

Deep Learning for Radiation Dose Estimation

Masterarbeit

Eingereicht von: Gutwein, Simon
Studiengang: Medizinische Strahlenwissenschaften
Matrikelnummer: 5419474
Betreuer: Prof. Dr. Daniela Thorwarth
Bearbeitungszeit: von 16.09.2021
bis 16.03.2022

Eberhard Karls Universität Tübingen

Universitätsklinik für Radioonkologie // Sektion für Medizinische Physik

Hoppe-Seyler-Straße 6, 72076 Tübingen

Abstract

Hello, here is some text without a meaning. This text should show what a printed text will look like at this place. If you read this text, you will get no information. Really? Is there no information? Is there a difference between this text and some nonsense like “Huardest gefburn”? Kjift – not at all! A blind text like this gives you information about the selected font, how the letters are written and an impression of the look. This text should contain all letters of the alphabet and it should be written in of the original language. There is no need for special content, but the length of words should match the language.

Dedication

Lorem ipsum dolor sit amet, consetetur sadipscing elitr, sed diam nonumy eirmod tempor invidunt ut labore et dolore magna aliquyam erat, sed diam voluptua. At vero eos et accusam et justo duo dolores et ea rebum. Stet clita kasd gubergren, no sea takimata sanctus est Lorem ipsum dolor sit amet. Lorem ipsum dolor sit amet, consetetur sadipscing elitr, sed diam nonumy eirmod tempor invidunt ut labore et dolore magna aliquyam erat, sed diam voluptua. At vero eos et accusam et justo duo dolores et ea rebum. Stet clita kasd gubergren, no sea takimata sanctus est Lorem ipsum dolor sit amet.

Declaration

I hereby declare that I have written the submitted Master's thesis without outside help. I have marked as such all statements taken verbatim or in spirit from other works, if applicable also from electronic media. I am aware that my Master's thesis in digital form will be checked for plagiarism if unmarked texts are used in whole or in part. A violation of the basic rules of scientific work is considered as an attempt to deceive or cheat and will result in the corresponding consequences according to the valid examination regulations of the Medical Radiation Sciences degree programme. For the purpose of comparing my Master's thesis with existing sources, I agree that my Master's thesis will be entered into a database where it will remain after my graduation. I do not cede any further rights to reprint and use it. I further declare that the thesis has not been the subject of any other examination procedure, neither in its entirety nor in essential parts, and that it has not been published.

Signature

Date

Contents

Abbildungsverzeichnis	vii
Tabellenverzeichnis	viii
List of Abbreviations	ix
1 Introduction	1
2 Medical Knowledge	3
2.1 Tumor Radiation	3
2.2 MR Linear Accelerator	4
2.3 Physical Fundamentals	5
2.4 Dose Calculation	6
3 Previous Work	7
4 Material and Methods	9
4.1 Network Architecture	9
4.2 Network Input & Output	9
4.2.1 Network Input	9
4.2.2 Network Output	11
4.3 Training Data Generation	11
4.4 Network Training	12
4.4.1 Dataloading	12
4.5 Evaluation Metrics	13
4.6 Hypotheses and Experiments	15
4.6.1 General Applicability	15
4.6.2 Poor Initial Translatability	15
4.6.3 Increased Robustness	15
4.6.4 Underlying Physics	16
4.7 Patient Data	16
5 Results	18
5.1 General Applicability	18
5.2 Poor Initial Translatability	19
5.3 Increased Robustness	19
5.4 Underlying Physics	21
6 Discussion	25
6.1 Single entity training	25
6.2 Mixed entity training	27
6.3 Dose deposition physics	28
6.4 Model improvement	29
7 Conclusion	32
8 Aknowledgement	33
9 Appendix	34

List of Figures

Figure 1:	CT image of a prostate cancer patient including the dose distribution of the respective treatment plan. Areas of high dose deposition are displayed in red and orange. Left and right femur head as well as rectum displayed in blue and orange and red respectively are displayed along the PTV displayed in green.	4
Figure 2:	Basic scheme of adapted network architecture with a downsampling ratio of 8 and a maximum depth of 512 in the bottleneck layer. Input dimensions: $5 \times W \times H \times D$; Output dimensions: $1 \times W \times H \times D$	9
Figure 3:	Exemplary input and output for a single segment of a prostate plan. (a): beam shape, (b): center beam line distance, (c): source distance, (d): CT, (e): radiological depth, (f): dose distribution,	11
Figure 4:	Dataloading scheme for memory efficient patch based dataloading for 3D training volumes. A subset of patches from one training volume is depicted in one color shade.	13
Figure 5:	Gamma test for two points r_{passed} and r_{failed} and the reference point r_m for a 2D (X,Y) space case. Orange and brown rings indicate distance and dose parameter acceptance margins respectively. The Gray sphere represents the set of all points that pass the gamma test. Green and Red arrow indicate a passed and failed gamma test for the given reference point r_m , dose and spatial parameter.	14
Figure 6:	dvh for prostate cancer test plan with gamma passrate of 99.36.	18
Figure 7:	Dose prediction example on isocentric slice on prostate cancer patient from test cohort with prostate data only trained model. From left to right: target dose; dose prediction; gamma map for isocentric slice with a gamma pass rate of 98.02%	19
Figure 8:	Dose prediction example on isocentric slice on breast cancer patient from test cohort with prostate data only trained model. From left to right: target dose; dose prediction; gamma map for respective slice with a gamma pass rate of 53.78%	19
Figure 9:	Gamma pass rate comparison for single segments (left) and summed up plans (right) between prostate-only and mixed entity trained models. Boxplot are displayed inside respective violin plots. Individual analysis of mixed entity set, composed of liver, mamma and H&N displayed on top plot. Significance level was assessed using Wilcoxon signed-rank test with significance level 0.05 and 0.01, 0.001 for *, ** and *** respectively.	21

Figure 10: Prediction accuracy of each segment with respect to field size for prostate only and mixed trained model. Relative occurrence of discretized fieldsize bin is displayed in the background.	22
Figure 11: Coronal, sagittal and transversal view of the dose distributions for 100, 200 and 300 px depth in the phantom volume	24
Figure 12: Exemplary images of target dose distribution of tangential segments on breast cancer patients with very small gamma-passrates (top) and the respective dose prediction from the prostate only trained model (bottom).	26
Figure 13: Top: Image of the isocentric slice from a prostate, H&N and breast cancer patient respectively. Bottom: Cross section through the patients anatomy at the marked position. HU value of water indicated with a dotted line.	27
Figure 14: Target, predicted and scaled predicted depth dose curve for water phantom positions at 100, 200 and 300 pixels depth, respectively, from left to right. Left: prostate-only trained model. Right: mixed entity trained model.	29

List of Tables

Table 1:	Patient data information for Prostate-Only as well as Mixed-Entity trained model and testing data set. Fieldsizes are given as mean (standard deviation).	17
Table 2:	Gamma passrates for segments as well as plans for prostate, liver, mamma, H&N and LN predicted with the prostate-only data trained model. Gamma test settings: 3 mm/3% lower percentage cutoff 10%	20
Table 3:	Gamma passrates for segments as well as plans for prostate, liver, mamma, H&N and LN predicted with the mixed data trained model. Gamma test settings: 3 mm/3% lower percentage cutoff 10%	20
Table 4:	Gamma passrate comparison of prostate-only and mixed entity trained models on the test dataset consisting of 306 prostate, 272 liver, 419 mamma, 379 head & neck and 659 lymphnode segments. Gamma test criteria: 3 mm/3% with a lower dose cutoff of 10%.	23
Table 5:	Gamma passrate comparison of prostate-only and mixed entity trained models on the test dataset consisting of 5 prostate, 5 liver, 5 mamma, 5 head & neck and 15 lymphnode treatment plans. Gamma test criteria: 3 mm/3% with a lower dose cutoff of 10%.	23
Table 6:	Gamma passrates for prostate only and mixed trained models for different water phantom phantom positions inside an air volume. Water phantom thickness was 200 pixels.	24

List of Abbreviations

CT	computer tomography	1
CV	computer vision	2
DD	dose deposition	1
DL	deep learning	2
DNA	deoxyribonucleic acid	3
DVH	dose volume histogram	18
ERE	electron return effect	5
GPU	graphics processing unit	12
H&N	head & neck	1
HPC	high-performance computing	12
HU	houndfield unit	10
IMRT	intensity modulated radio therapy	1
LN	lymphnodes	1
LSTM	long short term memory	7
MC	monte carlo	5
MLC	multi leaf collimator	3
MRI	magnet resonance imaging	1
MU	monitor units	6
OAR	organs at risk	1
PTV	planning target volume	3
RAM	random-access memory	12
SSD	source surface distance	15

1 Introduction

Cancer is one of the most contributing diseases leading to death under the age of 70 years in most countries [1]. A multitude of modalities exists for cancer treatment. Radiotherapy is a widespread and accessible method throughout the world [2] and is applied to a multitude of tumor entities such as prostate [3–5], breast [6–8], head & neck (H&N) [9–12], liver [13–17] and lymphnodes (LN) cancer [18–22]. On solid and delimited tumors it is mostly applied after surgical tumor resection. In a radiotherapy treatment an external irradiation source, in the form a linear accelerator is used to radiate the patient. Tumor cells have decreased repair capabilities compared to healthy cells when exposed to ionizing particles, such as photons. The application of radiation from photon beams in a fractionated manner results in a destruction of tumor cells, while minimizing the bad side effects on healthy tissue. To archive an optimal radiation dose inside the target volume (tumor cells) and to spare organs at risk (OAR) the photons are radiated from different angles around the patient. There exist multiple application modalities of radiotherapy that depend on the tumor entity, institutional guidelines and personal preference. With the intensity modulated radio therapy (IMRT) treatment modality, which is the most widespread, radiation is applied from multiple specific angles around the patient [23]. To adapt for different shapes of tumors the shape of the beam can be changed using lead apertures inside the accelerator head.

Individual treatment planning is needed to archive an optimum sparing of OAR and tumor dose coverage. The general workflow for treatment plan creation consists of computer tomography (CT) acquisition, tumor volume and OAR delineation and an iterative process of treatment planning between a medical physicist and an oncologist. This entire process can take up to multiple days until a convenient treatment plan is created.

A combination of a magnet resonance imaging (MRI) modality and a linear accelerator is currently used at our institution to enable live imaging during radio treatment. Current research effort is going towards the development of the needed software tools to enable an MR-only workflow for radio treatment planning, which makes the CT redundant in the treatment planning pipeline. Recent work involves the creation of pseudo CT images [24–26], delineation of OAR [27–29], plan optimization [30, 31], dose estimation [32] and . Ultimately this should lead to a supervised automated real-time treatment plan adaptation, which enables the treatment plan to be automatically adapted, when the patients anatomy changes between or during fractions. To enable this protocol multiple steps in the treatment planning pipeline needs to be adapted to be applicable in real-time. An important part in the pipeline consists of the dose deposition (DD) calculation, which is needed for the treatment planning software, as well as dose verification processes. State

of the art modality for accurate DD is a Monte-Carlo simulation [33]. In this simulation a multitude of particle histories is simulated inside the radiated volume. This simulation is based on physical processes that are very well described in literature. Downside of this process is that, due to the stochastic nature of the DD process of particles, a large number of particle histories is needed to yield a dose distribution with sufficient accuracy. Even in times of parallelization and GPU acceleration techniques this results in long simulation times. To enable the real time protocol involving the MR-linac, the need for a fast and accurate dose estimation tool is needed.

The use of deep learning (DL) has shown to be promising in a multitude of applications in computer vision (CV) tasks in medicine [34–38]. Due to the short inference times when applying DL to any given tasks it offers a promising base for a real time application of dose estimation in an MR-only automated treatment plan adaption workflow.

The aim of this paper is to improve the state of the art dose estimation tools involving DL and evaluate robustness aswell as the generalization capabilities. We do so by applying a 3D-UNet based DL solution to a multitude of tumor entities and treatment plans and evaluate its accuracy in dose prediction.

2 Medical Knowledge

2.1 Tumor Radiation

As mentioned in the previous section external tumor radiation is a widespread non-invasive modality for cancer treatment with or without tumor resection. It is or should be used for over 50% of all cancer patient once in the course of their illness [39]. The effectiveness of tumor radiation is based on the physical processes that cause biological damage to cells. Interaction of photons and secondary electrons release free radicals inside the tissue, which cause deoxyribonucleic acid (DNA) damage in the form of single- or double strand breaks. The tumor microenvironments conditions such as hypoxia, low pH and nutrient deficiency reparation mechanisms in tumor cells are down regulated compared to healthy tissue [40]. Since tumors are usually located in deeper regions of the body, it is unavoidable to irradiate healthy tissue as well. An example of a prostate tumor and its surrounding organs at risk is shown in fig. 1. Photons attenuation is described by an exponential decay, meaning that a photon beam can not be entirely shielded. This has crucial importance for the application of radio therapy for tumor irradiation. Due to the infinite range of photons, healthy tissue which is located behind the tumor volume, with respect to the beam direction, is also inevitably irradiated. The aim of a treatment plan optimization process is to achieve an optimal dose coverage over the entire target volume, while irradiating surrounding OAR as little as possible. Specific margins for dose coverage of tumor volumes as well as OAR are covered by the institutional standard operating procedure, which is oriented on current literature and medical studies. Therefore the planning software takes all delineated OAR in consideration and defines an optimal dose distribution. Optimizing the possible accelerator settings to achieve this dose distribution is the key step during treatment planning. The use of different defined radiation angles in combination with changing multi leaf collimator (MLC) shapes results in a sparing of OAR while achieving a high dose area with steep dose gradients at the planning target volume (PTV). To further decrease the bad side effects of irradiation of healthy tissue the plan is fractionated. In multiple sessions the tumor volume is irradiated benefiting from the increased reparation mechanisms of healthy tissue. The time between fractions is carefully chosen giving healthy tissue to repair some of the damage while the cells inside the tumor volume do not have enough time to repair crucial DNA damage resulting in a steady decrease of the tumor cell population.

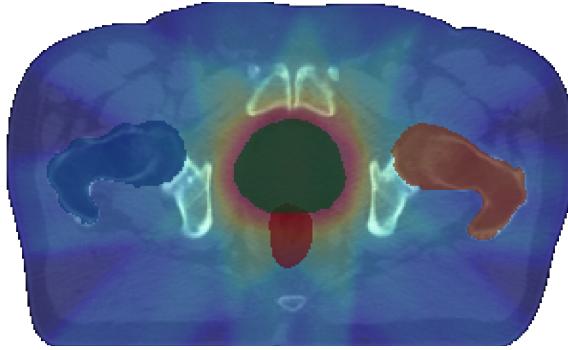


Figure 1: CT image of a prostate cancer patient including the dose distribution of the respective treatment plan. Areas of high dose deposition are displayed in red and orange. Left and right femur head as well as rectum displayed in blue and orange and red respectively are displayed along the PTV displayed in green.

2.2 MR Linear Accelerator

Basis for the treatment planning is a conventional CT taken as the first step in the treatment planning process. Planning on this CT is the base for each radiation session of all fractions. For some entities with moving organs such as sigmoid or bladder in the lower abdomen, this results in some uncertainties because it is assumed that the earliest acquired CT still represents the position of the tumor volume as well as the OAR. The Mr-Linac enables the acquisition of a MR image right before irradiation. In combination with the provided pipeline that allows for fast treatment plan adaption to shifts and shape morphing of tumor volumes or OAR these uncertainties can be reduced. The patient therefore receives a MR image before every fraction and the current treatmentplan is adapted based on the current positions and shapes of tumor and OAR. This is achieved by registration of the initial CT to the daily acquired MR image. Delineations of tumor volumes as well as OAR are registered and the treatmentplan optionally adapted.

Current research is going towards the further exploration of opportunities involving the MR-Linac. The vision behind the MR-Linac would be to adapt the treatment plan based on real-time image acquisition during the treatment. To enable this protocol all involved steps in the treatment planning pipeline need to be applicable in an MRI-only workflow, including delineation and dose deposition calculation. MRI imaging is not a quantitative imaging modality, meaning that the pixel values of MRI images are not correlated to the electron density of the tissue represented by that tissue. Dose deposition processes are dependant on particle energy as well as tissue density. Therefore a translation from qualitative MRI data to quantitative electron density information is needed to estimate deposited dose inside a volume.

An MRI only workflow is ultimately resulting in in-time adaption to intra-fraction movements such as unwanted patient movement as well as breathing motion potentially leading to a significant decrease in safety margins, which are currently needed to account for uncertainties during treatment.

Multiple steps in the planning pipeline need to be applied in real-time, which is currently not possible due to high calculation intensity. The estimation of dose deposition is a process that is involved in the step of iterative treatment plan optimization. Currently state of the art for precise dose deposition estimation is a monte carlo (MC) based solution. Downside of MC simulations is that they are very time consuming to achieve a sufficiently accurate result.

2.3 Physical Fundamentals

The magnetic field which is present at all times inside the MR-linac has a strength of 1.5 Tesla. Photons are not charged particles meaning that they are not affected by the magnet field. Secondary electrons are created by interaction of photons with matter. Since electrons are charged particles they obey the Lorentz force as first described by Lorentz in 1937 [41]. Electrons are the main contributor to the dose deposition. Therefore results a change in trajectory of electrons also in a change in dose deposition along the Lorentz force. Especially at the intersection of tissue to air cavities inside the human body the effect of the electric field is noticeable. It results in an increasing dose deposition on the surface of the tissue due to the electron return effect (ERE). The mean free pathlength of electrons is bigger in air resulting in an 180° degree turn of electrons and dose deposition at the tissue surface.

Base for the dose calculation algorithms in the form of MC is that basic interaction processes of particles are quantitatively described. The main contributing particles in the energy spectrum used at a linear accelerator are photons and electrons. Photons interact with matter in the form of inelastic interaction such as compton scattering or the photoeffect. For both interaction processes the entire or partial energy from the photon is transferred to an electron of the interaction sight. Due to the decreased mean free pathlength of electrons they deposit their entire energy in the directly surrounding tissue. Interaction processes are of stochastic nature. During the MC simulation parameters such as free pathlength, interaction process, radiation angle after an interaction are determined by the prior knowledge of the probability distributions of these physical processes.

2.4 Dose Calculation

As previously mentioned the main dose deposition algorithm we focus on is MC. There are multiple other dose deposition algorithms present in literature as well as clinical use, such as collapsed cone [42] or pencil beam algorithms [43] which is currently used as a primary dose estimation algorithms in the institutional planning software.

IMRT treatment plans are composed of individual segments. A segment consists of a specific gantry angle, MLC configuration and a predefined number of monitor units (MU). The amount of radiated photons is measured in MU and is measured in a constancy test for each linear accelerator. Summation of the dose deposited from all segments with their respected MU results in the dose distribution of the entire plan.

EGSnrc [44] is a MC based software tool which enables us to calculate dose distributions for single segments instead of the entire dose distribution of the plan, provided in the DICOM-dose file. Required for an accurate dose estimation is an accurate model of the accelerator head. Provided by previous work from Friedel et al. [45], the head model implemented in EGSnrc was used to create distributions for single segments. Each segment simulated with the EGSnrc software tool is normed to 100 MU which enables us to individually scale each segment to the radiated MU listed in the DICOM-plan file.

3 Previous Work

MR guided radiotherapy has received increased attention in the last years, with a multitude of studies and research areas involved. The problem of fast DD estimation is becoming increasingly important with the vision of online and even real-time radio treatment plan adaption. DD algorithms are involved in primary and secondary dose engines as well as dose verification processes. Research contribution goes towards the question of secondary dose engine or dose verification, because for both the specific accelerator parameters are given.

Different approaches towards the solution of the deposition involving deep learning have been used. Neishabouri et al. [46] applied a long short term memory (LSTM) in the application of DD for proton radiation. Input to the LSTM network was a sequence of 2D slices from the radiated volume. Each sequence represents the radiated volume of the proton beam. Due to the very limited range of protons inside matter and the small beam widening, volumes were small with $15 \times 15 \times 150$ voxels resulting in a field of view of $30 \times 30 \times 300$ mm² with an isotropic resolution of 2 mm per voxel. Each given slice was therefore of size 15×15 pixels. Network output was the respective dose distribution for the given input slice. To deviate from patient anatomies they used the approach of creating areas of increased density inside a $15 \times 15 \times 150$ voxels sized volume. By variation in position and size of the area the network was able to map DD processes to different densities translating to good results when applied to patient anatomies. Results show 98.57% gamma passrate (1%/3mm) for artificial cases and an average gamma passrate of 97.85% for patient test cases.

A combination of a 3D-UNet and convolutional LSTM networks as skip connections were used by Martinot et al. [47]. By approaching the problem as an active denoising problem, they used a set of noisy monte carlo simulations as an input to the 3D-convLSTM-UNet and the network yielded denoised dose distributions. Gamma passrates were 94.1% \pm 1.2% with 3%/3 mm as the criteria for 5 patient dose distributions.

Approaching the problem as an active denoising problem was also done by Neph et al. [48] that used 3 input channels consisting of patient CT, photon fluence map and undersampled dose of a single beamlet. By individual analysis and latter combination of all information they achieved to predict the residual dose needed for denoising. Mean absolute error was reduced to 0.106% from initial 25.7% of the undersampled beamlet. A significant increase in calculation time was also reached with 220 ms compared to the 380 s simulation time needed for a fully sampled beamlet.

Inclusion of fluence maps was also pursued by Fan et al. [49]. By projection of a 2D fluence map into the 3D patient anatomy they created 3D fluence volume from which a 3D UNet

like network make predictions of the deposited dose. Dose variation normalized to the prescribed dose was $0.17\% \pm 2.28\%$.

Kontaxis et al. [50] used a 3D-UNet with 5 3D Inputs combining information from the accelerator settings as well as the radiated volume into the input of the network. This enables them to predict single segments, which can be added up to entire treatment plans. Application was limited to lower abdomen cancer treatment plans with gamma passrates of 99.9 ± 0.3 for (3%/3mm) for prostate cancer plans. Dose estimation times were short with approximately 3 minutes for a treatment plan with 41 segments.

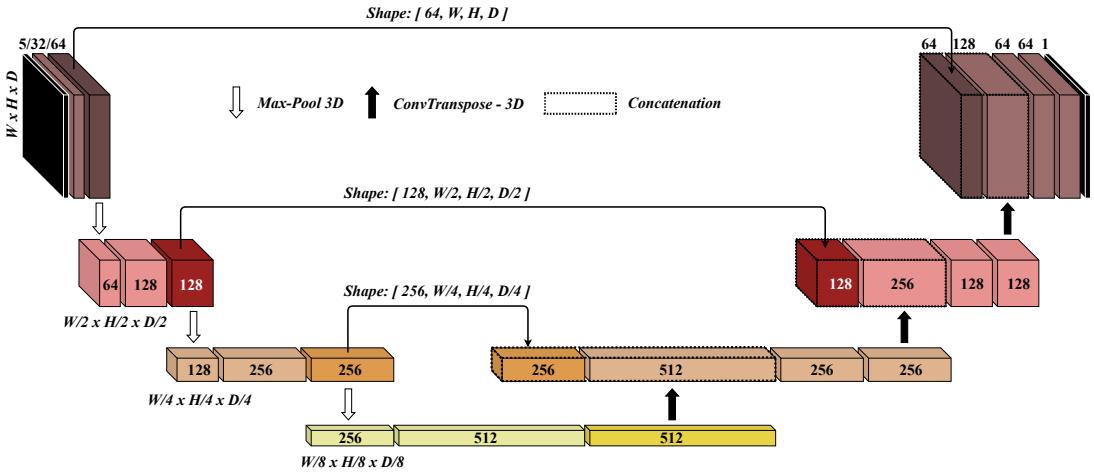


Figure 2: Basic scheme of adapted network architecture with a downsampling ratio of 8 and a maximum depth of 512 in the bottleneck layer. Input dimensions: $5 \times W \times H \times D$; Output dimensions: $1 \times W \times H \times D$

4 Material and Methods

4.1 Network Architecture

The network architecture (fig. 2) is based on a basic 3D-UNet implemented by Ronneberger et al. [51]. Hier war dazu schreiben was ein 3D-UNet ist und wie es genau funktioniert.

with slight adaptions in downsampling depth, due to memory limitations when using 3D data. We used a downsampling ratio of 8 and a maximum layer depth of 512 in the bottleneck connection. Skip connections are used after each convolutional steps before each pooling layer. To include a multitude of information from the patient aswell as the accelerator, we used five input volumes concatenated along the forth dimension, resulting in an input volume of size $5 \times \text{width } (W) \times \text{height } (H) \times \text{depth } (D)$. As the base resolution for all volumes we used the resolution of our CT scans with voxel dimensions of $1.1718 \times 1.1718 \times 3 \text{ mm}^3$ in the coronal, sagittal, and tanversal plane.

4.2 Network Input & Output

4.2.1 Network Input

The network expects five 3D volumes merged along their fourth dimension as resulting in an 4D input. These five 3D volumes (fig. 3) combine different spatial, anatomical

aswell as accelerator information into the training data. By doing so gantry aswell as accelerator head information can be directly translated into 3D space, which makes them interpretable for the network.

- (a) Beam Shape:** The trajectory of the beamshape into 3D space if of crucial importance. It is most important accelerator information for the network, combining leaf configurations as well as gantry angle into one volume. Projecting the beams eye view of the MLC shape into the target volume results in this masks. Information about which voxel is intersected or partial intersected by the beam. hier noch To additionaly account for the outputfactor the fieldsize of that specific segment is stored in each voxel intersected by the beam field. Voxel values of partially intersected voxels are scaled down by the percentage of the voxel which lies within the field. The output factor describes the physical process of increasing photon in-scatter into the central beamline for increasing fieldsizes.
- (b) Center Beam Line Distance:** The beam of a linear accelerator is best defined in the central beam line. hier sind wir FFF, also quadrat mit spitze, abstand von der mitte reduziert wert The minimum distance of each voxel from the central beam line accounting voxel dimensionality is saved in this mask. -> hier mal nochmal drüber reden
- (c) Source Distance:** The radiation pattern of photons from the accelerator head can be assumed to be the front of a spherical wave. The photon fluence is therefore decreasing with the square of the distance from the source. The source distance mask takes this physical relationship into account. Each voxel within the volume is assigned its distance to the source, taking into account the different voxel dimensions in the different spatial directions.
- (d) CT:** Dose deposition and interactions of particles depend on their energy as well as the electron density of the affected volume. As the initial energy of the radiated photons is assumed constant, the impact of the electron density of the volume (in this case the patients anatomy) is responsible to the different interaction processes of photons aswells as secondary electrons. houndfield unit (HU) of the patients anatomy aquired from a CT scan are stored in each voxel of the volume. Particles radiated from gantry angles below the treatment couch, interact partly inside the treatment couch, therefore it is included in the CT mask.

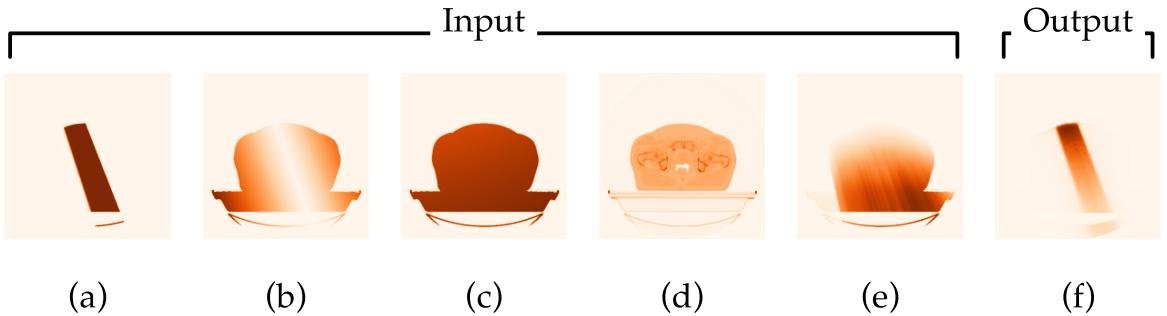


Figure 3: Exemplary input and output for a single segment of a prostate plan. (a): beam shape, (b): center beam line distance, (c): source distance, (d): CT, (e): radiological depth, (f): dose distribution,

(e) Radiological Depth: A photon beam that passes through the interaction medium loses its energy on its path. Therefore, the path that a particle follows through a medium is of highly relevant for the dose deposition effects that take place. The radiological depth combines the distance from the source as well as the interaction medium's density as well as prior areas irradiated. It is the sum of the path length through a medium multiplied by the density of intersected voxels. Therefore the same spatial depth in a dense material results in a higher radiological depth than the same depth in soft tissue.

Particles interact negligibly in air, therefore were all masks set to zero where the CT mask was 150 HU over the HU for air. Air or low density volumes inside the body of the patient were not affected by this step, to account for the ERE on the surface of air cavities inside the patient.

4.2.2 Network Output

The network combines the information from the five 3D volumes and yields a single volume with the respective dose distribution for the given input. Spatial dimensions for width, height and depth as well as voxel dimensions are preserved during inference. In fig. 3 (f) an exemplary dose distribution for the respective inputs (a) - (e) is given

4.3 Training Data Generation

All training data was generated from patient data taken from the institutional database of radiotherapy treatment plans. Information for all input masks are given in the CT

scan, the dose file aswell as the plan file. The patients anatomy for the CT masks is taken from the CT files and adjusted to the right slice thickness, due to different aquistion protocols using 2 mm or 3 mm slice spacing. Dose distribution used as training target for each segment were calculated using the EGSnrc open source software package provided under [44]. The work of Friedel et al. enabled us to accurately simulate single segment dose distributions for the MR-Linac. Simulation of all segments was dose using a remote high-performance computing (HPC) solution provided by the state of Baden-Württemberg. Coordinate system orientation of the patients anatomy as well as MLC and gantry positions are given in the dose and plan file. Simulation of 10^7 particles took around 4 hours on average. (**Welche Parameter sind von bedeutung**) The 3D input volumes to the network were calculated using in-house developed python scripts with the provided information from CT, dose and plan files. (GITHUB LINK?, hier noch näher drauf eingehen?) training data generation, dta e4xtraction, data precroressing blabla

4.4 Network Training

Four Nvidia GTX 2080 Ti, provided by the ML Cloud of the machine learning cluster from the university of Tübingen, were used to increase the GPU memory to 44 GB. This enabled a batchsize of 128 with a patch size of $32 \times 32 \times 32$ voxels resulting in a spatial field of view of $37.5 \times 37.5 \times 96 \text{ mm}^3$. After every epoch, the model performance was assessed on the validation set and the top 5 models were saved. The ADAM optimizer was used with 0.9, 0.99, 10^{-8} for β_1 , β_1 and ϵ respectively. An initial learning rate of 10^{-4} was used. If the validation loss did not decrease or the gamma passrate on the valdiation set did not increase over 50 epochs the learning rate was decreased by a factor of 10. When a learning rate of 10^{-6} was reached and no performance increase was noted over 50 epochs the training was stopped to prevent overfitting aswell as save computational recources.

4.4.1 Dataloading

Memory usage is a serious concern in the application of deep learning and espeacially when dealing with 3D data sets. In our case not only the memory usage in the graphics processing unit (GPU) but in the random-access memory (RAM) is of importance. For each training epoch, the network sees all present training data. In most applications all data can be loaded into the RAM and is then passed onto the GPU. A solution to this problem is to load data on the fly into the RAM and then process it. This is not applicable because of the size of the 3D volumes (multiple 100 MB per volume) that are used for training. We therefore developed a partial on the fly dataloading inspired by the *Queue* class from Pérez-García et al. [52] open source python libary *TorchIO*. classical

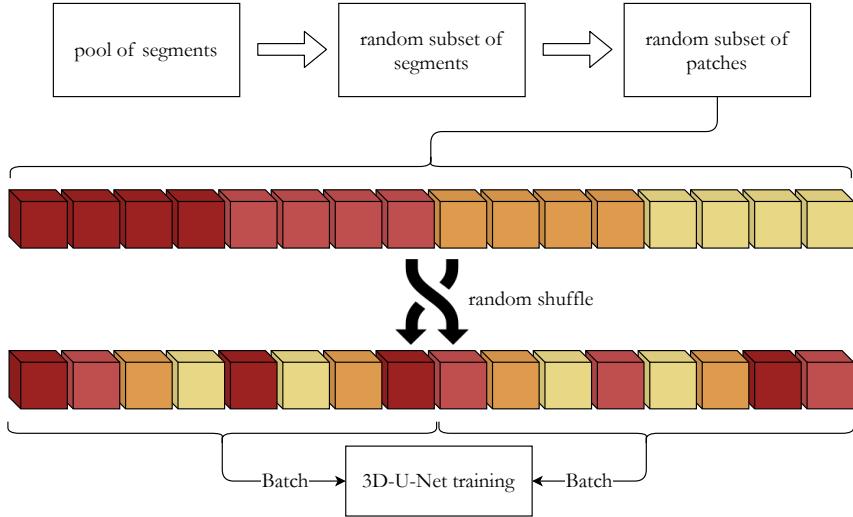


Figure 4: Dataloading scheme for memory efficient patch based dataloading for 3D training volumes. A subset of patches from one training volume is depicted in one color shade.

data augmentation in the form of verformungen, is not applicable due to aufgrund von complexitäts level und dass, kein direkter Zusammenhang von verformung und output nicht möglich ist. Due to the quantitative nature of our input data, we were not able to use data augmentation for our input data. To still achieve additional variation in training data we employed patch based training. A multitude of patches can therefore be extracted from the same volume, increasing the variety in training data while using a relatively small number of training volumes. To combine the semi-on-the-fly dataloading and the patch based training, multiple volumes are loaded into the RAM and then processed together. Fig. 4 shows a scheme of this dataloading approach. A fixed amount of training data is loaded from the randomized training data pool into the RAM. From this set of training volumes a subset of patches is extracted from each volume and stored in a queue of patches, which is then shuffled and batch wise passed onto the GPU and the network.

4.5 Evaluation Metrics

Conformality of dose distributions are clinically assessed using the gamma-index metric first introduced by Low et al. [53] in 1998. The evaluation metric composed of two parametric values that set the criteria for which dose distributions are analysed. Spatial deviations as well as deviations of dose are respected when analysing. Dose conformality is assessed by analyzing each individual voxel of a given dose distribution using the following equation:

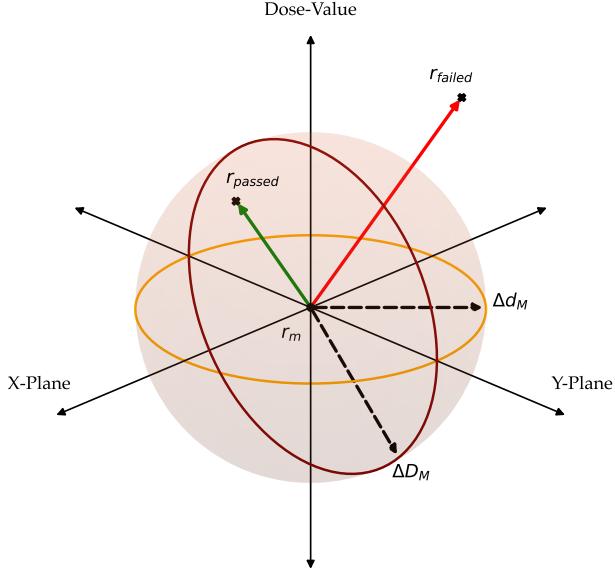


Figure 5: Gamma test for two points r_{passed} and r_{failed} and the reference point r_m for a 2D (X,Y) space case. Orange and brown rings indicate distance and dose parameter acceptance margins respectively. The Gray sphere represents the set of all points that pass the gamma test. Green and Red arrow indicate a passed and failed gamma test for the given reference point r_m , dose and spatial parameter.

$$\Gamma(r_m, r_c) = \sqrt{\frac{r^2(r_m, r_c)}{\Delta d_M^2} + \frac{\delta^2(r_m, r_c)}{\Delta D_M^2}} \quad (1)$$

where ΔD_M and Δd_M are the dose difference and spatial criterion respectively, in our case 3% and 3 mm. $r^2(r_m, r_c)$ and $\delta^2(r_m, r_c)$ are the squared spatial distance and dose difference from the reference point r_m to evaluation point r_c respectively. If $\Gamma \leq 1$ the criterion are passed. By evaluating the entire volume in that manner an overall gamma passrate can be calculated with the following formula.

$$\gamma = \frac{NoT(\Gamma \leq 1)}{NoT} \quad (2)$$

where NoT is the Number Of Test, in this case the number of voxels in the entire volume and $NoT(\Gamma \leq 1)$ is the number of tests that passed the gamma criterion following eq. 1. The equations eq. 1 and eq. 2 hold true for single aswell as multiple dimensions. A 2D example for a passed aswell as a failed gamma test is given in fig. 5.

4.6 Hypotheses and Experiments

4.6.1 General Applicability

Hypothesis

The network is able to learn the general dose deposition process from the given input taken from a multitude of radio treatment segments of one tumor entity. Dose predictions for segments of the same entity are accurate and robust.

Experiment

Training of the proposed 3D-UNet on a multitude of prostate cancer patient radio treatment segments. Dose conformaty assessment with the gamma passrate for each individual test segment aswell as entire test plans.

4.6.2 Poor Initial Translatability

Hypothesis

Different tumor sites vary drastically in segment shape, orientation and source surface distance (SSD). The SSD is, compared to other entities such as H&N, mostly constant for prostate cancer patients. The same holds true for tissue homogeneity aspects of the respective body regions, when comparing lower abdomen to e.g. H&N. We therefore assume that the network will not reach the previously achieved gamma passrates when testing on different tumor sites and patient anatomies.

Experiment

The from *4.6.1 General Applicability* trained network is then used for inference on additional testing data from a multitude of tumor entities aswell as patient anatomies to test its translational capabilities to a varying set of input characteristics.

4.6.3 Increased Robustness

Hypothesis

The inculcation of a broader variety of segments shapes, sizes and positions inside the patients anatomy, as well as differing body regions and therefore varying SSD values and tissue densities result in a better robustness of the networks prediction accuracy. while maintaining same training input size

Experiment

Network architecture aswell as training procedure of the in *4.6.1 General Applicability* trained network are kept, while adding data from liver, breast and H&N segments into

the pool of training data. The translational capabilities are then assed by testing the newly trained network on the entities that were included in the training data aswell as additional lymph node test cases. With this test design we investigate if the network performance decreases on prostate data by including new tumor sites into the training data and we assess its performance on seen aswell as unseen treatment plan data. We used lymph node plan data as the unseen test data because lymph nodes are present all over the human body, which makes this entity espeacially heterogeneous. Due to the small size of lymph nodes, segment fieldsize is therefore especially small, which we assume to pose the biggest challenge to the network.

4.6.4 Underlying Physics

Hypothesis

Inclusion of a wide variety of segment data into the training data set enables the network not only to predict accurate dose for patient anatomies, but any artifical tissue volume such as phantom volumes. We further assume that basic physical processes of the dose deposition such as dose depth profiles and penumbra of basis fields in phantoms aswell as the influence of the can be distance-squared law can be learned from this heterogeneous set of segment data.

Experiment

Artificial phantoms are used to investigate the physical accuracy of the dose predictions of the networks from *4.6.1 General Applicability* and *4.6.3 Increased Robustness* at various positions. Phantoms consist of a water volume placed inside a $600 \times 600 \times 600 \text{ mm}^3$ air volume at differing positions. With the default voxel dimensions of $1.1718 \times 1.1718 \times 3 \text{ mm}^3$ this results in a volume of shape $512 \times 512 \times 200 \text{ voxel}$ for the phantom. Comparing the target dose distribution of a $10 \times 10 \text{ cm}^2$ square field and 0° beam angle and the respective dose prediction yields information about the networks capability to learn the underlying physics.

4.7 Patient Data

All patient data used in the course of this study were previously treated at the MR-Linac in our institution. All treatment plans were created by a medical physicist in agreement with an oncologist. The training data consists of two main datasets. One is used for the training of the in *4.6.1 General Applicability* proposed network, which is only trained on prostate treatment data and consists of 45 treatment plans with a total number of 2342 segments. The second dataset consists a 15 treatment plans each for liver, breast

	Prostate-Only	Mixed-Entity			Testing	
Number of Patients	45	Liver	Mamma	Head & Neck	Prostate	Lymphnodes
Number of Segments	2342	819	656	929	720	659
Training / Validation / Testing	36/4/5	8/2/5	8/2/5	8/2/5	8/2/5	Only Testing
Fieldsizes Training/Validation	36.5 (17.9)	24.1 (18.4)	40.7 (28.5)	63.0 (50.5)	35.5 (18.0)	N/A
Fieldsize Testing	34.5 (15.8)	22.8 (14.4)	40.6 (38.0)	68.9 (53.6)	34.6 (15.9)	26.0 (25.6)

Table 1: Patient data information for Prostate-Only as well as Mixed-Entity trained model and testing data set. Fieldsizes are given as mean (standard deviation).

and H&N tumor sites with 819, 656 and 929 segments respectively. Additional 15 lymph nodes cancer treatment plans with a total of 659 segments were used only for testing. A comprehensive summary of all training and testing data is given in tab. 1 including training, validation, test split and information about the fieldsize distribution of training and test data. All patients gave their informed written consent to this study, which was approved by the local ethical committee (ethics approval No. 659/2017BO1).

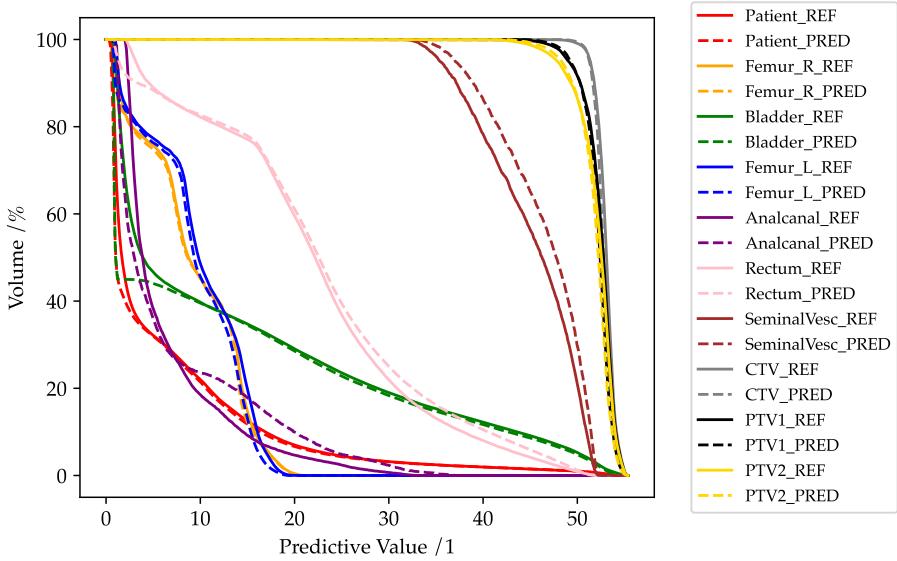


Figure 6: dvh for prostate cancer test plan with gamma passrate of 99.36.

5 Results

5.1 General Applicability

The training of the model was successful and took approximately 11 days until convergence. Training stopped at $6.7 \cdot 10^5$ due to no improvement in gamma and validation loss. On the test dataset consisting of 5 prostate patient plans and 306 segments in total the model reaches a gamma passrate of $94.35\% \pm 5.99\%$ (min: 58.02%, max: 99.96%) on all segments resulting in a passrate $96.16\% \pm 6.12\%$ (min: 85.33%, max: 99.96%) for the plans. A qualitative plan prediction sample is given in fig. 7, showing the isocentric slice of a target dose, the respective dose prediction and the gamma criterion evaluation. As shown the model is able to accurately predict the beamshape for single segments as well as the penumbra and beamwidening. The dose volume histogram (DVH) shown in fig. 6 gives crucial information about the applied dose inside target volumes as well as OAR for a test plan of the prostate patient cohort. A very good dose agreement for high as well as low dose regions is reached. Slight deviations in smaller structures such as rectum and seminal vesicles can be observed, most likely due to the larger contribution of single voxel to the entire volume.

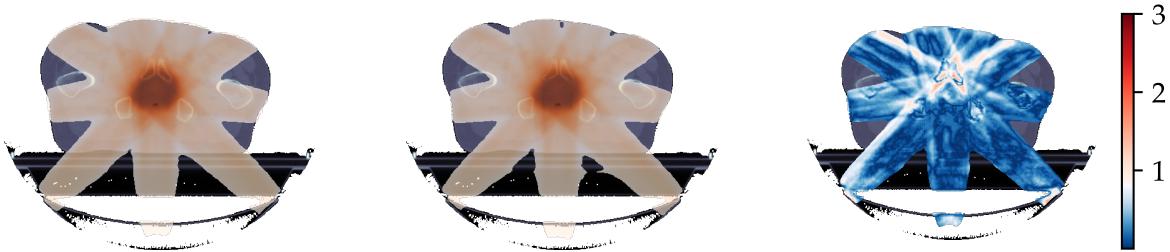


Figure 7: Dose prediction example on isocentric slice on prostate cancer patient from test cohort with prostate data only trained model. From left to right: target dose; dose prediction; gamma map for isocentric slice with a gamma pass rate of 98.02%



Figure 8: Dose prediction example on isocentric slice on breast cancer patient from test cohort with prostate data only trained model. From left to right: target dose; dose prediction; gamma map for respective slice with a gamma pass rate of 53.78%

5.2 Poor Initial Translatability

When applying the prostate only trained network on body entities deviating from the lower abdomen its accuracy decreases significantly. A summary over the prediction accuracy for all segments aswell as the summed up plans for all tumor entities is given in tab. 2. The model shows particular problems with breast cancer and H&N segments with a mean gamma passrate of $57.8 \pm 18.0\%$ (min: 1.9% max: 94.7%) and $60.5 \pm 11.8\%$ (min: 31.1% max: 91.5%). An exemplary case is given by fig. 8 showing low accuracy in the high dose aswell as lower dose areas of the volume.

5.3 Increased Robustness

The model training was succesful and took approximately 9 days until convergence. Training stopped at $5.0 \cdot 10^5$ due to no improvement in gamma and validation loss. Testing was done on prostate, liver, mamma and H&N entities, which were included in the training set as well as LN was was not included. A summary of the reached gamma passrates is given in tab. 3, showing significantly changed gamma passrates (p-level: 0.05) see

	Prostate		Liver		Mamma		H&N		LN		
Segments / Plans	/1	306	5	272	5	419	5	379	5	659	15
Mean		94.4	96.2	75.4	77.7	57.8	63.3	60.5	76.2	77.4	82.1
STD	/%	6.0	6.1	12.6	10.1	18.0	6.7	11.8	5.0	13.8	11.5
Min		58.0	85.3	38.9	67.6	1.9	53.8	31.1	69.5	36.0	63.8
Max		100	100	99.4	91.6	94.7	72.5	91.5	82.0	99.7	98.2

Table 2: Gamma rassrates for segments as well as plans for prostate, liver, mamma, H&N and LN predicted with the prostate-only data trained model. Gamma test settings: 3 mm/3% lower percentage cutoff 10%

	Prostate		Liver		Mamma		H&N		LN		
Segments / Plans	/1	306	5	272	5	419	5	379	5	659	15
Mean		93.5	97.9	89.8	97.6	76.9	87.5	80.1	90.7	85.2	90.7
STD	/%	6.3	2.8	7.1	4.6	12.8	6.6	10.9	6.4	10.7	9.6
Min		39.2	92.9	50.8	89.5	18.4	80.0	49.3	79.6	38.7	66.6
Max		99.8	100	99.1	100	98.5	94.5	97.4	96.3	99.9	100

Table 3: Gamma rassrates for segments as well as plans for prostate, liver, mamma, H&N and LN predicted with the mixed data trained model. Gamma test settings: 3 mm/3% lower percentage cutoff 10%

fig. 9 *left*. Mean gamma pass rate changes for single segments were -0.88, +14.4, +19.11, +19.65, +7.75 for prostate, liver, mamma, H&N and LN respectively. The mean gamma passrate shows no decrease in performance for the entire plan analysis. An increase of +1.69, +19.9, +24.2, +14.49, +8.59 was observed for prostate, liver, mamma, H&N and LN. A comparison between the results of the individual segments and the overall plans shows that, on average, the segments with the higher contributions to the irradiation plan are better predicted than less contributing ones.

To get a deeper insight into the networks method of operation and to enlarge the interpretability of the network, we analysed the the perfomance of the network regarding fieldsize and segment weights. The correlation between fieldsize and gamma passrate, aswell as between segment weight and gamma passrate are dispalyed in fig. 10. Relative occurence of segment sizes aswell as weights are displayed as a bar chart in the background of both plots. Around 82% segments have a fieldsize smaller than 60cm^2 that is also the fieldsize until network accuracy increases with increasing fieldsize for both the prostate-only and mixed entity trained model. Fieldsizes larger than 60cm^2 are not predicted as stable with fluctuating gamma passrates for both models. The prostate-only trained model expieriences a a larger drop in accuracy for fieldsizes bigger than 60cm^2 and accu-

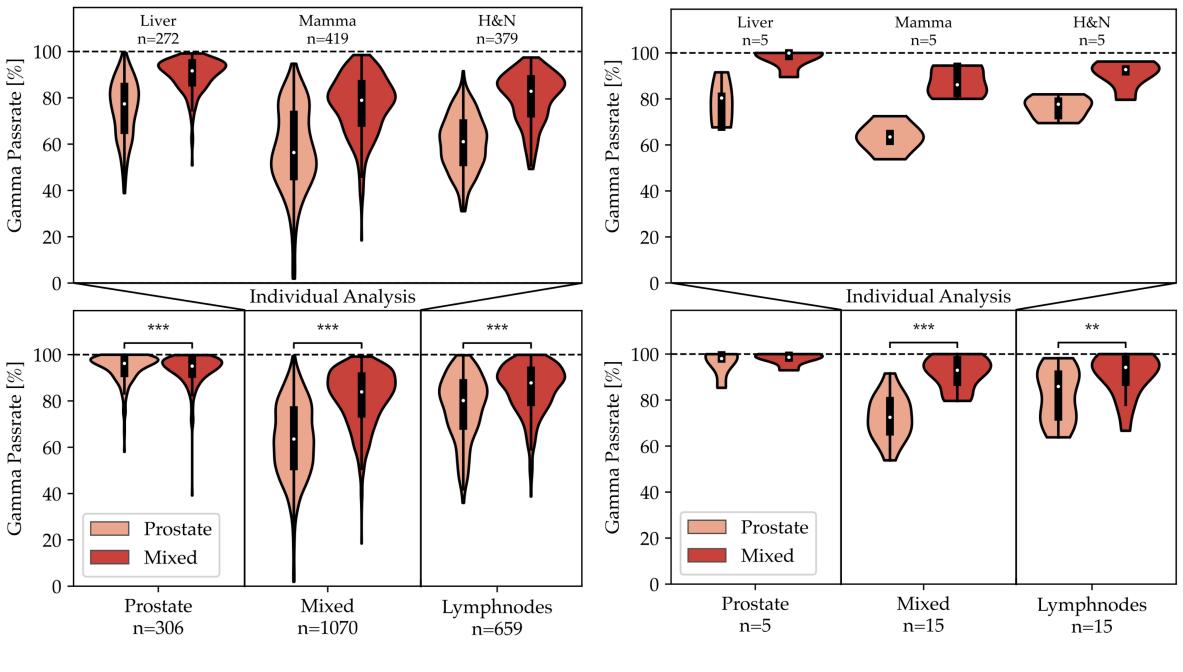


Figure 9: Gamma pass rate comparison for single segments (left) and summed up plans (right) between prostate-only and mixed entity trained models. Boxplot are displayed inside respective violin plots. Individual analysis of mixed entity set, composed of liver, mamma and H&N displayed on top plot. Significance level was assessed using Wilcoxon singed-rank test with significance level 0.05 and 0.01 ,0.001 for * , ** and *** respectively.

racy is steadily decreasing with increasing fieldsizes. It is observable that smaller fieldsizes are less robust and range from 1.9% to 100% and 18.5% to 99.9% for the prostate-only and the mixed entity model respectively on a fieldsize of 0 to 15cm^2 . Segments with weights under 0.2 make up for almost 80% of all segments. Gamma passrate analysis with respect to segment weight shows the same trend of increased accuracy and robustness with increasing weight values. A comprehensive and direct comparison between prostate-only and mixed entity model prediction accuracy for segment and plan predictions is given in tab. 4 and tab. 5 respectively.

5.4 Underlying Physics

Both the prostate-only aswell as the mixed entity model perform poorly regarding gamma passrates for all water phantom positions tested. The model accuracy is steadily decreasing with increasing SSD with gamma pass rates of 60.1%, 45.5%, 48.4%, and 56.1%, 27.7%, 19.2% for the prostate-only and mixed entity model at depths of 100 pixels, 200 pixels

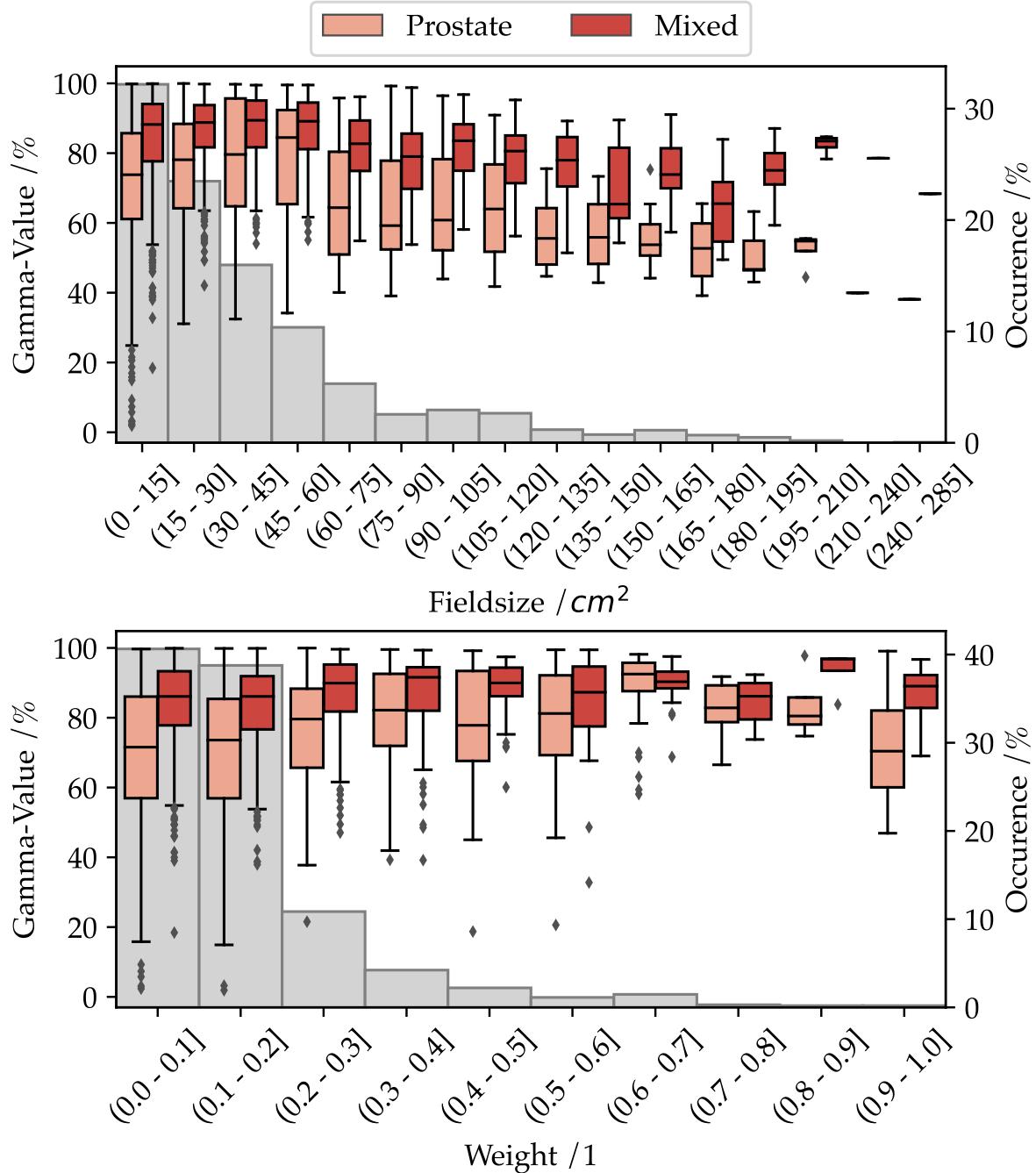


Figure 10: Prediction accuracy of each segment with respect to field size for prostate only and mixed trained model. Relative occurrence of discretized fieldsize bin is displayed in the background.

Model	Prostate		Liver		Mamma		H&N		LN	
	Prostate	Mixed								
Mean	94.3	93.5	75.4	89.8	57.8	76.9	60.5	80.1	77.4	85.2
Median	96.2	95.0	77.4	91.6	56.4	78.9	61.0	82.8	80.2	87.8
STD /%	6.0	6.3	12.6	7.0	18.0	12.8	11.8	10.9	13.8	10.7
Min	58.0	39.2	38.9	50.8	1.9	18.4	31.1	49.3	36.0	38.7
Max	100.0	99.8	99.4	99.1	94.7	98.5	91.5	97.4	99.7	99.9

Table 4: Gamma passrate comparison of prostate-only and mixed entity trained models on the test dataset consisting of 306 prostate, 272 liver, 419 mamma, 379 head & neck and 659 lymphnode segments. Gamma test criteria: 3 mm/3% with a lower dose cutoff of 10%.

Model	Prostate		Liver		Mamma		H&N		LN	
	Prostate	Mixed								
Mean	99.1	99.2	89.9	97.6	67.4	83.6	77.4	88.8	93.0	96.0
Median	99.6	99.7	91.8	97.4	66.2	83.7	78.8	90.9	92.8	98.8
STD /%	1.3	1.0	4.1	1.0	5.3	4.6	10.2	8.6	6.1	5.8
Min	96.8	97.5	84.0	96.6	60.6	77.0	62.5	76.0	75.3	77.9
Max	99.9	99.9	94.4	99.2	73.0	89.8	90.2	98.5	99.7	99.9

Table 5: Gamma passrate comparison of prostate-only and mixed entity trained models on the test dataset consisting of 5 prostate, 5 liver, 5 mamma, 5 head & neck and 15 lymphnode treatment plans. Gamma test criteria: 3 mm/3% with a lower dose cutoff of 10%.

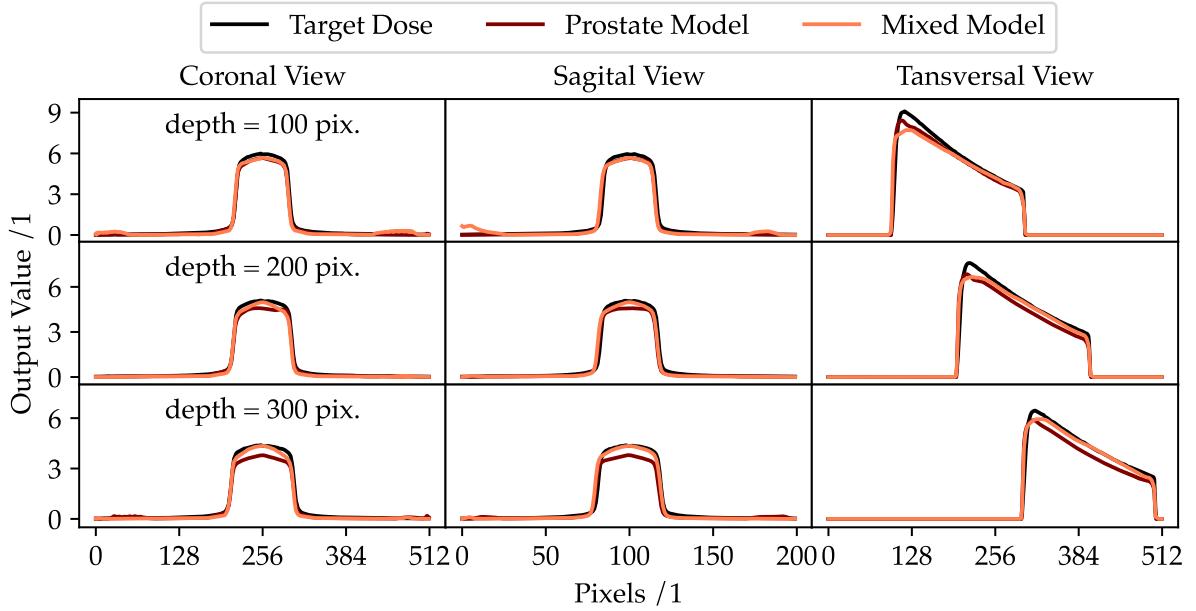


Figure 11: Coronal, sagittal and transversal view of the dose distributions for 100, 200 and 300 px depth in the phantom volume

Phantom position /pixel	100		200		300	
Model	Mixed	Prostate	Mixed	Prostate	Mixed	Prostate
Gamma Passrate /%	60.1	56.1	45.5	27.7	48.4	19.2

Table 6: Gamma passrates for prostate only and mixed trained models for different water phantom phantom positions inside an air volume. Water phantom thickness was 200 pixels.

and 300 pixels respectively. A summary of the accuracy is given in tab. 6. The transversal view in fig. 11 shows that the prostate model is better at representing the peak at the beginning of the phantom than the mixed model for a depth of 100 pixels. At higher phantom distances the prostate model resembles the qualitative dose distribution quite well, while failing at reaching the correct predictive value. The mixed model can not reach the correct peak height at the beginning but predicts the later curve very well in terms of height and curvature in the transversal view. Coronal and sagittal view both show that the penumbra is resembles very well by both models. At phantom depths of 200 and 300 pixels the shoulders of the dose profile drop too fast for the mixed trained model in the coronal view. Slight fluctuations at the phantom's edge can be observed at 100 pixel depth for the mixed model in the sagittal view.

6 Discussion

6.1 Single entity training

We have shown that training the proposed 3D-UNet with the given training algorithm and dataloading scheme can reach very good accuracy concerning the provided entity. It reached very high dose conformaty for single tested segments aswell as entire treatment plan dose distributions with mean gamma passrates of $94.4\% \pm 6.0\%$ and $96.2\% \pm 6.1\%$ respectively. DVH analysis, depicted in fig. 6, also shows very good agreement in high dose areas, such as CTV and PTV, aswell as OAR such as rectum or femur heads. Dose prediction is very accurate for the entirety of prostate plan segments with various fieldsizes, shapes, radiation angles and patient anatomies. The prediction was fast with inference times of 3 seconds per segment. This is a step in the direction of online plan adaption. With further developement and the usage of parallellization the inference time could be further reduced. Nonetheless, 3 seconds is a significant improvement over 4 hours of Monte Carlo simulation time on an HPC solution that already benefits from parallelization.

In contrast to other previous works such as Kontaxis et al. [50], that only included segments of five pre-defined beam angles into the training data, we did not limit us to specific beam angles but rather took the angles provided by the planning system. We further improved on their work by increasing the voxel size in sagital and coronal directions to increase the resolution of the dose distributions and to be able to achieve higher accuracy in high density gradients.

Testing against entities have shown the limits of this single entity trained approach. Rapidly decreasing gamma passrates with a lowest passrate of 1.9% for a breast cancer case were reached. Particularly low gamma-pass rates were obtained at tangential irradiation angles, where the radiation field passed only through the breast. Predicted values are particularly small for these segments. An example of three poorly predicted segments is depicted in fig. 12. We assume that due to the offset of the radiation target and the associated larger distance between the target and the central beam line, the effects of the distance mask for the central beam line are overweighted and thus reduce the predicted value.

This change in performance and accuracy is most probably occurring due to the increased complexity in testing data. The position of the prostate and the surrounding OAR is quite defined inside the humans anatomy irrespective of the individual patient. Therefore prostate cancer treatment plans have a very similar iso center, aswell as the majority of

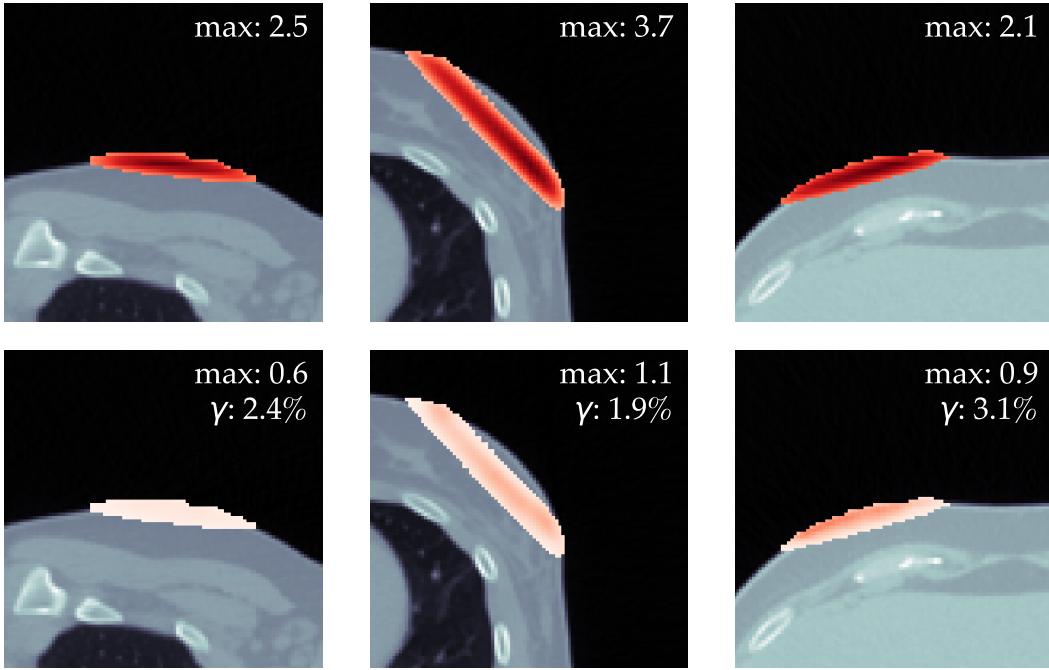


Figure 12: Exemplary images of target dose distribution of tangential segments on breast cancer patients with very small gamma-passrates (top) and the respective dose prediction from the prostate only trained model (bottom).

highly contributing segments have similar shape and sizes. In contrast H&N, LN or even mamma treatment plans have a wider range of segment shapes and fieldsizes (depicted in tab. 1: row fieldsizes). When looking at the anatomy of a patients lower abdomen the intersubject variation in SSD is small compared to an inter-entity comparison to H&N. In fig. 13 three anatomies from prostate, H&N and breast cancer are depicted, with the respective HU distribution in a slice through the isocentric slice. The variety in HU in a prostate cancer patient is comparable with the one of a water phantom, it has a ‘box-like’ shape, while crosssections from H&N and breast cancer show larger variation in height as well as width inside the patients anatomy. Therefore the impact of the distance-squared law is significantly less than on a H&N treatment plan. We assume that the network is not able to correctly map the impact of the squared distance-law on the dose distribution due to the small variation in SSD when using only prostate training data.

There exist a multitude of radiation protocols, which vary depending on the institutional guidelines as well as applications or accelerator type. These different treatment modalities have inherent a variety of different parameters, such as radiation angles, field sizes, and beam shapes. Therefore a robust dose prediction irrespective of the target volume, beam angle and MLC shapes is of crucial importance, and a careful verification process is needed.

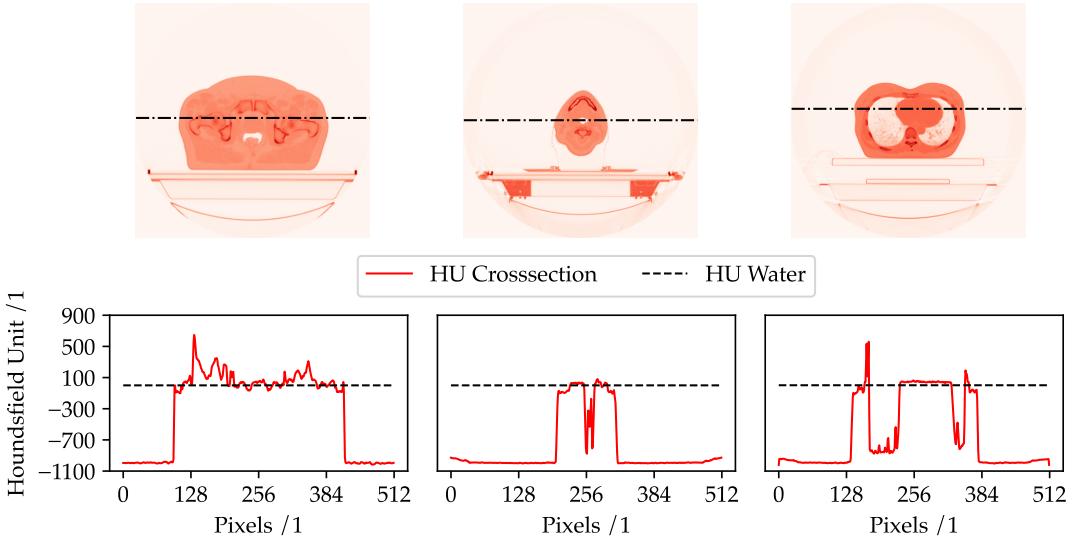


Figure 13: Top: Image of the isocentric slice from a prostate, H&N and breast cancer patient respectively. Bottom: Cross section through the patients anatomy at the marked position. HU value of water indicated with a dotted line.

6.2 Mixed entity training

The number of patches to convergence was lower for the mixed entity model than for the single entity model, at 5.0×10^7 compared to 6.7×10^7 patches. This suggests that the model learns the important aspects of dose deposition more quickly with the greater diversity of training data than with a relatively homogeneously distributed training data set. Model performance and dose prediction accuracy increased for the newly added entities as well as the LN segments as well as plans. Slight decreases in model accuracy regarding single segments of prostate treatment plans are noticed. This is most likely due to the increased generalization of the model and the decreased adaption to prostate plans only. Nevertheless the increased generalization leads to a better performance in the mean for the entire prostate plans compared to the prostate-only trained model. We have observed that the mean single segment prediction accuracies for all entities are smaller than the mean plan prediction accuracies. As depicted in the lower plot of fig. 10, the higher weighted segments tend to be predicted better for both models leading to the effect of better predicted plans.

Predictions for breast cancer segments are particularly bad for the prostate-only as well as the mixed-entity trained model. Beside the challenging anatomy of air cavities as well as tangential fields only partially hitting the patients body, the internal treatment protocol has changed between training and target dataset. Breast cancer treatment plans from the test dataset are significantly higher modulated, with 23.7 ± 5.6 and 83.8 ± 15.3 segments

per plan for old and new institutional guidelines respectively. This higher modulation results in more smaller segments and more complex MLC configurations. Nevertheless the newly adapted model showed promising results for all tested entities regarding entire plan dose distributions. As the model needs specific accelerator setting to predict a dose for a specific segment, this model is not usable as a primary dose engine for a radio treatment planning system. However due to its promising accuracy concerning DVH parameters and its short inference times, it may be promising as an online dose verification or a secondary dose engine after further development.

6.3 Dose deposition physics

Water phantoms are used clinically for dose verification and accelerator constancy testing. Water phantoms are not only the most basic model of a target volume, but dose distributions in them are very well described in literature. Due to the homogenous density distribution inside the phantom physical processes of dose distribution can be analyzed very well with a water phantom. We therefore choose a synthetic water phantom to assess the capability of the model to learn the underlying physics of the dose deposition process. Dose curves were resembled by the networks prediction quite well, dose peaks at the beginning of the phantom were present. The exponential dose decrease as well as the rapid dose drop to the zero level at the end of the phantom has been mapped well qualitatively. The prostate-only trained model showed particular problems reaching the correct predicted height. Scaling the curves to the correct maximum, as shown in fig. 14, gives a more comprehensible picture of the curvature for the prostate-only and mixed entity models. As depicted the qualitative curvature is closer resembled by the prostate model. When comparing different entity crosssections, as shown in fig. 13, the prostate density distribution is comparable to the one of a water phantom. Therefore it is probable that the model has not learned the underlying physics but predicts the qualitative dose distribution in a water phantom better due to the similarities between the density distribution of the lower abdomen and the one of a water phantom.

Both models have problems with the representation of the dose deposition peak at the beginning of the phantom. Taking this into account could explain why the dose from the tangential radiation fields represented in fig. 12 is underestimated in the breast cancer plans. The prediction accuracy of the first 100 pixels of the segment is quite poor, so segments whose radiological path is less than 100 pixels are predicted very poorly. Contrary to the first 100 pixels is the dose distribution well resembled by the mixed entity model. The majority of plans have their target volume and therefore the high dose volume located at higher depths. As the models prediction of dose curvature at greater depths become

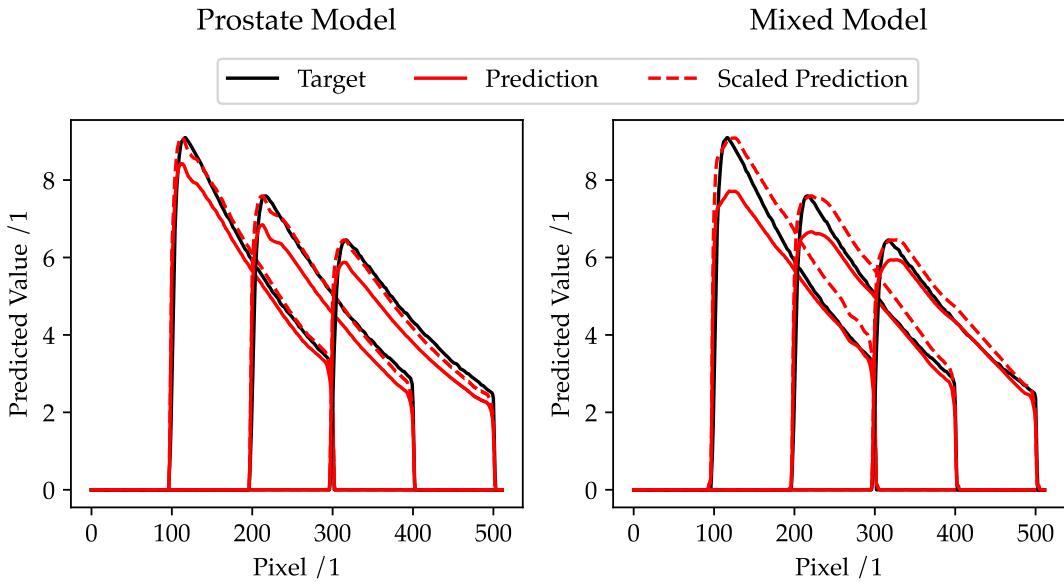


Figure 14: Target, predicted and scaled predicted depth dose curve for water phantom positions at 100, 200 and 300 pixels depth, respectively, from left to right. Left: prostate-only trained model. Right: mixed entity trained model.

more accurate and the gamma passrates only accounts for high dose areas, the gamma passrates increases for target volumes at higher depths.

6.4 Model improvement

The here proposed model showed promising results as a first step towards a fast, accurate and robust dose engine for an online treatment workflow using an MR-Linac. We have showed that the general concept of a spatial information informed 3D-UNet is applicable for the prediction of dose deposition processes in patients anatomy. Nevertheless the evaluation of dose prediction accuracy has shown outliers, especially on challenging tasks such as dose distributions in highly heterogeneous tissues such as air cavities as well as very small fields with short radiological paths inside the patients anatomy. The network has shown short estimation times with approximately 3 seconds per segment. To improve inference time, investigations towards the exploration of other applicable memory efficient network architectures are needed, as well as the utilization of a multi GPU prediction process at inference. In detail analysis of mask impact for every provided mask, as partially done by Kontaxis et al. in [50], might yield beneficial results for the interpretability of the networks performance as well as its limitations towards a robust and reliable dose prediction. As discussed in subsection 6.1, we assume that the center beam line distance, might have a negative effects regarding offset segments.

We have shown that incorporating more heterogenous training data improves the robustness and the prediction accuracy of the model. Our training data consisted of patient specific segments combined inside an entire radio treatment plan that was generated by the treatment planning software. This way we had no control over how segments were shaped or oriented. To achieve desired distributions for fieldsize and gantry angle and to gain control over the segment shapes and the respective iso center, we propose to further increase the training data variability by generating artificial segments. By using artificial segments, we can control the exact distributions of gantry angles aswell as fieldsizes over the entire training data pool, which might lead to an increased representation of otherwise underpresented segment shapes, sizes and areas of very uncommon iso centers.

Data from patients that were previously treated at the MR-linac in our institution is very limited. We therefore propose to use, in addition to arrtificial segments, the anatomy CTs of not only treated patients at the MR-Linac, but any patient CT. By doing so, we deviate from the small pool of cancer patient CTs from our radiotherapy department and increase the pool to any patient that has taken a CT on any given body region. This decreases the chance of the network overfitting to specific anatomic distribrutions of the given patient anatomy pool and contributes to the robustness and generalization process that is desired to be achieved. Implementing this approach would suit as a first proof of concept. Further data individualization can be achieved by deviating from the patient's anatomy and creating artificial target volumes with heterogeneous areas within the irradiated volume. The position, shape and density of these created areas inside the target volume, can be choosen based on the desired distribution.

Further development could go towards exploring an LSTM approach that provides 2.5-D or 3-D dose distributions in the direction of the irradiation angle, as suggested by Neishabouri et al. in her work on proton dose predictions for heterogeneous tissues [54]. Approaching the problem in this manner might lead to networks incorporating of previously seen tissues. As stated in subsubsection 4.2.1 the radiological depth is of crucial importance for the dose deposition processes on the particle path. Implementing a LSTM would account for the crucial information of prior passed tissues, accounting for the vanishing gradient problem. Further information incorporated into the model input, such as a vector mask to indicate the Lorentz force acting on the secondary particles, could increase in a more accurate dose deposition prediction.

In contrast to Kontaxis et al. [50] we decreased the voxel size of the target volume in the coronal and sagital plane. Longer predicion times of 3 seconds per segment were the

result.

The model shows short prediction times, with 3 seconds per segment, which is a first step towards an online plan adaption process involving the MR-linac.

- ausblick wenn diese dosisberechnung robust und auch schnell ist kann sie in der workflow am mr linac so eingebunden werde, dass man keine normale dosisberechnung mehr brauch und wir somit der adaptive bestrahlung einen schritt näher gekommen sind. - dafür müsste die interpretierbarkeit von ML in medizin erhöht werden, sonst keine Information darüber ob jetzt ein segment gut oder schlecht predichtet wird. - approach mit unsicherheit, auf paper verweisen. Uncertainty quantification

-> klinisch primär nicht nutzbar, wegen interpretierbarkeit und nicht gut genug, online dose verification oder second dose, phase 1 für planoptimierung anstatt penicbeam, penicbeam stellt mr feld nich dar und auch probleme bei dichtegradienten

- bezug nehmen auf die Werte, bzw auch ein paar DVH kriterien mal aufzeigen. (hab keine werte in Gy, brauch ich das überhaupt?) -> brauchen wir

- Dann darauf evtl eingehen, dass Feldgrößen von 53 mm feldlänge maximal in 32 Pixeln abgebildet werden kann, das ist auch der punkt ab dem besonders bei dem prostata model aber auch bei dem mixed model die performance abnimmt.

- eingehen auf die grafik mit den feldgrößen und die verteilung der feldgrößen, evlt. ins training evtl mehr kleine einbauen, damit diese besser werden. oder eher die felder berücksichtigen die mehr contribution zu einem plan haben (große?). normal verteilt? dann verteilung der winkel beachten?

7 Conclusion

Lorem ipsum dolor sit amet, consetetur sadipscing elitr, sed diam nonumy eirmod tempor invidunt ut labore et dolore magna aliquyam erat, sed diam voluptua. At vero eos et accusam et justo duo dolores et ea rebum. Stet clita kasd gubergren, no sea takimata sanctus est Lorem ipsum dolor sit amet. Lorem ipsum dolor sit amet, consetetur sadipscing elitr, sed diam nonumy eirmod tempor invidunt ut labore et dolore magna aliquyam erat, sed diam voluptua. At vero eos et accusam et justo duo dolores et ea rebum. Stet clita kasd gubergren, no sea takimata sanctus est Lorem ipsum dolor sit amet.

8 Aknowledgement

The authors acknowledge support by the state of Baden-Württemberg through bwHPC and the German Research Foundation (DFG) through grant no INST 39/963-1 FUGG (bwForCluster NEMO).

9 Appendix

Lorem ipsum dolor sit amet, consetetur sadipscing elitr, sed diam nonumy eirmod tempor invidunt ut labore et dolore magna aliquyam erat, sed diam voluptua. At vero eos et accusam et justo duo dolores et ea rebum. Stet clita kasd gubergren, no sea takimata sanctus est Lorem ipsum dolor sit amet. Lorem ipsum dolor sit amet, consetetur sadipscing elitr, sed diam nonumy eirmod tempor invidunt ut labore et dolore magna aliquyam erat, sed diam voluptua. At vero eos et accusam et justo duo dolores et ea rebum. Stet clita kasd gubergren, no sea takimata sanctus est Lorem ipsum dolor sit amet.

Literature

- [1] Freddie Bray, Mathieu Laversanne, Elisabete Weiderpass, and Isabelle Soerjomataram: *The ever-increasing importance of cancer as a leading cause of premature death worldwide*. In: *Cancer* 127.16 (Aug. 15, 2021), 3029–3030. ISSN: 0008-543X, 1097-0142. DOI: 10.1002/cncr.33587. URL: <https://onlinelibrary.wiley.com/doi/10.1002/cncr.33587> (visited on 11/17/2021).
- [2] Zaigham Abbas and Sakina Rehman: *An Overview of Cancer Treatment Modalities*. In: *Neoplasm*. Ed. by Hafiz Naveed Shahzad. InTech, Sept. 19, 2018. ISBN: 978-1-78923-777-1 978-1-78923-778-8. DOI: 10.5772/intechopen.76558. URL: <http://www.intechopen.com/books/neoplasm/an-overview-of-cancer-treatment-modalities> (visited on 11/17/2021).
- [3] Hans Geinitz, Frank B. Zimmermann, Reinhard Thamm, Andreas Schumertl, Raymond Busch, and Michael Molls: *3D conformal radiation therapy for prostate cancer in elderly patients*. In: *Radiotherapy and Oncology* 76.1 (July 2005), 27–34. ISSN: 01678140. DOI: 10.1016/j.radonc.2005.06.001. URL: <https://linkinghub.elsevier.com/retrieve/pii/S0167814005002082> (visited on 08/16/2021).
- [4] Tan Dat Nguyen et al.: *The curative role of radiotherapy in adenocarcinoma of the prostate in patients under 55 years of age: A rare cancer network retrospective study*. In: *Radiotherapy and Oncology* 77.3 (Dec. 2005), 286–289. ISSN: 01678140. DOI: 10.1016/j.radonc.2005.10.015. URL: <https://linkinghub.elsevier.com/retrieve/pii/S0167814005004779> (visited on 08/16/2021).
- [5] Tom Budiharto, Karin Haustermans, and Gyoergy Kovacs: *External Beam Radiotherapy for Prostate Cancer*. In: *Journal of Endourology* (), 10.
- [6] Joseph Ragaz et al.: *Adjuvant Radiotherapy and Chemotherapy in Node-Positive Premenopausal Women with Breast Cancer*. In: *The New England Journal of Medicine* (1997), 7.
- [7] Mario De Lena, Roberto Zucali, Giovanni Viganotti, Pinuccia Valagussa, and Gianni Bonadonna: *Combined Chemotherapy-Radiotherapy Approach in Locally Advanced (T3b-T4)Breast Cancer*. In: () , 7.
- [8] Carolyn Taylor et al.: *Estimating the Risks of Breast Cancer Radiotherapy: Evidence From Modern Radiation Doses to the Lungs and Heart and From Previous Randomized Trials*. In: *Journal of Clinical Oncology* 35.15 (May 20, 2017), 1641–1649. ISSN: 0732-183X. DOI: 10.1200/JCO.2016.72.0722. URL: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5548226/> (visited on 08/16/2021).

- [9] N. R. Datta, A. K. Bose, H. K. Kapoor, and S. Gupta: *Head and neck cancers: Results of thermoradiotherapy versus radiotherapy*. In: International Journal of Hyperthermia 6.3 (Jan. 1990), 479–486. ISSN: 0265-6736, 1464-5157. DOI: 10.3109/02656739009140944. URL: <http://www.tandfonline.com/doi/full/10.3109/02656739009140944> (visited on 08/16/2021).
- [10] S.A. Bhide and C.M. Nutting: *Advances in radiotherapy for head and neck cancer*. In: Oral Oncology 46.6 (June 2010), 439–441. ISSN: 13688375. DOI: 10.1016/j.oraloncology.2010.03.005. URL: <https://linkinghub.elsevier.com/retrieve/pii/S1368837510000941> (visited on 08/16/2021).
- [11] Pierre Castadot, John A. Lee, Xavier Geets, and Vincent Grégoire: *Adaptive Radiotherapy of Head and Neck Cancer*. In: Seminars in Radiation Oncology 20.2 (Apr. 2010), 84–93. ISSN: 10534296. DOI: 10.1016/j.semradonc.2009.11.002. URL: <https://linkinghub.elsevier.com/retrieve/pii/S1053429609000769> (visited on 08/16/2021).
- [12] Howard E. Morgan and David J. Sher: *Adaptive radiotherapy for head and neck cancer*. In: Cancers of the Head & Neck 5.1 (Dec. 2020), 1. ISSN: 2059-7347. DOI: 10.1186/s41199-019-0046-z. URL: <https://cancersheadneck.biomedcentral.com/articles/10.1186/s41199-019-0046-z> (visited on 08/16/2021).
- [13] Morten Høyer, Anand Swaminath, Sean Bydder, Michael Lock, Alejandra Méndez Romero, Brian Kavanagh, Karyn A. Goodman, Paul Okunieff, and Laura A. Dawson: *Radiotherapy for Liver Metastases: A Review of Evidence*. In: International Journal of Radiation Oncology*Biology*Physics 82.3 (Mar. 1, 2012), 1047–1057. ISSN: 0360-3016. DOI: 10.1016/j.ijrobp.2011.07.020. URL: <https://www.sciencedirect.com/science/article/pii/S0360301611030902> (visited on 08/16/2021).
- [14] Jörn Wulf, Ulrich Hädinger, Ulrich Oppitz, Wibke Thiele, Rea Ness-Dourdoumas, and Michael Flentje: *Stereotactic Radiotherapy of Targets in the Lung and Liver*: in: Strahlentherapie und Onkologie 177.12 (Dec. 2001), 645–655. ISSN: 0179-7158. DOI: 10.1007/PL00002379. URL: <http://link.springer.com/10.1007/PL00002379> (visited on 08/16/2021).
- [15] Joern Wulf, Matthias Guckenberger, Ulrich Haedinger, Ulrich Oppitz, Gerd Mueller, Kurt Baier, and Michael Flentje: *Stereotactic radiotherapy of primary liver cancer and hepatic metastases*. In: Acta Oncologica 45.7 (Jan. 2006), 838–847. ISSN: 0284-186X, 1651-226X. DOI: 10.1080/02841860600904821. URL: <http://www.tandfonline.com/doi/full/10.1080/02841860600904821> (visited on 08/16/2021).

- [16] Florian Sterzing, Thomas B. Brunner, Iris Ernst, Wolfgang W. Baus, Burkhard Greve, Klaus Herfarth, and Matthias Guckenberger: *Stereotactic body radiotherapy for liver tumors: Principles and practical guidelines of the DEGRO Working Group on Stereotactic Radiotherapy*. In: Strahlentherapie und Onkologie 190.10 (Oct. 2014), 872–881. ISSN: 0179-7158, 1439-099X. DOI: 10.1007/s00066-014-0714-1. URL: <http://link.springer.com/10.1007/s00066-014-0714-1> (visited on 08/16/2021).
- [17] Jacob S Witt, Stephen A Rosenberg, and Michael F Bassetti: *MRI-guided adaptive radiotherapy for liver tumours: visualising the future*. In: The Lancet Oncology 21.2 (Feb. 1, 2020), e74–e82. ISSN: 1470-2045. DOI: 10.1016/S1470-2045(20)30034-6. URL: <https://www.sciencedirect.com/science/article/pii/S1470204520300346> (visited on 08/16/2021).
- [18] Breast Cancer Expert Panel of the German Society of Radiation Oncology (DEGRO) et al.: *DEGRO practical guidelines for radiotherapy of breast cancer IV: Radiotherapy following mastectomy for invasive breast cancer*. In: Strahlentherapie und Onkologie 190.8 (Aug. 2014), 705–714. ISSN: 0179-7158, 1439-099X. DOI: 10.1007/s00066-014-0687-0. URL: <http://link.springer.com/10.1007/s00066-014-0687-0> (visited on 08/16/2021).
- [19] Haruo Matsushita, Keiichi Jingu, Rei Umezawa, Takaya Yamamoto, Yojiro Ishikawa, Noriyoshi Takahashi, Yu Katagiri, and Noriyuki Kadoya: *Stereotactic Radiotherapy for Oligometastases in Lymph Nodes—A Review*. In: Technology in Cancer Research & Treatment 17 (Jan. 2018), 153303381880359. ISSN: 1533-0346, 1533-0338. DOI: 10.1177/1533033818803597. URL: <http://journals.sagepub.com/doi/10.1177/1533033818803597> (visited on 08/16/2021).
- [20] John L. Mikell et al.: *Postoperative Radiotherapy is Associated with Better Survival in Non-Small Cell Lung Cancer with Involved N2 Lymph Nodes: Results of an Analysis of the National Cancer Data Base*. In: Journal of Thoracic Oncology 10.3 (Mar. 2015), 462–471. ISSN: 15560864. DOI: 10.1097/JTO.000000000000411. URL: <https://linkinghub.elsevier.com/retrieve/pii/S1556086415316543> (visited on 08/16/2021).
- [21] Dan Lundstedt, Magnus Gustafsson, Gunnar Steineck, David Alsadius, Agnetha Sundberg, Ulrica Wilderäng, Erik Holmberg, Karl-Axel Johansson, and Per Karlsson: *Long-term symptoms after radiotherapy of supraclavicular lymph nodes in breast cancer patients*. In: Radiotherapy and Oncology 103.2 (May 2012), 155–160. ISSN: 01678140. DOI: 10.1016/j.radonc.2011.12.017. URL: <https://linkinghub.elsevier.com/retrieve/pii/S0167814011007572> (visited on 08/16/2021).

- [22] Barbara Alicja Jereczek-Fossa, Sara Ronchi, and Roberto Orecchia: *Is Stereotactic Body Radiotherapy (SBRT) in lymph node oligometastatic patients feasible and effective?* In: Reports of Practical Oncology and Radiotherapy 20.6 (2015), 472–483. ISSN: 1507-1367. DOI: 10.1016/j.rpor.2014.10.004. URL: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4661354/> (visited on 08/16/2021).
- [23] Byungchul Cho: *Intensity-modulated radiation therapy: a review with a physics perspective.* In: Radiation Oncology Journal 36.1 (Mar. 31, 2018), 1–10. ISSN: 2234-1900, 2234-3156. DOI: 10.3857/roj.2018.00122. URL: <http://e-roj.org/journal/view.php?doi=10.3857/roj.2018.00122> (visited on 11/17/2021).
- [24] Xiao Han: *MR-based synthetic CT generation using a deep convolutional neural network method.* In: Medical Physics 44.4 (2017). _eprint: <https://aapm.onlinelibrary.wiley.com/doi/10.1002/mp.12155>. ISSN: 2473-4209. DOI: 10.1002/mp.12155. URL: <https://aapm.onlinelibrary.wiley.com/doi/abs/10.1002/mp.12155> (visited on 08/23/2021).
- [25] Jelmer M. Wolterink, Anna M. Dinkla, Mark H. F. Savenije, Peter R. Seevinck, Cornelis A. T. van den Berg, and Ivana Išgum: *Deep MR to CT Synthesis Using Unpaired Data.* In: *Simulation and Synthesis in Medical Imaging.* Ed. by Sotirios A. Tsaftaris, Ali Gooya, Alejandro F. Frangi, and Jerry L. Prince. Lecture Notes in Computer Science. Cham: Springer International Publishing, 2017, 14–23. ISBN: 978-3-319-68127-6. DOI: 10.1007/978-3-319-68127-6_2.
- [26] Anna M. Dinkla, Jelmer M. Wolterink, Matteo Maspero, Mark H.F. Savenije, Joost J.C. Verhoeff, Enrica Seravalli, Ivana Išgum, Peter R. Seevinck, and Cornelis A.T. van den Berg: *MR-Only Brain Radiation Therapy: Dosimetric Evaluation of Synthetic CTs Generated by a Dilated Convolutional Neural Network.* In: International Journal of Radiation Oncology*Biology*Physics 102.4 (Nov. 2018), 801–812. ISSN: 03603016. DOI: 10.1016/j.ijrobp.2018.05.058. URL: <https://linkinghub.elsevier.com/retrieve/pii/S0360301618309106> (visited on 08/23/2021).
- [27] Samaneh Kazemifar, Anjali Balagopal, Dan Nguyen, Sarah McGuire, Raquibul Hannan, Steve Jiang, and Amir Owraghi: *Segmentation of the prostate and organs at risk in male pelvic CT images using deep learning.* In: Biomedical Physics & Engineering Express 4.5 (July 23, 2018), 055003. ISSN: 2057-1976. DOI: 10.1088/2057-1976/aad100. URL: <https://iopscience.iop.org/article/10.1088/2057-1976/aad100> (visited on 08/23/2021).
- [28] Shujun Liang, Fan Tang, Xia Huang, Kaifan Yang, Tao Zhong, Runyue Hu, Shangqing Liu, Xinrui Yuan, and Yu Zhang: *Deep-learning-based detection and segmentation of organs at risk in nasopharyngeal carcinoma computed tomographic images for radiotherapy planning.* In: European Radiology 29.4 (Apr. 2019), 1961–1967. ISSN:

- 0938-7994, 1432-1084. DOI: 10.1007/s00330-018-5748-9. URL: <http://link.springer.com/10.1007/s00330-018-5748-9> (visited on 08/23/2021).
- [29] Dinggang Shen, Tianming Liu, Terry M. Peters, Lawrence H. Staib, Caroline Es-sert, Sean Zhou, Pew-Thian Yap, and Ali Khan, eds.: *Medical Image Computing and Computer Assisted Intervention – MICCAI 2019: 22nd International Conference, Shenzhen, China, October 13–17, 2019, Proceedings, Part II*. Vol. 11765. Lecture Notes in Computer Science. Cham: Springer International Publishing, 2019. ISBN: 978-3-030-32244-1 978-3-030-32245-8. DOI: 10.1007/978-3-030-32245-8. URL: <https://link.springer.com/10.1007/978-3-030-32245-8> (visited on 08/23/2021).
- [30] Jiawei Fan, Jiazhou Wang, Zhi Chen, Chaosu Hu, Zhen Zhang, and Weigang Hu: *Automatic treatment planning based on three-dimensional dose distribution predicted from deep learning technique*. In: Medical Physics 46.1 (Jan. 2019), 370–381. ISSN: 00942405. DOI: 10.1002/mp.13271. URL: <http://doi.wiley.com/10.1002/mp.13271> (visited on 03/16/2021).
- [31] Zhiqiang Liu, Jiawei Fan, Minghui Li, Hui Yan, Zhihui Hu, Peng Huang, Yuan Tian, Junjie Miao, and Jianrong Dai: *A deep learning method for prediction of three-dimensional dose distribution of helical tomotherapy*. In: Medical Physics 46.5 (May 2019), 1972–1983. ISSN: 0094-2405, 2473-4209. DOI: 10.1002/mp.13490. URL: <https://onlinelibrary.wiley.com/doi/abs/10.1002/mp.13490> (visited on 03/16/2021).
- [32] Umair Javaid, Kevin Souris, Damien Dasnoy, Sheng Huang, and John A. Lee: *Mitigating inherent noise in Monte Carlo dose distributions using dilated U-Net*. In: Medical Physics 46.12 (Dec. 2019), 5790–5798. ISSN: 0094-2405, 2473-4209. DOI: 10.1002/mp.13856. URL: <https://onlinelibrary.wiley.com/doi/10.1002/mp.13856> (visited on 08/24/2021).
- [33] Keyvan Jabbari: *Review of fast Monte Carlo codes for dose calculation in radiation therapy treatment planning*. In: Journal of Medical Signals & Sensors 1.1 (2011), 73. ISSN: 2228-7477. DOI: 10.4103/2228-7477.83522. URL: <http://www.jmssjournal.net/text.asp?2011/1/1/73/83522> (visited on 11/18/2021).
- [34] Andre Esteva, Brett Kuprel, Roberto A. Novoa, Justin Ko, Susan M. Swetter, Helen M. Blau, and Sebastian Thrun: *Dermatologist-level classification of skin cancer with deep neural networks*. In: Nature 542.7639 (Feb. 2, 2017), 115–118. ISSN: 0028-0836, 1476-4687. DOI: 10.1038/nature21056. URL: <http://www.nature.com/articles/nature21056> (visited on 11/18/2021).

- [35] Jie-Zhi Cheng, Dong Ni, Yi-Hong Chou, Jing Qin, Chui-Mei Tiu, Yeun-Chung Chang, Chiun-Sheng Huang, Dinggang Shen, and Chung-Ming Chen: *Computer-Aided Diagnosis with Deep Learning Architecture: Applications to Breast Lesions in US Images and Pulmonary Nodules in CT Scans*. In: Scientific Reports 6.1 (Apr. 2016), 24454. ISSN: 2045-2322. DOI: 10.1038/srep24454. URL: <http://www.nature.com/articles/srep24454> (visited on 11/18/2021).
- [36] Mark Cicero, Alexander Bilbily, Errol Colak, Tim Dowdell, Bruce Gray, Kuhan Perampaladas, and Joseph Barfett: *Training and Validating a Deep Convolutional Neural Network for Computer-Aided Detection and Classification of Abnormalities on Frontal Chest Radiographs*. In: Investigative Radiology 52.5 (May 2017), 281–287. ISSN: 1536-0210, 0020-9996. DOI: 10.1097/RLI.0000000000000341. URL: <https://journals.lww.com/00004424-201705000-00004> (visited on 11/18/2021).
- [37] Osamu Iizuka, Fahdi Kanavati, Kei Kato, Michael Rambeau, Koji Arihiro, and Masayuki Tsuneki: *Deep Learning Models for Histopathological Classification of Gastric and Colonic Epithelial Tumours*. In: Scientific Reports 10.1 (Dec. 2020), 1504. ISSN: 2045-2322. DOI: 10.1038/s41598-020-58467-9. URL: <http://www.nature.com/articles/s41598-020-58467-9> (visited on 11/18/2021).
- [38] Ryan Poplin, Avinash V. Varadarajan, Katy Blumer, Yun Liu, Michael V. McConnell, Greg S. Corrado, Lily Peng, and Dale R. Webster: *Prediction of cardiovascular risk factors from retinal fundus photographs via deep learning*. In: Nature Biomedical Engineering 2.3 (Mar. 2018), 158–164. ISSN: 2157-846X. DOI: 10.1038/s41551-018-0195-0. URL: <http://www.nature.com/articles/s41551-018-0195-0> (visited on 11/18/2021).
- [39] Geoff Delaney, Susannah Jacob, Carolyn Featherstone, and Michael Barton: *The role of radiotherapy in cancer treatment: Estimating optimal utilization from a review of evidence-based clinical guidelines*. In: Cancer 104.6 (Sept. 15, 2005), 1129–1137. ISSN: 0008-543X, 1097-0142. DOI: 10.1002/cncr.21324. URL: <https://onlinelibrary.wiley.com/doi/10.1002/cncr.21324> (visited on 11/18/2021).
- [40] Lan-ya Li, Yi-di Guan, Xi-sha Chen, Jin-ming Yang, and Yan Cheng: *DNA Repair Pathways in Cancer Therapy and Resistance*. In: Frontiers in Pharmacology 11 (Feb. 8, 2021), 629266. ISSN: 1663-9812. DOI: 10.3389/fphar.2020.629266. URL: <https://www.frontiersin.org/articles/10.3389/fphar.2020.629266/full> (visited on 11/18/2021).
- [41] H. A. Lorentz: *Versuch Einer Theorie der Electrischen und Optischen Erscheinungen in Bewegten Körpern*. In: *Collected Papers: Volume V*. Ed. by H. A. Lorentz. Dordrecht: Springer Netherlands, 1937, 1–138. ISBN: 978-94-015-3445-1. DOI: 10.

- 1007/978-94-015-3445-1_1. URL: https://doi.org/10.1007/978-94-015-3445-1_1 (visited on 11/19/2021).
- [42] Anders Ahnesjö: *Collapsed cone convolution of radiant energy for photon dose calculation in heterogeneous media: Photon dose calculation*. In: Medical Physics 16.4 (July 1989), 577–592. ISSN: 00942405. DOI: 10.1118/1.596360. URL: <http://doi.wiley.com/10.1118/1.596360> (visited on 06/22/2021).
- [43] R. Mohan, C. Chui, and L. Lidofsky: *Differential pencil beam dose computation model for photons: Differential pencil beam computation for photons*. In: Medical Physics 13.1 (Jan. 1986), 64–73. ISSN: 00942405. DOI: 10.1118/1.595924. URL: <http://doi.wiley.com/10.1118/1.595924> (visited on 06/22/2021).
- [44] nrc-cnrc/EGSnrc. original-date: 2012-11-28T19:26:41Z. Nov. 2, 2021. URL: <https://github.com/nrc-cnrc/EGSnrc> (visited on 11/04/2021).
- [45] M. Friedel, M. Nachbar, D. Mönnich, O. Dohm, and D. Thorwarth: *Development and validation of a 1.5 T MR-Linac full accelerator head and cryostat model for Monte Carlo dose simulations*. In: Medical Physics 46.11 (Nov. 2019), 5304–5313. ISSN: 0094-2405, 2473-4209. DOI: 10.1002/mp.13829. URL: <https://onlinelibrary.wiley.com/doi/abs/10.1002/mp.13829> (visited on 03/16/2021).
- [46] Ahmad Neishabouri, Niklas Wahl, Andrea Mairani, Ullrich Köthe, and Mark Bangert: *Long short-term memory networks for proton dose calculation in highly heterogeneous tissues*. In: Medical Physics (Mar. 11, 2021), mp.14658. ISSN: 0094-2405, 2473-4209. DOI: 10.1002/mp.14658. URL: <https://onlinelibrary.wiley.com/doi/10.1002/mp.14658> (visited on 03/16/2021).
- [47] Sonia Martinot, Norbert Bus, Maria Vakalopoulou, Charlotte Robert, Eric Deutsch, and Nikos Paragios: *High-Particle Simulation of Monte-Carlo Dose Distribution with 3D ConvLSTMs*. In: *Medical Image Computing and Computer Assisted Intervention – MICCAI 2021*. Ed. by Marleen de Bruijne, Philippe C. Cattin, Stéphane Cotin, Nicolas Padoy, Stefanie Speidel, Yefeng Zheng, and Caroline Essert. Vol. 12904. Series Title: Lecture Notes in Computer Science. Cham: Springer International Publishing, 2021, 499–508. ISBN: 978-3-030-87201-4 978-3-030-87202-1. DOI: 10.1007/978-3-030-87202-1_48. URL: https://link.springer.com/10.1007/978-3-030-87202-1_48 (visited on 11/20/2021).
- [48] Ryan Neph, Yangsibo Huang, Youming Yang, and Ke Sheng: *DeepMCDose: A Deep Learning Method for Efficient Monte Carlo Beamlet Dose Calculation by Predictive Denoising in MR-Guided Radiotherapy*. In: arXiv:1908.04437 [physics] (Aug. 12,

- 2019). arXiv: 1908.04437. URL: <http://arxiv.org/abs/1908.04437> (visited on 08/02/2021).
- [49] Jiawei Fan, Lei Xing, Peng Dong, Jiazhou Wang, Weigang Hu, and Yong Yang: *Data-driven dose calculation algorithm based on deep U-Net*. In: Physics in medicine and biology 65.24 (Dec. 22, 2020), 245035. ISSN: 0031-9155. DOI: 10.1088/1361-6560/abca05. URL: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7870566/> (visited on 11/20/2021).
- [50] C. Kontaxis, G. H. Bol, J. J. W. Lagendijk, and B. W. Raaymakers: *DeepDose: Towards a fast dose calculation engine for radiation therapy using deep learning*. In: Physics in Medicine & Biology 65.7 (Apr. 2020). Publisher: IOP Publishing, 075013. ISSN: 0031-9155. DOI: 10.1088/1361-6560/ab7630. URL: <https://doi.org/10.1088/1361-6560/ab7630> (visited on 08/23/2021).
- [51] Olaf Ronneberger, Philipp Fischer, and Thomas Brox: *U-Net: Convolutional Networks for Biomedical Image Segmentation*. In: arXiv:1505.04597 [cs] (May 18, 2015). arXiv: 1505 . 04597. URL: <http://arxiv.org/abs/1505.04597> (visited on 11/03/2021).
- [52] Fernando Pérez-García, Rachel Sparks, and Sébastien Ourselin: *TorchIO: A Python library for efficient loading, preprocessing, augmentation and patch-based sampling of medical images in deep learning*. In: Computer Methods and Programs in Biomedicine 208 (Sept. 2021), 106236. ISSN: 01692607. DOI: 10.1016/j.cmpb.2021.106236. URL: <https://linkinghub.elsevier.com/retrieve/pii/S0169260721003102> (visited on 11/03/2021).
- [53] Daniel A. Low, William B. Harms, Sasa Mutic, and James A. Purdy: *A technique for the quantitative evaluation of dose distributions*. In: Medical Physics 25.5 (May 1998), 656–661. ISSN: 00942405. DOI: 10.1118/1.598248. URL: <http://doi.wiley.com/10.1118/1.598248> (visited on 03/16/2021).
- [54] Ahmad Neishabouri, Niklas Wahl, Ulrich Köthe, and Mark Bangert: *Long short-term memory networks for proton dose calculation in highly heterogeneous tissues*. In: arXiv:2006.06085 [physics] (June 10, 2020). arXiv: 2006 . 06085. URL: <http://arxiv.org/abs/2006.06085> (visited on 03/16/2021).

Appendix