Pathway, showcasing the propagation of features and the correlations among slices up to time step t , with details on input and output dimensions and activation layers for each module.

As depicted in Figure 2.8, the CC-LSTM model incorporates two sets of convolutional operations, enhancing its processing capabilities. The initial set pre-processes each slice before it is introduced to the temporal encoder, as shown in figure 2.9a. This initial convolutional stage involves two layers of CNNs, each followed by leaky ReLU activation functions designed to condense the spatial dimensions effectively. The first layer utilizes a 3×3 kernel with a stride of three in both axes, efficiently reducing the spatial dimensions and producing a series of lower-dimensional feature maps, denoted as $f1$. Subsequent to this non-linearity, a second convolutional layer is applied to extract high-level, abstract spatial features, outputting a further reduced set of feature maps, denoted as $f2$, while preserving the original spatial dimensions.

To maintain the coherence of spatial features—a challenge not adequately met by previous models that employed flattening layers—we have adopted a variant of the ConvLSTM architecture. This model is tailored for many-to-many

sequence modeling and incorporates convolutional mechanisms instead of the traditional linear gating functions in standard LSTMs. This structural modification enhances the propagation of spatial features along the temporal axis, employing strategies initially developed for applications such as precipitation nowcasting (Shi et al., 2015). This innovative approach ensures that spatial relationships are preserved and effectively integrated throughout the temporal sequence processing, optimizing the model's performance for complex spatial-temporal tasks.

2.3.2 LSTM-133 Architecture

The LSTM-133 model closely follows the deterministic LSTM model previously described in Section 2.1.2, with the primary difference being in the input dimensions. The increased input dimensions in the deterministic LSTM model led to a loss of coherence and a corresponding increase in the number of parameters, which compromised the learning procedure.

To address this issue, the LSTM-133 model leverages the BEV domain formulation of the problem and the favourable $O(1)$ complexity inference performance of LSTM models, allowing superfast run times with an autoregressive estimation of the dose distributions. The new extraction scheme introduced in the LSTM-133 model employs a finer resolution longitudinally in the BEV (1 mm), where dose gradients are highest, and a coarser resolution laterally (3 mm). This approach maintains the same number of parameters as the deterministic LSTM model, thereby preserving learning efficiency.

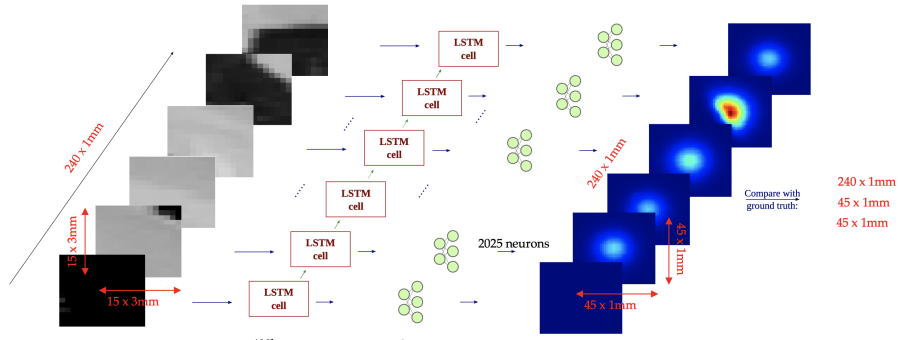


Figure 2.10: Illustration of the LSTM-133 architecture. The model follows closely the parameterization and the information flow described in Figure 2.2.

In this approach, although the input will have an anisotropic resolution of [1 mm, 3 mm, 3 mm], the dose distributions can still be generated in the desired isotropic 1 mm resolution by increasing the number of neurons in the output layer. The importance of generating dose distributions at this resolution will

be discussed in Section 2.3.5. This problem formulation and architecture design are presented in Figure 2.10.

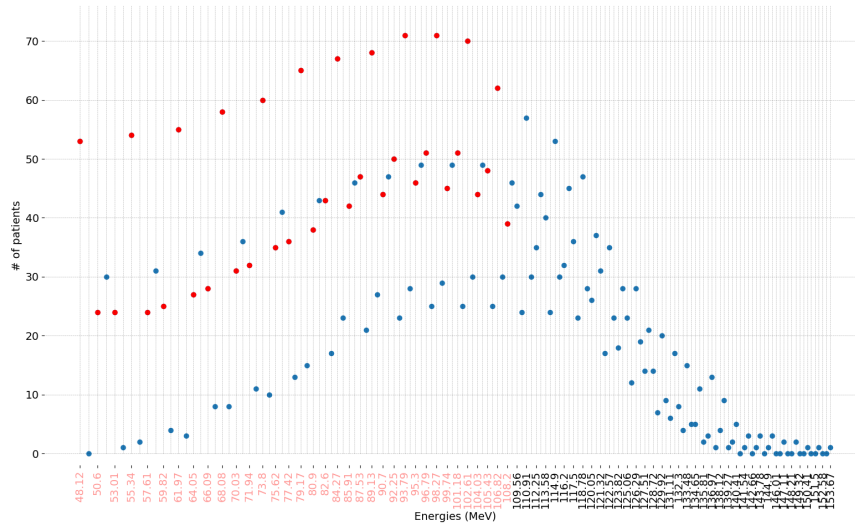
The benefit of parameterization introduced in the LSTM-133 architecture can be quantified using equation 11. Assuming an equal number of hidden neurons as the input dimension, which has empirically shown to be necessary, the deterministic LSTM would require approximately 5 million parameters for a $m = 45$ input dimension. In contrast, LSTM-133 will require only 800 000 learnable parameters by delegating the upscaling of dose distributions to the fully connected output layer outside the LSTM cell, rendering the learning procedure more efficient.

Specifically, the LSTM-133 model extracts input data with a higher longitudinal resolution to capture detailed variations in dose distribution along the beam path, while the lateral resolution is reduced to ensure that the model remains computationally manageable. This balanced approach allows the LSTM-133 model to effectively manage the increased input dimensions without compromising performance, thereby improving the overall accuracy and efficiency of proton dose calculations.

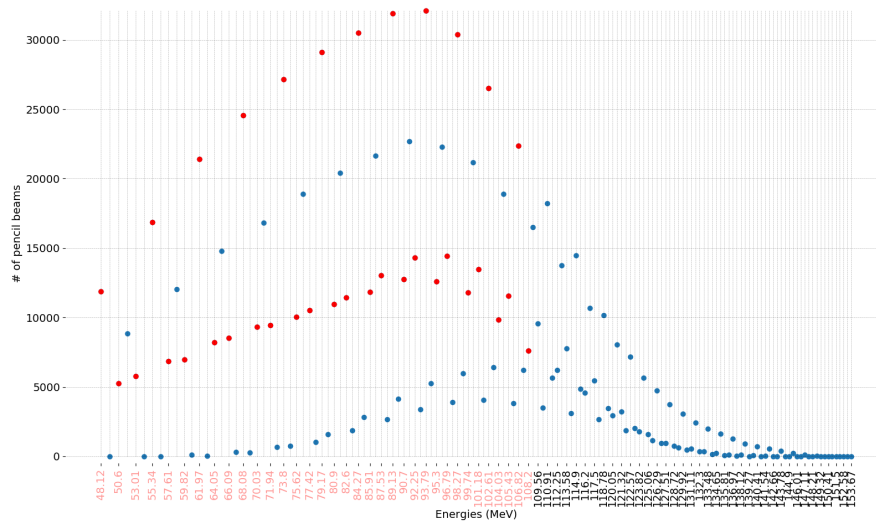
2.3.3 Data Preparation

This study further evaluates the model's performance using two distinct datasets, addressing the limitations identified in our previous work described in earlier sections. The earlier dataset was confined to pencil beams from a single patient, with randomly selected gantry angles and PB spot locations. In this study, a clinically relevant dataset was curated, which simulates realistic scenarios by restricting gantry and couch angles to those commonly used in prior treatment plans. This dataset includes a cohort of 90 patients with low-grade gliomas (LGG) in the head region, treated at the HIT center between 2010 and 2015 using actively scanned beams. The head region, with its significant tissue variations, provides an adequate challenge to assess the model's robustness. Overall statistics of the patient cohort are summarized in Table 2.1.

HIT offers 255 commissioned energies ranging from 48.12 MeV to 221.06 MeV. The total number of energies incorporated in this cohort is 131 unique energies. This means that although the cohort encompasses almost one million unique pencil beams, a specialized sampling scheme is necessary to ensure the availability of each energy in terms of the number of PBs and the variability of patients. To this end, an exploratory statistical overview of the available data was conducted, depicted in Figure 2.11. In order to sam-



(a) Number of patients per commissioned energy.



(b) Number of PBs/spots per commissioned energy.

Figure 2.11: Statistical overview of the number of patients and pencil beams available in the cohort. Red dots represent the energies sampled in this study.

| Statistic | Value |
|--|---------|
| Number of patients | 90 |
| Number of fields | 166 |
| Number of unique gantry angles | 9 |
| Number of unique couch angles | 42 |
| Number of unique energies incorporated | 131 |
| Number of unique pencil beam spots | 999,146 |

Table 2.1: Statistics of the patient cohort.

ple a subset of energies that are frequently used across all patients in the cohort and that collectively allow for a comprehensive dose evaluation for a select group of test patients, a sampling scheme that satisfies the above criteria while minimizing the number of unique energies required for training must be devised, ensuring a scalable time frame for the project.

The sampling strategy was adjusted based on the availability of PBs in each IES across all patients, ensuring a well representation of clinical scenarios. The result of this sampling scheme is a subset of a total of 35 unique energies, collectively enabling dose calculations for 7 test patients. As depicted in Figure 2.11, the selected energies all have at least 4000 unique spots and 20 unique patients per energy, ensuring a wide variance of informative and challenging data. Seven patients were selected to ensure a challenging testing scenario. Figure 2.12 outlines the unique impinging angles and how the selected test patients are distributed among those angles.

Finally, in order to have a streamlined model training and evaluations, in this phase, the study focuses on three distinct energies from the selected 35, representing short-, mid-, and high-range dose distributions (48.1 MeV, 79.1 MeV, and 102.6 MeV, respectively). Once again, for the representative energies to have a wide range of variance, a statistical analysis of unique patients, couch angles, and gantry angles per energy was carried out and illustrated in Figure 2.13. This approach tests the model's adaptability across different energy ranges and underscores their interchangeable nature through an AI perspective. The coordinates of PBs and their impinging delivery angles were randomly sampled from clinically delivered treatment plans.

The ground truth for each PB's dose distribution was simulated using FLUKA with 10^5 histories. A representative sample is reported in Figure 2.14, illustrating the dose distribution of single pencil beams before and after extraction over the CT grid and post-extraction in the BEV. To minimize interpolation errors and information loss, the doses were scored directly on the CT grid, and

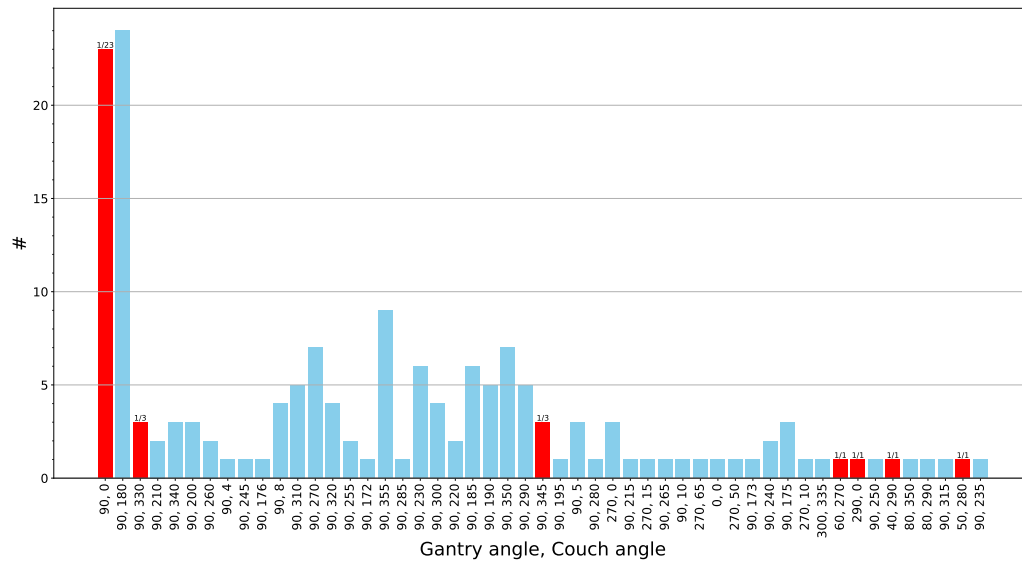


Figure 2.12: Histogram of incorporated gantry and couch angles in the cohort. Reds are the selected test patients.

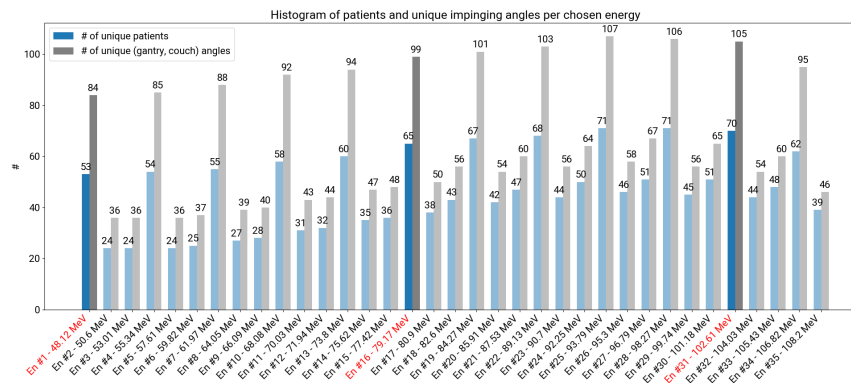


Figure 2.13: Histogram of unique patients, couch angles, and gantry angles per trained energy

the cube extractions were carried out with an isotropic resolution of 1 mm, with $l = 210$ voxels longitudinally and $m = 45$ voxels laterally.

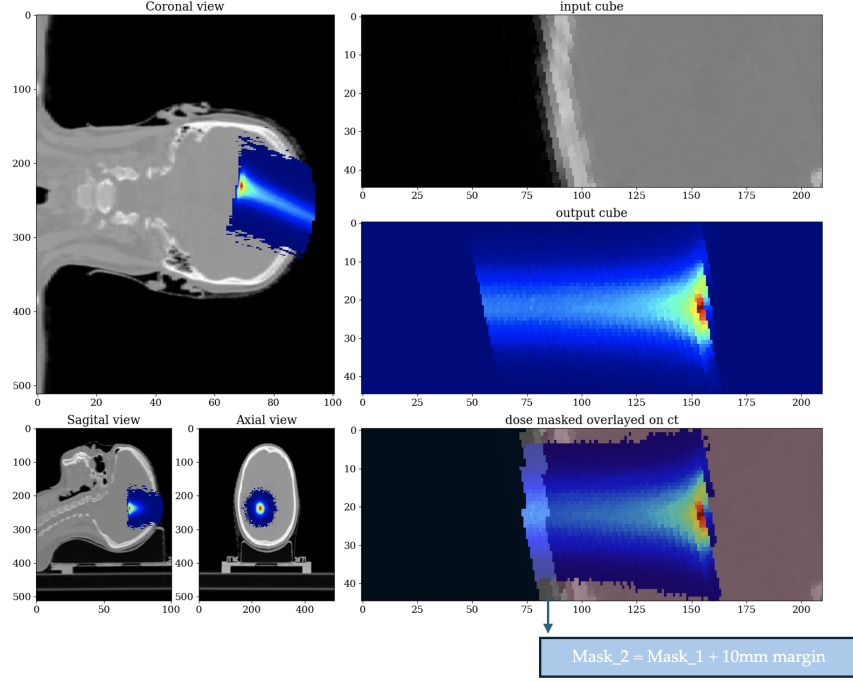


Figure 2.14: This figure provides a comprehensive view of the dose distribution simulation for single pencil beams. The coronal view is shown in the top left, demonstrating the dose distribution within the patient. As well as the sagittal view (bottom left) and axial view (bottom middle) which provide additional views on the dose distribution within the patient. The input RSP cube is displayed in the top right, and the output dose distribution cube, post-extraction, is shown in the middle right. Finally, the bottom right panel overlays the dose distribution on the CT grid, illustrating the 2 stage masking scheme incorporated in this study.

To compare the performance of the model against the current state-of-the-art DoTA, the model was also trained using a publicly available dataset previously employed to train the DoTA model, as documented by Pastor-Serrano et al. (2022). DoTA employs a parameterized energy model, with its dataset showing a semi-continuous energy increase in 0.001 MeV steps, spanning from 70 to 220 MeV. Analysis of DoTA's dataset revealed an uneven distribution of training samples, predominantly between 70 and 140 MeV, and less densely populated from 140 to 220 MeV. To ensure sufficient data for training, the focus was exclusively on the lower energy segment, selecting three distinct energy ranges: 70-71 MeV for low-range PB, 105-106 MeV for mid-range PB, and 139-140 MeV for high-range PBs. This comparative analysis is pivotal for validating the training feasibility of the model on an external dataset and benchmarking it against existing architectures.

2.3.4 *Model Evaluation Update*

In developing a novel deep-learning model, selecting informative metrics is crucial. These metrics are necessary for guiding the iterative process of incorporating or excluding specific deep learning techniques and during hyperparameter tuning. This becomes especially critical in modeling the physics of proton transport within a data-driven framework, where the objective is to achieve a model that performs well across diverse scenarios. The medical physics and artificial intelligence fields provide distinct metrics, each rigorously evaluated to establish a framework guiding our development. Given our task's interdisciplinary nature, precisely quantifying each modification's impact within the deep learning architecture is challenging. Due to the inherent complexity and high-dimensional optimization space, training deep learning models often resembles a stochastic search for a satisfactory (local) minimum. This unpredictability contrasts sharply with the deterministic nature of analytical models in proton physics, where minor adjustments can significantly alter the model's behavior. Furthermore, during the evaluation phase, the stochastic nature of trained models can cause the spatial distribution of errors in pencil beam dose distributions to obscure one another, making the task of condensing these complex outputs into a single, practical metric notably challenging.

This study incorporated various metrics from both fields to fully monitor our model's behavior and visually inspect PB dose distributions in challenging, heterogeneous scenarios. Our comprehensive assessment strategy involved evaluating individual metrics and considering the models' overall performance across multiple dimensions. This multi-faceted approach ensured that the models were thoroughly tested and validated, providing confidence in their clinical applicability. From the field of AI, two metrics were utilized for our regression task: MSE, which also serves as the loss function in our framework, emphasizing larger errors, and MAE, providing a general evaluation of the average magnitude of errors irrespective of their direction. While MAE and MSE quantify the average discrepancy between two datasets, they do not account for the spatial distribution of errors, which is crucial in radiotherapy dose verification. For this purpose, Gamma analysis (Low et al., 1998) was used, enabling a quantitative comparison of two dose distributions by considering both DD and distance-to-agreement criteria, leading to a spatially resolved evaluation of the conformity between dose distributions.

However, Gamma analysis is typically used to compare two homogeneous full-field dose distributions, so criteria must be adapted for comparing two PBs with heterogeneous dose distributions. The DD criterion, defined by the percentage of the prescribed dose, becomes problematic as this value is not

available when comparing two PBs. To address this, a ‘local’ gamma analysis is performed where the DD of each voxel is evaluated based on a local neighborhood around each voxel. To reduce noise in regions with low dose, a dose cut-off set by a percentage of the maximum dose was applied. To ensure that our test sets were representative of the entire patient cohort, we meticulously curated diverse samples encompassing a wide range of anatomical and clinical scenarios. This included selecting test beams that traverse various tissue densities, anatomical complexities, and regions with significant heterogeneities. By doing so, we aimed to capture the full spectrum of potential challenges encountered in clinical practice.

All previous metrics were evaluated within the PB domain, that is, in the BEV. However, the ultimate goal of this research is to achieve high accuracy when transforming the PB dose distribution back to the original dose grid. In this phase, interpolation errors are introduced, and dose distributions are scaled by the fluence of the corresponding spot of the PB in the treatment plan. The *Peak Deviation Percentage* (PDP) has been introduced to address this complex transformation. This metric evaluates the precision of regression models in estimating the peak values within a dataset and is particularly relevant when assessing the model’s performance across different domains—PB and the original dose grid. The PDP is calculated using the following equation:

$$\text{PDP} = \left(\frac{|\max(\hat{y}) - \max(y)|}{\max(y)} \right) \times 100\% \quad (12)$$

where $\max(\hat{y})$ represents the maximum predicted value from the model, and $\max(y)$ denotes the maximum actual value observed in the ground truth.

2.3.5 Estimating Full-Field SOBPDose Distributions

Thus far, the primary focus has been the improvement of accuracy and run times in the BEV. However, as indicated in the aims of this study, the objective is to generate dose for an entire proton therapy treatment plan. To this end, the transition from the BEV domain to the CT grid is necessary. Specifically, it is essential to estimate the dose of each pencil beam using the trained network, back-project these estimates from the BEV domain to the CT grid, and accumulate them to form the planned SOBPDose. Figure 2.15 demonstrates the integration procedure, from preprocessing to postprocessing and the eventual dose on CT grid.

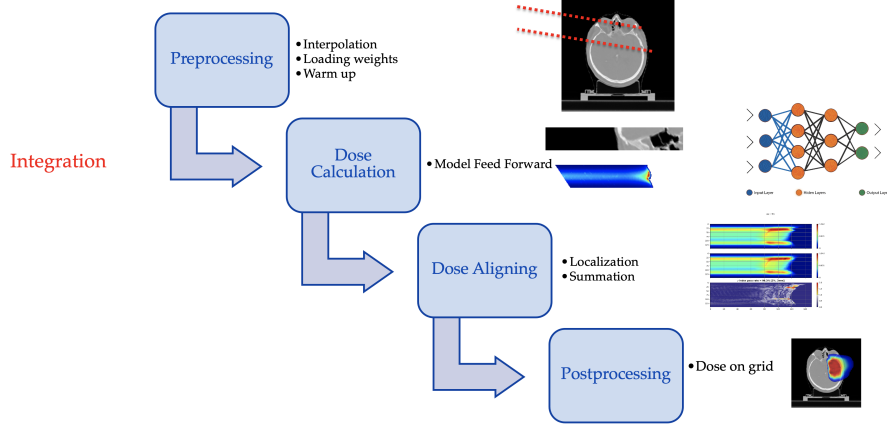


Figure 2.15: The outlook of integration procedure in transition from PB domain to CT grid

Extraction Resolution: 1 mm vs 2 mm vs 3 mm

An important design parameter in this study was the extraction resolution in the longitudinal axis (in the direction of the beam), which consequently defines the resolution of the generated dose distributions. This is because RNN models are recursive and autoregressive, generating the dose for each input slice iteratively and producing the next lateral dose slice using the previous dose slice and the current CT input slice.

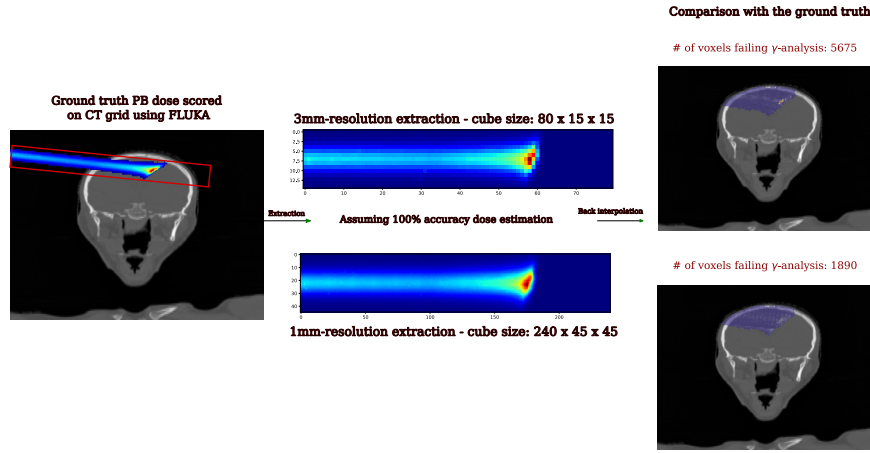


Figure 2.16: a comparison between 1 mm and 3 mm isotropic extraction schemes.

It is important to note that this is a restriction when adhering to pre-implemented architectures in deep learning frameworks, specifically PyTorch in this study. The motivation for using pre-implemented architectures is the significant run time advantage due to optimized built-in performance compared to custom-designed architectures. The performance superiority of pre-

implemented architectures stems from the parallel processing of batches. In contrast, custom architectures, such as the CC-LSTM used in this study, require different input/output dimensions and must define the cell architecture manually, leading to increased run times.

In the feasibility study outlined in Section 2.1.1, an isotropic resolution of 2 mm resulted in extraction dimensions of 15 voxels laterally and 150 voxels longitudinally for each patient setup. Subsequently, as described in Section 2.3.3, an isotropic 1 mm resolution was adopted, leading to extraction dimensions of 45 voxels laterally and 210 voxels longitudinally. While the increased lateral dimension of the model input was addressed in the previous section, the rationale for choosing a finer 1 mm resolution has not been explained.

The choice of a 1 mm resolution is justified for two primary reasons. First, it prevents the apparent loss of information. As illustrated in Figures 2.17, a comparison between 1 mm and 3 mm isotropic extraction schemes shows that the finer resolution results in significantly fewer failed voxels during global gamma analysis. In this example involving an oblique impinging proton beam, the number of failed voxels was nearly three times lower when opting for the finer 1 mm resolution.

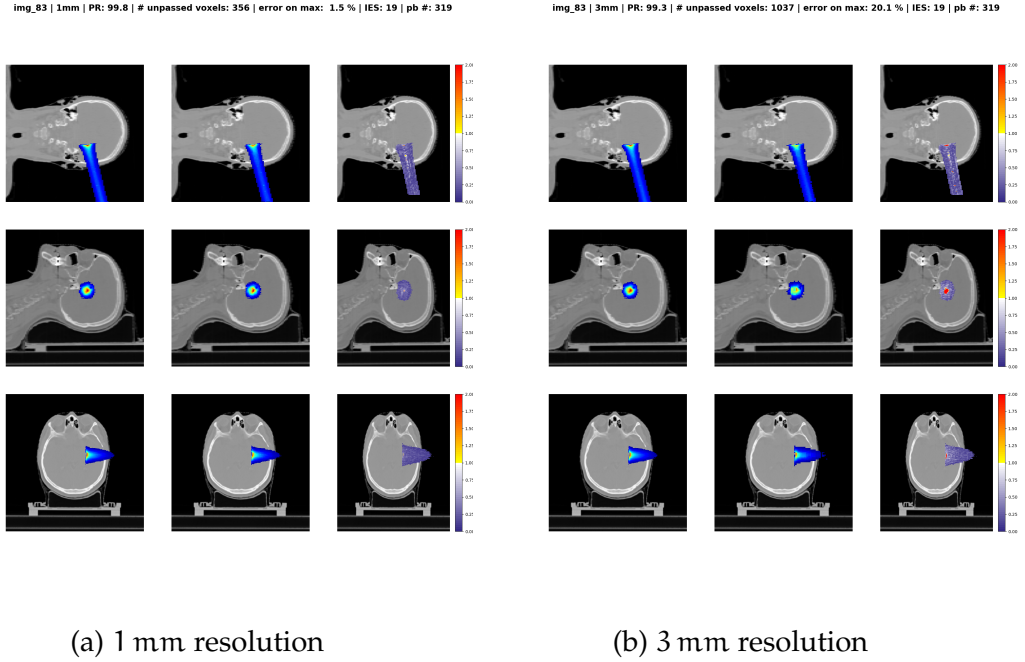


Figure 2.17: Comparison between (a) 1 mm and (b) 3 mm resolution extraction and back-interpolation, assuming 100 % model accuracy.

Second, the 1 mm resolution facilitates the construction of SOBPs from constituent PBs. Each PB dose in the BEV is estimated using the pre-trained network corresponding to the incorporated energy, multiplied by the fluence

of the corresponding spot, and then projected in a predefined cube in the BEV on the GPU.

This cube in the BEV will recursively add the dose distribution of each PB to itself until the IES dose distribution is calculated in the BEV. Depending on the objective of re-calculation, whether to calculate the entire SOBP or only a single IES, this cube can then be added to other IES dose distributions in the BEV or transformed on the fly, on the GPU, to the CT grid by a single interpolation operation.

However, when opting for a 2 mm resolution, the aforementioned projection of the PB cube in the predefined cube would require an intermediate interpolation step. For instance, at HIT, the lateral spot spacing is 3 mm, and to add the PBs while maintaining their designated spatial spot would necessitate an intermediate interpolation step from 2 mm to 1 mm or 3 mm, in order for them to be aligned and accumulated efficiently.

As described earlier in the problem parameterization of this study, given the independent nature of each IES dose calculation task, one could assign each corresponding model and the corresponding data to the designated GPU based on the GPU infrastructure availability. Figure 2.18 demonstrates this approach.

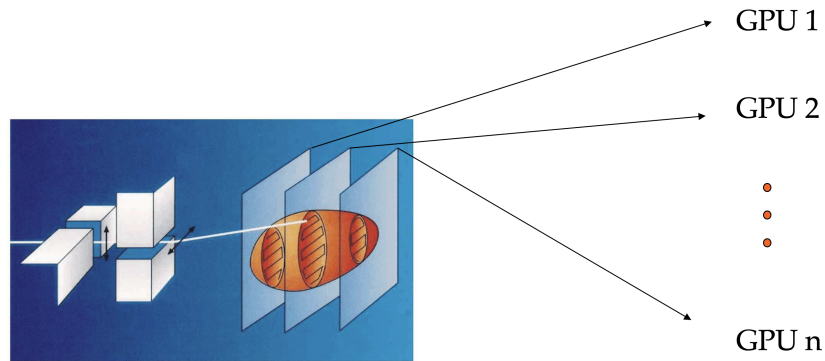


Figure 2.18: multiple GPU setup for AI-based dose calculation.

RESULTS

3.1 FEASIBILITY OF RNN-BASED MODELS FOR FULL-FIELD DOSE ESTIMATION

The results of two experiments designed to test the generalization of the trained network to other unseen patients and distinct energies are reported. Table 3.1 outlines the network’s performance in dose estimation for an unseen patient dataset. Table 3.2 summarizes the outcomes for three separate networks trained specifically for initial energies of 67.85 MeV, 104.25 MeV, and 134.68 MeV. Representative challenging samples from the low-range and high-range datasets are shown in Figure 3.1, highlighting the model’s robustness across different energy values.

Table 3.1: γ -index analysis for five different lung cancer patients [1 %, 3 mm]). The model was trained on Patient 0 (Experiment 1 - 1) (Neishabouri et al., 2021b).

| | Mean (%) | Std (%) | Min (%) | Max (%) |
|-----------|----------|---------|---------|---------|
| Patient 0 | 98.50 | 1.00 | 93.93 | 99.82 |
| Patient 1 | 98.27 | 0.97 | 94.66 | 99.65 |
| Patient 2 | 98.35 | 1.30 | 94.35 | 99.78 |
| Patient 3 | 98.45 | 1.10 | 94.51 | 99.60 |
| Patient 4 | 96.71 | 3.01 | 81.66 | 99.61 |
| Patient 5 | 97.47 | 1.87 | 87.82 | 99.61 |

Table 3.3 lists the average run times for estimating the dose for the 5 above-mentioned patients for a single pencil beam, for MC, PB, and LSTM dose calculation. The MC simulations were performed with Topas on a calculation node with 28 virtual CPUs on an Openstack¹ cluster. For the trained network, the run times were measured for two systems with different GPUs. Depend-

¹ <https://www.openstack.org/>

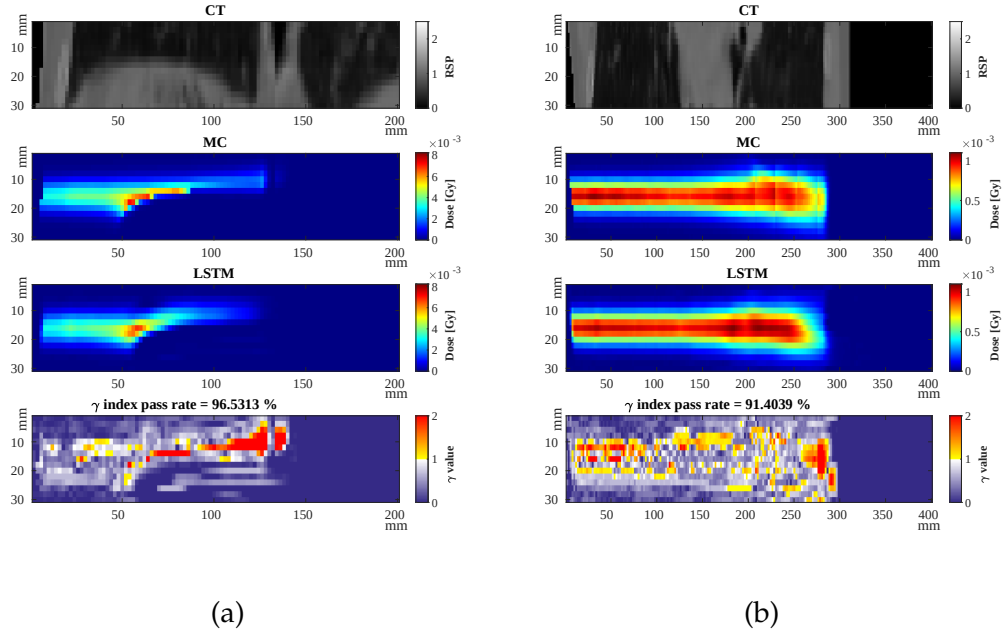


Figure 3.1: Dose estimation results for representative samples from two energy ranges: (a) the lower energy set (67.85 MeV) and (b) the higher energy set (134.68 MeV). Each sub-figure follows the layout presented in Figure 1.8. Note the varying longitudinal axis ranges between the sub-figures (Neishabouri et al., 2021b).

Table 3.2: γ -index analysis on datasets with three distinct energies of proton pencil beams (Experiment 1 - 2) (Neishabouri et al., 2021b).

| Energy (MeV) | Mean (%) | Std (%) | Min (%) | Max (%) |
|--------------|----------|---------|---------|---------|
| 67.85 | 98.56 | 1.30 | 95.35 | 99.79 |
| 104.25 | 97.74 | 1.48 | 92.57 | 99.74 |
| 134.68 | 94.51 | 2.99 | 85.37 | 99.02 |

ing on the facilitated hardware, we measure average run times of 6 ms to 23 ms for the LSTM approach.

The MC simulations had an average run time of 1160 s, performed with $\sim 2.5 \times 10^6$ histories on average. Note that these run times included the time required to send the input CT cube for each pencil beam from CPU to GPU and vice versa for the yielded dose cube. However, in applications such as adaptive RT which requires repetitive online dose estimations, the input CT cubes can be prepared and sent to the GPU in advance. Consequently, the only relevant run times would be the network feed forward, i.e., matrix multiplication operations run times, reported to be 1.5 ms to 2.5 ms for the two facilitated hardware stacks.

Table 3.3: Comparative analysis of execution times between MC calculations, LSTM predictions, and the PB algorithm. Noted times (in parentheses) represent the direct computational time excluding data transfer between CPU and GPU (Neishabouri et al., 2021b).

| | MC ^a | LSTM ^b | LSTM ^c | PB ^d |
|----------------------|-----------------|-------------------|-------------------|-----------------|
| Average run time (s) | 1159.5 | 0.023 (0.0025) | 0.006 (0.0015) | 1.025 |

Computational node, 28 VCPUs, 64 GB RAM.

Intel Core i7-6700 3.4 GHz - Nvidia GTX 970 - 64 GB RAM.

Intel Xeon W-2135 3.7 GHz - Nvidia Quadro RTX 6000 - 64 GB RAM.

maRad software toolkit: Intel Xeon W-2135 3.7 GHz - 64 GB RAM.

3.2 BAYESDOSE EVALUATIONS OUTCOMES

The initial evaluation of the BayesDose model's predictive accuracy was conducted using phantom data, similar to the previous section, focusing on individual pencil beams within the test set (Table 3.4).

The distribution of voxels outside the $n\sigma$ range for the entire test set is tabulated for each sigma level from one to five, as presented in Table 3.5.

Table 3.6 summarizes the comparative performance of BayesDose and its deterministic counterpart, using the first lung patient's data. Both algorithms demonstrated exceptional performance on the test dataset, achieving average γ pass-rates exceeding 99.5 %, with only negligible differences in their MSE and MAE values.

On the other hand, the lowest dosimetric accuracy observed in a test sample is shown in Figure 3.6, where BayesDose underestimated the extent of dose

Table 3.4: Comparison of γ -index results, MAE, and MSE for different BayesDose model configurations against MC simulations in a phantom study (Experiment 2 - 1).

| | Deterministic | Bayesian |
|-----------------------------|------------------------|------------------------|
| Mean γ pass-rate [%] | 97.81 | 97.93 |
| SD γ pass-rate [%] | 1.90 | 1.87 |
| Min γ pass-rate [%] | 87.65 | 88.93 |
| Max γ pass-rate [%] | 100 | 100 |
| MAE [Gy] | 2.42×10^{-6} | 2.70×10^{-6} |
| MSE [Gy ²] | 2.17×10^{-10} | 2.31×10^{-10} |

Table 3.5: Proportion of voxels deviating from MC simulation beyond specified confidence intervals in the phantom study.

| Confidence bound | | | | | | |
|------------------|---------------|---------------|---------------|---------------|--------|---------|
| $\pm 1\sigma$ | $\pm 2\sigma$ | $\pm 3\sigma$ | $\pm 4\sigma$ | $\pm 5\sigma$ | 5–95 % | 20–80 % |
| 26.4 | 9.50 | 4.07 | 1.89 | 0.94 | 14.0 | 34.0 |

Table 3.6: γ -index analysis [1 %, 3 mm] (top) and MAE and MSE comparisons among BayesDose and the deterministic LSTM, against the MC simulation (bottom), in the lung patient case study (Experiment 2 - 2)

| | Deterministic | BayesDose |
|------------------------|---------------|-----------|
| γ analysis | LSTM | BLSTM |
| Mean [%] | 99.60 | 99.70 |
| SD [%] | 0.62 | 0.60 |
| Min [%] | 95.89 | 95.02 |
| Max [%] | 100 | 100 |
| MAE [Gy] | 1.62e-5 | 1.46e-5 |
| MSE [Gy ²] | 3.14e-9 | 3.70e-9 |

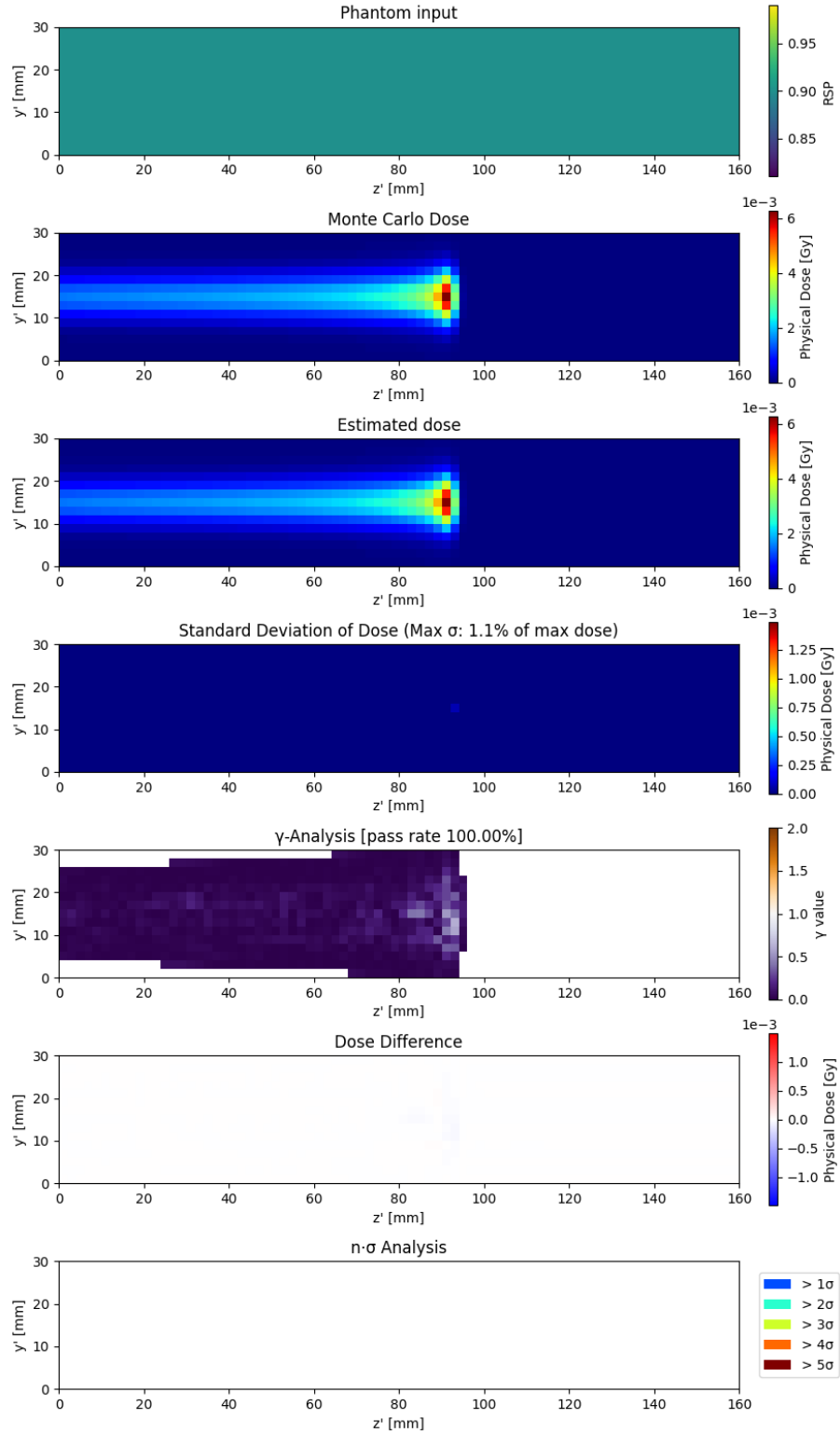


Figure 3.2: Optimal prediction by the Bayesian LSTM network achieving a 100 per cent γ pass-rate in the phantom case. Displayed sequentially are the input cube, MC ground truth, averaged BayesDose prediction across the ensemble, the standard deviation of the prediction, γ pass-rate, the differential between the MC dose and the prediction, dose discrepancy relative to the standard deviation, highlighting voxels with deviations greater than $n\sigma$. This format is standardized for all subsequent individual prediction displays.

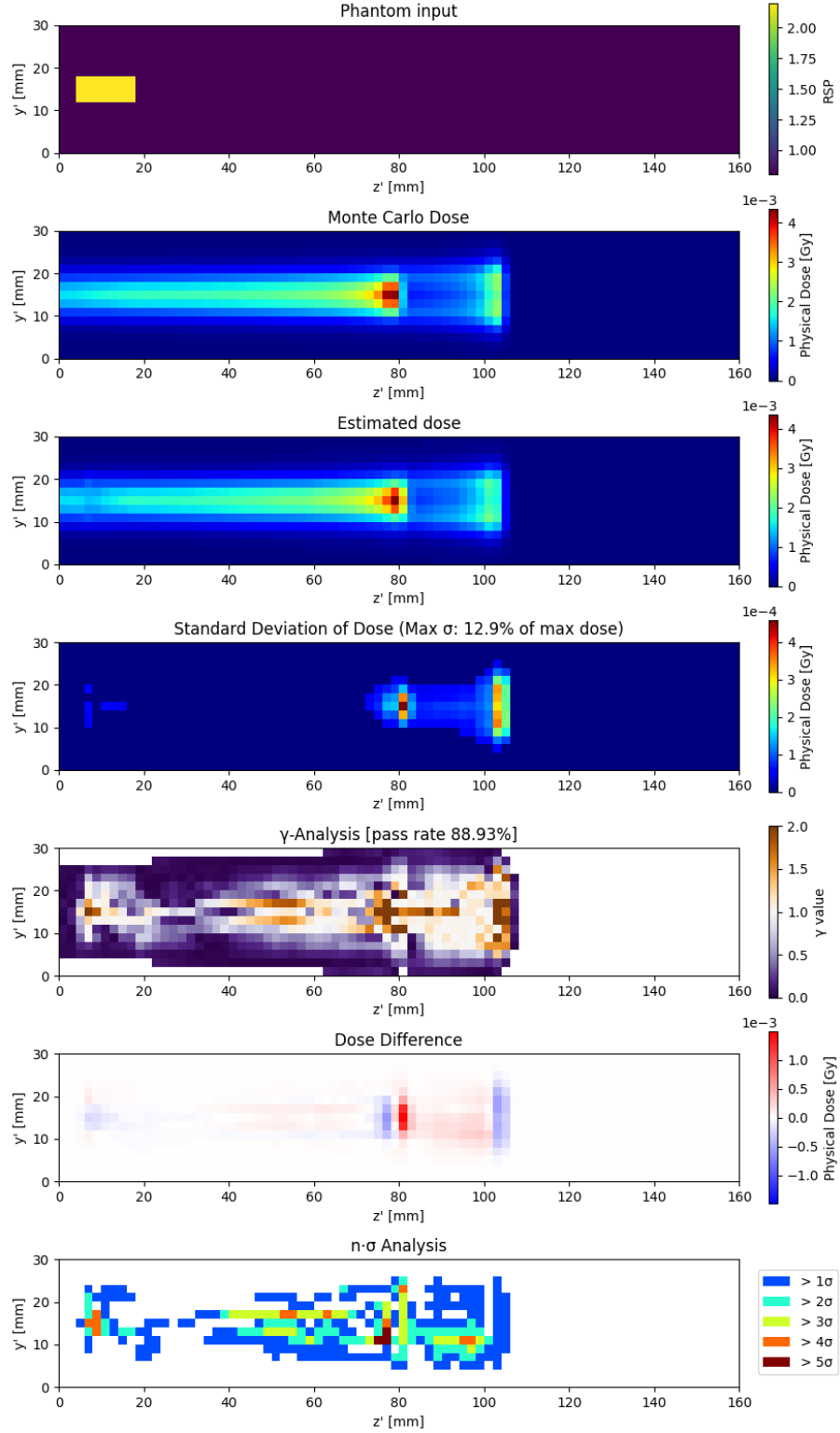


Figure 3.3: Least effective prediction by the Bayesian LSTM network, marked by the lowest γ pass-rate at 88.93 percent in the phantom case, coupled with one of the highest standard deviations observed across predictions. The layout of this figure mirrors that established in 3.2.

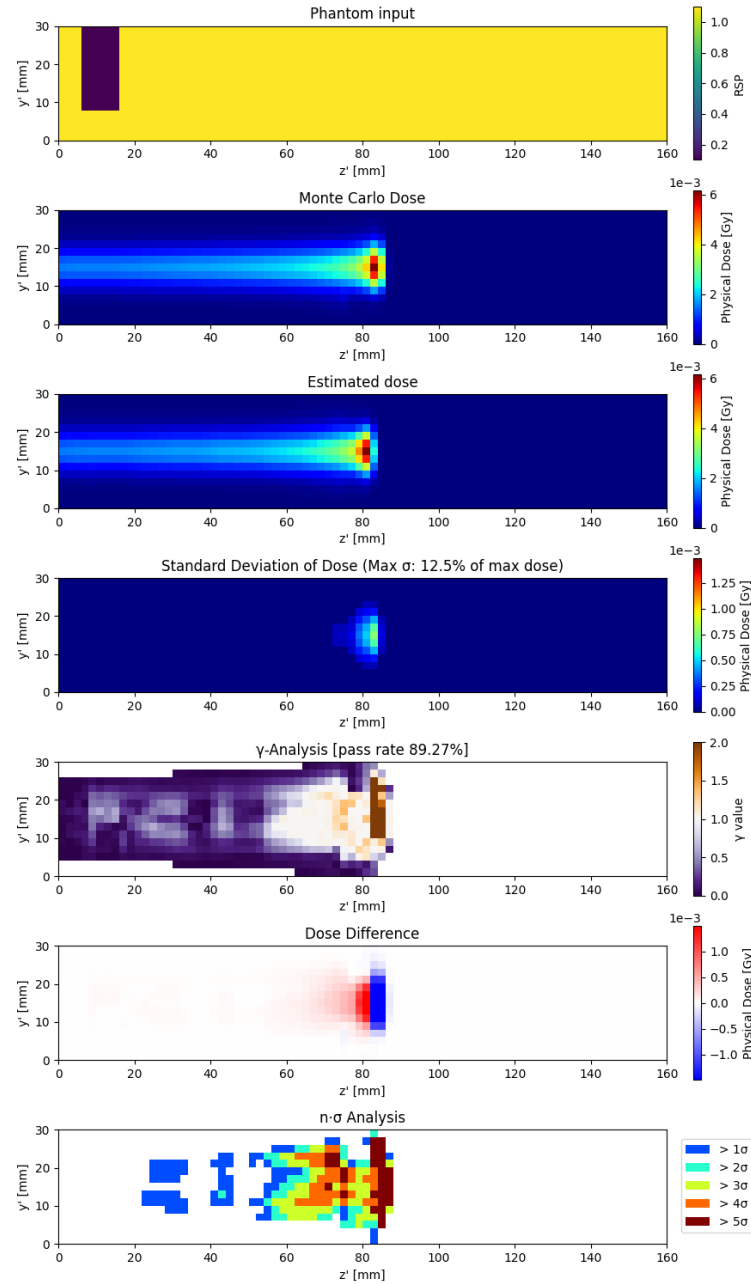


Figure 3.4: Most significant quantification of uncertainty for the phantom case, where the highest proportion of voxels exhibited substantial deviations (31.1 % $> 3\sigma$). The layout of this figure is consistent with that of 3.2.

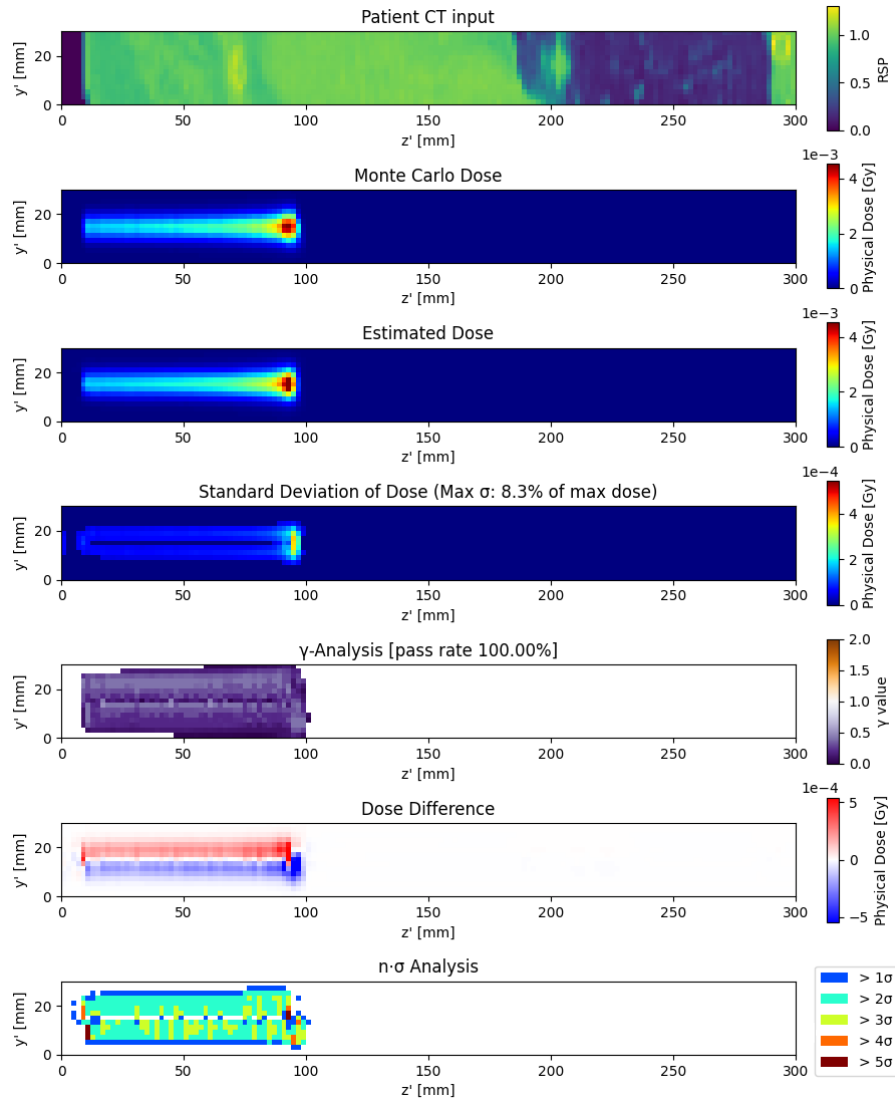


Figure 3.5: Depiction of the most precise prediction by BayesDose on the lung patient dataset, achieving a 100 percent γ pass-rate. This visualization underscores the model's effectiveness with actual patient data.

Table 3.7: Percentage of voxels where the average prediction deviates from the MC simulation outside specified confidence intervals, calculated using both standard deviations and empirical percentiles, for the patient data.

| Confidence bound | | | | | | |
|------------------|-----------|-----------|-----------|-----------|--------|---------|
| 1σ | 2σ | 3σ | 4σ | 5σ | 5–95 % | 20–80 % |
| 21.3 | 11.7 | 6.30 | 2.97 | 1.46 | 15.0 | 24.5 |

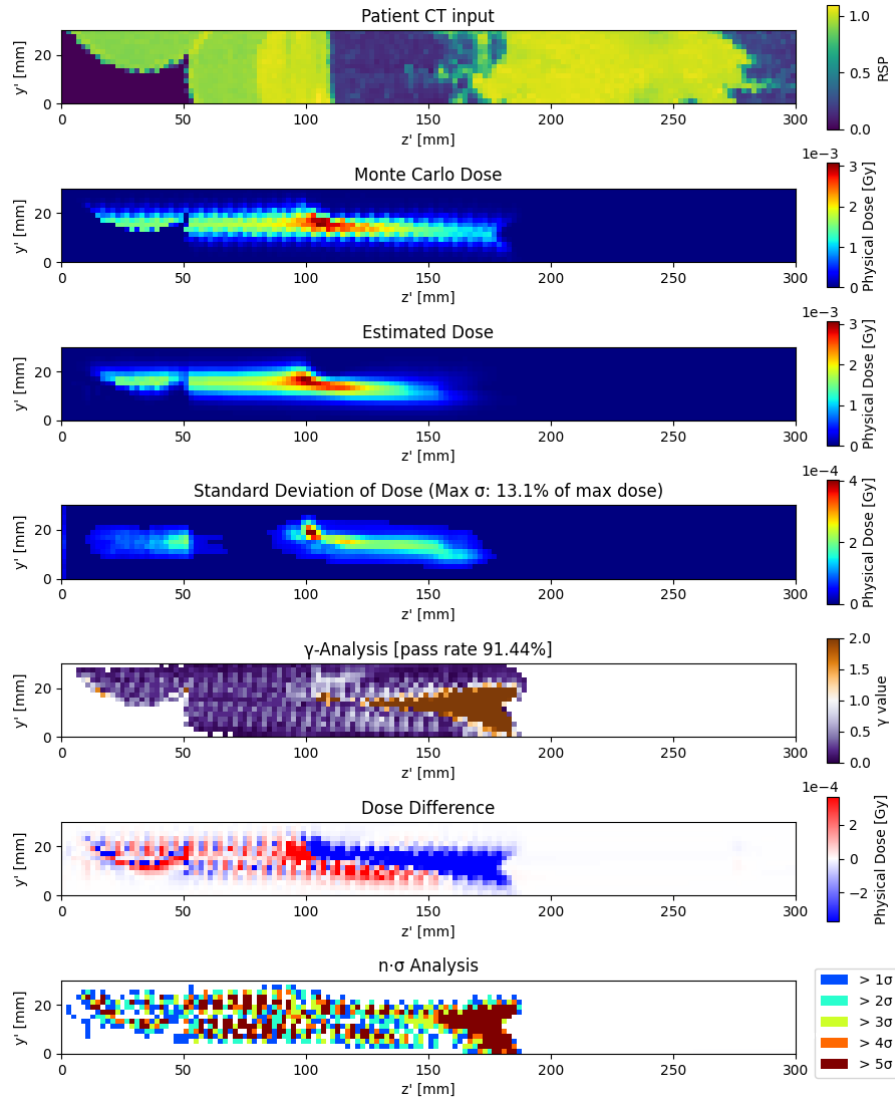


Figure 3.6: Least accurate BayesDose prediction for the patient dataset, displaying the lowest γ pass-rate at 88.93 percent and a notably high standard deviation. This figure's arrangement mirrors that of Figure 3.2.

Table 3.8: γ -index analysis ([1 %, 3 mm]) for three proton beamlets with varying initial energies (Experiment 2 - 3).

| E (MeV) | Mean(%) | SD(%) | Min(%) | Max(%) |
|---------|---------|-------|--------|--------|
| 67.85 | 99.45 | 0.92 | 94.16 | 100 |
| 104.25 | 99.59 | 0.78 | 91.44 | 100 |
| 134.68 | 99.00 | 1.13 | 92.47 | 100 |

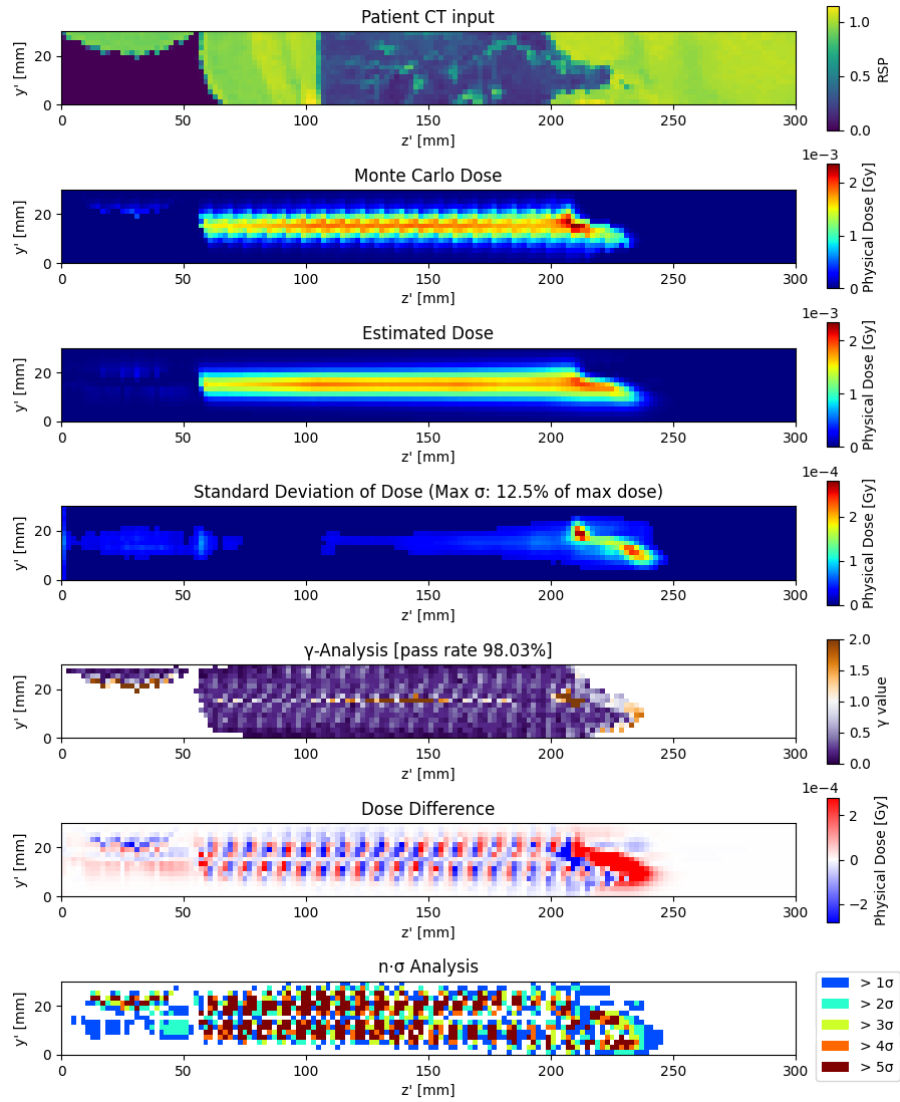


Figure 3.7: Most substantial quantification of uncertainty in the patient dataset, showing the largest fraction of voxels exceeding significant deviations (15.99percent $> 3\sigma$). This figure follows the layout of Figure 3.2.

Table 3.9: γ -index evaluation ([1 %, 3 mm]) for five lung cancer patients, showcasing the performance post-training on Patient 0 (Experiment 2 - 4).

| | Mean(%) | SD(%) | Min(%) | Max(%) |
|-----------|---------|-------|--------|--------|
| Patient 0 | 99.59 | 0.78 | 91.44 | 100 |
| Patient 1 | 96.33 | 4.04 | 73.01 | 99.98 |
| Patient 2 | 99.29 | 1.02 | 90.21 | 100 |
| Patient 3 | 99.12 | 1.07 | 94.82 | 100 |
| Patient 4 | 99.32 | 0.87 | 94.05 | 100 |
| Patient 5 | 98.04 | 2.09 | 88.27 | 100 |

dispersion beyond the Bragg peak. This region is associated with very high uncertainty, with standard deviations reaching up to 13.1 % of the maximum dose. The interpolation artifacts significantly influence the incorrect voxel dose predictions, especially noticeable in the large number of voxels deviating more than 5σ .

Across the entire patient dataset, the average percentage of voxels deviating more than $n\sigma$, detailed in Table 3.7, was consistently higher than in the phantom experiments. This suggests that BayesDose may offer less conservative uncertainty predictions when applied to patient data. Notably, a considerable number of voxels deviating beyond 5σ , as illustrated in the prediction example with the maximum amount of voxels showing a dose difference greater than $n\sigma$, are predominantly due to interpolation artifacts in the training data.

BayesDose was evaluated across two additional proton energy levels. BayesDose maintained a high level of accuracy across all tested energies. The details of this performance are captured in Table 3.8, which notes that the mean γ pass rate for all energies exceeded 99 %.

Additionally, the analysis of BayesDose across five additional lung patients shows a performance consistency similar to the initial deterministic model. As outlined in Table 3.9, however, Patient 1 and Patient 5 had notably poorer performance compared to others.

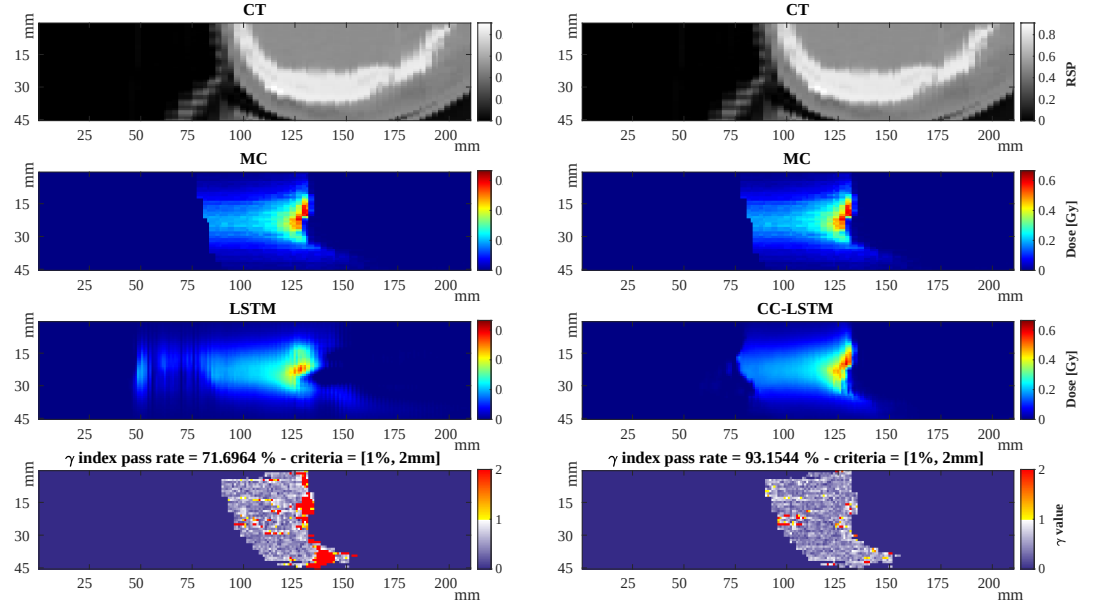
To address the notably poor performance of Patient 1, both in terms of dosimetric results and the quality of uncertainty prediction, the model underwent retraining on Patient 5. This patient was selected due to having the second most extensive HU range, just after Patient 1, who had the broadest HU range. Table 3.10 illustrates significant enhancements in dosimetric accuracy for Pa-

tient 1, bringing it in line with other patients after retraining for 10 epochs on Patient 5.

Table 3.10: γ -index analysis ([1 %, 3 mm]) for lung cancer patients before and after 10 epochs of transfer learning on Patient 5 (Experiment 2 - 5).

| | Patient 1 | Patient 2 | Patient 3 | Patient 4 |
|-------------------|-----------|-----------|-----------|-----------|
| Mean(%) before TL | 96.33 | 99.29 | 99.12 | 99.32 |
| Mean(%) after TL | 97.88 | 99.13 | 99.27 | 99.3 |

3.3 CC-LSTM: A NEW BENCHMARK



(a) LSTM model dose estimation.

(b) CC-LSTM model dose estimation.

Figure 3.8: Dose estimation comparison for the worst-case scenario: (a) LSTM model versus (b) the newly developed CC-LSTM model.

This section examines the CC-LSTM model's accuracy compared to our previous model and the current state-of-the-art (Pastor-Serrano and Perkó, 2022), alongside its performance across three distinct energy ranges. Our analysis utilizes a newly compiled dataset, incorporating patient cases previously treated at the HIT facility with clinical proton base data.

Table 3.11 presents the results comparing the performance of the CC-LSTM model against the previously published LSTM-based model for proton pencil beams at an energy of 79.17 MeV. Figure 3.8 illustrates the previous model's loss of spatial coherence over the temporal axis and its difficulty in accurately predicting dose distribution changes in the presence of significant heterogeneities within proton trajectories. Figure 3.8 also demonstrates the CC-LSTM model's capability in accurately estimating dose distributions in highly heterogeneous scenarios within the Bragg peak region. Figure 3.9 evaluates the model's capability to adapt to immediate heterogeneities within the Bragg peak region. Figure 3.10 highlights the CC-LSTM's precise estimation of dose deformation when faced with significant heterogeneity at the beginning of the PB trajectory.

Table 3.11: Evaluative comparison of an established network (LSTM) and an advanced model (CC-LSTM) against a MC simulation benchmark at an initial energy of 79.17 MeV.

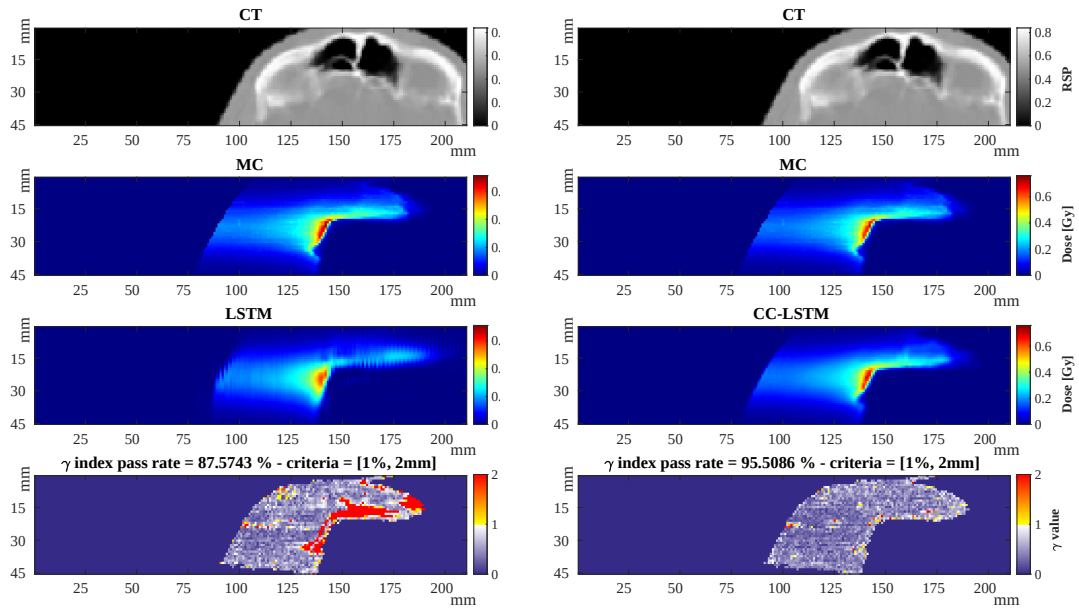
| γ -index analysis ([1 %, 2 mm]) | | | | |
|--|----------|---------|---------|---------|
| Model | Mean (%) | Std (%) | Min (%) | Max (%) |
| LSTM | 92.54 | 4.42 | 71.69 | 98.31 |
| CC-LSTM | 97.21 | 0.84 | 92.37 | 99.81 |

| Other Metrics | | | |
|---------------|-----------------------|------------------------|---------|
| Model | MAE (Gy) | MSE (Gy ²) | PDP (%) |
| LSTM | 2.74×10^{-3} | 2.02×10^{-4} | 10.8 |
| CC-LSTM | 1.1×10^{-3} | 2.41×10^{-5} | 3.9 |

After validating the superior performance of the CC-LSTM model over the previous LSTM model, Table 3.12 presents the results from training the CC-LSTM using the same procedures for both low- and high-energy range protons (49.21 MeV and 108.17 MeV, respectively) across their respective test sets.

Next, to evaluate and benchmark the performance of CC-LSTM against the current state-of-the-art Transformers model (DoTA), we trained CC-LSTM with the publicly available dataset from Pastor-Serrano and Perkó (2022). Training DoTA on our data resulted in unsatisfactory outcomes, underscoring the challenging training procedures associated with self-attention blocks in Transformers.

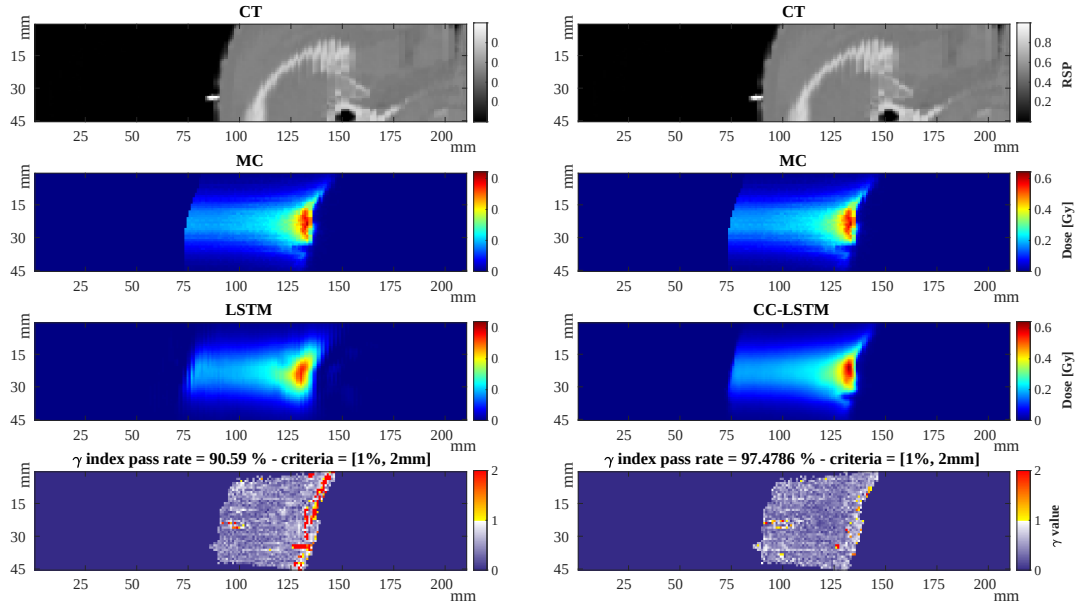
Table 3.14 presents the outcome of this comparison using the introduced metrics, demonstrating CC-LSTM's superiority across all specified metrics, even when trained on its own dataset. This is notable despite the model encountering an approximate 1 cm range uncertainty, a consequence of the semi-continuous energy variation within a 1 MeV energy span in each training set (specific to the data prepared by Pastor-Serrano and Perkó (2022)). We highlight a sample where CC-LSTM outperforms DoTA in estimating the PB



(a) LSTM model dose estimation.

(b) CC-LSTM model dose estimation.

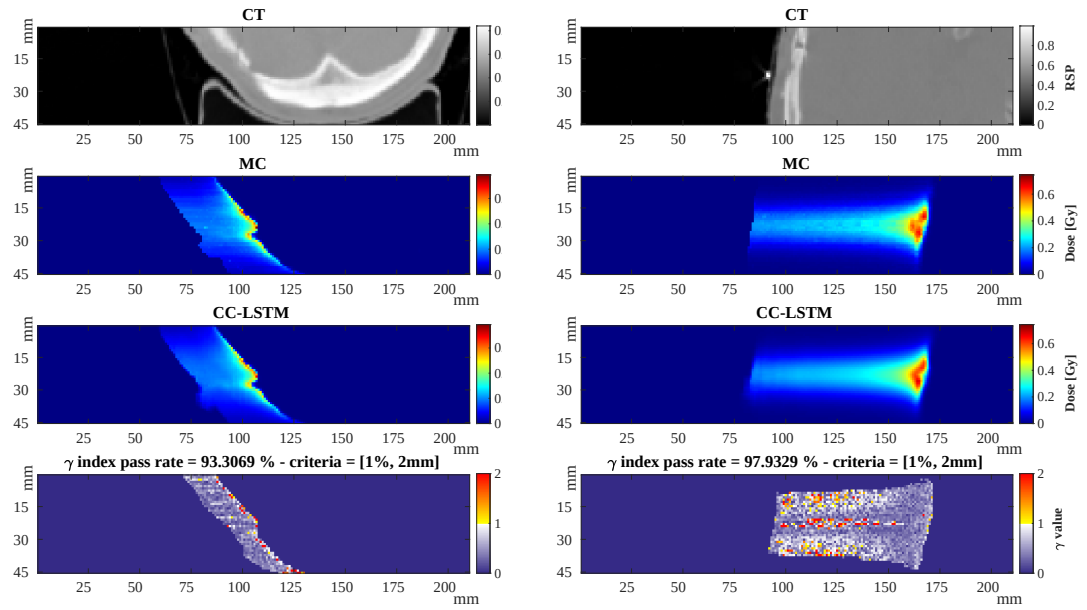
Figure 3.9: Dose estimation comparison for a representative sample: (a) LSTM model versus (b) the newly developed CC-LSTM model.



(a) LSTM model dose estimation.

(b) CC-LSTM model dose estimation.

Figure 3.10: Dose estimation results for a representative sample: (a) LSTM model versus (b) the newly designed CC-LSTM model.



(a) Low-range model dose estimation. (b) High-range model dose estimation.

Figure 3.11: Dose estimation results for a representative sample: (a) LSTM model versus (b) the output of the newly designed CC-LSTM model.

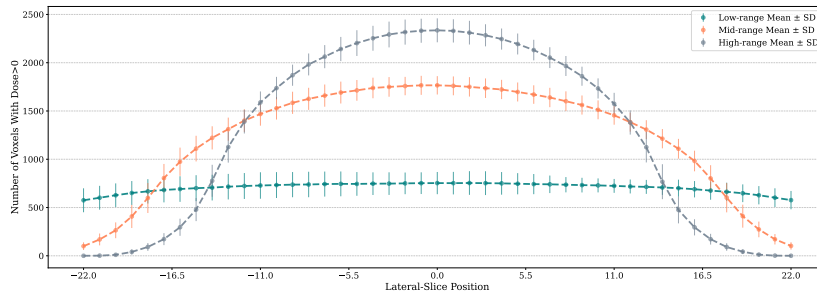


Figure 3.12: Distribution of the average number of dose voxels with dose values greater than zero.

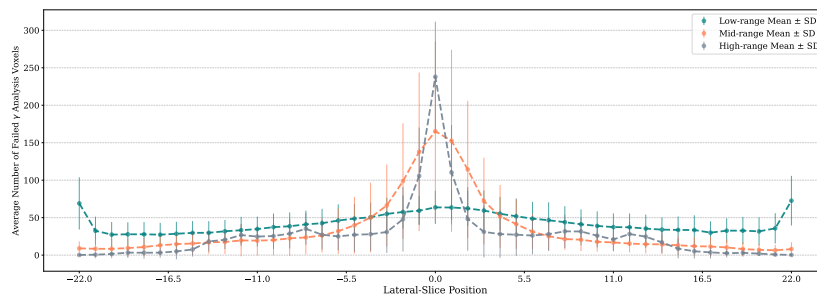
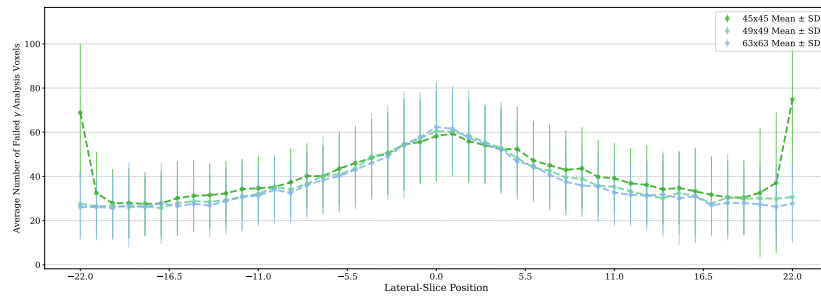
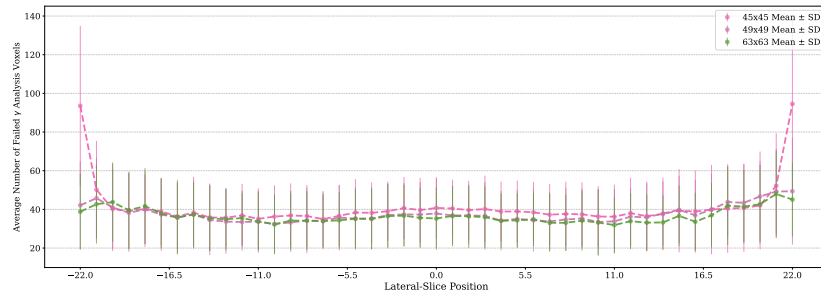


Figure 3.13: Lateral distribution of 'failed' voxels across the lateral axis.



(a) Average lateral distribution of 'failed' voxels across the lateral axis (axis a).



(b) Average lateral distribution of 'failed' voxels across the lateral axis (axis b).

Figure 3.14: The average lateral distribution of 'failed' voxels across the lateral axis for both axes (a) and (b).

Table 3.12: Local γ -index analysis ([1 % , 2 mm]), MAE, and MSE of the CC-LSTM model compared to MC calculations for the three representative energies

| γ -index analysis ([1 % , 2 mm]) | | | | |
|---|----------|---------|---------|---------|
| Energy | mean (%) | std (%) | min (%) | max (%) |
| low-range | 94.28 | 4.42 | 71.69 | 98.31 |
| mid-range | 97.21 | 0.84 | 92.37 | 99.81 |
| high-range | 97.83 | 0.91 | 92.65 | 98.64 |

| Other Metrics | | | |
|---------------|----------------------|------------------------|---------|
| Model | MAE (Gy) | MSE (Gy ²) | PDP (%) |
| low-range | 7.5×10^{-4} | 1.66×10^{-5} | 5.19 |
| mid-range | 1.1×10^{-3} | 2.41×10^{-5} | 3.9 |
| high-range | 1.8×10^{-3} | 4.46×10^{-5} | 5.8 |

range, as depicted in Figure 3.15. No instances were identified where DoTA notably outperformed CC-LSTM.

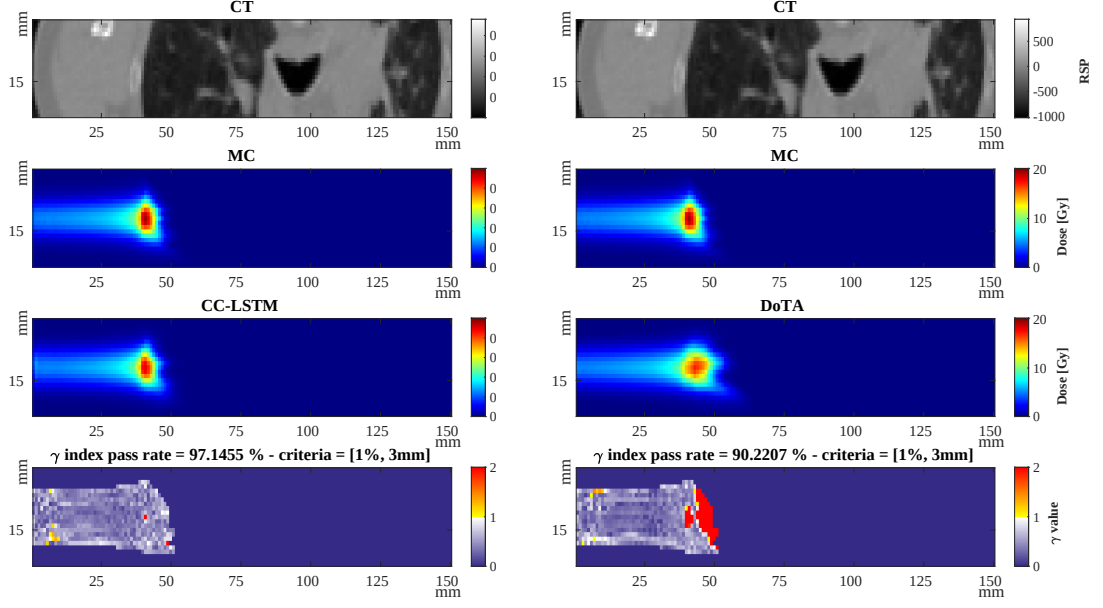
Table 3.13: Evaluation of the CC-LSTM model performance relative to ground truth data from MC simulations at a low-range initial energy of 79.17 MeV, employing three unique cube parametrizations along the lateral axis.

| γ -index analysis ([1 %, 2 mm]) | | | | | Other Metrics | |
|--|----------|---------|---------|---------|----------------------|------------------------|
| lateral shape | mean (%) | std (%) | min (%) | max (%) | MAE (Gy) | MSE (Gy ²) |
| 45 × 45 | 94.06 | 1.83 | 86.1 | 96.8 | 8.9×10^{-4} | 2.4×10^{-5} |
| 49 × 49 | 94.59 | 1.73 | 87.6 | 96.7 | 8.3×10^{-4} | 2.3×10^{-5} |
| 63 × 63 | 94.73 | 1.58 | 88.2 | 96.9 | 8.2×10^{-4} | 1.4×10^{-5} |

Table 3.14: Comparison of CC-LSTM (left) against DoTA (right) using gamma index analysis for three distinct PB energies.

| γ -index analysis ([1 %, 2 mm]) | | | | | |
|--|--------------|-------------|----------|------|-----------------|
| Energy | CC-LSTM (%) | | DoTA (%) | | Improvement (%) |
| | mean | std | mean | std | |
| low-range | 99.8 | 0.43 | 98.8 | 1.23 | +1.0 |
| mid-range | 99.75 | 0.57 | 98.65 | 1.5 | +0.9 |
| high-range | 99.62 | 0.92 | 97.58 | 4.0 | +2.08 |

| Other Metrics | | | | |
|---------------|--|----------------------|---|-----------------------|
| Model | MAE (Gy) | | MSE (Gy ²) | |
| | CC-LSTM | DoTA | CC-LSTM | DoTA |
| low-range | 1.3×10^{-3} | 3.2×10^{-2} | 5.0×10^{-5} | 5.3×10^{-2} |
| mid-range | 8.9×10^{-4} | 3.6×10^{-2} | 1.43×10^{-5} | 5.7×10^{-2} |
| high-range | 1.2×10^{-3} | 4.4×10^{-2} | 1.57×10^{-5} | 4.70×10^{-2} |



(a) Dose estimation by CC-LSTM model. (b) Dose estimation by DoTA model.

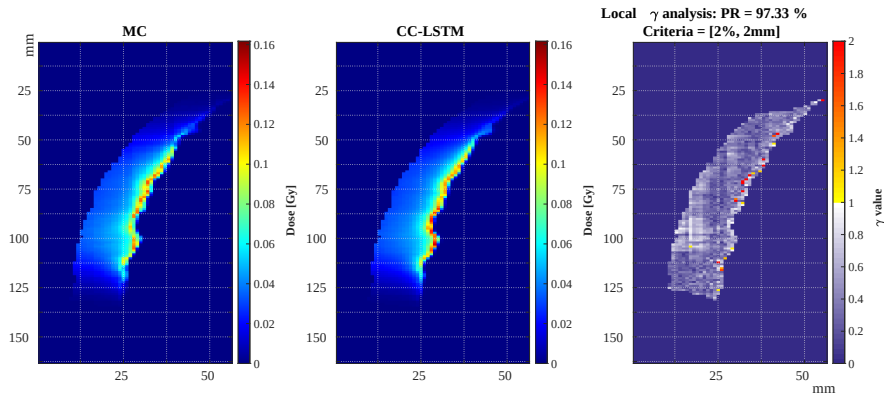
Figure 3.15: Comparison of dose estimation for a sample case: (a) CC-LSTM model and (b) DoTA model.

3.4 FULL-FIELD DOSE ESTIMATIONS

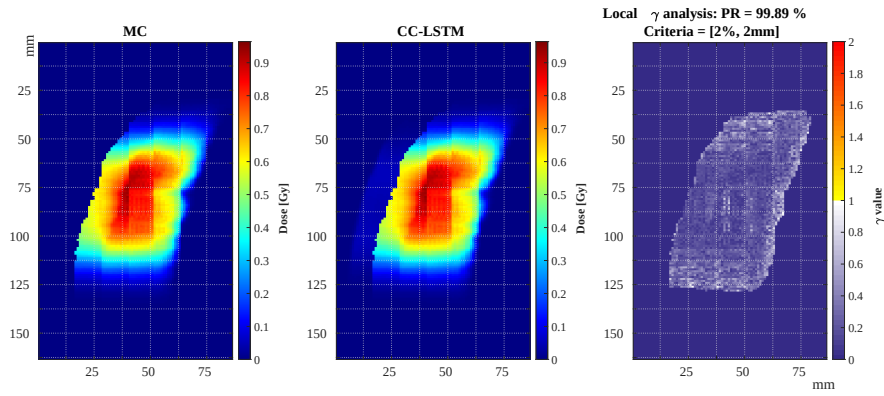
In this section, the CC-LSTM's performance in estimating the dose of a single IES and eventually the entire SOBP, for realistic patient cases were studied.

3.4.1 Estimating IES Dose Distributions

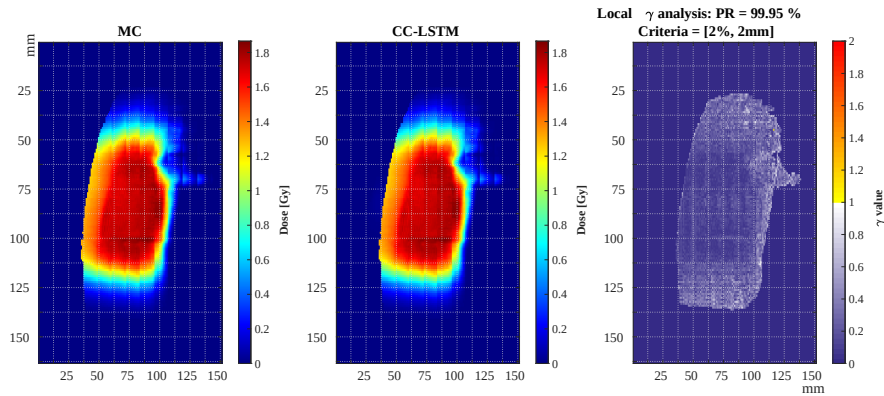
Here, we perform comparisons within the pencil beam domain and through realistic patient cases, evaluating IES dose distributions for the chosen energy ranges. Figure 3.16 compares the IES dose distribution of the three representative energies with the ground truth MC simulation. For the two mid-range and high-range energies, the previous IES's dose distributions were generated from the ground-truth MC dose distributions, and only the corresponding energy IES were simulated using the trained CC-LSTM. This method allows gamma analysis to effectively compare the two cubes, enabling us to focus exclusively on the dose distribution generated by the model for the current energy.



(a) Low-range energy IES dose distribution.



(b) Mid-range energy IES dose distribution.



(c) High-range energy IES dose distribution.

Figure 3.16: Comparison of IES dose distributions for low-range, mid-range, and high-range energies in the BEV. Each energy level's current IES is simulated using the CC-LSTM model, while previous IESs are generated from the ground truth MC dose distributions. This approach allows direct evaluation of the CC-LSTM's effectiveness in estimating the current IES dose distribution.

3.4.2 SOBPDose Estimations

Finally, having demonstrated the capability of CC-LSTM in accurately estimating an entire IES dose distribution, we began scaling up the designed architecture to train for all the defined energies in the curated dataset. The box plot presented in Figure 3.17 summarizes the outcome of performing γ analysis with the same strict [1 %, 2 mm] local criterion used earlier, when training all the defined energies with the unified parametrization established previously.

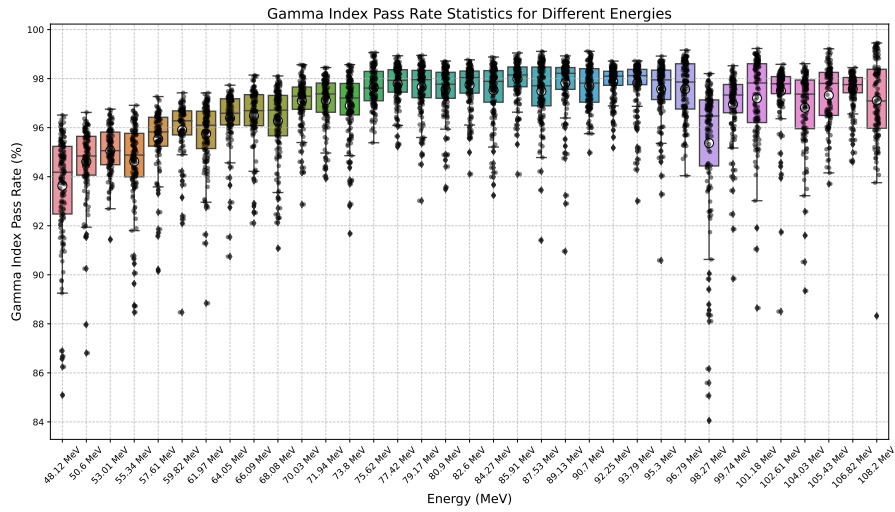
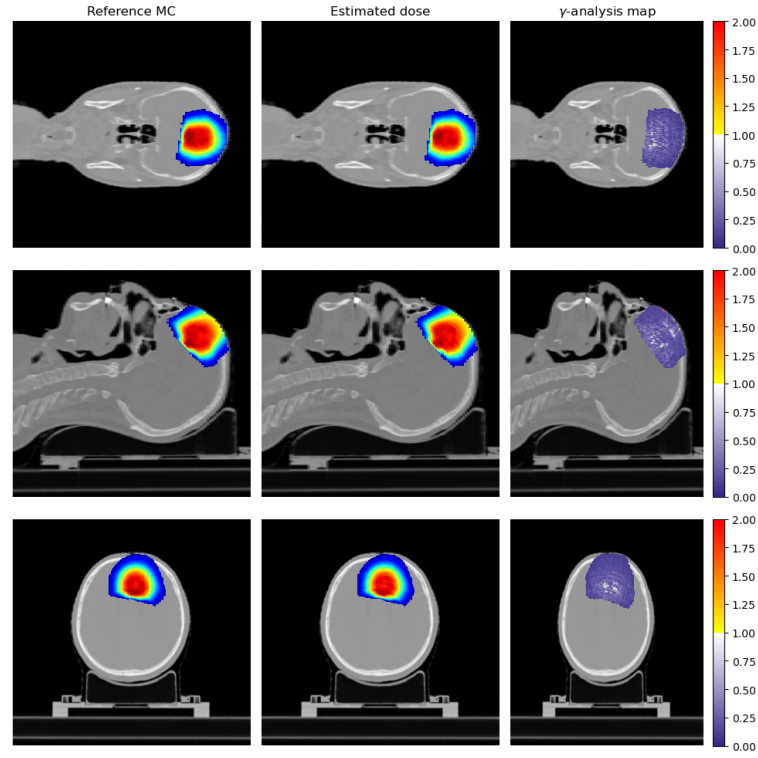


Figure 3.17: γ analysis ([1%, 2 mm], local) pass rates of all defined energies.

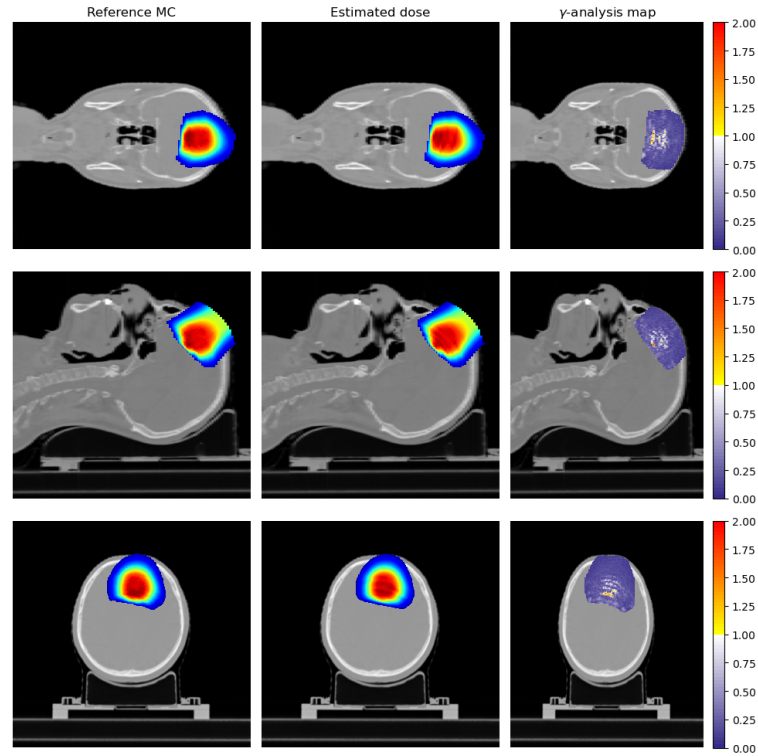
Having prepared specialized models trained for each of the defined 35 energies, and following the workflow described in Section 2.3 to accumulate the dose for an entire field, we generated the dose distribution for the 7 unseen patients. Table 3.15 summarizes the results of the γ -analysis performed on all these test patients.

As reported in Table 3.15, patient Ho3476, shown in Figure 3.18, is the only case where both networks estimated the overall dose distribution with exceptional accuracy, each achieving over 99.5 %. Figure 3.19 compares the Dose Volume Histogram (DVH) comparison from both models against the ground truth Monte Carlo distribution.

Patient Ho2520 is a case where both CC-LSTM and LSTM-133 achieve one of their lowest γ -index pass rates, as depicted in Figure 3.20, with the DVH comparison shown in Figure 3.22. It is also the case where the two models exhibit the highest performance discrepancy, highlighting CC-LSTM's capability in handling challenging scenarios with heterogeneous geometries. Given the apparent discrepancy in the accumulated dose in the left inner ear, as shown in



(a) Dose estimation by CC-LSTM model.



(b) Dose estimation by LSTM-133 model.

Figure 3.18: γ analysis comparison ([2 %, 2 mm]; global) of dose estimation for an unseen, real-patient (H03476) case with 1537 PBs, calculated via (a) CC-LSTM model (γ PR: 99.5) and (b) LSTM-133 model (γ PR: 99.7), against the ground truth MC simulations. The 3D cube views (coronal, sagittal, and axial) are shown for the regions where the MC dose is maximum.

Table 3.15: γ -index evaluation ([2 %, 2 mm]; global) of CC-LSTM and LSTM-133 models on all test patient cases, against the ground truth MC simulations.

| Patient ID | Pbs (#) | IES (#) | γ index pass rate (%) | |
|------------|---------|---------|------------------------------|----------|
| | | | CC-LSTM | LSTM-133 |
| Ho2651 | 4131 | 35 | 99.4 | 96.8 |
| Ho2794 | 4999 | 28 | 98.3 | 96.2 |
| Ho2348 | 4155 | 33 | 99.4 | 95.9 |
| Ho2593 | 11 135 | 32 | 99.7 | 97.0 |
| Ho3037 | 9656 | 35 | 99.4 | 97.2 |
| Ho3476 | 1537 | 21 | 99.5 | 99.7 |
| Ho2520 | 8686 | 33 | 98.4 | 95.0 |

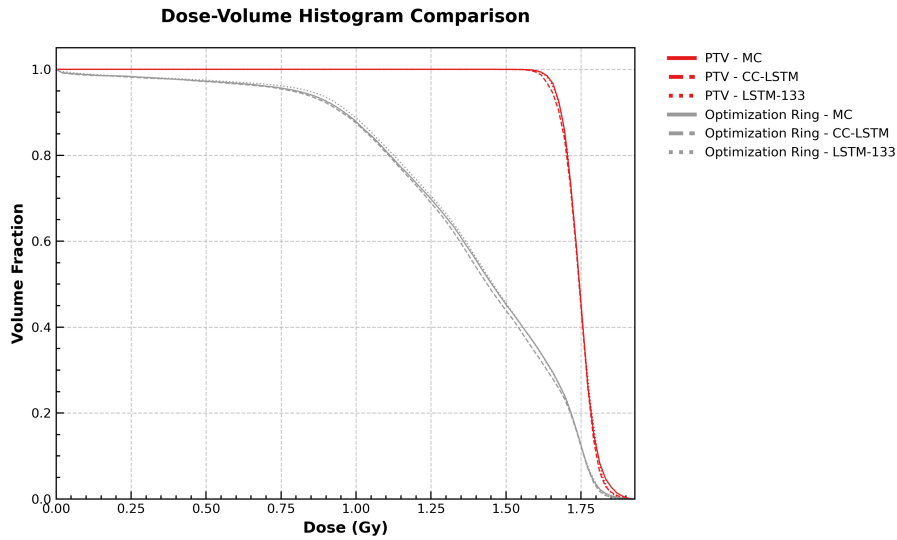


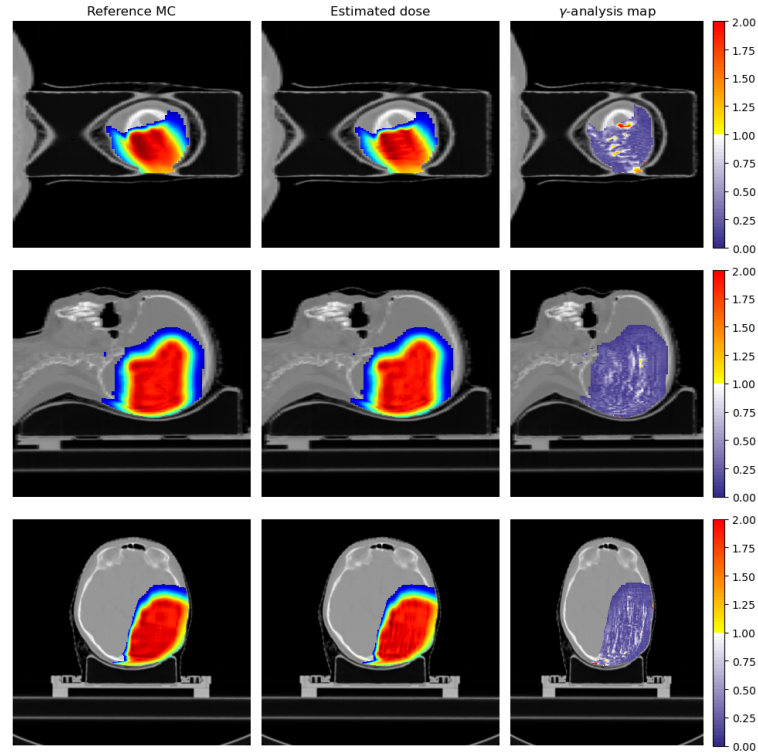
Figure 3.19: DVH comparison for patient Ho3476, illustrating the performance of MC, CC-LSTM, and LSTM-133 models for both the Planning Target Volume (PTV) and Optimization Ring.

Figure 3.22, we outline the OAR in Figure 3.21 to cross-check the γ -analysis with the discrepancy reported in the DVH analysis.

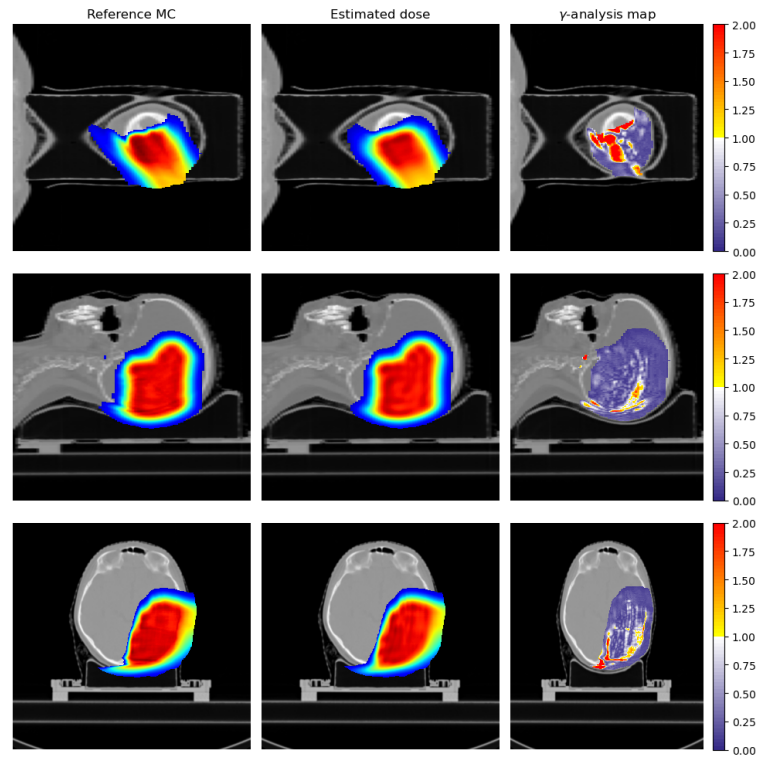
Finally, the dose distribution for patient Ho2348 is another example where, similar to the case of patient Ho2520 in Figure 3.20, CC-LSTM demonstrates exceptional capabilities in estimating dose distributions for highly heterogeneous geometry cases. Figure 3.23 illustrates this by showcasing the dose distribution comparison of CC-LSTM and LSTM-133 against MC in the bony structures of the face and nose area.

Given its proximity to the bony and heterogeneous structures, Figure 3.24 presents the DVH comparison of the left optical nerve to highlight the discrepancies between the two models' performance. The LSTM-133 model reported a 1.8 % overdose to the OAR, while CC-LSTM closely matched the ground truth, with only a 0.7 % overdose to the organ.

To avoid the high dosimetric values of the target and other organs overshadowing the comparison, we focused on the DVH for the left optic nerve to better illustrate the differences between the two models. The full DVH reports, along with three view γ index plots of test patients reported in Table 3.15 can be found in Appendix 6.

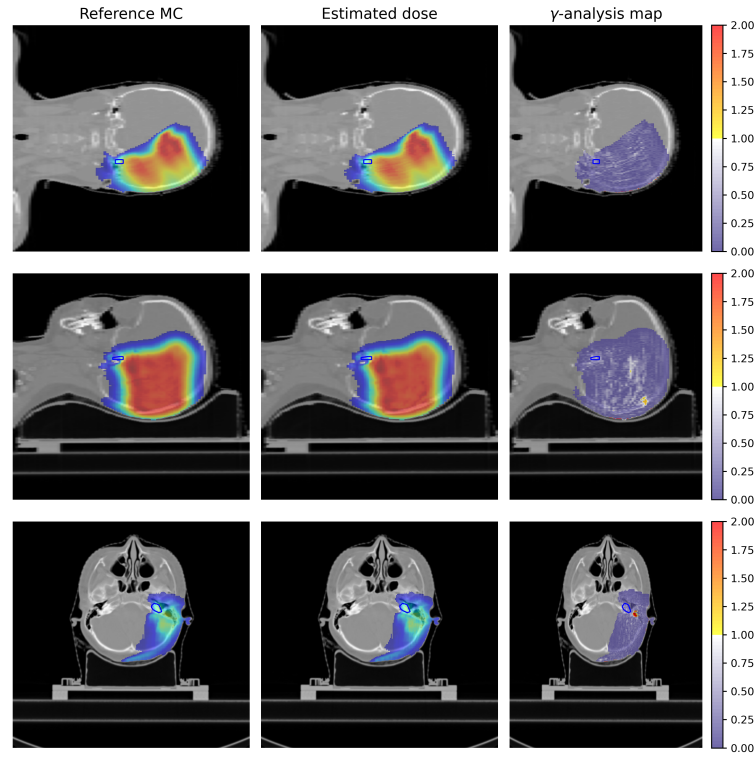


(a) Dose estimation by CC-LSTM model.

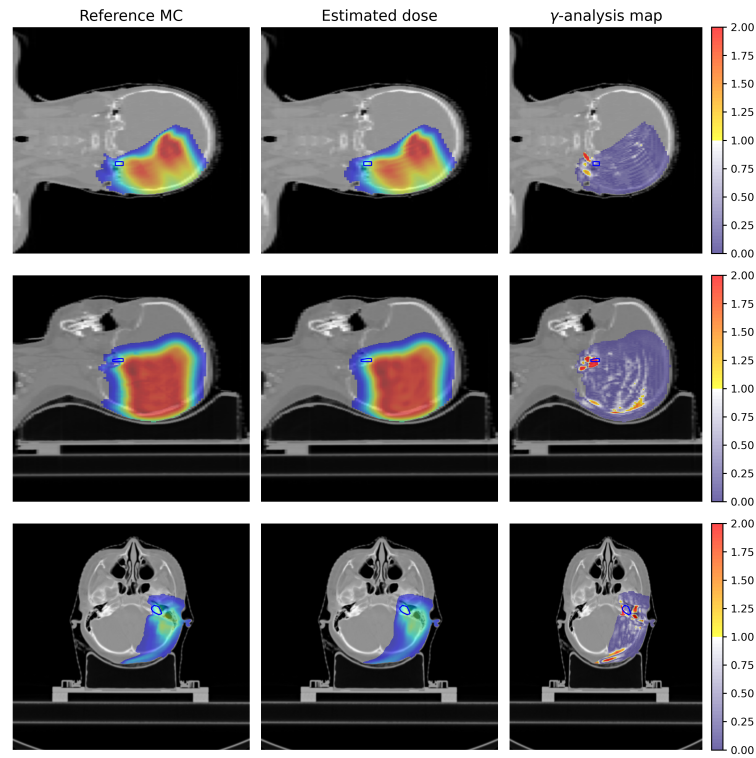


(b) Dose estimation by LSTM-133 model.

Figure 3.20: γ analysis comparison ([2 %, 2 mm]; global) of dose estimation for an unseen, real-patient (Ho2520) case with 8686 PBs, calculated via (a) CC-LSTM model (γ PR: 98.4) and (b) LSTM-133 model (γ PR: 95.0), against the ground truth MC simulations. The 3D cube views (coronal, sagittal, and axial) are shown for the regions where the MC dose is maximum.



(a) Dose estimation by CC-LSTM model.



(b) Dose estimation by LSTM-133 model.

Figure 3.21: γ analysis comparison ([2 %, 2 mm]; global) of dose estimation for an unseen, real-patient (Ho2520) case with 8686 PBs, calculated via (a) CC-LSTM model (γ PR: 98.4) and (b) LSTM-133 model (γ PR: 95.0), against the ground truth MC simulations. The 3D cube views (coronal, sagittal, and axial) are displayed, with the left inner ear visible and contoured in blue.

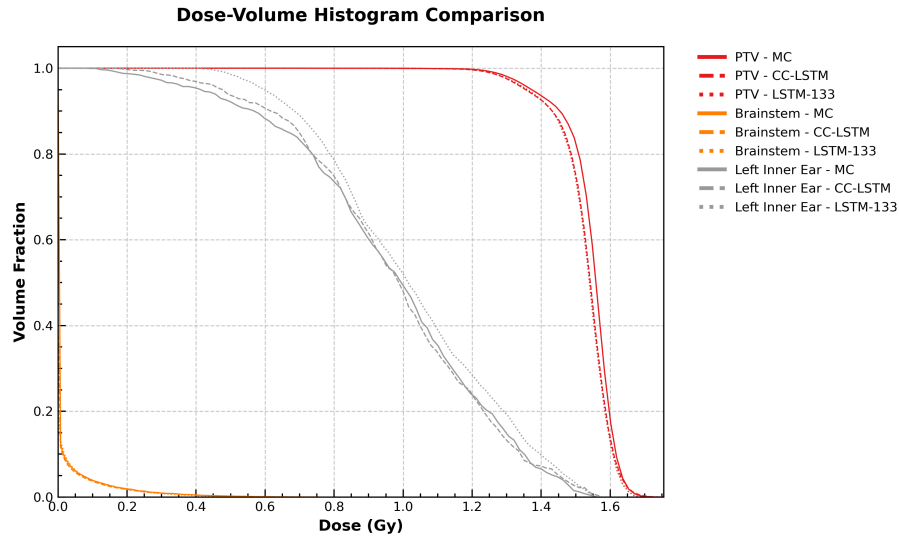
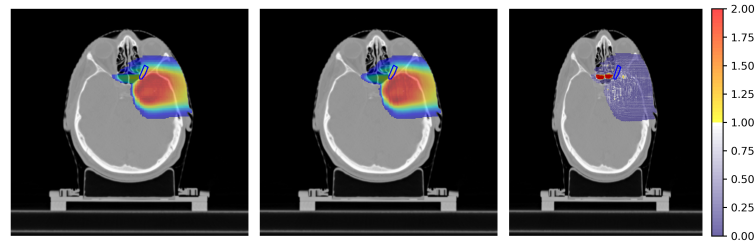
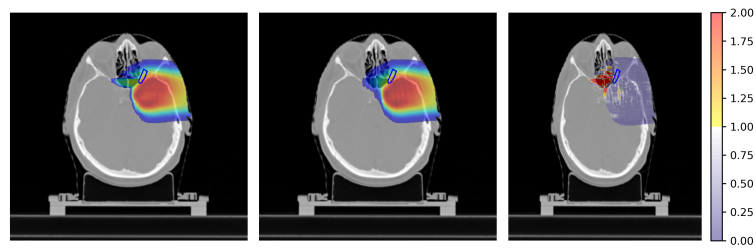


Figure 3.22: DVH comparison for patient Ho2520, illustrating the performance of the MC, CC-LSTM, and LSTM-133 models for the target PTV, as well as two OARs: the left inner ear and the brainstem.



(a) Dose estimation by CC-LSTM model.



(b) Dose estimation by LSTM-133 model.

Figure 3.23: γ analysis comparison ([2 %, 2 mm]; global) of dose estimation for an unseen, real-patient (Ho2348) case with 4155 PBs, calculated via (a) CC-LSTM model (γ PR: 99.4) and (b) LSTM-133 model (γ PR: 95.9), against the ground truth MC simulations. The presented slice is manually selected to highlight the left optical nerve (contoured in blue).

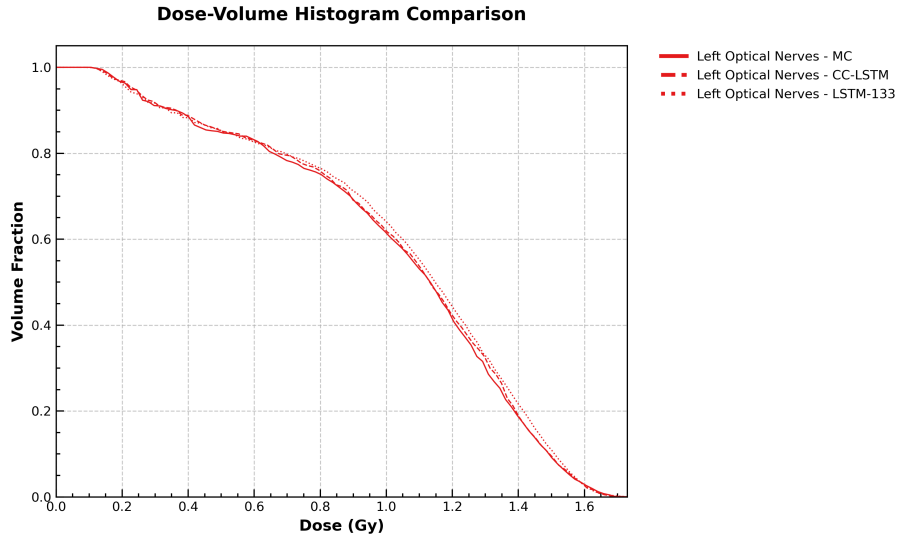


Figure 3.24: DVH comparison for patient Ho2348, illustrating the performance of the CC-LSTM and LSTM-133 models for the OAR in proximity to highly heterogeneous geometry, specifically the left optical nerve, as shown in Figure 3.23.

3.5 RUN TIME ANALYSIS

Having demonstrated the superiority of CC-LSTM over the earlier model and the current state-of-the-art in generating both single PB dose distributions and full IES dose distributions, as well as its generalizability to unseen test cases and real patient scenarios, the run times of the CC-LSTM are analyzed. As previously mentioned, the objective of this study is to estimate the physical dose appropriate for real-time adaptive APT. In this section, the model's run times for representative samples are reported and compared to the earlier model and the current state-of-the-art. Additionally, it will be demonstrated how CC-LSTM can be custom-designed for each energy and the run time advantages gained from such modifications will be highlighted. Finally, it will be shown how these run times are achievable in real patient case scenarios.

Table 3.16 outlines the run times of CC-LSTM, along with other factors affecting run times, using the default hyperparameters to achieve the results reported in Table 3.12.

Table 3.13, shows that modifying the data input shape could alleviate failing voxels at the peripheries to achieve higher accuracy. The mono-energetic approach not only provides flexibility in tailoring parameterization to optimize accuracy but also allows customization of the model's parameterization or data I/O shapes based on the desired run times or specific energy require-

Table 3.16: CC-LSTM run times vs its predecessor and DoTA

| setup | | | | | run time |
|---------|-------------------------|---------------------------|---------------------------|-----------|--------------|
| Model | Inference Batch Size | # of Learn. Parameters | Data I/O Shape | Framework | Mean (ms) |
| CC-LSTM | 48 | 424 353 | $210 \times 45 \times 45$ | PyTorch | 5.6 |
| LSTM | 64 | 8 072 525 | $210 \times 45 \times 45$ | PyTorch | 2.6 |
| CC-LSTM | 128 | 162 209 | $150 \times 23 \times 23$ | PyTorch | 1.6 |
| DoTA | 32 | 12 632 809 | $150 \times 24 \times 24$ | Keras | 5.6* |

Table 3.17: CC-LSTM run times when introducing customized parameterization in the longitudinal axis

| setup | | | | | run time |
|---------|-------------------------|---------------------------|---------------------------|-----------|--------------|
| Model | inference batch size | # of learn. parameters | Data I/O shape | Framework | Mean (ms) |
| CC-LSTM | 48 | 424 353 | $210 \times 45 \times 45$ | PyTorch | 5.6 |
| CC-LSTM | 48 | 424 353 | $160 \times 45 \times 45$ | PyTorch | 4.2 |
| CC-LSTM | 64 | 424 353 | $130 \times 45 \times 45$ | PyTorch | 3.6 |

ments. Table 3.17 illustrates this by presenting various configurations of the default design and their impact on run time improvements.

For low-range energy sets, as Figure 3.11 suggests, a straightforward modification of the parameterization can be achieved by defining a shorter longitudinal clipping. Since the 48.12 MeV PB won't have ranges higher than the extracted ≈ 13 cm used in this analysis, this adjustment reports a 35 % improvement in run times, outlined in Table 3.18.

Table 3.18: CC-LSTM run times to generate the dose distribution of an entire IES (*: showcased in sub-figures outlined in 3.16)

| setup | | | | run time |
|-----------------|-------------------|---------------------------|------------|--------------------|
| Energy (MeV) | PBs in IES (#) | Data I/O shape | batch size | per IES (s) |
| 48.1 MeV* | 349 | $210 \times 45 \times 45$ | 64 (32) | 0.93 (1.35) |
| 79.1 MeV* | 400 | $210 \times 45 \times 45$ | 64 (32) | 1.11 (1.62) |
| 102.6 MeV* | 34 | $210 \times 45 \times 45$ | 64 (32) | 0.16 (0.25) |
| 79.1 MeV | 400 | $160 \times 45 \times 45$ | 64 (32) | 0.87 (1.27) |
| 48.1 MeV | 349 | $160 \times 45 \times 45$ | 64 (32) | 0.73 (1.05) |
| 48.1 MeV | 349 | $130 \times 45 \times 45$ | 64 (32) | 0.6 (0.86) |

Having demonstrated the potential of CC-LSTM in delivering dose estimation for real-time IES dose estimations in Table 3.18, and by incorporating the trained energies reported in Figure 3.17, the run times for carrying out SOBP dose estimations, with accuracies reported in table 3.15, are outlined in table 3.19.

Table 3.19: Comparison of CC-LSTM and LSTM-133 models on all test patient cases, highlighting the γ -index pass rates ([2 %, 2 mm]; global) and average run times (mean \pm std) based on 5 measurements.

| Patient ID | Pbs (#) | IES (#) | γ index pass rate (%) | | run time (s) (mean \pm std) | |
|------------|---------|---------|------------------------------|----------|-------------------------------|----------------|
| | | | CC-LSTM | LSTM-133 | CC-LSTM | LSTM-133 |
| Ho2651 | 4131 | 35 | 99.4 | 96.8 | 41.7 \pm 0.4 | 11.4 \pm 0.3 |
| Ho2794 | 4999 | 28 | 98.3 | 96.2 | 44.5 \pm 0.3 | 13.8 \pm 0.1 |
| Ho2348 | 4155 | 33 | 99.4 | 95.9 | 39.2 \pm 0.1 | 11.8 \pm 0.3 |
| Ho2593 | 11135 | 32 | 99.7 | 97.0 | 79.7 \pm 0.1 | 30.9 \pm 0.7 |
| Ho3037 | 9656 | 35 | 99.4 | 97.2 | 81.6 \pm 0.4 | 26.2 \pm 0.3 |
| Ho3476 | 1537 | 21 | 99.5 | 99.8 | 18.1 \pm 0.2 | 4.5 \pm 0.4 |
| Ho2520 | 8686 | 33 | 98.4 | 95.0 | 67.4 \pm 0.5 | 23.9 \pm 0.4 |

DISCUSSION

This study begins by extending the preliminary feasibility work presented in Section 1.4, followed by an investigation into upscaling the proposed method for full-field dose estimation. It was demonstrated that RNN-based networks, specifically LSTM networks, can accurately predict three-dimensional dose distributions of mono-energetic PBs, showing high agreement with ground truth MC simulations and generalizing well to unseen patient cases and varying energies. Furthermore, the runtimes achieved make these models suitable for incorporation into real-time adaptive frameworks, which is the primary objective of this research.

However, the crucial aspect of quantifying uncertainty in dose predictions had not yet been addressed. Therefore, we proposed the B-LSTM based model (BayesDose). Unlike deterministic models, BayesDose incorporates probabilistic elements, allowing us to measure uncertainty across different samples and scenarios. This capability is important for clinical translation, as it provides insights into the confidence levels of the model's predictions, enhancing the robustness and trustworthiness of the dose calculations.

While BayesDose quantifies uncertainty in dose estimation, it was not primarily designed to demonstrate accuracy improvements. However, the insights gained from uncertainty quantification highlighted the areas where our initial framework could be enhanced. The original model encountered difficulties in accurately predicting the wide dose distribution of pencil beams, especially in complex anatomical regions.

To address these challenges, LSTM-133 and CC-LSTM models were developed, which were specifically designed to handle the wide lateral range of dose distributions encountered in previously delivered treatment plans. By taking into account the underlying physics of proton interactions, the model sizes were significantly reduced, leading to an increase in their computational efficiency. These architectural designs improved accuracy and achieved the computational run time necessary for real-time APT.

FEASIBILITY STUDY: LESSONS LEARNED

The proton dose estimation problem was approached as a supervised learning task, aiming to map input CT cubes (RSP distribution) to their corresponding 3D dose distributions. Various ANN architectures and deep learning models were investigated, with a particular focus on LSTM networks, favored for their ability to capture spatial dependencies and handle sequential data effectively. The training process incorporated an updated dataset of dose distributions generated through MC simulations, with different initial energies impinging on random spots across the lung area, serving as the ground truth.

The feasibility study revealed three additional key findings. Firstly, although trained on samples extracted from a single patient, the model's learning generalized well to other test samples from unforeseen patients. Secondly, when trained on samples with distinct energies, different from the prior energy, the performance generalized for both short-range PBs and long-range PBs. Tables 3.1 and 3.2, as well as Figure 3.1, demonstrate this both in terms of visual inspection and the devised metrics.

Last but not least, the LSTM model demonstrated superior performance in computational efficiency for dose estimation, achieving run times faster by orders of magnitude compared to MC simulations (Table 3.3), as well as other tested architectures, such as 3D-CNN (Neishabouri et al., 2021a) and Transformer-based architectures. The millisecond run time performance of LSTM models highlights their potential to provide real-time dose calculations, which are essential for adaptive proton therapy.

However, several limitations were identified. First, the dataset incorporated was not sampled from clinically delivered plans, which could affect the learning task considerably. Moreover, the LSTM model struggled to accurately estimate the dose of PBs with a wide lateral range, resulting in a loss of coherence in the dose predictions (Figure 3.8a).

UNCERTAINTY QUANTIFICATION

Using our proposed approach, the BNNs incorporated in BayesDose facilitated uncertainty estimates for dose calculation. Unlike traditional deterministic models that when trained, output a single predicted value, BNNs incorporate probabilistic elements, allowing the model to produce a distribution of possible outcomes. This probabilistic approach enabled the quantification of

uncertainty and provided insights into the confidence of the proposed models in estimating dose distributions in various scenarios.

By incorporating Bayesian inference, BayesDose evaluates the uncertainty associated with each dose prediction, offering a range of potential dose distributions rather than a single deterministic estimation. Both BayesDose and its deterministic variant have demonstrated robust capabilities for dose calculation, achieving pass-rates greater than 97.81 %. The most accurate predictions occurred in scenarios where cuboid heterogeneities had minimal impact on the dose distribution, leading to only slight distortions of the Bragg peak. The network consistently identified these instances as cases with low output uncertainty. An instance of such high accuracy from BayesDose, depicted in Figure 3.2, shows that the maximum standard deviation was only 1.1 % relative to the maximum dose, with no voxel deviating more than 3σ from the predicted values.

The instance with the lowest γ pass-rate in the test set, 88.93 %, is detailed in Figure 3.3. This example highlights significant discrepancies between the predicted and the ground truth dose distributions, primarily occurring in high dose regions coupled with high uncertainty. Notably, the standard deviation from BayesDose in this scenario reached up to 12.9 % relative to the maximum dose, marking it as one of the highest observed in the experiment. This level of deviation underscores a less reliable prediction compared to instances such as those depicted in Figure 3.2.

This approach highlights areas where the model is confident (Figures 3.2 in the phantom case study, and 3.5 in the patient case study, both reporting 100 % γ -index pass rates) and regions where predictions are less certain (Figures 3.3 in the phantom case study, and 3.6 in the patient case study, both reporting worst dose estimation accuracies). The ability to quantify uncertainty is particularly valuable in medical physics, where the uncertainty maps provided by BayesDose allow clinicians to assess the reliability of the dose calculations and identify potential risks for OARs, leading to informed decisions. This capability was demonstrated in the Section 3.2, where the model's uncertainty estimates correlated with the complexity of the anatomical regions being treated, indicating areas where further review or alternative planning might be necessary.

The distribution of voxels outside the $n\sigma$ range for the entire test set is presented in Table 3.5. These values facilitate a comparison with Gaussian assumptions about the probability mass within a confidence interval. These metrics consistently fall short of expectations for the phantom cases, suggesting an overestimation of uncertainty. Figure 3.4 presents a scenario where the largest proportion of voxels exceeded the 3σ deviation threshold, under-

scoring significant prediction errors. The elevated percentage of significantly deviating voxels, as shown in Figure 3.4, stems from a systematic underestimation of the proton path length, which consequently misplaces the Bragg peak and high dose values distally. This misplacement indicates the model's failure to accurately capture the extended downstream dependency induced by the entrance cavity in this specific instance.

While BayesDose is not yet the primary dose calculation scheme, it can effectively be used to flag scenarios with large standard deviations and high uncertainty, e.g. samples illustrated in Figures 3.4 and 3.7 exhibiting significant quantification of uncertainty. Regions with high uncertainty can indicate potential errors or limitations in the model, prompting using models with lower uncertainty. Section 3.3 outlined a case study where custom parameterization was incorporated for low energies, showcasing the favorable characteristics of the proposed mono-energetic approach in tailoring the design according to the complexity of the samples. Similar analysis can be carried out using BayesDose, flagging samples with high uncertainty and prompting the workflow to opt for models with higher confidence in estimating dose distributions in heterogeneous scenarios. Ultimately, for estimating the dose distribution of an entire IES (or SOBP), an optimization task can be defined based on the available models, their uncertainty in estimating dose distributions, and finally their computational efficiency, aiming to minimize the run times while maximizing the dosimetric accuracy.

Moreover, BayesDose replicated the generalization behavior of the deterministic model when tested via the earlier benchmarking procedure for generalizability, reporting similar performance to the deterministic LSTM model. A similar performance discrepancy was observed for two of the test patients (Patient 1 and 5), which was attributed to the range of input values, i.e. RSP map values, where in those cases the RSP range of values were beyond those observed in Patient 0, where the training was performed.

In the regulatory landscape of medical devices and treatment planning systems, demonstrating robust uncertainty quantification can support compliance with strict safety standards. It provides evidence that the model's predictions are accurate and reliable under varying conditions. The uncertainty maps generated by BayesDose could thus play a critical role in validating the model for clinical use, ensuring it meets the necessary safety and efficacy requirements.

CC-LSTM: THE STATE-OF-THE-ART

Following the proposed approach in Sections 1.4 and 2.1, Pastor-Serrano and Perkó (2022) have implemented a Transformer-based architecture, achieving state-of-the-art accuracy. Additionally, they employed a parametric input for particle energy, generating dose distributions across a semi-continuous range of therapeutic energies. However, we argue that despite their recent popularity and strong performance, Transformers are not ideally suited for this task. Their prominence is due to training parallelism, enabling handling of large language corpora. Yet, they suffer from inefficient inference due to their $O(N)$ complexity per step (Sun et al., 2023), a critical drawback given our study's primary aspiration of real-time inference. This issue is further exacerbated in cases of larger input/output dimensions, as showcased in this study.

Moreover, while acknowledging the scientific value of mixed-input capability of these models, we contend against a one-for-all model approach. By concentrating the learning objectives on mono-energetic pencil beams, the complexity of the learning task is significantly reduced to mainly involve the patient's geometry. The decision to exclude energy parameterization from the problem formulation is based on several strategic considerations:

- 1) Within the realm of IMPT, each IES is treated sequentially and distinctly. This operational characteristic implies that for real-time dose recalculations, there is no necessity to handle varying energy pencil beams simultaneously. Thus, introducing energy parameterization into the model would complicate it unnecessarily and divert from the primary optimization goals.
- 2) Pencil beams exhibit markedly different characteristics across various energies due to underlying physical phenomena such as range straggling and distinct scattering profiles. For instance, beams at lower energy levels are characterized by pronounced scattering, leading to a broader lateral dose distribution, while those at higher energies tend to have a narrower focus but an extended Bragg peak. These significant physical variations justify the need for tailored machine-learning strategies for each energy level, especially when considering the application across different types of cancer treatments.
- 3) In scenarios where the complete field dose distribution needs to be calculated, the cumulative dose is essentially the sum of the individual IES dose distributions. This non-overlapping nature of data lends itself well to parallel processing across multiple GPUs, thereby enhancing computational efficiency in proportion to the available hardware resources.

When a clinically justifiable dose cut-off is incorporated, the extent of a mono-energetic pencil beam's dose distribution both laterally and longitudinally can be estimated. Therefore, the learning task employs a restricted representation instead of the entire CT/Dose grid, focusing on a clipped area that aligns with the beam's eye view to encapsulate the complete extent of the pencil beam's dose projection. This method is depicted in Figure 2.14 for initial energy configurations in an LGG patient case, illustrating the energy-specific clipping dimensions determined by the behavior of the pencil beam across its operational range.

In contrast to Transformers, RNNs, particularly in their advanced forms, such as LSTM networks, handle sequential data with a fixed-cost operation per element, independent of sequence length. The recent decline in the usage of these models is primarily attributed to their need to process the entire sequence of input data sequentially and condense it into a final hidden state. This requirement can complicate decoding, especially in natural language processing, where maintaining the context of lengthy sentences is crucial. However, the application of RNNs in particle therapy offers unique advantages. Unlike in language processing, where the entire sentence must be provided to generate the output, the unidirectional high-energy travel of protons means that to accurately estimate the dose distribution at each time step, only the input up to and including the current time step needs to be considered. Consequently, the common drawbacks of RNN-based networks do not hinder their effectiveness in proton therapy applications.

Building on the foundation set by previous research, the CC-LSTM model development significantly improves the precision of dose estimation for particle therapy. The performance of CC-LSTM model against the previously established LSTM model demonstrated higher accuracy in almost all samples within the test set as summarized in Table 3.11. For instance, the *local* gamma index pass rate ($[1\%, 2\text{mm}]$; $D < 0.03D_{\text{max}} \rightarrow D = 0$) has reported an approximate 5 percentage points increase.

Important to note here is the varying scale of values in γ analysis and the outcome pass rates, given the different dimensions on which the pass rates are evaluated, compared to full-field γ analysis. For a better understanding of the performance superiority, note CC-LSTM's worst γ -index pass rate of 92.37%, which is almost equal to what the prior LSTM model reported on average. The ability to handle worst-case scenarios effectively proves the robustness and accuracy of the CC-LSTM model. Finally, given the training was carried out via an MSE loss function for both models, it is worth noting CC-LSTM's MSE decreases by an order of magnitude, on average, when measured against the prepared test datasets. This performance highlights CC-LSTM's superior

capability in avoiding overfitting and thereby it's improved generalizability to unseen test cases.

CC-LSTM achieved high performance when tested on the LGG dataset prepared for this study as well. As outlined in Section 2.3.3, prepared specifically for HIT (simulated via FLUKA MC), this dataset is considerably more complex than the data prepared in the feasibility study (simulated via Topas) for its wider lateral range, finer extraction resolution, and extreme heterogeneities exhibited in the head region. However, given that this dataset is sampled from previously delivered treatment plans, the desirable outcome suggests enhanced performance of the model in real patient scenarios.

Moreover, as previously discussed, the shift from low-range energy PB to high-range energy PB represents a fundamental change in the task from an AI perspective (Figure 3.11). Similarly, the outcomes of each energy test set may vary even when trained using the same procedure. Performance differences across various energies can be related to several factors, including test set sampling schemes, as each model is tested on a specific test set. However, a notable decline in performance is observed in the gamma analysis outcomes for the low-range energy set. This highlights the benefits of the proposed mono-energetic approach, which allows for unique handling of each energy based on the overall behavior of that energy range.

Low-range energy PBs exhibit a considerably wider lateral extent compared to other sets. This is due to the slow speed of these low-range PBs and their increased scattering in the skull region. Additionally, because of the low focus of proton beams at HIT, more voxels with a dose greater than zero are observed at the periphery of the PB cubes (Figure 3.12). When examining the average lateral distribution of *failed* voxels across the lateral axis, as shown in Figure 3.13, failures for low-range energy PBs predominantly occur at the periphery of the lateral axis of the PB cube. This indicates the model's difficulty in accurately estimating the dose distribution in the margins of the output cube. This observation stems from the fact that the interpolated cubes encounter interpolation artifacts at the periphery, and since low-energy PBs have a wider lateral scattering, these artifacts result in higher discrepancies at the periphery.

To this end, two alternative parameterizations were performed, using 49 voxels and 63 voxels in the lateral axis, resulting in both input \mathcal{G}_i and output cubes \mathcal{D}_i of shape $\mathbb{R}_+^{1 \times 49 \times 49}$ and $\mathbb{R}_+^{1 \times 63 \times 63}$, instead of the original $\mathbb{R}_+^{1 \times 45 \times 45}$. Figure 3.14 demonstrates how the wider cube extraction completely alleviates the failures of voxels in the peripheries, with performance enhancements outlined in Table 3.13, showing improvements in accuracy across all metrics for the low-range energy set.

The tailoring of the model outlined earlier can be practiced in different mechanisms of the model. For instance, as it was pointed out earlier, the proposed workflow can store variants of CC-LSTMs, each trained for certain complex scenarios, and determine what variant to incorporate based on the uncertainties provided by Bayesdose. This is feasible owing to the efficient and reduced number of learnable parameters in CC-LSTMs (Table 3.16).

Performance Against the State of the Art

The superior performance of the CC-LSTM model extends beyond comparisons with previous work, demonstrating significant advantages over the current state-of-the-art Transformer-based model, DoTA. In every metric incorporated, CC-LSTM consistently outperformed DoTA, with particularly pronounced differences in MSE and MAE, where discrepancies in error magnitude were notable (Table 3.14). A direct and fair comparison was conducted using the publicly available dataset by Pastor-Serrano and Perkó (2022). By training CC-LSTM on the same dataset used for DoTA without any alterations in preprocessing or postprocessing steps, we ensured a fair evaluation, further substantiating CC-LSTM's enhanced performance in particle therapy dose estimation.

Evaluating CC-LSTM Models for Full IES Plan Dose Distribution

Thus far, our accuracy assessments have focused on comparisons between individual PBs in the BEV, normalized per fluence proton. However, in realistic clinical scenarios, the dose distributions for a planned treatment are scaled by the fluence of the corresponding spot, which can amplify or obscure PB errors when evaluated and accumulated collectively. To address this, the evaluation was extended to showcase the performance of CC-LSTM models in generating the dose distribution of an entire IES plan for the representative energy ranges, as illustrated in Figure 3.16. As shown, CC-LSTM's performance is not limited to the PB domain but extends to the realistic scenario of estimating an entire IES dose distribution. This performance is especially important for APT, where prompt estimation of IES dose distribution allows for online APT within the 4 s to 6 s time window available during therapy, as is the case at HIT.

Evaluating Proposed Models for Full-Field Dose Distribution

To generate the dose for the seven unseen test patients introduced in Section 2.3.3, CC-LSTM training over all 35 defined energies had to be carried out. The box plot represented in Figure 3.17 demonstrates the generalizability of the high accuracy outlined earlier in this study. There is an apparent performance decline in the lower energies, attributed to the interpolation artifact phenomena explained earlier in low-dose PBs.

However, tailoring each energy to its corresponding dataset was not carried out for this study, and the reported gamma analysis in Figure 3.17 has been performed using a unified parameterization. Moreover, the performance of LSTM-133 was not reported due to limited space and to provide a clearer representation of the CC-LSTM, which is primarily designed to increase accuracy while maintaining superfast run times. Nonetheless, LSTM-133 was designed primarily for run time advantages, and its accuracy report in the PB domain was omitted. However, the performance of LSTM-133 is reported when evaluating the dose for entire field, comparing it with CC-LSTM and the ground truth MC simulations, as shown in e.g. figure 3.18.

Table 3.15 presents the results of the γ -index pass rate analysis for all test patients. The CC-LSTM model demonstrates robust performance, achieving over 98% γ -index pass rates under strict criteria ($[1\%, 2\text{ mm}]$), $D < 0.01D_{\text{max}} \Rightarrow D = 0$, across all unseen patient cases. While the LSTM-133 model generally shows lower accuracy compared to CC-LSTM, it nonetheless exhibits a high level of agreement with the MC dose across all patients. This discrepancy is less apparent when performing DVH analysis, where both models show a close matching with the ground truth MC for target and OARs, across all patients.

Patient Ho3476 is the only case where LSTM-133 marginally outperforms CC-LSTM in terms of γ -index analysis. Upon investigating this case (Figure 3.18), it becomes evident that Ho3476 is one of the few cases where the isocenter is located centrally in the brain region, resulting in nominal cases that are less challenging compared to heterogeneous cases where the dose distribution profile is significantly distorted. In such cases, LSTM-133 can effectively evaluate the dose due to minimal anatomical variations. Furthermore, given the added computational efficiency of LSTM-133 over CC-LSTM, by incorporating probabilistic models such as BayesDose, one could perform preliminary uncertainty quantification and employ LSTM-133 for cases with low uncertainty, thereby benefiting from the runtime savings that LSTM-133 offers. As a clinically relevant endpoint, the DVH plot for this patient shows a close

match between the two proposed models and the MC evaluations as well (Figure 3.19).

It's important to reiterate that these patients were not seen by the trained models in any capacity. For example, Patient Ho3476, discussed earlier, is one of three patients whose gantry and couch angles (GA, CA) were not present in the training set (GA: 50°, CA: 280°; see Figure 2.12). Patient Ho2520 (GA: 90°, CA: 330°), on the other hand, while sharing GA and CA with two other patient cases in the cohort, posed a significant challenge for both models in accurately estimating the dose distribution. This is because the tumor's location causes the impinging angles to result in a tangential traversal of the PBs through the skull, a challenging testing scenario we encountered throughout this study. However, as depicted in Figure 3.20, CC-LSTM robustly estimates the dose distribution, closely matching the ground truth MC dose distribution, while LSTM-133 fails to accurately estimate the dose in the marginal areas of the plan. In terms of DVH comparison, both models underestimate the dose to the target PTV, while CC-LSTM shows improved estimation of the dose for OARs. This outcome is further supported by Figure 3.21, where it is evident that the OAR is critically located within a heterogeneous geometry. LSTM-133 clearly struggles in this region, with a considerable number of voxels in the corresponding OAR failing the γ -analysis test, while CC-LSTM shows a close match with the ground truth MC.

Patient Ho2348 (GA: 90°, CA: 0°) presents an interesting case. The impinging angle is one of the most common in the cohort, with 22 other patients sharing the same angle. Additionally, due to the right impinging angles, the interpolation step minimally affects the final dose distribution. However, similar to Patient Ho2520, the tumor's location significantly compromises the final dose distribution estimated by the models, with LSTM-133 reporting its second worst γ -index pass rate. Figure 3.24 presents the DVH comparison for the left optical nerve, where, similar to the case of Ho2520, the OAR shows discrepancies between the two models when compared to the ground truth MC. Figure 3.23 illustrates the OAR's location and how each model estimated the dose in these regions. Once again, CC-LSTM shows a clear advantage over the LSTM-133 model.

Finally, the performance of the two proposed models was evaluated for the remaining patients: Ho2794 (GA: 290°, CA: 0°), Ho3037 (GA: 40°, CA: 290°), Ho2593 (GA: 90°, CA: 345°), and Ho2651 (GA: 60°, CA: 270°). All patients exhibited favorable γ -index pass rates and DVH comparisons, particularly for the CC-LSTM model, which consistently demonstrated superior performance across these cases.

Given the accuracy reported for these patients, it can be concluded that the objectives of this study were achieved in terms of accuracy. The next step is to assess whether such dose estimations can be performed end-to-end within real-time APT time frames, as will be discussed in the following section.

Real Time Adaptive Dose Calculation

Throughout this study, the development of the CC-LSTM model was primarily governed by run time considerations, ensuring that the model not only delivers high accuracy but also adheres to super-fast run time constraints. As detailed earlier, the selection of modules and parameter setups was optimized to meet these requirements without compromising performance. The results have demonstrated that CC-LSTM meets these objectives, handling complex dose calculations with sub-second speed. Given the accuracy reported in section 3.3 and figure 3.16, one of the pivotal findings, reported in Table 3.18, demonstrates that a full IES dose calculation, consisting of 349 PBs, can be executed on a desktop PC computation infrastructure *in under one second*, with accuracies that closely match those obtained through the gold-standard MC simulation. This achievement implies that CC-LSTM can operate within the stringent real-time adaptive radiotherapy run times and, at a minimum, acts as an effective safeguard, mitigating the risk of incorrect radiation delivery during treatment. Table 3.16 contrasts CC-LSTM's run time and model design to DoTA, where it is highlighted that CC-LSTM's superior accuracy reported in table 3.14 is achieved with approximately 1.3 % number of learnable parameters, resulting in more than threefold reduction in run time.

The run times mentioned above were achieved while evaluating CC-LSTM with the default parameter set chosen for this study. However, the adaptations we have detailed can be employed not only to enhance precision but also to optimize both the accuracy and speed, tailored to the specific requirements of the task at hand. This flexibility allows the model to be finely adjusted to strike the desired balance between performance and computational efficiency in various clinical scenarios. Table 3.17 demonstrates such adaptation in its simplest form. Here, the shorter longitudinal range required for low-energy PBs does not necessitate the same longitudinal shape as higher-energy PBs. Simply adapting the input/output parameterization of the problem has been shown to improve run times for low-energy PBs by up to 35 %. This analysis was also performed for estimating IES dose distribution, where sub-second run times were reported (Table 3.18).

Finally, the dose distribution for the seven patients was carried out with high accuracy and low run times, in the scale of seconds, as reported in Table 3.19.

Considering the strict criteria in choosing the seven patients to set aside as test patients, which involved selecting cases with unique impinging angles (Figure 2.12), the reported results highlight the potential of AI-based dose calculation in both delivering accurate and fast dose distributions. The speed and accuracy achieved are sufficient to enable incorporation into real-time adaptive proton therapy, thus fulfilling the key objectives of this study.

Outlook: Multi-GPU setup

We plan to further investigate factors affecting CC-LSTM's speed and accuracy, and examine hardware configurations, including multi-GPU setups, and their advantageous characteristics for our task, given the inherently parallel nature of this task. Finally, we will explore the feasibility of scaling the under-one-second dose estimation of a single IES, reported in this study, to a multi-GPU setup, to achieve sub-second full-field dose estimation run times. This scaling will include streamlining all necessary processes in the workflow, such as preprocessing and cube extractions, interpolations and dose cube alignments, data I/O, and subsequent post-processing to generate the dose distribution on the desired dose grid. Our aim is to optimize the entire workflow for both speed and accuracy because in the realm of real-time adaptive proton therapy, every millisecond counts.

SUMMARY

This doctoral research investigated the application of Artificial Neural Networks (ANNs) and deep learning techniques for estimating proton dose distributions in particle therapy. The study focused on evaluating the suitability of ANNs as a dose estimation method. The primary objectives were: 1) To develop and evaluate ANN models, particularly Long Short-Term Memory (LSTM) networks, capable of capturing the spatial dependence and heterogeneity of patient anatomy in mapping CT images to dose distributions. 2) To quantify dose prediction uncertainties using Bayesian LSTM models. 3) To investigate the ability of ANNs to meet the goal of real-time adaptive proton therapy by creating a dose estimation engine and models.

The initial feasibility study demonstrated that LSTM networks could effectively learn the supervised task of proton dose estimation, correlating spatio-temporal features between input CT images and simulated ground truth Monte Carlo dose distributions. It was shown, that the LSTM networks outperformed other variants of Recurrent Neural Networks-based models in terms of accuracy and computational efficiency and exhibited generalization capabilities for different energies and patient anatomies. The millisecond runtime per pencil beam suggested the potential of these models to generate full-field dose distributions for real-time adaptive proton therapy.

To address the frequently voiced concern of ‘explainability’ and to quantify model prediction accuracy for clinical translation, the Bayesian LSTM model, BayesDose, was developed. By incorporating probabilistic elements, BayesDose enabled the assessment of prediction uncertainties, providing confidence intervals and identifying potential sources of error.

Based on the lessons learned from the feasibility study, a custom dataset was created to investigate the deep learning models in a real patient scenario, with previously delivered treatment plans in the head region exhibiting a full range of heterogeneities, from air cavities to dense bony structures. Due to the challenges posed by the new dataset and the wide proton beam in HIT, the initial LSTM models were unable to correctly learn the spatio-temporal features. To address these challenges, two custom-designed, physics-informed variations of LSTM were proposed: LSTM₁₃₃ and CC-LSTM. LSTM₁₃₃ facilitated the training of higher dimensional inputs and outputs, while CC-LSTM was designed from scratch, tailored specifically to the characteristics

of proton interactions, significantly improving accuracy and computational efficiency. CC-LSTM processed the input in three steps: 1) dimensionality reduction of the input using strided convolutions, followed by spatial feature extraction using two Convolutional Neural Network (CNN) layers 2) spatial feature fed to a ConvLSTM cell, that by updating its cell state and hidden state, propagates the spatial features temporally, and 3) dose distribution prediction using a three layer CNN backend. This model achieved the computational speed necessary for real-time adaptive proton therapy (APT) and outperformed the current state-of-the-art in terms of accuracy and run times, setting a new benchmark for deep learning-based particle therapy dose calculation. Moreover, this thesis involved a comprehensive development aspect that allowed forward calculation of an entire field, end-to-end, via only the treatment plan, and the corresponding patient CT acquisition. This forward calculation operate in four steps, 1) The RSP cubes are extracted from the CT based on the treatment plan 2) the BEV dose distributions of all pencil beams are inferred in one feed forward of the model 3) these dose distribution are then accumulated based on the plan, on the GPU, in the BEV, and 4) The dose distribution is back-interpolated to the CT grid with the desired resolution. The design and evaluation of methods, models, workflows, and code libraries generated in this study have facilitated fast dose calculations for APT. The proposed dose calculation engine could carry out dose distribution of iso-energy surfaces in *under one second* on a desktop computational system, with highly conformal accuracy comparing to the ground truth MC simulations. Such performance ensures that dose estimations could be performed rapidly enough to adapt to dynamic changes in patient anatomy during treatment, given the availability of in-room monitoring systems such as cone beam CT, vision RT, MRI, and read from the available in-room imaging systems. This way, the proposed engine can evaluate updated dose distributions based on discrepancies in dose delivery read from machine's guiding systems such as beam application and monitoring system or based on intra- and inter-fractional changes in the patient anatomy.

ZUSAMMENFASSUNG

Diese Doktorarbeit untersuchte die Anwendung von Künstlichen Neuronalen Netzwerken (ANNs) und Deep-Learning-Techniken zur Berechnung von Protonendosisverteilungen in der Partikeltherapie. Die Studie konzentrierte sich auf die Bewertung der Eignung von ANNs als Dosisberechnungsmethode. Die Hauptziele waren: 1) Die Entwicklung und Bewertung von ANN-Modellen, insbesondere Long Short-Term Memory (LSTM)-Netzwerken, die in der Lage sind, die räumliche Abhängigkeit und Heterogenität der Patientenanatomie bei der Abbildung von CT-Bildern auf Dosisverteilungen zu erfassen. 2) Die Quantifizierung von Vorhersageunsicherheiten und Modellfehlern bei ANN-basierten Dosisberechnungen unter Verwendung von Bayesian LSTM-Modellen. 3) Die Untersuchung der Fähigkeit von ANNs, das Ziel der Echtzeit-adaptiven Protonentherapie durch die Erstellung einer Dosisberechnungseingine und entsprechender Modelle zu erreichen.

Die erste Machbarkeitsstudie zeigte, dass LSTM-Netzwerke die überwachte Aufgabe der Protonendosisberechnung effektiv erlernen konnten, indem sie räumlich-zeitliche Merkmale zwischen Eingabe-CT-Bildern und simulierten Ground-Truth-Monte-Carlo (MC)-Dosisverteilungen korrelierten. Die LSTM-Netzwerke übertrafen andere Varianten von Recurrent Neural Network (RNN)-basierten Modellen in Bezug auf Genauigkeit und Recheneffizienz und zeigten Generalisierungsfähigkeiten für unterschiedliche Energien und Patientenanatomien. Die Millisekunden-Laufzeit pro Pencil-Beam deutete auf das Potenzial dieser Modelle hin, Gesamtdosisverteilungen für die Echtzeit-adaptive Protonentherapie zu erzeugen.

Um die Bedenken hinsichtlich der "Erklärbarkeit" und Quantifizierung der Vorhersagegenauigkeit von Modellen für die klinische Übersetzung anzugehen, wurde das Bayesian LSTM-Modell BayesDose entwickelt. Durch die Einbeziehung probabilistischer Elemente ermöglicht BayesDose die Bewertung von Vorhersageunsicherheiten, indem es Konfidenzintervalle bereitstellt und potenzielle Fehlerquellen identifiziert.

Basierend auf den Erkenntnissen aus der Machbarkeitsstudie wurde ein maßgeschneiderter Datensatz erstellt, um die Deep-Learning-Modelle in einem realen Patientenszenario mit zuvor durchgeführten Behandlungsplänen in der Kopfregeion zu untersuchen, die eine Vielzahl von Heterogenitäten aufwiesen, von Lufthohlräumen bis hin zu dichten Knochenstruk-

turen. Aufgrund der Herausforderungen, die sich durch den neuen Datensatz und den breiten Protonenstrahl am Heidelberg Ion Therapy Center (HIT) ergaben, konnten die ursprünglichen LSTM-Modelle die räumlich-zeitlichen Merkmale nicht korrekt erlernen. Um diese Herausforderungen anzugehen, wurden zwei maßgeschneiderte Variationen von LSTM Modellen vorgeschlagen welche physikalische Informationen enthalten: LSTM₁₃₃ und Convolutional Conv Long Short-Term Memory (CC-LSTM). LSTM₁₃₃ erleichterte das Training von höherdimensionalen Eingaben und Ausgaben, während CC-LSTM von Grund auf neu entwickelt wurde und speziell auf die Eigenschaften von Protoneninteraktionen zugeschnitten war, wodurch die Genauigkeit und Recheneffizienz deutlich verbessert wurde. CC-LSTM verarbeitete die Eingabe in drei Schritten: 1) Dimensionsreduzierung der Eingabe mittels gestaffelter Faltungen, gefolgt von räumlicher Merkmalextraktion unter Verwendung von zwei Convolutional Neural Network (CNN)-Schichten, 2) Räumliche Merkmale werden an eine Convolutional LSTM (ConvLSTM)-Zelle gefüttert, die durch Aktualisierung ihres Zellzustands und versteckten Zustands die räumlichen Merkmale zeitlich fortpflanzt, und 3) Vorhersage der Dosisverteilung unter Verwendung eines dreischichtigen CNN-Backends. Dieses Modell erreichte die für die Echtzeit-adaptive Protonentherapie erforderliche Rechengeschwindigkeit und übertraf den aktuellen Stand der Technik in Bezug auf Genauigkeit und Laufzeit, wodurch ein neuer Maßstab für die Deep-Learning-basierte Dosisberechnung in der Partikeltherapie gesetzt wurde. Zudem wurde in dieser Arbeit die End-to-End-Dosisberechnung eines vollständigen Feldes allein anhand eines Bestrahlungsplans und eines entsprechenden Patienten-CTs demonstriert. Diese Dosisberechnung erfolgte in vier Schritten: 1) Die Relative Stopping Power (RSP)-Würfel wurden aus dem CT basierend auf dem Behandlungsplan extrahiert, 2) die Beam's Eye View (BEV)-Dosisverteilungen aller Pencil-Beams wurden in einem Vorwärtsdurchlauf des Modells abgeleitet, 3) diese Dosisverteilungen wurden dann basierend auf dem Plan auf der Graphics Processing Unit (GPU) im BEV akkumuliert und 4) die Dosisverteilung wurde mit der gewünschten Auflösung auf das CT-Gitter zurückinterpoliert. Die Entwicklung und Bewertung von Methoden, Modellen, Workflows und Code-Bibliotheken, die in dieser Studie erstellt wurden, haben schnelle Dosisberechnungen für die Adaptive Proton Therapy ermöglicht. Die vorgeschlagene Dosisberechnungseingine konnte die Dosisverteilung von Iso-Energieflächen in unter einer Sekunde auf einem Desktop-Computersystem durchführen, mit einer hochkonformen Genauigkeit im Vergleich zu den Ground-Truth-MC-Simulationen. Diese Leistung stellt sicher, dass Dosisberechnungen schnell genug durchgeführt werden können, um sich an dynamische Änderungen der Patientenanatomie während der Behandlung anzupassen, vorausgesetzt, es sind In-Raum-Überwachungssysteme wie Cone-Beam Computed Tomography, Vision-RT oder Magnetic Resonance Imaging. Auf diese Weise kann

die vorgeschlagene Engine aktualisierte Dosisverteilungen basierend auf Abweichungen bei der Dosisabgabe, die von Maschinenführungssystemen wie Beam Application and Monitoring System (BAMS) ausgelesen werden, oder basierend auf intra- und interfraktionellen Änderungen der Patientenanatomie bewerten.

REFERENCES LIST

- Agostinelli, S. et al. (2003). **GEANT4: A Simulation toolkit**. In: *Nucl. Instrum. Meth. A* 506, pp. 250–303. DOI: [10.1016/S0168-9002\(03\)01368-8](https://doi.org/10.1016/S0168-9002(03)01368-8).
- Akiba, Takuya, Shotaro Sano, Toshihiko Yanase, Takeru Ohta, and Masanori Koyama (2019). **Optuna: A Next-generation Hyperparameter Optimization Framework**. In: *Proceedings of the 25rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*.
- An, Y., J. Shan, S. Patel, W. Wong, S. Schild, X. Ding, M. Bues, and W. Liu (2017). **Robust intensity-modulated proton therapy to reduce high linear energy transfer in organs at risk**. In: *Medical Physics* 44 (12), pp. 6138–6147. DOI: [10.1002/mp.12610](https://doi.org/10.1002/mp.12610).
- Barragán-Montero, Ana et al. (May 2022). **Towards a safe and efficient clinical implementation of machine learning in radiation oncology by exploring model interpretability, explainability and data-model dependency**. In: *Physics in Medicine & Biology* 67.11, 11TR01. ISSN: 0031-9155. DOI: [10.1088/1361-6560/ac678a](https://doi.org/10.1088/1361-6560/ac678a). URL: <https://dx.doi.org/10.1088/1361-6560/ac678a> (visited on 02/06/2023).
- Blundell, Charles, Julien Cornebise, Koray Kavukcuoglu, and Daan Wierstra (May 21, 2015). *Weight Uncertainty in Neural Networks*. arXiv. DOI: [10.48550/arXiv.1505.05424](https://doi.org/10.48550/arXiv.1505.05424). arXiv: 1505.05424 [cs, stat]. URL: <http://arxiv.org/abs/1505.05424> (visited on 09/25/2022).
- Böhlen, TT, Francesco Cerutti, MPW Chin, Alberto Fassò, Alfredo Ferrari, P Garcia Ortega, Andrea Mairani, Paola R Sala, George Smirnov, and Vasilis Vlachoudis (2014). **The FLUKA code: developments and challenges for high energy and medical applications**. In: *Nuclear data sheets* 120, pp. 211–214.
- Elfwing, Stefan, Eiji Uchibe, and Kenji Doya (2018). **Sigmoid-weighted linear units for neural network function approximation in reinforcement learning**. In: *Neural Networks* 107, pp. 3–11.
- Esposito, Piero (2020). *BLiTZ - Bayesian Layers in Torch Zoo (a Bayesian Deep Learning library for Torch)*. GitHub. URL: <https://github.com/piEsposito/blitz-bayesian-deep-learning/>.
- Gal, Yarin (2016). “Uncertainty in Deep Learning.” University of Cambridge.
- Garapati, Sankeerth S. et al. (Sept. 2017). **Urinary bladder cancer staging in <sc>CT</sc> urography using machine learning**. In: *Medical Physics* 44.11,

- 5814–5823. ISSN: 2473-4209. DOI: [10.1002/mp.12510](https://doi.org/10.1002/mp.12510). URL: <http://dx.doi.org/10.1002/mp.12510>.
- Gers, Felix A., Jürgen A. Schmidhuber, and Fred A. Cummins (Oct. 2000). **Learning to Forget: Continual Prediction with LSTM**. In: *Neural Comput.* 12.10, 2451–2471. ISSN: 0899-7667. DOI: [10.1162/089976600300015015](https://doi.org/10.1162/089976600300015015). URL: <https://doi.org/10.1162/089976600300015015>.
- Gers, Felix A. and Jürgen Schmidhuber (2000). **Recurrent nets that time and count**. In: *Proceedings of the IEEE-INNS-ENNS International Joint Conference on Neural Networks. IJCNN 2000. Neural Computing: New Challenges and Perspectives for the New Millennium 3*, 189–194 vol.3.
- Graves, Alex and Jürgen Schmidhuber (July 2005). **Frameworkwise phoneme classification with bidirectional LSTM and other neural network architectures**. In: *Neural Networks* 18.5-6, pp. 602–610. DOI: [10.1016/j.neunet.2005.06.042](https://doi.org/10.1016/j.neunet.2005.06.042). URL: <https://doi.org/10.1016/j.neunet.2005.06.042>.
- Hack, Berkin, Katy Piddock, Susannah Stanway, Kirsty Balachandran, Ruth Board, Neha Chopra, David Okonji, Bhawna Sirohi, Richard Sullivan, and Katie Wakeham (May 2019). **Cancer control in low- and middle-income countries: time for action**. In: *Journal of the Royal Society of Medicine* 112.6, 213–217. ISSN: 1758-1095. DOI: [10.1177/0141076819843653](http://dx.doi.org/10.1177/0141076819843653). URL: <http://dx.doi.org/10.1177/0141076819843653>.
- Holcomb, George W. (Mar. 1999). **Minimally invasive surgery for solid tumors**. In: *Seminars in Surgical Oncology* 16.2, 184–192. ISSN: 1098-2388. DOI: [10.1002/\(sici\)1098-2388\(199903\)16:2<184::aid-ssu9>3.0.co;2-9](http://dx.doi.org/10.1002/(sici)1098-2388(199903)16:2<184::aid-ssu9>3.0.co;2-9). URL: [http://dx.doi.org/10.1002/\(sici\)1098-2388\(199903\)16:2<184::aid-ssu9>3.0.co;2-9](http://dx.doi.org/10.1002/(sici)1098-2388(199903)16:2<184::aid-ssu9>3.0.co;2-9).
- Hong, Linda, Michael Goitein, Marta Buccioli, Robert Comiskey, Bernard Gottschalk, Skip Rosenthal, Chris Serago, and Marcia Urie (1996). **A pencil beam algorithm for proton dose calculations**. In: *Physics in Medicine and Biology* 41.8, pp. 1305–1330. DOI: [10.1088/0031-9155/41/8/005](https://doi.org/10.1088/0031-9155/41/8/005). URL: <https://doi.org/10.1088/0031-9155/41/8/005>.
- Jia, Xun, Jan Schümann, Harald Paganetti, and Steve B Jiang (2012). **GPU-based fast Monte Carlo dose calculation for proton therapy**. In: *Physics in Medicine & Biology* 57.23, p. 7783.
- Jiang, H and H Paganetti (2004). **Adaptation of geant4 to Monte Carlo dose calculations based on CT data: Monte Carlo dose calculations based on CT data**. In: *Medical physics* 31.10, pp. 2811–2818.
- Jäkel, Oliver, Gerhard Kraft, and Christian P. Karger (2022). **The history of ion beam therapy in Germany**. In: *Zeitschrift für Medizinische Physik* 32.1, pp. 6–22. ISSN: 0939-3889. DOI: <https://doi.org/10.1016/j.zemedi.2021.11.003>. URL: <https://www.sciencedirect.com/science/article/pii/S0939388921001082>.

- Kamada, T. et al. (2015). **Carbon ion radiotherapy in japan: an assessment of 20 years of clinical experience**. In: *The Lancet Oncology* 16 (2), e93–e100. DOI: [10.1016/s1470-2045\(14\)70412-7](https://doi.org/10.1016/s1470-2045(14)70412-7).
- Keall, Paul, Per Poulsen, and Jeremy T Booth (2019). **See, think, and act: real-time adaptive radiotherapy**. In: *Seminars in radiation oncology*. Vol. 29. 3. Elsevier, pp. 228–235.
- Kingma, Diederik P. and Jimmy Ba (2014). *Adam: A Method for Stochastic Optimization*. arXiv: [1412.6980](https://arxiv.org/abs/1412.6980) [cs.LG].
- (Jan. 29, 2017). *Adam: A Method for Stochastic Optimization*. arXiv. DOI: [10.48550/arXiv.1412.6980](https://doi.org/10.48550/arXiv.1412.6980). arXiv: [1412.6980](https://arxiv.org/abs/1412.6980) [cs]. URL: <http://arxiv.org/abs/1412.6980> (visited on 09/17/2022).
- Koh, Benjamin et al. (Aug. 2023). **Patterns in Cancer Incidence Among People Younger Than 50 Years in the US, 2010 to 2019**. In: *JAMA Network Open* 6.8, e2328171–e2328171. ISSN: 2574-3805. DOI: [10.1001/jamanetworkopen.2023.28171](https://doi.org/10.1001/jamanetworkopen.2023.28171). eprint: https://jamanetwork.com/journals/jamanetworkopen/articlepdf/2808381/koh_2023_oa_230813_1691512977.55832.pdf. URL: <https://doi.org/10.1001/jamanetworkopen.2023.28171>.
- Kontaxis, C, G H Bol, J J W Lagendijk, and B W Raaymakers (2020). **DeepDose: Towards a fast dose calculation engine for radiation therapy using deep learning**. In: *Physics in Medicine & Biology* 65.7, p. 75013. DOI: [10.1088/1361-6560/ab7630](https://doi.org/10.1088/1361-6560/ab7630). URL: <https://doi.org/10.1088/1361-6560/ab7630>.
- Laboratories, RaySearch (2017). **RayStation 6 reference manual**. In:
- Loeffler, J. and M. Durante (2013). **Charged particle therapy—optimization, challenges and future directions**. In: *Nature Reviews Clinical Oncology* 10 (7), pp. 411–424. DOI: [10.1038/nrclinonc.2013.79](https://doi.org/10.1038/nrclinonc.2013.79).
- Low, D. A., W. B. Harms, S. Mutic, and J. A. Purdy (May 1998). **A technique for the quantitative evaluation of dose distributions**. In: *Medical Physics* 25.5, pp. 656–661. ISSN: 0094-2405. DOI: [10.1118/1.598248](https://doi.org/10.1118/1.598248). pmid: [9608475](https://pubmed.ncbi.nlm.nih.gov/9608475/).
- Lysakovski, Peter, Alfredo Ferrari, Thomas Tessonier, Judith Besuglow, Benedikt Kopp, Stewart Mein, Thomas Haberer, Jürgen Debus, and Andrea Mairani (Nov. 2021). **Development and Benchmarking of a Monte Carlo Dose Engine for Proton Radiation Therapy**. In: *Frontiers in Physics* 9, p. 741453. DOI: [10.3389/fphy.2021.741453](https://doi.org/10.3389/fphy.2021.741453).
- Lysakovski, Peter, Benedikt Kopp, Thomas Tessonier, Stewart Mein, Alfredo Ferrari, Thomas Haberer, Jürgen Debus, and Andrea Mairani (2024). **Development and validation of MonteRay, a fast Monte Carlo dose engine for carbon ion beam radiotherapy**. In: *Medical Physics* 51.2, pp. 1433–1449. DOI: <https://doi.org/10.1002/mp.16754>. eprint: <https://aapm.onlinelibrary.wiley.com/doi/pdf/10.1002/mp.16754>.

- 1002/mp.16754. URL: <https://aapm.onlinelibrary.wiley.com/doi/abs/10.1002/mp.16754>.
- MacKay, David J. C. (May 1992). **A Practical Bayesian Framework for Backpropagation Networks**. In: *Neural Computation* 4.3, pp. 448–472. ISSN: 0899-7667, 1530-888X. DOI: [10.1162/neco.1992.4.3.448](https://doi.org/10.1162/neco.1992.4.3.448). URL: <https://direct.mit.edu/neco/article/4/3/448-472/5654> (visited on 03/02/2023).
- Martinot, Sonia, Norbert Bus, Maria Vakalopoulou, Charlotte Robert, Eric Deutsch, and Nikos Paragios (2021). **High-Particle Simulation of Monte-Carlo Dose Distribution with 3D ConvLSTMs**. In: *Medical Image Computing and Computer Assisted Intervention – MICCAI 2021*. Ed. by prefix=de useprefix=true family=Bruijne given=Marleen, Philippe C. Cattin, Stéphane Cotin, Nicolas Padoy, Stefanie Speidel, Yefeng Zheng, and Caroline Essert. Lecture Notes in Computer Science. Cham: Springer International Publishing, pp. 499–508. ISBN: 978-3-030-87202-1. DOI: [10.1007/978-3-030-87202-1_48](https://doi.org/10.1007/978-3-030-87202-1_48).
- Mein, Stewart, Thomas Tessonier, Benedikt Kopp, Semi Harrabi, Amir Abdollahi, Jürgen Debus, Thomas Haberer, and Andrea Mairani (Feb. 2021). **Spot-Scanning Hadron Arc (SHArc) Therapy: A Study With Light and Heavy Ions**. In: *Advances in Radiation Oncology* 6.3, p. 100661. ISSN: 2452-1094. DOI: [10.1016/j.adro.2021.100661](https://doi.org/10.1016/j.adro.2021.100661). URL: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC8010580/> (visited on 06/05/2024).
- Mizumoto, Masashi et al. (2021). **Proton beam therapy for children and adolescents and young adults (AYAs): JASTRO and JSPHO Guidelines**. In: *Cancer Treatment Reviews* 98, p. 102209. ISSN: 0305-7372. DOI: <https://doi.org/10.1016/j.ctrv.2021.102209>. URL: <https://www.sciencedirect.com/science/article/pii/S0305737221000578>.
- Neishabouri, A., P. Salome, S. Mein, J. Debus, and A. Mairani (Nov. 2021a). **Toward AI-Driven Proton Dose Calculation: Development and Evaluation of 3D and 2D Sequential Neural Network Design**. In: *International Journal of Radiation Oncology*Biophysics*Physics* 111.3, e116. ISSN: 0360-3016. DOI: [10.1016/j.ijrobp.2021.07.528](https://doi.org/10.1016/j.ijrobp.2021.07.528). URL: <http://dx.doi.org/10.1016/j.ijrobp.2021.07.528>.
- Neishabouri, Ahmad, Niklas Wahl, Andrea Mairani, Ullrich Köthe, and Mark Bangert (Apr. 2021b). **Long short-term memory networks for proton dose calculation in highly heterogeneous tissues**. In: *Medical Physics* 48.4, pp. 1893–1908. ISSN: 0094-2405, 2473-4209. DOI: [10.1002/mp.14658](https://doi.org/10.1002/mp.14658). URL: <https://onlinelibrary.wiley.com/doi/10.1002/mp.14658> (visited on 09/15/2021).
- Newhauser, W. (Jan. 2009). **International Commission on Radiation Units and Measurements Report 78: Prescribing, Recording and Reporting Proton-beam Therapy**. In: *Radiation Protection Dosimetry* 133.1, 60–62. ISSN: 1742-3406. DOI: [10.1093/rpd/ncp005](https://doi.org/10.1093/rpd/ncp005). URL: <http://dx.doi.org/10.1093/rpd/ncp005>.

- Paganetti, H. (2012). **Range uncertainties in proton therapy and the role of monte carlo simulations.** In: *Physics in Medicine and Biology* 57 (11), R99–R117. DOI: [10.1088/0031-9155/57/11/r99](https://doi.org/10.1088/0031-9155/57/11/r99).
- Paganetti, Harald and Thomas Bortfeld (Jan. 2005). **Proton Beam Radiotherapy - The State of the Art.** In: *Med Phys* 32.
- Paganetti, Harald, Hongyu Jiang, Katia Parodi, Roelf Slopesma, and Martijn Engelsman (2008). **Clinical implementation of full Monte Carlo dose calculation in proton beam therapy.** In: *Physics in Medicine & Biology* 53.17, p. 4825.
- Paganetti, Harald, Pablo Botas, Gregory C Sharp, and Brian Winey (2021). **Adaptive proton therapy.** In: *Physics in Medicine & Biology* 66.22, 22TR01. DOI: [10.1088/1361-6560/ac344f](https://doi.org/10.1088/1361-6560/ac344f). URL: <https://dx.doi.org/10.1088/1361-6560/ac344f>.
- Pakela, Julia M., Antje Knopf, Lei Dong, Antoni Rucinski, and Wei Zou (Mar. 2022). **Management of Motion and Anatomical Variations in Charged Particle Therapy: Past, Present, and Into the Future.** In: *Frontiers in Oncology* 12, p. 806153. ISSN: 2234-943X. DOI: [10.3389/fonc.2022.806153](https://doi.org/10.3389/fonc.2022.806153). URL: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC8959592/> (visited on 04/26/2024).
- Parodi, K, A Mairani, S Brons, B G Hasch, F Sommerer, J Naumann, O Jäkel, T Haberer, and J Debus (2012). **Monte Carlo simulations to support start-up and treatment planning of scanned proton and carbon ion therapy at a synchrotron-based facility.** In: *Physics in Medicine & Biology* 57.12, p. 3759. DOI: [10.1088/0031-9155/57/12/3759](https://doi.org/10.1088/0031-9155/57/12/3759). URL: <https://dx.doi.org/10.1088/0031-9155/57/12/3759>.
- Pascanu, Razvan, Tomás Mikolov, and Yoshua Bengio (2012). **Understanding the exploding gradient problem.** In: CoRR abs/1211.5063. arXiv: [1211.5063](https://arxiv.org/abs/1211.5063). URL: <http://arxiv.org/abs/1211.5063>.
- Pastor-Serrano, Oscar and Zoltán Perkó (May 21, 2022). **Millisecond speed deep learning based proton dose calculation with Monte Carlo accuracy.** In: *Physics in Medicine & Biology* 67.10, p. 105006. ISSN: 0031-9155, 1361-6560. DOI: [10.1088/1361-6560/ac692e](https://doi.org/10.1088/1361-6560/ac692e). URL: <https://iopscience.iop.org/article/10.1088/1361-6560/ac692e> (visited on 05/31/2022).
- Pastor-Serrano, Oscar, Peng Dong, Charles Huang, Lei Xing, and Zoltán Perkó (2023). **Sub-second photon dose prediction via transformer neural networks.** In: *Medical Physics* 50.5, pp. 3159–3171. ISSN: 2473-4209. DOI: [10.1002/mp.16231](https://doi.org/10.1002/mp.16231). URL: <https://onlinelibrary.wiley.com/doi/abs/10.1002/mp.16231> (visited on 12/21/2023).
- Pattyn, Alexander, Karl Kratkiewicz, Naser Alijabbari, Paul L. Carson, Peter Littrup, J. Brian Fowlkes, Nebojsa Duric, and Mohammad Mehrmohammadi (2022). **Feasibility of ultrasound tomography-guided localized mild hyperthermia using a ring transducer: Ex vivo and in silico studies.** In: *Medical Physics* 49.9, pp. 6120–6136. DOI: <https://doi.org/10.1002/mp.15829>. eprint: <https://aapm.onlinelibrary.org/doi/abs/10.1002/mp.15829>.

- wiley.com/doi/pdf/10.1002/mp.15829. URL: <https://aapm.onlinelibrary.wiley.com/doi/abs/10.1002/mp.15829>.
- Perl, J., J. Shin, J. Schumann, B. Faddegon, and H. Paganetti (Nov. 2012). **TOPAS: an innovative proton Monte Carlo platform for research and clinical applications**. In: *Medical Physics* 39.11, pp. 6818–6837. ISSN: 0094-2405. DOI: [10.1118/1.4758060](https://doi.org/10.1118/1.4758060). PMID: [23127075](https://pubmed.ncbi.nlm.nih.gov/23127075/).
- Rosenblatt, Eduardo, Joanna Izewska, Yavuz Anacak, Yaroslav Pynda, Pierre Scalliet, Mathieu Boniol, and Philippe Autier (Feb. 2013). **Radiotherapy capacity in European countries: an analysis of the Directory of Radiotherapy Centres (DIRAC) database**. In: *The Lancet Oncology* 14.2, e79–e86. ISSN: 1470-2045. DOI: [10.1016/S1470-2045\(12\)70556-9](https://doi.org/10.1016/S1470-2045(12)70556-9). URL: [http://dx.doi.org/10.1016/S1470-2045\(12\)70556-9](http://dx.doi.org/10.1016/S1470-2045(12)70556-9).
- Salvat, Francesc, José Fernández-Varea, and Josep Sempau (Jan. 2007). **Penelope. A code system for Monte Carlo simulation of electron and photon transport**. In: *NEA Data Bank, Workshop Proceeding, Barcelona*, pp. 4–7.
- Schaffner, Barbara, Eros Pedroni, and Antony Lomax (Jan. 1, 1999). **Dose calculation models for proton treatment planning using a dynamic beam delivery system: an attempt to include density heterogeneity effects in the analytical dose calculation**. In: *Physics in Medicine and Biology* 44.1, pp. 27–41. ISSN: 0031-9155. DOI: [10.1088/0031-9155/44/1/004](https://doi.org/10.1088/0031-9155/44/1/004). URL: <http://stacks.iop.org/0031-9155/44/i=1/a=004?key=crossref.5b31b0fc0fb8d274bd73e0975ca69b84> (visited on 11/15/2016).
- Schiavi, A, M Senzacqua, S Pioli, A Mairani, G Magro, S Molinelli, M Ciocca, G Battistoni, and V Patera (2017). **Fred: a GPU-accelerated fast-Monte Carlo code for rapid treatment plan recalculation in ion beam therapy**. In: *Physics in Medicine & Biology* 62.18, p. 7482. DOI: [10.1088/1361-6560/aa8134](https://doi.org/10.1088/1361-6560/aa8134). URL: <https://dx.doi.org/10.1088/1361-6560/aa8134>.
- Shi, Xingjian, Zhourong Chen, Hao Wang, Dit-Yan Yeung, Wai kin Wong, and Wang chun Woo (2015). *Convolutional LSTM Network: A Machine Learning Approach for Precipitation Nowcasting*. arXiv: [1506.04214](https://arxiv.org/abs/1506.04214).
- Shorten, Connor and Taghi M. Khoshgoftaar (July 2019). **A survey on Image Data Augmentation for Deep Learning**. In: *Journal of Big Data* 6.1. DOI: [10.1186/s40537-019-0197-0](https://doi.org/10.1186/s40537-019-0197-0). URL: <https://doi.org/10.1186/s40537-019-0197-0>.
- Smith, Leslie N. (Mar. 2017). **Cyclical Learning Rates for Training Neural Networks**. In: *2017 IEEE Winter Conference on Applications of Computer Vision (WACV)*, pp. 464–472. DOI: [10.1109/WACV.2017.58](https://doi.org/10.1109/WACV.2017.58).
- Smith, Leslie N and Nicholay Topin (2019). **Super-convergence: Very fast training of neural networks using large learning rates**. In: *Artificial intelligence and machine learning for multi-domain operations applications*. Vol. 11006. SPIE, pp. 369–386.
- Soukup, Martin, Matthias Fippel, and Markus Alber (Oct. 2005). **A pencil beam algorithm for intensity modulated proton therapy derived from Monte Carlo sim-**

- ulations. In: *Physics in Medicine and Biology* 50.21, pp. 5089–5104. ISSN: 0031-9155. DOI: [10.1088/0031-9155/50/21/010](https://doi.org/10.1088/0031-9155/50/21/010). URL: <https://doi.org/10.1088/0031-9155/50/21/010> (visited on 09/20/2022).
- Sun, Yutao, Li Dong, Shaohan Huang, Shuming Ma, Yuqing Xia, Jilong Xue, Jianyong Wang, and Furu Wei (2023). *Retentive Network: A Successor to Transformer for Large Language Models*. arXiv: [2307.08621](https://arxiv.org/abs/2307.08621) [cs.CL].
- Tessonnier, T, A Mairani, S Brons, P Sala, F Cerutti, A Ferrari, T Haberer, J Debus, and K Parodi (2017). **Helium ions at the heidelberg ion beam therapy center: comparisons between FLUKA Monte Carlo code predictions and dosimetric measurements**. In: *Physics in Medicine & Biology* 62.16, p. 6784. DOI: [10.1088/1361-6560/aa7b12](https://doi.org/10.1088/1361-6560/aa7b12). URL: <https://dx.doi.org/10.1088/1361-6560/aa7b12>.
- Tessonnier, Thomas, Tiago Marcelos, Andrea Mairani, Stephan Brons, and Katia Parodi (2016). **Phase Space Generation for Proton and Carbon Ion Beams for External Users' Applications at the Heidelberg Ion Therapy Center**. In: *Frontiers in Oncology* 5. ISSN: 2234-943X. DOI: [10.3389/fonc.2015.00297](https://doi.org/10.3389/fonc.2015.00297). URL: <http://dx.doi.org/10.3389/fonc.2015.00297>.
- Tessonnier, Thomas et al. (2023). **Commissioning of Helium Ion Therapy and the First Patient Treatment With Active Beam Delivery**. In: *International Journal of Radiation Oncology*Biophysics* 116.4, 935–948. ISSN: 0360-3016. DOI: [10.1016/j.ijrobp.2023.01.015](https://doi.org/10.1016/j.ijrobp.2023.01.015). URL: <http://dx.doi.org/10.1016/j.ijrobp.2023.01.015>.
- Unkelbach, Jan et al. (2018). **Robust Radiotherapy Planning**. In: *Physics in Medicine and Biology*. DOI: [10.1088/1361-6560/aae659](https://doi.org/10.1088/1361-6560/aae659).
- Vaswani, Ashish, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Lukasz Kaiser, and Illia Polosukhin (2023). *Attention Is All You Need*. arXiv: [1706.03762](https://arxiv.org/abs/1706.03762).
- Wieser, H P, P Hennig, N Wahl, and M Bangert (2017). **Analytical probabilistic modeling of {RBE}-weighted dose for ion therapy**. In: *Physics in Medicine & Biology* 62.23, pp. 8959–8982. DOI: [10.1088/1361-6560/aa915d](https://doi.org/10.1088/1361-6560/aa915d). URL: <https://doi.org/10.1088/1361-6560/aa915d>.
- Wilson, Robert R. (1946). **Radiological Use of Fast Protons**. In: *Radiology* 47.5. PMID: 20274616, pp. 487–491. DOI: [10.1148/47.5.487](https://doi.org/10.1148/47.5.487). eprint: <https://doi.org/10.1148/47.5.487>. URL: <https://doi.org/10.1148/47.5.487>.
- Wood, Doran, Sila Çetinkaya, Harsha Gangammanavar, Weiguo Lu, and Jing Wang (2022). **On the Value of a Multistage Optimization Approach for Intensity-Modulated Radiation Therapy Planning***. In: *Physics in Medicine and Biology*. DOI: [10.1088/1361-6560/ac7a8a](https://doi.org/10.1088/1361-6560/ac7a8a).
- Wu, Chao, Dan Nguyen, Yixun Xing, Ana Barragan Montero, Jan Schuemann, Haijiao Shang, Yuehu Pu, and Steve Jiang (Mar. 2021). **Improving proton dose calculation accuracy by using deep learning**. In: *Machine Learning: Science and Technology* 2.1,

p. 015017. ISSN: 2632-2153. DOI: [10.1088/2632-2153/abb6d5](https://doi.org/10.1088/2632-2153/abb6d5). URL: <http://dx.doi.org/10.1088/2632-2153/abb6d5>.

Zhao, Jun, Xianwei Wu, Ying Xing, Yongqiang Li, and Zhi Chen (2022). **Technical note: A method to evaluate the effect of scanning beam delivery error on 3D dose and its utilization on carbon ion radiotherapy for prostate cancer.** In: *Medical Physics* 50.2, 1228–1236. ISSN: 2473-4209. DOI: [10.1002/mp.16109](https://doi.org/10.1002/mp.16109). URL: <http://dx.doi.org/10.1002/mp.16109>.

APPENDIX

In this section, we present the remaining γ -analysis comparisons of dose estimation for the unseen test patients listed in Table 3.15. Additionally, we include the DVH comparisons to highlight how closely the outcomes of the introduced models align with the ground truth MC dose delivery for both the target and the OARs.

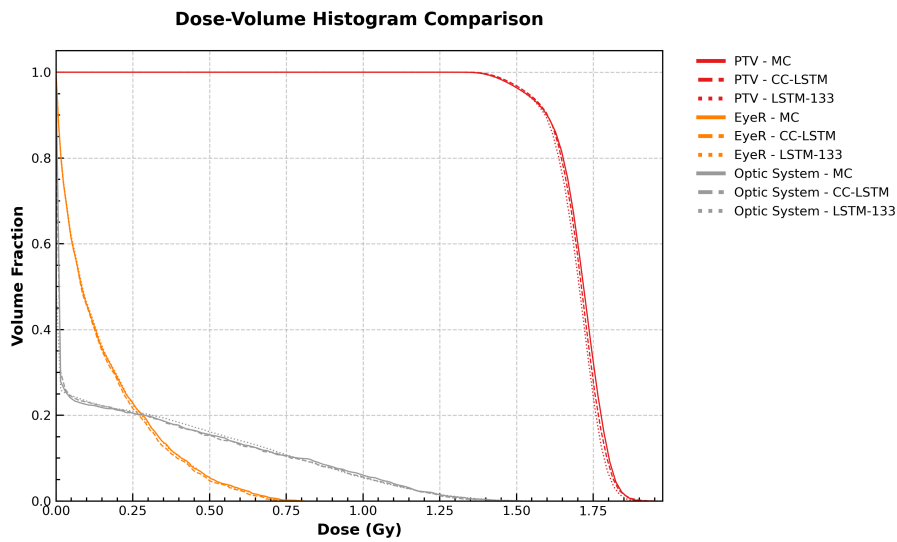
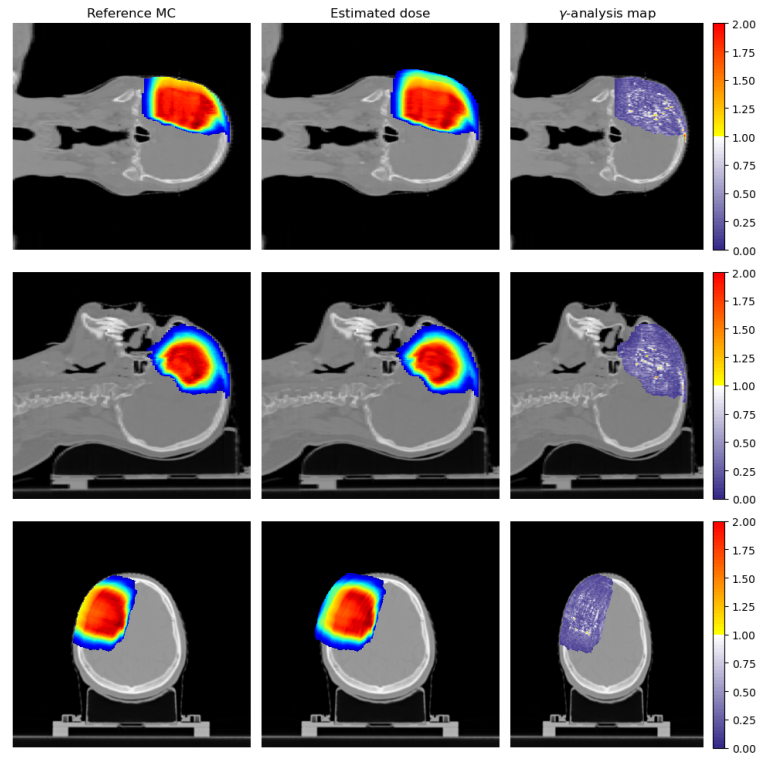
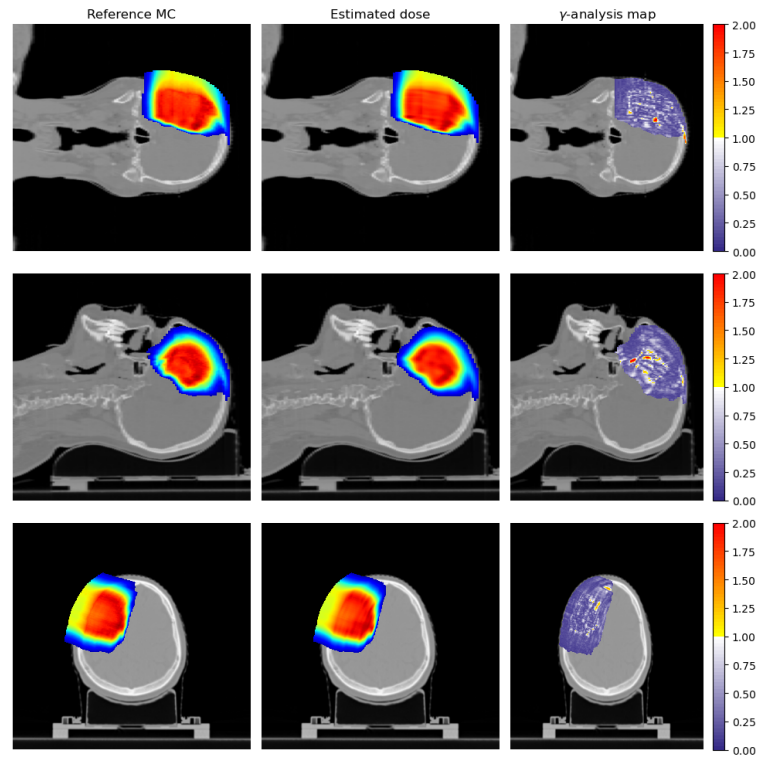


Figure 6.1: DVH comparison for patient Ho2794, illustrating the performance of MC, CC-LSTM, and LSTM-133 models.

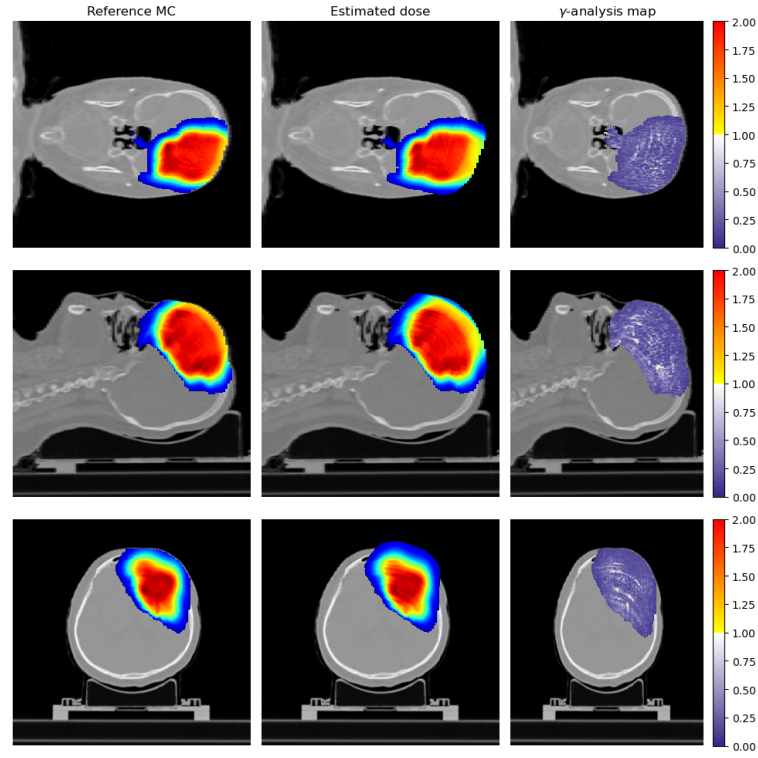


(a) Dose estimation by CC-LSTM model.

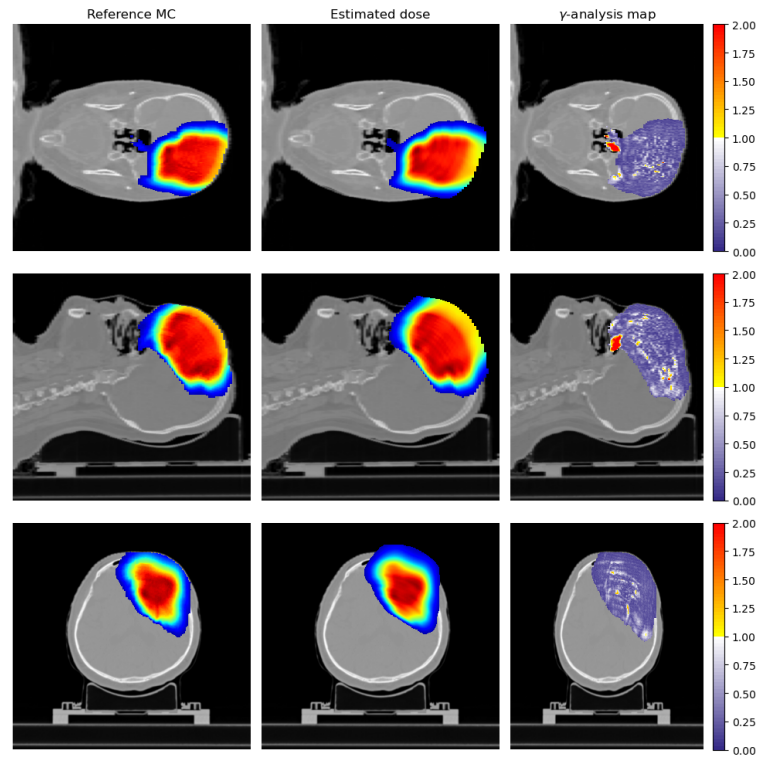


(b) Dose estimation by LSTM133 model.

Figure 6.2: γ analysis comparison ([2 %, 2 mm]; global) of dose estimation for an unseen, real-patient (Ho2794) case with 4999 PBs, calculated via (a) CC-LSTM model (γ PR: 98.3) and (b) LSTM133 model (γ PR: 96.2), against the ground truth MC simulations. The 3D cube views (coronal, sagittal, and axial) are shown for the regions where the MC dose is maximum.



(a) Dose estimation by CC-LSTM model.



(b) Dose estimation by LSTM133 model.

Figure 6.3: γ analysis comparison ([2 %, 2 mm]; global) of dose estimation for an unseen, real-patient (H03037) case with 9656 PBs, calculated via (a) CC-LSTM model (γ PR: 99.4) and (b) LSTM133 model (γ PR: 97.2), against the ground truth MC simulations. The 3D cube views (coronal, sagittal, and axial) are shown for the regions where the MC dose is maximum.

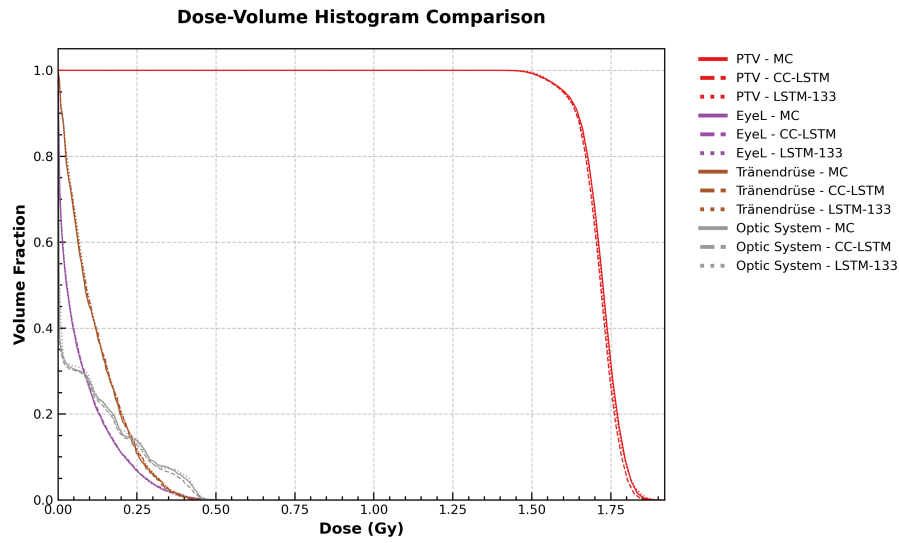


Figure 6.4: DVH comparison for patient H03037, illustrating the performance of MC, CC-LSTM, and LSTM-133 models.

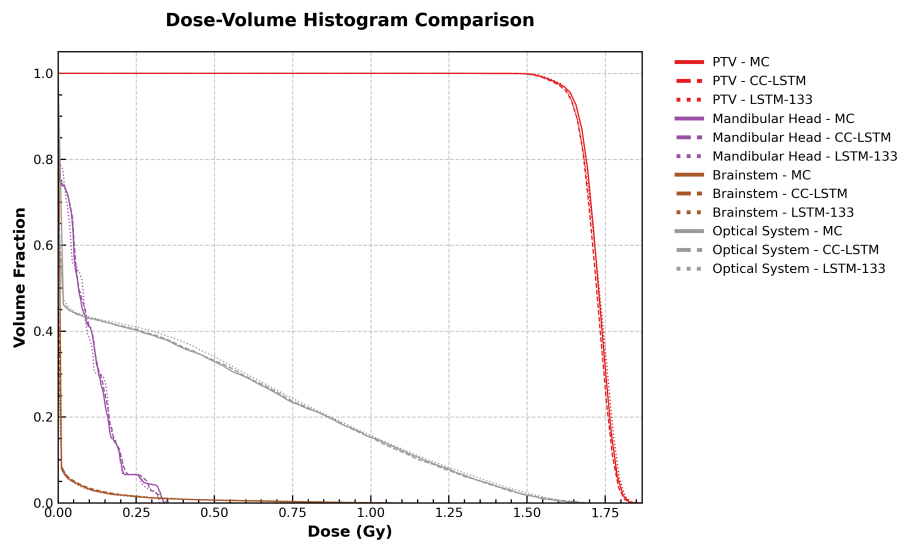
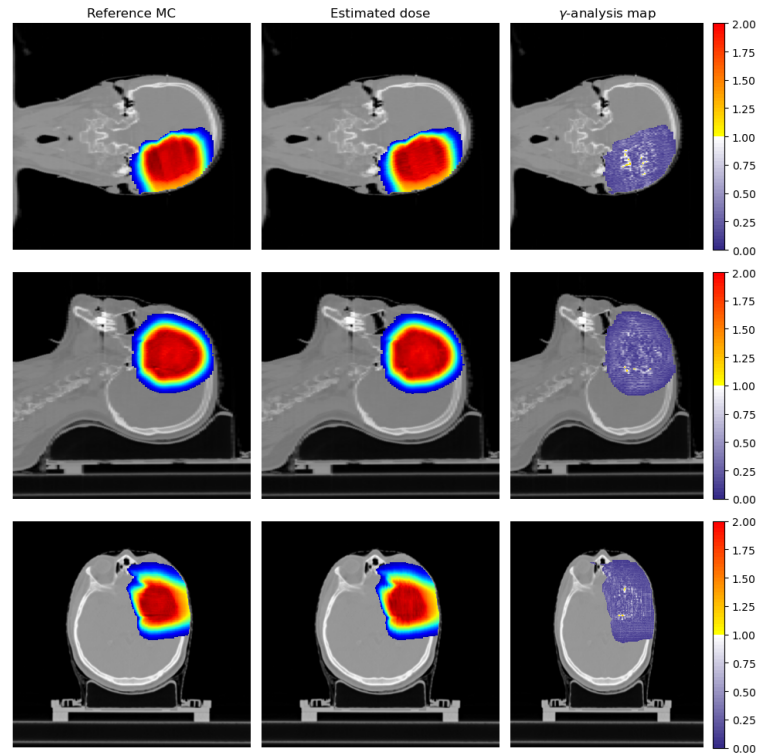
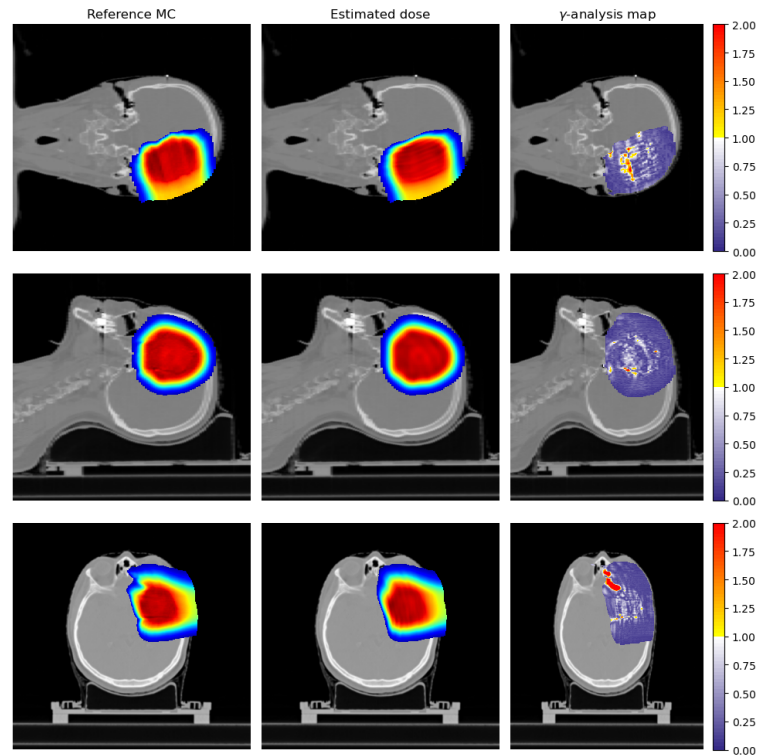


Figure 6.5: DVH comparison for patient H02593, illustrating the performance of MC, CC-LSTM, and LSTM-133 models.

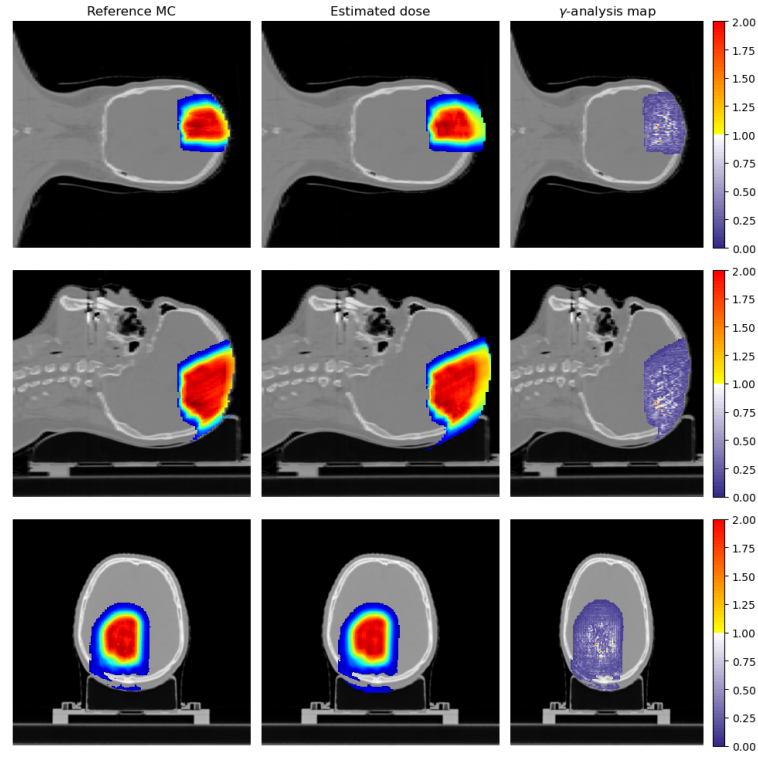


(a) Dose estimation by CC-LSTM model.

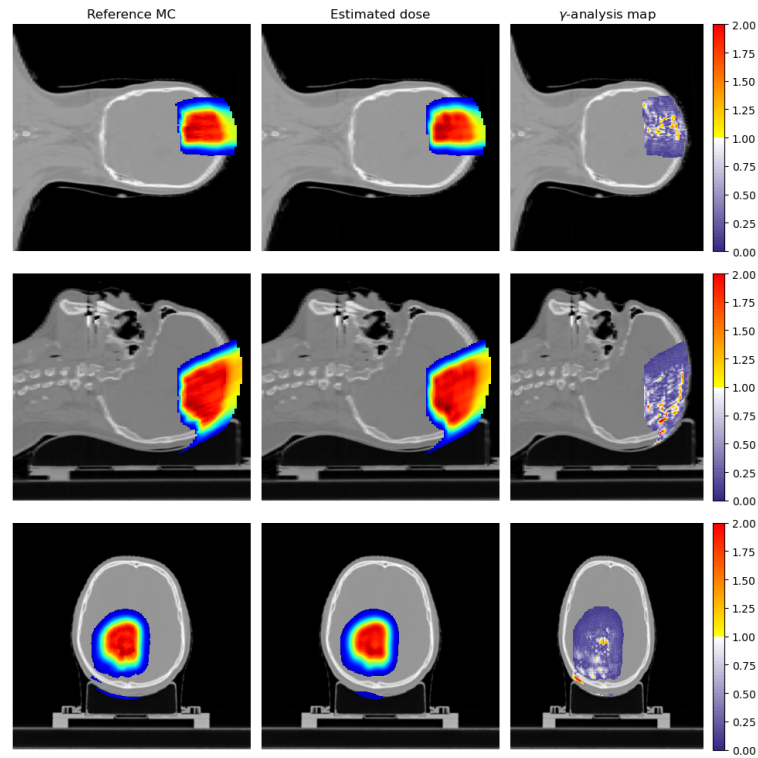


(b) Dose estimation by LSTM133 model.

Figure 6.6: γ analysis comparison ([2 %, 2 mm]; global) of dose estimation for an unseen, real-patient (Ho2593) case with 11135 PBs, calculated via (a) CC-LSTM model (γ PR: 99.7) and (b) LSTM133 model (γ PR: 97.0), against the ground truth MC simulations. The 3D cube views (coronal, sagittal, and axial) are shown for the regions where the MC dose is maximum.



(a) Dose estimation by CC-LSTM model.



(b) Dose estimation by LSTM133 model.

Figure 6.7: γ analysis comparison ([2 %, 2 mm]; global) of dose estimation for an unseen, real-patient (Ho2651) case with 4131 PBs, calculated via (a) CC-LSTM model (γ PR: 99.4) and (b) LSTM133 model (γ PR: 96.8), against the ground truth MC simulations. The 3D cube views (coronal, sagittal, and axial) are shown for the regions where the MC dose is maximum.

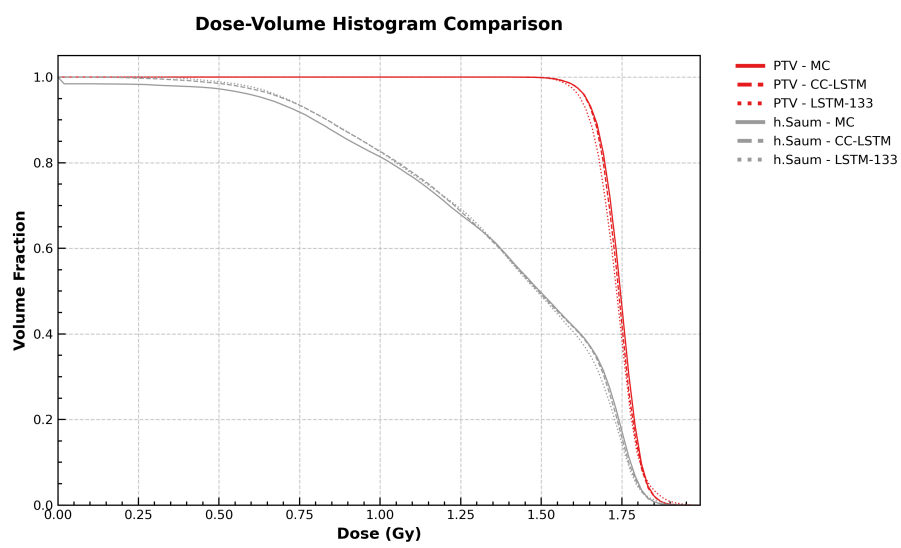


Figure 6.8: DVH comparison for patient Ho2651, illustrating the performance of MC, CC-LSTM, and LSTM-133 models.

PERSONAL CONTRIBUTION TO DATA ACQUISITION/ASSESSMENT AND PERSONAL PUBLICATIONS

This work was funded by the German Cancer Research Center (DKFZ). It originated during my master's thesis within the "Scientific Computing" program at the Faculty of Mathematics and Computer Science at Heidelberg University. The research problem, initially undefined, was systematically explored through various parameterizations. The sequence-to-sequence approach, particularly the mono-energetic pencil beam method, was developed as a novel parameterization strategy. The idea, methodology, data preparation, and implementation were all exclusively my work, guided through brainstorming sessions with my supervisor at the time, Dr. Mark Bangert.

Following the successful results from the sequence-to-sequence parametrization, I started my PhD thesis to expand this approach into an end-to-end, fully AI-based secondary dose engine capable of generating dose distributions for adaptive proton therapy, intended for use at the Heidelberg Ion-Therapy (HIT) facility. A subset of the data used in the feasibility check of Phase 1 of this PhD thesis was initially generated during my master's thesis. However, the generalizability data, which includes results for five additional patients previously treated at HIT and two other commissioned energies at HIT, were produced as part of my PhD work. Although the LSTM architecture was employed in my master's thesis, the computational framework and training procedures have been extensively refined in this PhD, yielding to the results reported in Section 3.1.

The refined methodologies and framework's generalizability to other patients and energies were validated and published in peer reviewed journal. This publication significantly influences the results and the textual content of Phase 1 of this PhD thesis:

Neishabouri, A., Wahl, N., Mairani, A., Köthe, U., & Bangert, M. (2021). **Long short-term memory networks for proton dose calculation in highly heterogeneous tissues.** *Medical Physics*. <https://doi.org/10.1002/mp.14658>

Phase 2 of this study, focusing on the quantification of uncertainty, originated from my idea and proposals, including the innovative use of Bayesian Neural Networks. This phase was executed by Luke Voß for his bachelor's thesis, under my supervision and that of Dr. Niklas Wahl. The outcomes of this col-

laborative effort have been documented in a publication currently undergoing the second round of review in a peer-reviewed journal:

Voss, L., Neishabouri, A. (joint first authorship), Ortkamp, T., Mairani, A., & Wahl, N. (2023). **BayesDose: Comprehensive proton dose prediction with model uncertainty using Bayesian LSTMs. Medical Physics, Under Revision.**

Phase 3 of this study represents my concerted efforts to fulfill the principal objective: the development of a fully AI-based dose engine. This phase utilized low-grade glioma patient data previously treated at the HIT facility, curated specifically for this project. Additionally, a comprehensive code library was developed to automate the generation of MC simulations and facilitate their integration into a deep learning framework.

The outcomes of this phase have led to two publications. The first presents the core AI model developed in this research:

Neishabouri, A., Bauer, J., Abdollahi, A., Debus, J., & Mairani, A. (2025). **Real-time adaptive proton therapy: An AI-based spatio-temporal mono-energetic dose calculation model (CC-LSTM).** Computers in Biology and Medicine, 188(109777), 109777. doi:10.1016/j.compbiomed.2025.109777

The second, currently in preparation, details the dose engine and evaluates its performance in clinical and real-time scenarios using the core AI model introduced in the previous publication:

Neishabouri, A., Abdollahi, A., Debus, J., Mairani, A., **Every Millisecond Counts: sub-second fully AI-based dose calculation engine - To be submitted to International Journal of Radiation Oncology.Biology.Physics (RED).**

Besides the first authorship papers outlined above, I contributed to the following publications below:

Radonić, D., Xiao, F., Wahl, N., Voß, L., Neishabouri, A., Delopoulos, N., Marschner, S., Corradini, S., Belka, C., Dedes, G., Kurz, C., & Landry, G. (2024). **A long short-term memory network-based proton dose calculation method in a magnetic field. Physics in Medicine and Biology. Under revision.**

Xiao, F., Radonić, D., Kriechbaum, M., Wahl, N., Neishabouri, A., Delopoulos, N., Parodi, K., Corradini, S., Belka, C., Kurz, C., Landry, G., & Dedes, G. (2024). **Prompt gamma emission prediction using a long short-term memory network. Physics in Medicine and Biology. Under Revision.**

Finally, the followings are my scientific contributions in forms of oral presentations or posters:

Poster at ASTRO Conference - A. Neishabouri, P. Salome, S. Mein, A. Mairani (2021). **Performance comparison of 3D-CNN and LSTM networks for proton dose calculation**

Oral Presentation at PTCOG Conference - A.Neishabouri, Mein, S., Debus, J., Mairani, A., (2022). **Superfast Proton-Ion Dose Engine Based on Recurrent Neural Networks (SPIDER).**

Oral Presentation at PTCOG Conference - A.Neishabouri, Mein, S., Debus, J., Mairani, A., (2023). **Toward Real-Time Adaptive Proton Therapy: A Particle Therapy Custom-Designed Deep Learning Model.**

Oral Presentation at PTCOG Conference - A.Neishabouri, Mairani, A., (2024). **Toward Real-Time Adaptive Proton Therapy: A Spatio-Temporal Physics-Informed Mono-Energetic Model.**

CURRICULUM VITAE

PERSONAL INFORMATION

Name: Ahmad Neishabouri

Date of Birth: 24/04/1988

Place of Birth: Damascus, Syria

RESEARCH EXPERIENCE AND EDUCATION

November 2019 – Present – PhD Student, **DKFZ and University of Heidelberg**

Heidelberg, Germany – Proton Dose Calculation using Artificial Neural Networks

January 2019 – September 2019 – Research Internship/ Master Thesis, **DKFZ**

Heidelberg, Germany – Particle Dose Calculation using Artificial Neural Networks

April 2017 – January 2019 – Master of Science, **Heidelberg University**

Heidelberg, Germany – Scientific Computing with application in Medical Physics

September 2010 – February 2013 – Master of Science, **Tehran Science & Research Branch of IAU**

Tehran, Iran – Nuclear Reactor Engineering

September 2006 – September 2010 – Bachelor of Science, **Tehran Science & Research Branch of IAU**

Tehran, Iran – Nuclear Reactor Engineering

ACKNOWLEDGEMENTS

I want to first thank Prof. Dr. Andrea Mairani for the initial spark he provided when I presented my basic outcomes and his interest in building an AI-based dose engine. Just as he is a great father at home, he has also been a great supervisor and friend to me. The path to carry out this research, due to its novelty, was like a hilly marathon with no signs and no people to cheer you on. However, Andrea was always there to cheer me up and to guide me when it was frustrating to find the path. After all, we didn't just finish this work together; we also completed an Ironman 70.3 together, so we knew how to work and endure alongside each other. I would also like to extend my appreciation to Prof. Dr. Jürgen Debus and Prof. Dr. Amir Abdollahi for providing the scientific environment that allowed us to pursue our research freely and based on our interests. Their guidance and comments throughout this research and in the Thesis Advisory Committee meetings were invaluable.

Moreover, I cannot forget and must acknowledge the contribution of Dr. Mark Bangert in the ideation of this work. Mark has provided me the space to explore and grow my interest in both Medical Physics and AI. Through his insightful brainstorming, we planted the seed that led to this PhD thesis.

Next, I would like to extend my appreciation to my brilliant and loving colleagues and group mates, with whom we pursued our zest for finding the answers. I specifically want to thank Dr. Benedikt Kopp, Dr. Stewart Mein, Dr. Thomas Tessonier, Judith Besuglow and Dr. Friderike Longarino for their senior assistance and guidance when I was new at HIT, and for sharing their codes and implementations without hesitation whenever I asked. I also thank Dr. Peter Lysakovski for walking the path together, helping each other, brainstorming ideas, and spending many weekends working on annoying bugs and saving frogs stuck at the HIT garden (no pun intended)! I am grateful to my beloved coworkers: Filipa Baltazar, Celine Karle, Aaron Osburg, Luisa Rank, Hailey Ahn, Ozan Dogan, Yasmin Hamad, Selver Pepić, and Maria Plana Herranz for all the good times we had in the office. We were more than just coworkers; We supported each other through every challenge and celebrated every win.

Throughout the pursuit of this research, I experienced both profoundly challenging and immensely rewarding moments. Saleh Eskandarian, my steadfast friend, was there for me before and during this journey. Saleh, you have my deepest acknowledgments and gratitude. Without your support, complet-

ing this work might not have been possible. Additionally, I met Dr. Patrick Salomé and Diana Mindroc during my first days in Germany. I deeply appreciate and value your friendship. Patrick was also a great scientific companion. Together, we explored and ideated AI applications across various areas of Medical Physics. Moreover, I am grateful for his thorough and prompt reviews of my work, which refined this thesis with his expert grasp of the field. I look forward to what we can achieve together in the future, for I am sure we will, us the Authentic Stiffers!

Reflecting on my journey, I've come to appreciate the profound impact of forming relationships and friendships. Though not all have lasted, each has enriched my understanding and contributed to my personal growth. True connections are often forged in our most vulnerable moments, and I am grateful for these experiences. To this end, I would like to also thank those who I did not name here, deliberately or inadvertently. Your involvement has not gone unnoticed, and your support has been truly appreciated.

Last but certainly not least, I wish to express my deepest gratitude to my family, to whom this work is dedicated, especially to my father, Mohsen Neishabouri. My passion for science began with your phrase you used to repeat whenever you wanted to teach me math: "Dars e Shirin e Riazī" (the sweet lessons of mathematics). You and my beloved mother, Zaynab, have always made our education a top priority, a commitment that has shaped me from childhood to this day. I am deeply thankful for all the sacrifices you have made towards this end. Know that any contributions I make to the scientific community are also a reflection of your influence.

I extend the same gratitude to my brother, Amir Neishabouri, and his wonderful wife, Somaye Heidari, as well as my younger brother, Yousef Neishabouri. Despite the distance, you never let me feel as though I was abroad. You looked after my interests back home when my studies kept me away, and always welcomed me with open arms during my visits, preserving the sweet feeling of home.

Thank you all.

Yours sincerely,
Ahmad Neishabouri

EIDESSTATTLICHE VERSICHERUNG

1. Bei der eingereichten Dissertation zu dem Thema:

Proton Dose Calculation using Artificial Neural Networks

handelt es sich um meine eigenständig erbrachte Leistung.

2. Ich habe nur die angegebenen Quellen und Hilfsmittel benutzt und mich keiner unzulässigen Hilfe Dritter bedient. Insbesondere habe ich wörtlich oder sinngemäß aus anderen Werken übernommene Inhalte als solche kenntlich gemacht.
3. Die Arbeit oder Teile davon habe ich bislang nicht an einer Hochschule des In- oder Auslands als Bestandteil einer Prüfungs- oder Qualifikationsleistung vorgelegt.
4. Die Richtigkeit der vorstehenden Erklärungen bestätige ich.
5. Die Bedeutung der eidesstattlichen Versicherung und die strafrechtlichen Folgen einer unrichtigen oder unvollständigen eidesstattlichen Versicherung sind mir bekannt. Ich versichere an Eides statt, dass ich nach bestem Wissen die reine Wahrheit erkläre und nichts verschwiegen habe.

Ort, Datum

Unterschrift Doktorand