Title

Automated Patient Safety Management and Quality Control in Radiation Therapy

Author

Charters, John

Publication Date

2024

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA

Los Angeles

Automated Patient Safety Management and Quality Control in Radiation Therapy

A dissertation submitted in partial satisfaction of the requirements for the degree

Doctor of Philosophy in Physics and Biology in Medicine

by

John Austin Charters

2024

ABSTRACT OF THE DISSERTATION

Automated Patient Safety Management and Quality Control in Radiation Therapy

by

John Austin Charters
Doctor of Philosophy in Physics and Biology in Medicine
University of California, Los Angeles, 2024
Professor James Michael Lamb, Chair

The research aims presented in this dissertation are centered on broad themes of improving automation and error detection in radiation oncology. The first aim was to create a stereoscopic radiographic generator for the ExacTrac image-guidance system. Our methodology enables medical physicists to compute geometric parameters for ray tracing based on values contained in ExacTrac configuration logs. Rigid registrations to ground-truth radiographs were performed using medical imaging software, and the results demonstrated sub-millimeter accuracy.

The second aim was to create deep-learning models to automatically detect off-by-one vertebral body misalignments with on-board planar imaging. Thoracic and abdominal radiotherapy plans were retrieved from our clinical servers using the DICOM networking protocol. Pairs of digital and treatment radiographs were organized according to beam energy and orientation. Realistic off-by-one misalignments were systematically produced. Convolutional neural networks were trained to classify whether such radiographic pairs are aligned or misaligned. Given a desired 95% model specificity, the orthogonal kilovoltage model achieved a sensitivity of 99%. The models established an independent review process for setup error

incidents over large retrospective datasets, which are nearly impossible to review by hand. Of particular emphasis, an instance of a previously unnoticed off-by-one setup error at our institution was found.

The third aim was to create automated algorithms to evaluate the quality of prostate radiotherapy treatment plans. Quality metrics included number of days to plan approval, target margins, presence of fiducial markers, and prescribed radiation dose. The automated measurements were compared with values determined manually in clinical software, and high accuracy was obtained. Furthermore, deep-learning models for auto-contouring prostate bed target volumes were created. Refined models were developed using a novel data-driven approach to separate contours based on anterior and posterior convexity.

Quality outliers were flagged for retrospective human review. Among the cases reviewed, a previously unnoticed mistake was identified, where a prostate patient treated at our institution was overexposed by 2 Gy. The automated algorithms were applied on treatment plans from our institution and from hospitals in the greater community, which allowed us to assess the existing range of standards of care in clinical practice.

The dissertation of John Austin Charters is approved.

Minsong Cao

Melissa Ghafarian

Daniel Abraham Low

Drew Moghanaki

John Paul Neylon

James Michael Lamb, Committee Chair

University of California, Los Angeles

2024

DEDICATION

To my grandfather, Dr. John J. Gallucci, M.D.

Contents

List of Figures

List of Tables

Acknowledgments

My achievements in graduate school originated from an outstanding undergraduate education at the University of Notre Dame. I would like to thank all of my undergraduate mathematics and physics instructors, notably my mathematics advisor Dr. Dennis Snow. The honors mathematics curriculum I completed remains foundational to my approach to problem-solving and has driven my lifelong desire for learning.

I was overjoyed to have been accepted into the Physics and Astronomy Department at UCLA, and I would like to thank Dr. Michael Gutperle for initially reaching out regarding my acceptance and inspiring me to select UCLA during my orientation visit. I received an exceptional graduate education at UCLA in physics and mathematics, and I feel particularly fortunate to have spent a year studying under theoretical physicist Dr. Eric D'Hoker.

My professional aspirations in medical physics motivated me to reconsider a graduate program accredited by the Commission on Accreditation of Medical Physics Education Programs (CAMPEP). I am grateful to Dr. Mike McNitt-Gray and Dr. Nzhde Agazaryan for introducing me to the Physics and Biology in Medicine (PBM) Interdepartmental Program at the David Geffen School of Medicine at UCLA. Dr. McNitt-Gray was welcoming and encouraging in my decision to transfer programs. Dr. Agazaryan allowed me to audit his radiation therapy course while still enrolled as a physics graduate student, which further solidified my decision. I would also like to acknowledge Stephanie Krilov for assisting and supporting me during that time.

It has been an honor to have been mentored by my doctoral advisor Dr. Jim Lamb, and he deserves the utmost credit in my preparation for this dissertation milestone. Throughout my medical physics PhD journey, Dr. Lamb guided my growth as a clinical scientist. He regularly monitored my PhD progress and continually motivated me throughout my research efforts. Dr. Lamb provided tremendous assistance with clinical projects, even during the height of the

COVID-19 pandemic. Throughout the past five years, I have developed a fond appreciation of an interdisciplinary education in biomedical engineering, artificial intelligence, applied mathematics, and scientific programming. The PBM graduate coursework and my lineup of independent research projects have provided me highly valuable knowledge in an increasingly technological world.

Additionally, it has been a privilege collaborating with and learning from a remarkable cohort of graduate students. I would like to thank Dr. Lamb's research group: Dishane Luximon, Rachel Petragallo, Yasin Abdulkadir, and Justin Hink. I would also like to thank my doctoral committee members for their involvement in my research projects: Dr. Minsong Cao, Dr. Melissa Ghafarian, Dr. Dan Low, Dr. Drew Moghanaki, and Dr. Jack Neylon. For their wonderful management of the PBM Program, I am grateful to Reth Im and Alondra Correa Bautista.

I feel incredibly honored to have been accepted by the UCLA Department of Radiation Oncology into the Medical Physics Residency Program. I am grateful and enthusiastic about this next phase of my medical physics career at UCLA Health. I look forward to my clinical training and continuing the American Board of Radiology (ABR) certification process.

Finally, I am immensely appreciative of my parents Jeff and Gina, my siblings Drew and Audrey, and my extended family for their ongoing love and support. I am forever grateful to my maternal grandfather Dr. John Gallucci for his influence over my career path. His decades of service as a radiation oncologist and his dedication to the lives of cancer patients continues to be the primary source of my professional inspiration.

Preface

Chapter 1 is based on the following publication:

John A. Charters, Pascal Bertram, and James M. Lamb. "Offline generator for digitally reconstructed radiographs of a commercial stereoscopic radiotherapy image-guidance system". In: *Journal of Applied Clinical Medical Physics* 23.3 (2022). DOI: 10.1002/acm2.13492

Chapter 2 is based on the following publication:

John A. Charters, Dishane Luximon, Rachel Petragallo, Jack Neylon, Daniel A. Low, and James M. Lamb. "Automated detection of vertebral body misalignments in orthogonal kV and MV guided radiotherapy: application to a comprehensive retrospective dataset". In: *Biomedical Physics and Engineering Express* 10.2 (2024). DOI: 10.1088/2057-1976/ad2baa

Chapter 3 is based on the following manuscript in preparation for publication:

John A. Charters, Yasin Abdulkadir, Justin Hink, Melissa Ghafarian, Jack Neylon, and James M. Lamb. "Automated community review of prostate cancer quality of care".

Chapter 4 is based on the following manuscript in preparation for publication:

John A. Charters, Yasin Abdulkadir, Dan Ruan, and James M. Lamb. "Intact prostate and prostate bed target volume auto-contouring for radiotherapy plan quality review".

The research projects comprising this dissertation were supported by the Agency for Healthcare Research and Quality (AHRQ) under grant number R01 HS026486. The principal investigator (PI) was Dr. James M. Lamb. All medical data analyzed in this dissertation are protected under the Health Insurance Portability and Accountability Act (HIPAA) Privacy Rule.

Vita

EDUCATION

University of California, Los Angeles — Los Angeles, CA

M.S. in Physics and Biology in Medicine — *December 2021*

M.S. in Physics — *December 2019*

University of Notre Dame — Notre Dame, IN

B.S. in Honors Mathematics — *May 2018*

B.S. in Honors Physics — *May 2018*

CERTIFICATIONS

American Board of Radiology

Therapeutic Medical Physics Initial Qualifying Part I Exam — *December 2020*

PUBLICATIONS

John A. Charters et al. "Automated detection of vertebral body misalignments in orthogonal kV and MV guided radiotherapy: application to a comprehensive retrospective dataset". In: *Biomedical Physics and Engineering Express* 10.2 (2024). DOI: 10.1088/2057-1976/ad2baa

John A. Charters et al. "Dosimetric evaluation of respiratory gating on a 0.35-T magnetic resonance-guided radiotherapy linac". In: *Journal of Applied Clinical Medical Physics* 23.9 (2022). DOI: 10.1002/acm2.13666

John A. Charters, Pascal Bertram, and James M. Lamb. "Offline generator for digitally reconstructed radiographs of a commercial stereoscopic radiotherapy image-guidance system". In: *Journal of Applied Clinical Medical Physics* 23.3 (2022). DOI: 10.1002/acm2.13492

Dishane C. Luximon et al. "Development and interinstitutional validation of an automatic vertebral-body misalignment error detector for cone-beam CT-guided radiotherapy". In: *Medical Physics* 49.10 (2022), pp. 6410–6423. DOI: 10.1002/mp.15927

SELECTED CONFERENCES

John A. Charters et al. "Automated patient positioning error detection with orthogonal kV and unpaired MV treatment radiographs". In: *AAPM 65th Annual Meeting & Exhibition*. Houston, TX: American Association of Physicists in Medicine, July 2023

John A. Charters et al. "Dosimetric evaluation of respiratory gating on a 0.35-T magnetic resonance guided radiotherapy linac". In: *AAPM 64th Annual Meeting & Exhibition*. Washington, DC: American Association of Physicists in Medicine, July 2022

John A. Charters et al. "Improving DWI spatial resolution with respiratory motion modeling on an interleaved bSSFP-DWI sequence". In: *AAPM 63rd Annual Meeting & Exhibition*. Virtual: American Association of Physicists in Medicine, July 2021

John A. Charters, Amar Kishan, and James M. Lamb. "Application of super-resolution reconstruction to prostate 0.35 T MRI scans". In: *AAPM 63rd Annual Meeting & Exhibition*. Virtual: American Association of Physicists in Medicine, July 2021

Chapter 1

ExacTrac stereoscopic DRR rendering

1.1 Introduction

In image-guided radiotherapy (IGRT), ExacTrac (Brainlab, Munich, Germany) is a system designed for spatially precise treatments such as stereotactic body radiotherapy (SBRT) and intracranial stereotactic radiosurgery (SRS) [53]. The system consists of a pair of stereoscopic floor-mounted kV X-ray sources and ceiling-mounted flat-panel detectors. ExacTrac is typically installed in conjunction with C-arm gantry radiotherapy machines. Each flat-panel detector has a field-of-view (FOV) of approximately 10 cm at the linear accelerator (linac) isocenter. Since ExacTrac is not directly attached to the C-arm, geometric calibration for gravity-induced mechanical flex [66] is not required. At our institution, ExacTrac is implemented with Varian TrueBeam STx and Novalis Tx radiotherapy linacs (Varian Medical Systems, Inc., Palo Alto, CA), predominantly in the treatment of targets that are rigidly registered to bony anatomy such as intracranial, spinal, and bone-metastatic tumors.

ExacTrac is coupled to an infrared camera device that enables precise positioning of a custom carbon-fiber treatment couch with six degrees of freedom (i.e., three translations and three rotations) in order to align a radiotherapy target with its planned location. Translational and rotational shifts are optimized by minimizing a gradient-based cross-correlation

similarity measurement between a given radiograph and digitally reconstructed radiographs (DRRs) derived from the planning computed tomography (CT) scan [67].

Our medical physics research group has previously investigated algorithms for detecting patient positioning errors from IGRT in a variety of treatment sites, including along the vertebral column [64, 51]. Algorithm training required simulating errors by matching clinically-acquired X-ray projections of a given vertebral body with DRR projections of an adjacent, *incorrect* vertebral body. The ExacTrac offline preparation and review station is capable of simulating spinal alignments to the wrong vertebral body, but it is not capable of exporting the incorrectly optimized DRRs. Thus, our research efforts required generating DRRs that would accurately replicate those produced by the ExacTrac system, given arbitrary 6D CT transformations.

In this chapter, we present two stereoscopic DRR generators, one written from scratch and one using image processing libraries from the open-source Insight Segmentation and Registration Toolkit (ITK) [75]. We describe how to obtain ExacTrac rendering quantities and convert them to suitable parameters for offline DRR generation with ray tracing. Finally, we validate the geometric accuracy of our programs by comparing generated DRRs to reference ExacTrac DRRs of objects in treatment positions (which *are* exportable) using landmark analysis and rigid registrations.

1.2 Linac coordinate systems

In radiation oncology, a volumetric CT imaging procedure is typically stored as an axial sequence of planar image files saved in Digital Imaging and Communications in Medicine (DICOM) format [65]. We will frequently reference attributes found in the DICOM Standard, published by the National Electrical Manufacturers Association (NEMA) [3]. Assume that all of our DICOM data uses patient-centric Cartesian coordinates in a head-first supine (HFS) orientation, whereby when facing the linac, $+x$ is right, $+y$ is down, and $+z$

is forward. Patient-centric coordinates are referred to as the reference coordinate system (RCS) in the DICOM Standard. The ExacTrac rendering process assumes machine-centric coordinates that are standardized by the International Electrotechnical Commission (IEC), whereby when facing the linac, $+x$ is right, $+y$ is forward, and $+z$ is up. The change-of-basis transformation is given by

$$\mathbf{v}_{\text{IEC}} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & -1 & 0 \end{pmatrix} \mathbf{v}_{\text{RCS}}. \tag{1.1}$$

An example of the requisite clinical apparatus in IEC coordinates is sketched in Figure 1.1. Given an X-ray tube and its corresponding flat-panel detector, the *source-to-image distance* (SID) provides the distance between the X-ray point source and the detector center. The *source-to-object distance* (SOD) provides the distance between the X-ray point source and the linac isocenter. By symmetry of the linac and ExacTrac machines, both tube-detector pairs share equivalent SIDs and SODs. For each tube-detector pair, define the *central beamline* as the beamlet path beginning at the X-ray point source and ending at the detector center. The angle between both central beamlines is referred to as the *crossing angle* φ. Last, define the *oblique plane* as the unique plane containing both central beamlines. The angle of incline between the oblique plane and the floor is referred to as the *oblique plane angle* θ.

1.3 DRR rendering from first principles

Suppose the radiographic detectors are divided into uniform grids of pixels with some predetermined pixel spacing. In order to render a radiographic projection, we may conceptually divide the X-ray beam into distinct beamlets, one for each pixel. The radiological path of a beamlet is the distance that the radiation travels weighted by linear attenuation coefficients of the materials that it passes through. To compute radiological paths, we wrote a Siddon-Jacobs ray tracing program following the methods in [90, 42]. See Appendix A for details

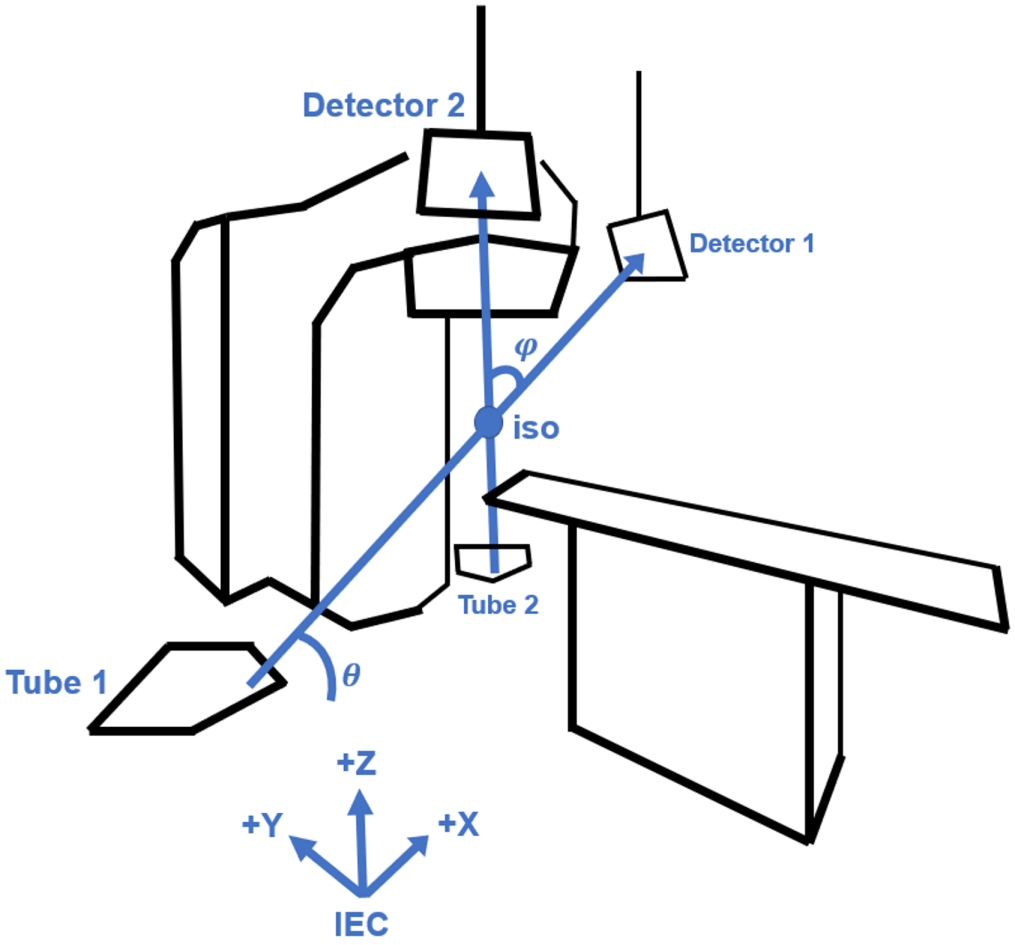


Figure 1.1: Sketch of an ExacTrac IGRT system installed in a radiation oncology treatment room. The X-ray central beamlines intersect at the linac isocenter.

on the Siddon-Jacobs algorithm.

Tracing a beamlet through a volumetric CT (more generally, any line segment through a rectangular cuboid) requires not only knowledge of all voxel coordinates, but also a parameterization of the beamlet. It is assumed that each voxel is small enough to represent a homogeneous tissue with a well-defined linear attenuation coefficient. Let us systematically explain both steps, starting with the CT voxel coordinates. Throughout, we will fix a background IEC coordinate system (x, y, z) where the isocenter coincides with the origin.

1.3.1 DICOM affine formula

The DICOM Standard includes an image plane module that defines attributes for pixel spacing, image orientation, image position, slice thickness, and slice location [22]. Following the DICOM Standard, we construct a homogeneous matrix that converts voxel indices into physical RCS locations,

$$\begin{pmatrix} P_x \\ P_y \\ P_z \\ 1 \end{pmatrix} = \begin{pmatrix} X_x\Delta_i & Y_x\Delta_j & Z_x\Delta_k & S_x \\ X_y\Delta_i & Y_y\Delta_j & Z_y\Delta_k & S_y \\ X_z\Delta_i & Y_z\Delta_j & Z_z\Delta_k & S_z \\ 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} i \\ j \\ k \\ 1 \end{pmatrix}, \tag{1.2}$$

where

- $\mathbf{P} = (P_x, P_y, P_z)$ is the RCS location of voxel (row, column, slice) $= (i, j, k)$ [mm];
- $\mathbf{S} = (S_x, S_y, S_z)$ is the RCS location of the front upper-left voxel $(0, 0, 0)$ [mm];
- $\mathbf{X} = (X_x, X_y, X_z)$ is the row direction cosines vector;
- $\mathbf{Y} = (Y_x, Y_y, Y_z)$ is the column direction cosines vector;
- $\mathbf{Z} = (Z_x, Z_y, Z_z)$ is the slice direction cosines vector;
- Δ_i is the row pixel spacing [mm/pix];
- Δ_j is the column pixel spacing [mm/pix];
- Δ_k is the slice thickness [mm/slice].

Equation 1.2 is known as the *DICOM affine formula*. Voxel indexing conventionally starts at zero. The $\mathbf{X}$ and $\mathbf{Y}$ direction cosines vectors are found in the image orientation attribute, whereas the $\mathbf{Z}$ direction cosines vector is found by taking the difference of image position attributes $\mathbf{S}$ between DICOM files for two adjacent CT slices.

It is important to note that the image position attribute $\mathbf{S}$ is defined with respect to the center of initial voxel $(0, 0, 0)$. However, our implementation of the Siddon-Jacobs ray tracing algorithm expects us to find the locations of the volume borders. We anticipate that the front upper-left corner of the CT volume should be provided by the DICOM affine matrix in Equation 1.2 acting on indices $(i, j, k) = (-0.5, -0.5, -0.5)$.

Next, we need to subtract the RCS location of the isocenter. To achieve this, we referenced the associated radiotherapy plan (RT Plan) DICOM file, specifically the isocenter position attribute found in the RT beams module of the RT Plan [24]. Any 6D translational fusion correction is added here, if present.

The opposite corner (back lower-right corner) of the CT volume is found by simply adding the vector $(N_x\Delta_x, N_y\Delta_y, N_z\Delta_z)$, where $\mathbf{N} = (N_x, N_y, N_z) \in \mathbb{N}^3$ are the matrix dimensions. Similarly, any parallel plane dividing the volume can be found by substituting $\mathbf{N}$ with indices (i, j, k). A straightforward application of Equation 1.1 converts volume border locations to an IEC reference frame.

1.3.2 Beamlet parameterization

The second step is to derive expressions for the coordinates of all beamlet endpoints, namely all detector pixels coordinates together with the two X-ray focal points. By symmetry, it suffices to analyze a single tube-detector pair, our convention being the tube residing on the left side of the linac and its detector on the opposite side. Then the central beamline has direction

$$\mathbf{d} = \mathbf{R}_{\mathbf{x}}\left(\frac{\pi}{2} - \theta\right)\begin{pmatrix} 1 \\ 0 \\ \tan(\frac{\pi-\varphi}{2}) \end{pmatrix}, \tag{1.3}$$

where $\mathbf{R}_{\boldsymbol{\alpha}}(\beta)$ denotes a counterclockwise rotation matrix about vector $\boldsymbol{\alpha}$ by angle β. Normalizing to obtain a unit direction $\widehat{\mathbf{d}}$ effectively parameterizes the beamlet by path length.

Thus, the focal point $\mathbf{f}$ satisfies

$$\mathbf{0} = \mathbf{f} + \mathrm{SOD} \cdot \widehat{\mathbf{d}}$$

and the detector center $\mathbf{c}$ satisfies

$$\mathbf{c} = \mathbf{f} + \mathrm{SID} \cdot \widehat{\mathbf{d}}.$$

All beamlets start at location $\mathbf{f}$.

Assume that the central beamline is orthogonal to the plane of the flat-panel receptive field. Let $\mathbf{m}$ and $\mathbf{n}$ denote the detector coordinate vectors in the horizontal and vertical directions, respectively (cf., direction cosines of the CT volume). In our methodology, expressions for $\mathbf{m}$ and $\mathbf{n}$ in terms of IEC coordinates involve an intermediate vector

$$\mathbf{d}_{xy} = \mathbf{R}_{\mathbf{z}}\left(\frac{\pi}{2}\right) \mathbf{P}_{xy}\, \mathbf{d}, \tag{1.4}$$

where

$$\mathbf{P}_{xy} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{pmatrix}$$

is the natural projection matrix onto the plane $z = 0$. The desired vertical component is then given by

$$\mathbf{n} = \mathbf{R}_{\mathbf{d}_{xy}}\left(\frac{\pi}{2}\right) \mathbf{d}. \tag{1.5}$$

To construct a right-handed detector coordinate system, we can immediately set

$$\widehat{\mathbf{m}} = \widehat{\mathbf{n}} \times \widehat{\mathbf{d}}. \tag{1.6}$$

Suppose a 6D rotational fusion correction $\boldsymbol{\psi} = (\psi_x, \psi_y, \psi_z)$ is present. Then all the vectors $\mathbf{f}, \mathbf{c}, \widehat{\mathbf{m}}$, and $\widehat{\mathbf{n}}$ need to be transformed by a rotation operator $\mathbf{R}_{\boldsymbol{\psi}}$ associated with $\boldsymbol{\psi}$.

Since rotation operators are generally non-commutative, we specify an ordering convention

$$\mathbf{R}_{\boldsymbol{\psi}} \equiv \mathbf{R}_{\mathbf{y}}(\psi_y)\mathbf{R}_{\mathbf{z}}(\psi_z)\mathbf{R}_{\mathbf{x}}(\psi_x) \tag{1.7}$$

consistent with the ExacTrac system.

For simplicity, consider detector pixel indices (m, n) relative to the center pixel located at $\mathbf{c}$. Let Δ_m and Δ_n be the detector pixel spacing resolutions. Then a beamlet incident on pixel (m, n) ends at location $\mathbf{c} + m\Delta_m\widehat{\mathbf{m}} + n\Delta_n\widehat{\mathbf{n}}$.

All relevant parameters for ray tracing have now been determined. Therefore, we have shown that using in-room measurements only, the minimal set of parameters for stereoscopic DRR rendering are as follows:

- Source-to-image distance (SID);
- Source-to-object distance (SOD);
- Oblique plane angle θ;
- Central beamline crossing angle φ.

An implicit assumption is the orientation of flat-panel direction cosines with respect to the central beamlines, which is provided by Equations 1.4, 1.5, and 1.6. Using SID and SOD measurements from the ExacTrac clinical user guide [6], as well as θ and φ measurements from [52], we managed to generate DRRs that visually replicated the ground-truth ExacTrac DRRs to a high degree of geometrical accuracy. Furthermore, the methodology presented in the next section *perfectly* replicates the geometry of the ExacTrac system.

1.4 ExacTrac configurations

For every imaged subject, the ExacTrac system records data including the CT volume, DRRs for each treatment fraction, treatment logs, and configuration logs. A treatment log

includes not only information from the treatment planning system (TPS), but also 6D X-ray correctional shifts for alignment verification. A configuration log, saved in initialization (INI) file format, records a wealth of settings.

One such configuration setting characterizes the projectional geometry for DRR rendering. We search for the "FlatPanel" section of the INI file and extract the values of the keys "MLinToFlat1" and "MLinToFlat2", corresponding to detectors 1 and 2, respectively. These values are represented as comma-separated lists of numbers. It turns out that the first number is always zero, hence it should be disregarded. In order to interpret the remaining numbers as a linear transformation, we rearrange them into a 3×4 matrix. Row-major order is presumed, so that each row is completely filled before proceeding to the next row. Henceforth, we refer to this constructed matrix as the *renderer matrix*, as it encapsulates all the information for DRR generation. The renderer matrices for both detectors are expressed in IEC coordinates.

Given CT slices stored as DICOM files, we used a built-in ITK function that imports the CT dataset into a nearly raw raster data (NRRD) file format [75]. After loading the CT volume, we subtracted out the location of the isocenter found in the RT Plan DICOM file, as described earlier. This translation effectively moves the CT isocenter to the linac isocenter origin. Suppose that an abstract DRR renderer camera is initialized on the origin, facing the IEC $+y$ direction. We seek linear transformations that place the camera on the X-ray focal points, oriented for 2D flat-panel viewing. To achieve this goal, we turn to projectional geometry and the notion of perspective projections [1]. It will be helpful to regard linear operators as 4×4 matrices acting on real projective space $\mathbb{P}^3$.

1.4.1 Perspective projections

In this subsection, we build a projective linear operator $\mathbf{M} \in \mathrm{PGL}(4, \mathbb{R})$ that represents DRR rendering [35]. Our general decomposition is

$$\mathbf{M} = \mathbf{S}_{\mathrm{NDC}}\mathbf{S}_{\mathrm{det}}\mathbf{PRT}. \tag{1.8}$$

The first (rightmost) matrix is a translation $\mathbf{T}$, which sends the abstract renderer camera from the origin to an X-ray tube focal point. Given a translation vector $\mathbf{f} = (f_x, f_y, f_z)$, we have

$$\mathbf{T} = \begin{pmatrix} 1 & 0 & 0 & -f_x \\ 0 & 1 & 0 & -f_y \\ 0 & 0 & 1 & -f_z \\ 0 & 0 & 0 & 1 \end{pmatrix}. \tag{1.9}$$

In order to aim the camera at the overhead panels, we need to apply a rotation operator $\mathbf{R}$, which is a homogenized version of the transposed flat-panel direction cosines $\widehat{\mathbf{m}} = (m_x, m_y, m_z)$, $\widehat{\mathbf{n}} = (n_x, n_y, n_z)$, and $\widehat{\mathbf{d}} = (d_x, d_y, d_z)$. We have

$$\mathbf{R} = \begin{pmatrix} m_x & m_y & m_z & 0 \\ n_x & n_y & n_z & 0 \\ d_x & d_y & d_z & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}. \tag{1.10}$$

The product of the rotation and translation operators is known as the localization matrix $\mathbf{L} = \mathbf{RT}$.

Once the camera translations and rotations are applied, we may express a projectional

calibration matrix $\mathbf{P}$ in its standard form

$$\mathbf{P} = \begin{pmatrix} \text{SID} & 0 & 0 & 0 \\ 0 & \text{SID} & 0 & 0 \\ 0 & 0 & \text{SID} & 0 \\ 0 & 0 & 1 & 0 \end{pmatrix}. \tag{1.11}$$

A detector scaling matrix $\mathbf{S}_{\text{det}}$ is required to convert camera coordinates into detector coordinates. In our ExacTrac system, the detectors have a standard 200×200 mm^2 FOV and 512×512 pix^2 resolution after imposing 2×2 binning. It follows that the pixel spacings are $\Delta_m = \Delta_n = 0.3906 \, \frac{\text{mm}}{\text{pix}}$. If the imaging data does not undergo 2×2 binning, then the pixel spacings should be divided by two accordingly. We also have parameters p_m and p_n that determine where the upper-left corner of the DRR is located with respect to the detector center. The general form of a detector scaling matrix is

$$\mathbf{S}_{\text{det}} = \begin{pmatrix} \Delta_m^{-1} & 0 & 0 & p_m \\ 0 & \Delta_n^{-1} & 0 & p_n \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}. \tag{1.12}$$

Finally, it is often convenient that the renderer convert to normalized device coordinates (NDC), for the purpose of mapping a frustum onto the unit cube. In DRR rendering, we impose a simplified NDC scaling matrix $\mathbf{S}_{\text{NDC}}$ for a frustum that depends on the detector FOV and SID. The desired matrix is

$$\mathbf{S}_{\text{NDC}} = \begin{pmatrix} \frac{2}{W} & 0 & 0 & -1 \\ 0 & \frac{2}{H} & 0 & -1 \\ 0 & 0 & \frac{1}{\text{SID}} & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \tag{1.13}$$

where detector width W and height H are expressed in units of pixels.

When both scaling matrices are applied to the standard calibration matrix $\mathbf{P}$, we arrive at

$$\mathbf{P}_{\text{NDC}} = \mathbf{S}_{\text{NDC}}\mathbf{S}_{\text{det}}\mathbf{P} = \begin{pmatrix} \frac{2\cdot\text{SID}}{W\Delta_m} & 0 & \frac{2p_m}{W} - 1 & 0 \\ 0 & \frac{2\cdot\text{SID}}{H\Delta_n} & \frac{2p_n}{H} - 1 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 \end{pmatrix}. \tag{1.14}$$

Observe that the third and fourth rows are identical, which we expect since depth information is lost in two-dimensional projections. Thus, the fourth row may be removed to yield a 3×4 matrix, agreeing with the renderer matrices constructed from the ExacTrac configuration settings.

One last technical issue in rendering is referred to as the w_c *condition*, which states that homogeneous coordinates of a rendered object must be positive. If not, then the imaging plane and object points lie on opposite sides of the camera. A simple and efficient way to verify the w_c condition is to act on the linac isocenter with $\mathbf{M}$ and check the sign of the homogeneous element. In order to always satisfy the w_c condition, one ought to multiply $\mathbf{M}$ by the sign of its last entry.

1.4.2 ExacTrac renderer matrices

We now have enough information to extract the necessary geometric quantities for DRR generation from the provided ExacTrac renderer matrices $\mathbf{M} = \mathbf{M}_{3\times4}$. It turns out that in our system, $\mathbf{M}$ is already denormalized, so that $\mathbf{S}_{\text{NDC}}$ may be removed from our analysis. Consider the 3×3 submatrix of $\mathbf{M}$ consisting of its first three rows and first three columns. This submatrix is easily seen to be a product of projectional and rotational transformations.

Any real square matrix X admits a factorization $X = RQ$, where R denotes an upper-triangular (projection) matrix and Q denotes an orthogonal (rotation) matrix. In Appendix B, we explain how to perform such an RQ decomposition via a Gram-Schmidt procedure.

For further information on the closely-related, but more familiar, QR decomposition and related numerical methods, see [8].

Neglecting the $\mathbf{S}_{\text{NDC}}$ factor, the detector-scaled projection is

$$\mathbf{P}_{\text{det}} = \mathbf{S}_{\text{det}}\mathbf{P} = \begin{pmatrix} \frac{\text{SID}}{\Delta_m} & 0 & p_m & 0 \\ 0 & \frac{\text{SID}}{\Delta_n} & p_n & 0 \\ 0 & 0 & \text{SID} & 0 \\ 0 & 0 & 1 & 0 \end{pmatrix}, \tag{1.15}$$

which corresponds to 'R' in the RQ decomposition. It follows that $\text{SID} = \mathbf{P}_{\text{det}}^{(1,1)}\Delta_m$ mm, $p_m = \mathbf{P}_{\text{det}}^{(1,3)}$ pix, and $p_n = \mathbf{P}_{\text{det}}^{(2,3)}$ pix.

Focal point locations are derived from the full localization operator $\mathbf{L} = \mathbf{L}_{4\times4}$, which takes the form

$$\mathbf{L} = \begin{pmatrix} m_x & m_y & m_z & -\mathbf{f}\cdot\widehat{\mathbf{m}} \\ n_x & n_y & n_z & -\mathbf{f}\cdot\widehat{\mathbf{n}} \\ d_x & d_y & d_z & -\mathbf{f}\cdot\widehat{\mathbf{d}} \\ 0 & 0 & 0 & 1 \end{pmatrix}. \tag{1.16}$$

Besides the homogeneous element, the fourth column of $\mathbf{L}$ is precisely the fourth column of $\mathbf{M}$. Meanwhile, the 3×3 submatrix of $\mathbf{L}$ (equivalently, of the rotation operator $\mathbf{R}$), given by removing the fourth row and fourth column, corresponds to 'Q' in the RQ decomposition. Therefore, $\mathbf{f}$ is determined by inverting a simple system of linear equations, which may be carried out with a linear algebra computer package.

Finally, our ITK ray tracing implementation requires inputting the DRR image origin $\mathbf{p}$ in physical units, which we found by computing

$$\mathbf{p} = \mathbf{f} + \mathbf{L}_{3\times3}^{\text{T}} \begin{pmatrix} -p_m\Delta_m \\ -p_n\Delta_n \\ \text{SID} \end{pmatrix}. \tag{1.17}$$

These geometric quantities extracted from $\mathbf{M}$ may be sent into either our DRR generator written from first principles or our DRR generator written with modified ITK functions. Note that our ITK program uses RCS coordinates, in which case we have to apply the inverse of the change-of-basis in Equation 1.1.

1.5 Results

We evaluated our methodology on data from ExacTrac versions 6.1.1, 6.2.1, and 6.2.3, which comprise the versions in clinical use at our institution since 2014. Our data consisted of one SBRT phantom and two example patients, each imaged in treatment positions so that their 6D fusion parameters would be known and readily acquired from ExacTrac treatment logs.

Figure 1.2 shows the output of our DRR generators for the SBRT phantom and one of the two patients in our analysis. Note that only one flat-panel imager is presented here, although stereoscopic DRR pairs were always created. For computing radiological paths with line integrals, we assumed a linear relationship between CT numbers and voxel intensities (the actual mapping displayed on the ExacTrac workstation is proprietary to Brainlab). We fixed a minimum CT number threshold of 100 HU, whereby any material beneath this threshold was treated as air and did not contribute to beam attenuation. This cutoff helped to accentuate the bones by reducing soft tissue contrast. We also used a mass energy absorption coefficient of water subject to a 70 keV X-ray beam, namely [49]

$$\left(\frac{\mu_{\text{en}}}{\rho}\right)_{\text{water}} = 2.8935 \times 10^{-2}\ \frac{\text{cm}^2}{\text{g}}, \quad \rho_{\text{water}} = 1\ \frac{\text{g}}{\text{cm}^3}.$$

1.5.1 Rigid registrations

First, we performed rigid image registrations using the *elastix* software package [62] for an independent assessment of systematic offsets between our stereoscopic DRRs and the treatment-position reference DRRs from the ExacTrac system. Rigid registrations were con-

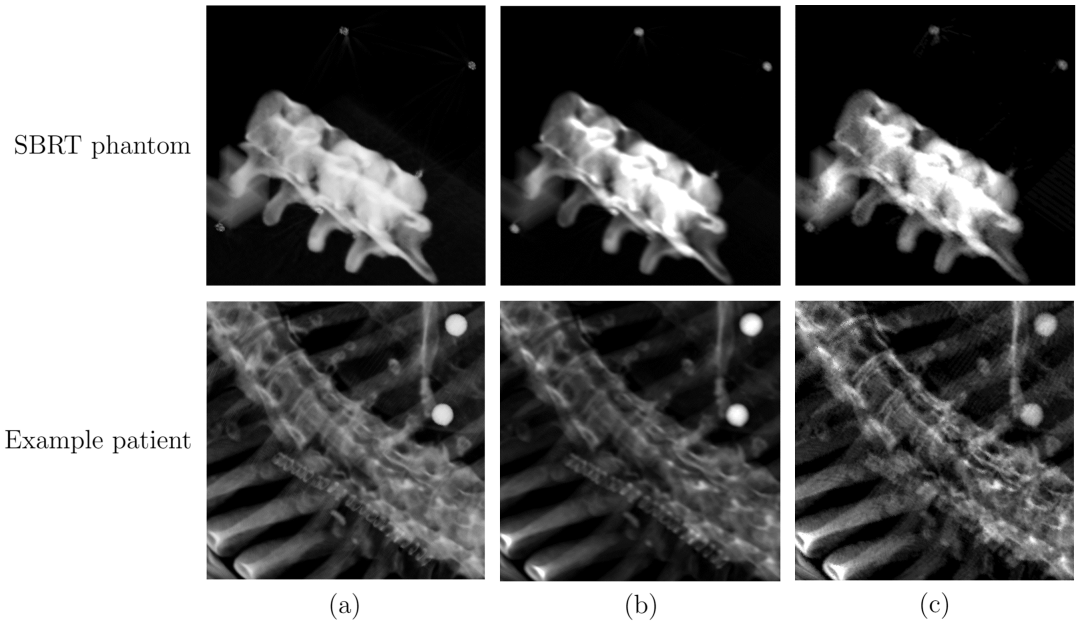


Figure 1.2: DRR generator output for an SBRT phantom and a patient treated at our clinic. (a) Ground-truth DRRs produced by the ExacTrac system. (b) DRRs generated by our program written with modified ITK functions. (c) DRRs generated by our program written from scratch.

figured with a multi-resolution Euler transform, which depends on three global parameters: vertical translations t_y, horizontal translations t_x, and counterclockwise rotations about the image center θ_z. Results on the optimized *elastix* rigid registration parameters are presented in Table 1.1. We found that rigid rotations and translations are negligible, with $|\overline{\theta_z}| = 0$ rad, $|\overline{t_x}| = 0.35$ mm, and $|\overline{t_y}| = 0.18$ mm.

1.5.2 Landmark analysis

Second, we performed a landmark analysis to quantify the geometric accuracy of our generators. For every datapoint, we displayed the corresponding DRRs side-by-side for comparison. Across the three images, we selected a total of 20 distinct landmark points-of-interest (POIs) placed at readily identifiable regions, such as bright spots or anatomical boundaries, at our discretion. We were cognizant to choose POIs that were approximately uniformly spaced about the imaging FOV and characterized the full extent of the imaged anatomy. Then, we

Parameter	SBRT phantom				Patient A				Patient B			
	1-pre	2-pre	1-post	2-post	1-pre	2-pre	1-post	2-post	1-pre	2-pre	1-post	2-post
θ_z [rad]	0	0	0	0	0	0	0	0	0	-0.01	0	0
t_x [mm]	0.19	0.28	0.40	0.44	0.06	0.74	0.19	0.59	0.48	0.27	0.54	0.24
t_y [mm]	0.17	0.09	0.07	0.25	-0.08	-0.05	-0.36	-0.14	0.16	0.07	0.17	0.16

(a)

Parameter	SBRT phantom				Patient A				Patient B			
	1-pre	2-pre	1-post	2-post	1-pre	2-pre	1-post	2-post	1-pre	2-pre	1-post	2-post
θ_z [rad]	0	0	0	0	0.01	0	0	0	0	0	0	0
t_x [mm]	0.27	0.15	0.38	0.28	0.09	0.68	0.35	0.46	0.31	0.40	0.47	0.27
t_y [mm]	0.34	0.23	0.36	0.41	0.12	-0.15	-0.19	-0.08	0.28	-0.05	0.23	0.17

(b)

Table 1.1: Rigid registration parameters with respect to the ground-truth stereoscopic DRRs produced by the ExacTrac system, before and after 6D fusion corrections. (a) DRRs generated by our program written with modified ITK functions. (b) DRRs generated by our program written from scratch.

computed displacements between corresponding POIs in physical units. Results on the mean μ and standard deviation σ of all landmark displacements are presented in Table 1.2. We found that the average mean displacement is 1.15 mm, which is certainly within a reasonable margin of error for manual POI selection on a voxelized imaging grid.

1.6 Discussion

Perspective projections are elegantly described as a sequential application of linear operators on $\mathbb{P}^3$. Given an object in an ambient space with fixed origin, the goal is to find an operator that sends object points to an imaging plane along pinhole camera rays. First, we localized the camera position through translation and rotation matrices. Then, we applied a standard calibration matrix that projects 3D object points to a 2D imaging plane. A detector scaling operator was required to achieve a desired radiographic resolution and upper-left corner image location. In general, rendering is restricted to a frustum, so that points outside the frustum are clipped. An NDC conversion mapped the frustum to a unit cube. Dividing out by the homogeneous coordinates yielded DRR pixel locations.

	SBRT phantom				Patient A				Patient B			
Statistic	1-pre	2-pre	1-post	2-post	1-pre	2-pre	1-post	2-post	1-pre	2-pre	1-post	2-post
μ [mm]	0.76	0.87	1.03	0.91	0.87	0.90	1.01	1.10	0.87	0.86	1.07	0.83
σ [mm]	0.58	0.79	0.74	0.50	0.53	0.62	0.67	0.68	0.64	0.44	0.53	0.53

(a)

	SBRT phantom				Patient A				Patient B			
Statistic	1-pre	2-pre	1-post	2-post	1-pre	2-pre	1-post	2-post	1-pre	2-pre	1-post	2-post
μ [mm]	1.42	1.16	1.30	1.60	1.61	1.46	1.60	1.48	1.62	1.15	1.10	1.00
σ [mm]	0.75	0.60	0.77	0.91	0.80	0.84	0.61	0.63	1.25	0.55	0.57	0.68

(b)

Table 1.2: Landmark analysis of the geometric accuracy with respect to the ground-truth stereoscopic DRRs produced by the ExacTrac system, before and after 6D fusion corrections. Displacement statistics are computed over 20 manually-selected landmark points. (a) DRRs generated by our program written with modified ITK functions. (b) DRRs generated by our program written from scratch.

Our DRR algorithms were not designed to handle imaging artifacts such as streaking. Although it is preferable to view artifact-free DRRs, artifacts are acceptable when developing models for misalignment detection research, so long as their presence or absence is consistent across training and inference data.

Finally, note that radiation therapists will often perform multiple X-ray verifications prior to irradiating a patient. In such cases, the applied 6D fusions should be input sequentially into the DRR generator. Large rotations of a patient (i.e., couch kicks) such as 90° or 180° may be unreported in the treatment logs, hence some experimentation may be required by clinical researchers to replicate the desired DRRs.

1.7 Conclusion

Offline DRR production is important in IGRT analysis, particularly for studying the effects of arbitrary 6D couch shifts. One potential application of the research presented in this chapter is toward building mathematical models for radiotherapy patient positioning and error detection on the ExacTrac system. In the next chapter, we proceed to investigate

automated error detection methods about the vertebral column, focusing instead on on-board orthogonal planar imaging.

Algorithms for projectional rendering were derived from first principles and were shown to successfully reproduce ExacTrac stereoscopic DRRs. Our landmark analysis and rigid registration results demonstrate an excellent level of spatial accuracy between the DRRs generated by our algorithms and the known DRRs with 6D couch corrections. All parameters necessary for ExacTrac stereoscopic DRR rendering are stored in the ExacTrac configuration logs and treatment files within a radiotherapy treatment session.

Chapter 2

Automated patient setup error detection

2.1 Introduction

On-board image guidance has become the modern standard for patient alignment during high-precision radiation therapy treatments. Image-guided radiotherapy (IGRT) is most frequently performed with either cone-beam computed tomography (CBCT) scans or projectional radiographic imaging [78]. In the former case, three-dimensional CBCTs are registered to the planning CT. In the latter case, bony anatomy from treatment radiographs are registered to bony anatomy from digitally reconstructed radiographs (DRRs) derived from the planning CT. The implementation of image-guidance has improved tumor control and reduced normal tissue toxicity, especially among high-risk populations [7].

Gantry-mounted radiographic imagers for IGRT reduce the frequency of treatment errors compared to non-IGRT approaches [87]. However, it is well-known that patient localization errors still occur during intensity-modulated radiotherapy (IMRT) and three-dimensional conformal radiotherapy (3D-CRT) procedures [34, 47, 50, 74, 82, 92, 99]. A recent review by the Radiation Oncology Incident Learning System (RO-ILS) of nearly 400 high-priority

inter-institutional treatment events found that about 19% of events had either incorrect setup instructions or incorrect couch shifts [37].

A report by the French Nuclear Safety Authority (ASN) illustrated the significance of vertebral body setup errors in IGRT treatments. Among 40 such IGRT incidents reviewed, 36 originated from planar radiographs, as compared with 4 from CBCT scans. The primary factor resulting in vertebral body setup errors was difficulty in differentiating between adjacent vertebrae. Contributing factors included poor image quality, longitudinal matching using non-discriminating landmarks, and excessively small collimation [4]. Related findings have been reported in the context of spinal surgeries [36]. In modern radiotherapy departments, extensive effort is directed toward reducing the probability of such IGRT errors through methods like identification of workflow failure modes and adherence to safety guidelines and checklists.

Computer automation and interlocks are widely believed to be the most effective methods of error prevention [50]. Previous studies from our medical physics research group have shown that image similarity measures calculated between IGRT images and their corresponding setup images can distinguish between patients correctly aligned and patients incorrectly aligned in a variety of treatment sites [64, 51]. These approaches involved machine learning models with hand-curated features. More recently, artificial intelligence (AI) based on deep learning has exhibited state-of-the-art performance in automated quality assurance and error detection in radiation oncology [59, 71, 81, 84].

Convolutional neural networks (CNNs) are a subset of deep learning architectures that offer highly successful models for medical image classification and segmentation [9, 18, 69]. Our medical physics research group recently developed CNN-based models for detecting misalignments to adjacent vertebrae (off-by-one errors) with CBCT scans [72] and ExacTrac stereoscopic images [83]. According to a current practice patterns survey, planar kV and MV radiographic imaging represent the second and third most frequently used IGRT technologies after CBCT, being used by 22 – 67% of institutions depending on disease site [78]. Compared

to CBCT images, planar kV and MV images provide less soft-tissue information but are acquired more rapidly, hence they are more suitable for palliative treatments. The ExacTrac stereoscopic system records oblique image pairs with a narrower field-of-view (FOV) of 10×10 cm^2. By comparison, the IGRT dataset considered in this study consists of orthogonally-paired anterior-posterior (AP) and lateral (LAT) radiographs produced from gantry-mounted imagers with a larger FOV of 30×40 cm^2. The study on ExacTrac treated each radiograph from a stereoscopic image pair completely separately, and our research group is making progress to improve the stereoscopic misalignment accuracy by incorporating our offline DRR generator from Chapter 1.

In this chapter, we develop CNN-based deep learning models for detecting off-by-one vertebral body setup errors in orthogonal kV or MV radiographic images. Due to relatively lower image contrast, larger treatment FOV, and lower alignment precision, the detection of off-by-one misalignments in radiographs is a more challenging classification problem and presents a significant gap in the literature on preventing IGRT errors with deep learning. We develop and test our error detection models on data from thoracic and abdominal radiotherapy treatments performed at our institution between 2011 – 2021. Our investigation not only involves training models on AP and LAT orientations separately, but also training a model that combines information from both projection orientations simultaneously. Our goal is to increase patient safety by developing these error detection models as potential automated interlocks in the radiotherapy workflow. The objective is to achieve at least a 95% true positive rate of error detection given a false positive rate of less than 5%. Finally, we perform a retrospective error analysis of our institutional dataset by examining false positives in the training data.

2.2 Data collection

The technical standard for storing and transferring medical information is Digital Imaging and Communications in Medicine (DICOM) [3]. The DICOM Standard defines not only a file format, but also a network communication protocol built over the Internet protocol (TCP/IP). In order to efficiently acquire DICOM data from the clinical database at our institution, we used the open-source *pynetdicom* package that implements the DICOM protocol [33].

The DICOM Standard organizes medical information according to the hierarchy [32]

$$\text{PATIENT} \supset \text{STUDY} \supset \text{SERIES} \supset \text{OBJECT}.$$

Every radiotherapy patient is identified by their medical record number (MRN). A patient may undergo multiple studies, i.e., courses of treatment (e.g., treatment of independent tumors or tumor relapse). Each study has a unique identifier (UID). Within a study, there exist multiple series' comprising the radiotherapy data, each having their own series UID. An object represents an individual DICOM file and is not always equivalent to its series (e.g., a volumetric CT scan is characterized by one series with one series UID, but its planar image slices are stored as separate objects with their own object UIDs).

Our institution's radiation oncology clinical database is maintained by the ARIA Radiation Oncology Information System (Varian Medical Systems, Inc., Palo Alto, CA). We queried ARIA for patients treated at our clinic with study dates between the years 2011 – 2021. Given a patient's MRN, a list of studies and series' may be acquired. The relevant series modalities that we queried for this investigation are:

- *RTPlan* for treatment planning information [30];
- *RTImage* for treatment radiographs and setup DRRs [25];
- *RTReg* for spatial registration between the treatment radiograph and setup DRR

frames of reference [21].

When an RTPlan series modality was encountered during the query, we applied plan name filters for the following anatomical sites:

- thoracic and lumbar vertebrae (specifically, T2 – L2);
- esophagus;
- ribs;
- lungs, including left upper lobe (LUL), left lower lobe (LLL), right upper lobe (RUL), and right lower lobe (RLL);
- abdomen, including stomach / mid-stomach and pancreas.

These filters were facilitated by our institution's rigorous adherence to an anatomy-based plan labeling nomenclature.

After obtaining a list of all RTImage and RTReg objects within the same studies as the filtered RTPlans, an association between treatment radiographs and their setup DRRs was established. Henceforth, we will colloquially refer to treatment radiographs as X-rays. For on-board imaging acquired with only one orientation (typically AP), there should be two RTImage object UIDs contained in the RTReg, one for the X-ray and one for the DRR. For on-board imaging acquired with both AP and LAT orientations, there should be four RTImage object UIDs accordingly. In general, the order in which these UIDs are itemized in an RTReg object is not fixed. Thus, we relied on matching RTImage labels with case-insensitive substrings. Specifically, the substrings "DRR" or "setup" indicated that the RTImage belonged to a DRR, and otherwise the RTImage belonged to an X-ray. If the substrings "kV" or "MV" were not found in an X-ray RTImage label, then we categorized the energy level based on the peak kilovoltage output (kVp) of the X-ray generator from the RTImage exposure sequence. The orientations of the images were automatically determined by filtering for the following case-insensitive substrings:

- Anteroposterior (AP) orientation: "AP", "PA", "_0_", "_180_";
- Lateral (LAT) orientation: "LAT", "_90_", "_270_".

2.2.1 Data selection

In order to ensure appropriate data selection, the retrieved DRR and X-ray images were manually reviewed. We systematically removed several cases from the dataset due to the following circumstances, listed in order of prevalence:

- FOV was lateral to spinal column;
- FOV was centered about sacrum or coccyx;
- FOV was centered about cranium or cervical vertebrae;
- Patient was positioned feet-first supine (FFS) or feet-first prone (FFP) instead of head-first supine (HFS) or head-first prone (HFP);
- Gantry angle deviated by at least 5° from integer multiples of 90°;
- Contrast was too low to interpret the image.

The remaining datapoints were subdivided according to X-ray energy level (kV, MV, or both) and image orientation (AP, LAT, or both). This subdivision allowed for model training, validation, and testing to be performed on medical images with similar features. The final dataset sizes are listed in Table 2.1. Besides a relatively small handful of exceptions, every kV AP radiograph was paired with an orthogonal kV LAT radiograph. We did not find any orthogonal MV pairings. About 20% of the MV images were paired with kV images, which we sorted into a dual-energy dataset. However, the amount of dual-energy datapoints was insufficient for model training, so this category was ignored in our analysis.

Dataset	Number of images	Number of patients
kV AP – LAT	2252	254
MV AP	1629	250
dual-energy AP – LAT	326	47

Table 2.1: On-board planar spinal imaging dataset summary. Datasets are categorized by image orientation and X-ray energy level. The datasets in the first two rows were used for error detection model development.

2.3 Clinically aligned DRRs

We developed a straightforward algorithm to transform a CT simulation DRR onto the space of a particular treatment fraction radiograph according to DICOM metadata. Registration transformations between frame of reference UIDs are located in the RTReg objects. These transformations are stored as sequences of 16 numbers that may be readily converted into 4×4 matrices using a row-major ordering. Given an RTReg object, one of the two transformation matrices is the identity matrix, implying that its attached frame of reference UID belongs to the primary RTImage. In general, the other transformation matrix is nontrivial, and its attached frame of reference UID belongs to the secondary RTImage. We noticed that primary RTImages always represented treatment radiographs (i.e., X-rays) and secondary RTImages always represented CT simulation DRRs.

A nontrivial transformation matrix $\mathbf{M}$ acts on its associated DRR isocenter $\mathbf{I}$, expressed in homogeneous HFS coordinates, as $\mathbf{I}' = \mathbf{MI}$. Isocenter correction factors in the AP and LAT orientations are given by

$$\Delta\mathbf{I}_{\mathrm{AP}} = \begin{pmatrix} \sigma(I'_x - I_x) \\ I'_z - I_z \end{pmatrix}, \quad \Delta\mathbf{I}_{\mathrm{LAT}} = \begin{pmatrix} \sigma(I'_y - I_y) \\ I'_z - I_z \end{pmatrix}.$$

Here σ is a parity signature for the gantry angle θ found in the RTImage exposure sequence. For AP-oriented images, $\sigma = +1$ when $\theta = 0^\circ$ and $\sigma = -1$ when $\theta = 180^\circ$. For LAT-oriented images, $\sigma = +1$ when $\theta = 90^\circ$ and $\sigma = -1$ when $\theta = 270^\circ$. If the patient was positioned

HFP instead of HFS, then simply introduce another negative sign $\sigma \mapsto -\sigma$.

Next, an in-plane X-ray image receptor translation correction factor was applied to the primary X-ray image position. Then an SID correction factor, namely the ratio between primary and secondary SIDs, was applied to the secondary DRR image position and pixel spacing.

Finally, we resampled the DRR onto the X-ray grid, following Equation 1.2. We defaulted to using bilinear interpolation for computational efficiency, although bicubic or higher-order interpolations are certainly acceptable.

2.4 Semi-automated misalignment generation

In order to train an AI model to classify off-by-one vertebral body setup errors, it is necessary to provide sufficient data with *error* (true positive) and *no error* (true negative) class labels. We assumed that all images retrieved from our clinical database were correctly aligned, i.e., belonged to the *no error* class. Therefore, we needed to devise a methodology to artificially misalign the retrieved treatment radiographs (i.e., X-rays) with respect to their DRRs by one vertebral body. Such artificial misalignments should accurately simulate potential setup errors made by radiation therapists. In a typical clinical workflow, radiation therapists view a blended overlay of the DRR and X-ray in a single window and drag the DRR until the vertebral bodies are visually aligned, thus determining a couch shift to be applied.

Our semi-automated off-by-one misalignment generation methodology is explained below. Note that we did not re-render the misaligned DRRs because the angular divergences were observed to be within 2°, which we assumed were negligible compared to the rotational alignment precision used for palliative treatments.

2.4.1 Superior-inferior and left-right misalignments

Given an AP DRR and X-ray image pair, we first performed approximate misalignments at off-by-one vertebral body locations. Landmark points of interest (POIs) were selected on either the spinous processes or vertebral foramens of adjacent vertebrae. If contrast was particularly low, a rib was used in place of its attached vertebra. Displacements between the POIs allowed us to resample the DRR with, say, bilinear interpolation.

Subsequently, translational shifts were optimized by maximizing a local zero-normalized cross-correlation (ZNCC),

$$\max_{\tau \in A} (f \star g)(\tau) = \max_{\tau \in A} \int_B \mathrm{d}t \, \overline{f(t)} g(t + \tau), \tag{2.1}$$

between the zero-normalized gradients f and g of the X-ray and DRR, respectively. Here, we restricted the search area A to a 1×1 cm^2 region. A manually-selected bounding polygon B about the vertebral column was found to be helpful for appropriately fusing the vertebrae. The value of $\tau \in A$ that maximized the ZNCC enabled a fine-tuned resampling of the DRR. For computational purposes, we discretized all ZNCC integrals with step sizes of 1 mm.

2.4.2 Anterior-posterior misalignments

The superior-inferior translations computed in the previous section were applied to the corresponding LAT DRR and X-ray image pair, if they existed. By definition, the left-right translations were always orthogonal to the LAT images, hence they were ignored altogether. Then, we performed a similar ZNCC maximization procedure using Equation 2.1, this time searching along a 2 cm anterior-posterior line A. We manually delineated another bounding polygon B in the lateral orientation.

2.5 Manual misalignment generation

In clinical practice, radiation therapists manually align secondary DRRs to primary treatment radiographs. Due to the fact that our research project involved a large database of images, manually misaligning all images for model development was impractical. However, in order to assess the accuracy of our models in realistic clinical situations, we designed a smaller test dataset from our list of test patients with true positives generated manually in MIM (MIM Software, Inc., Cleveland, OH).

For this manual test subset, only a single delivered treatment fraction for each patient was considered. By convention, the retrieved RTImage DICOM objects were ordered according to their content data and content time. Thus, we presume that for the smaller testing subset, the first delivered treatment fraction was chosen. See Table 2.2 below. Observe that for both the kV and MV datasets, the number of images in the manual testing subset is three times the number of testing patients. The factor of three arises because the original retrieved DRR and X-ray RTImages were assumed to be aligned, and our misalignment procedure operates both superiorly and inferiorly. By the same argument, one can verify that the total numbers of images in Table 2.2 equal three times the numbers of images in Table 2.1 (and that the numbers of patients add up correctly).

We retrospectively confirmed the initial clinical AP and LAT fusions between the DRR and X-ray image pairs in MIM, adjusting image contrast to enhance visibility of the vertebrae. In a MIM fusion window, the AP and LAT DRRs were manually translated up and down by one vertebral body. Then the newly-created misaligned images were exported in DICOM format.

2.6 DenseNet architecture

Extensions to conventional convolutional neural network (CNN) architectures include residual convolutional networks (ResNets) [43] and densely-connected convolutional networks

(DenseNets) [46]. In this investigation, we propose a DenseNet model to differentiate between correctly aligned and incorrectly aligned patients imaged with kV or MV planar radiographs. A schematic of our model is depicted in Figure 2.1. Setup DRR and treatment radiograph image pairs are pre-processed with zero-normalized image gradients and combined into multichannel tensors. The tensors are then input into multi-layered, densely connected CNN. If both AP and LAT orientations are available, the output tensors are concatenated. A sequence of global average pooling (GAP) and fully connected (FC) layers reduces the tensorial information down to two classifier neurons. A softmax function converts these output neurons into a probability density function (PDF).

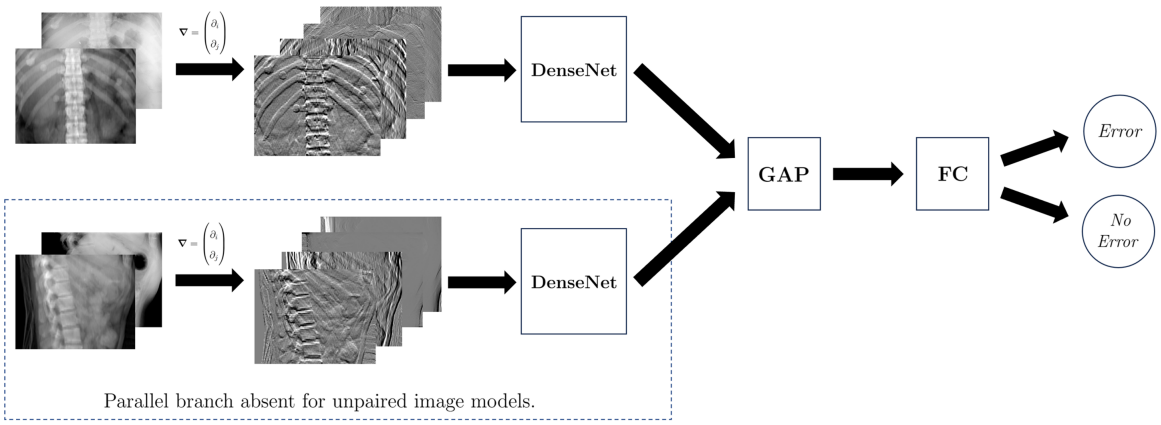


Figure 2.1: Our proposed DenseNet-based model architecture for off-by-one vertebral body setup error detection.

2.7 Model development

Neural network models for this project were developed in the MATLAB programming environment (MathWorks, Natick, MA). Training was implemented on an NVIDIA RTX A5000 graphics processing unit (GPU) (Nvidia Corporation, Santa Clara, CA). All patient MRNs from the dataset were carefully split into separate groups for model training, validation, and testing. The exact data subset sizes are listed in Table 2.2. We spatially downsampled every image to a conventional size of 300×300 pix^2 using bicubic interpolation. By inspection,

this size was sufficiently small for rapid deep learning computations without significantly reducing the original number of rows and columns. As mentioned in the model architecture section above, we pre-processed the images by taking image gradients, i.e., discrete partial derivatives in the horizontal and vertical directions. All partial derivatives were computed using matrix convolutions with kernel

$$\omega = \frac{1}{2}\begin{pmatrix} 1 & 0 & -1 \end{pmatrix}$$

and vanishing boundary conditions. This pre-processing operation was motivated by our preliminary studies on linear discriminant analysis (LDA) [40]. As anticipated, image gradient pre-processing indeed improved the model training accuracy and efficiency. After concatenating the partial derivative images of both the DRR and X-ray, the input tensor has four channels. Each channel was zero-normalized separately.

	kV AP – LAT		MV AP	
Data subset	Number of images	Number of patients	Number of images	Number of patients
Training	4917	186	3408	181
Validation	1047	38	792	37
Testing (semi-automated)	792	30	687	32
Testing (manual)	90	30	96	32

Table 2.2: Training, validation, and testing subset sizes for the on-board planar spinal imaging dataset. Observe that the number of images refers to the number of DRR and X-ray image pairs.

2.7.1 Training

We applied the Adam optimization algorithm for stochastic gradient descent [61] with a default hyperparameter of $\epsilon = 10^{-8}$. Other training hyperparameters included a batch size of 64, a learning rate of $\lambda = 10^{-3}$, and a maximum number of 30 epochs before concluding training sessions.

Our optimization objective was to minimize a standard cross-entropy loss function

$$H = -\sum_{i=1}^{2} \omega_i p_i(x) \log q_i(x), \tag{2.2}$$

where ω_i are the weights of the class labels. Here, p denotes the ground-truth PDF and q denotes the model-predicted PDF. Observe that the training dataset was imbalanced, since for every approved aligned image, we generated two off-by-one misaligned images (one superior, one inferior). To rebalance the training dataset, the class weights in the loss function H were set to $\omega_1 = 1/3$ and $\omega_2 = 2/3$ for class labels *error* and *no error*, respectively.

For kV beam energies, individual models were trained on the AP- and LAT-oriented data. Then a paired kV AP – LAT model was trained via a transfer learning approach. We defined two parallel branches of the DenseNet-based architecture (see Figure 2.1) by initializing their weights and biases according to those of the trained unpaired models. On the contrary, we trained the final FC layer from scratch, hence we increased its weight and bias learning rates by a factor of 10 accordingly.

For MV beam energies, one model was trained on AP-oriented data.

2.7.2 Testing

The performance of the trained error detection models were evaluated on the testing subsets withheld from model training. As described above, testing subsets were created with true positives produced algorithmically as well as manually to obtain greater assurance that the models work in realistic clinical settings.

The artificial output neurons of all networks returned a PDF of classification predictions. The actual classification decision was given by whichever neuron had the greatest probability. These decisions could be directly compared to the ground-truth labels of the testing subset datapoints.

A receiver operating characteristic (ROC) curve [60] was assembled for each dataset

evaluated. Briefly, an ROC curve describes how the model classification decisions vary as a function of a threshold $T \in [0, 1]$. Such thresholds indicate that a datapoint X belongs to a particular class, say the *error* class, if the probability given by the *error* neuron $p(X)$ satisfies $p(X) \geq T$. The area under the ROC curve (AUC) was used as our metric for the success of the classifier. We used trapezoidal numerical integration to calculate AUCs. Due to the rarity of clinical setup errors, it was desirable to select a high specificity on the ROC curves to minimize disruption of the clinical workflow. We chose a fixed specificity of 95% for model testing and a target sensitivity of at least 95%.

2.8 Complete database error search

After our paired kV AP – LAT model was developed, we processed all clinically aligned paired kV imaging datapoints as a safety analysis tool for retrospective error hunting. Note that due to the large volume of data, we deemed the task of visually verifying every clinically applied registration to be impractical. The small handful of images flagged by our error detection algorithm were visually analyzed in our clinical offline review software in the ARIA ecosystem.

2.9 Results

An example set of orthogonal kV treatment radiographs together with clinically aligned and artificially misaligned DRRs is depicted in Figure 2.2.

The ROC curves of the testing subset are displayed in Figure 2.3, with quantitative diagnostic results summarized in Table 2.3.

Out of the 2252 clinically aligned kV image pairs in the database, 16 were flagged by the orthogonal kV classifier. After reviewing these cases in ARIA, we confirmed that only one kV radiograph was truly misaligned. This case, illustrated in Figure 2.4, was reported to our institution's incident learning system. Therefore, the per-fraction off-by-one vertebral body

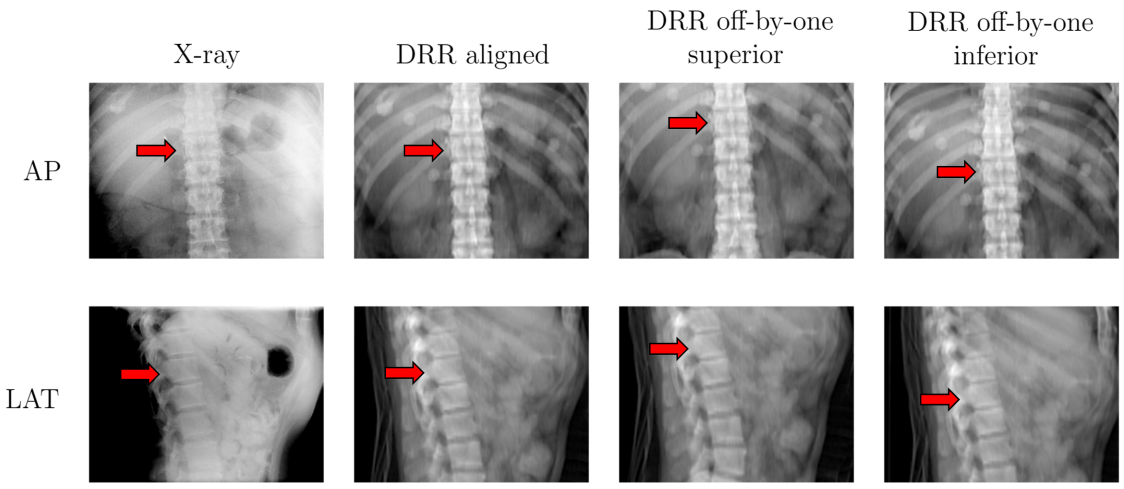


Figure 2.2: Example set of orthogonally-paired kV radiographs. The clinically aligned DRRs were shifted superiorly and inferiorly by one vertebral body, maximizing a local cross-correlation image similarity metric.

error rate with paired kV images at our institution was determined to be 0.044%. Assuming Poisson statistics [10], a 95% confidence interval about this error rate is $[0.0022\%, 0.21\%]$. Common features of the remaining incorrectly flagged 15 cases include low contrast, spinal implants, and abnormal spinal curvature (e.g., scoliosis, kyphosis, or lordosis).

For our most accurate model on orthogonally-paired kV radiographs, it was instructive to analyze the cases that the model misclassified. We selected a classification threshold T belonging to the point on the ROC curve with specificity 96.4% and sensitivity 98.2%. Out of the 90 cases in the testing subset, only 2 were misclassified, with one false positive and one false negative. See Figure 2.5. We observed that the false positive case had extremely low contrast, which invariably creates challenges for the neural networks, as noted for the LAT orientations and MV beam energies. The AP radiograph was also observed to be slightly rotated with respect to its DRR. The false negative case was characterized by a large AP radiographic field that included the pelvis (which perhaps should have been excluded in the first place during our data selection process). This patient also had a chest catheter. Fortunately, misaligned catheters and iliac crests would be easily noticeable by radiation therapists.

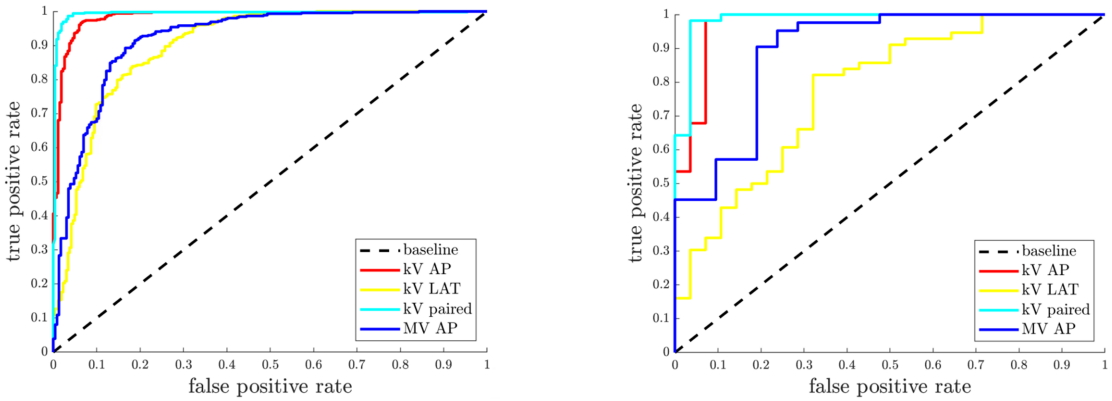


Figure 2.3: ROC curves for the DenseNet classifier applied to the testing patient dataset left out during model training. (a) All available treatment fractions with simulated errors generated semi-automatically. (b) One treatment fraction per patient with simulated errors generated manually.

2.10 Discussion

Maximizing a cross-correlation similarity metric implies that the input tensor channels were as challenging as possible for the neural network to distinguish. It is possible, however, that improved results would arise from more data augmentations, specifically by randomly translating the DRR about the local cross-correlation maxima.

Our preliminary LDA studies indicated a strong favoring of vertical (superior-inferior) partial derivative features as opposed to horizontal (left-right). On a small sample of training datapoints, we obtained an AUC of about 90% and a classifier accuracy of about 80% on the optimal point on the ROC curve. These results are promising, but also provide evidence to the fact that deep learning outperforms machine learning.

Observe that the ROC curves obey a similar trend in both the semi-automated and manual testing subsets, which is reassuring. The classifier performed the lowest on the kV LAT images, although it was significantly better than a baseline random classifier. This result is likely due to the very low contrast in laterally-acquired images. For this reason, kV LAT images in our clinic were almost always acquired with a paired orthogonal AP image for

Testing dataset	AUC	Model sensitivity	
		90% specificity	95% specificity
kV AP	0.98	**0.97**	0.94
kV LAT	0.90	0.73	0.40
kV AP – LAT	0.99	**1.00**	**0.99**
MV AP	0.92	0.68	0.51

(a)

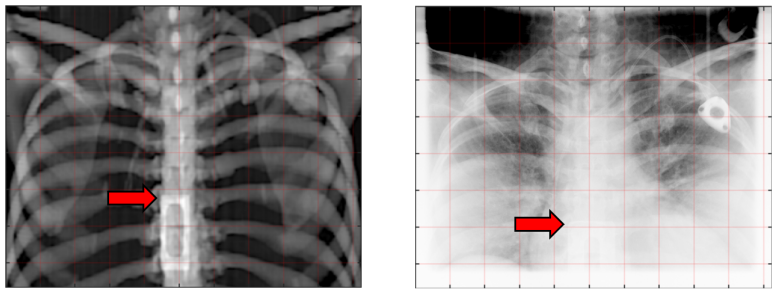
Testing dataset	AUC	Model sensitivity	
		90% specificity	95% specificity
kV AP	0.97	**0.98**	0.68
kV LAT	0.78	0.34	0.30
kV AP – LAT	0.99	**0.98**	**0.98**
MV AP	0.90	0.57	0.45

(b)

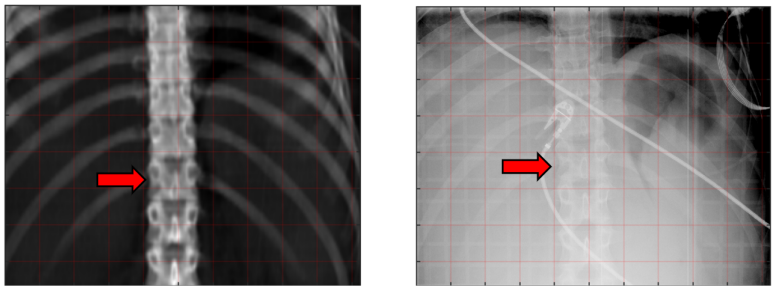
Table 2.3: Classifier diagnostics extracted from the ROC curves in Figure 2.3. Sensitivity values in bold exceed our original goal of 95%. (a) All available treatment fractions with simulated errors generated semi-automatically. (b) One treatment fraction per patient with simulated errors generated manually.

these anatomical sites. Also note that there were no instances of unpaired MV LAT images retrieved from our database. The second-lowest performance is for the MV AP dataset, and this was once again expected on account of the low contrast for MV imaging. The kV AP dataset was classified with remarkable accuracy. Superior contrast on the vertebral bodies and surrounding anatomy compared to the MV and kV LAT likely account for this improvement. Finally, as anticipated, the combined orthogonal kV data sent through separate branches of a parallel DenseNet architecture performed the best. Additional information provided by the lateral direction yielded a small but noticeable improvement in the classifier accuracy for both semi-automated and manual datasets.

Limitations of our study include the fact that misalignments used for training were created algorithmically rather than manually. It is conceivable that a more realistic error detection model could be developed if every misaligned training datapoint was translated manually, as opposed to one fraction per testing patient. A future direction of this research



(a)

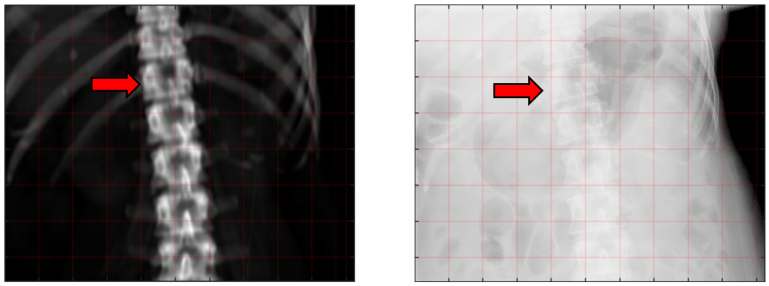


(b)

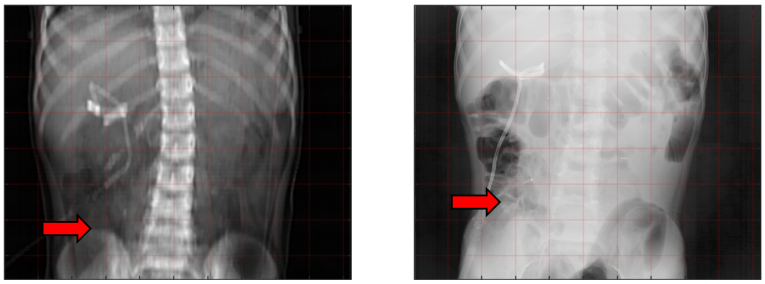
Figure 2.4: Real clinical cases that were flagged by our kV AP – LAT model and confirmed as misalignments in our treatment planning system. (a) An off-by-one vertebral body error, deemed to be a clinically significant error. (b) A misalignment off by about 1.5 cm in the superior-inferior direction, deemed to be clinically insignificant.

could involve extending to a multi-institutional dataset.

The extent to which low image contrast decreases the predictive accuracy of the models was not quantified in this investigation. A future improvement could involve expert medical physicists providing a quantitative scoring of the images based on the visual contrast and anatomical detail available for alignment. Our research could also benefit from training the model using a third category where the presence of an off-by-one error is undetermined. This would potentially provide more insight into the frequency of treatment radiographs with poor contrast as well as how the model performance is correlated with contrast.



(a)



(b)

Figure 2.5: The only testing cases that were misclassified by our kV AP – LAT model. (a) False positive. The treatment radiograph had relatively poor contrast. Here the intensity windowing was adjusted to improve visibility. The treatment radiograph was also observed to be slightly rotated with respect to its DRR. (b) False negative. The presence of a chest catheter and pelvic bones within a large FOV may have contributed to an incorrect label.

2.11 Conclusion

We successfully trained a densely-connected convolutional neural network to detect off-by-one vertebral body misalignments for radiotherapy patients positioned with planar radiographic setup images. Our models achieved high AUCs when applied to the testing subsets. The kV AP and kV AP – LAT models achieved our objective goal of at least a 95% true positive rate given a false positive rate of less than 5%. A low false positive rate ensures that the clinical workflow is minimally disrupted, which is desired since patient setup errors are infrequently encountered. This level of accuracy demonstrates a potential workflow in which AI models would analyze images in real time at a treatment machine. If a misalignment is detected, an overridable interlock would be imposed to alert radiation therapists to double-check the

alignment before shifts are sent to the treatment couch. We used the kV AP – LAT model to retrospectively search our institution's radiation oncology clinical database to determine a per-fraction off-by-one vertebral body setup error of 0.044% $[0.0022\%, 0.21\%]$ over the period 2011 – 2021. Our results confirm the overall safety of planar radiographic IGRT.

Chapter 3

Automated review of prostate RT treatment planning quality

3.1 Introduction

One of the greatest challenges in the broad clinical implementation of intensity-modulated radiotherapy (IMRT) has been to assess objective, quantitative measurements of treatment quality. In recent years, the number of radiotherapy patients and the complexity of treatment planning has increased. The outcome of radiotherapy is known to be correlated with physician experience and institutional volume [80, 89]. Given the same patient, treatment goals, and delivery technique, one may obtain different optimized plans. In other words, there exists a range of *standards of care* in clinical practice.

Notably, the standards of care in radiation oncology have been less widely analyzed in community-based medical centers than in academic centers [95, 70, 38]. Recent efforts to create large treatment planning databases [101, 63] open up the possibility of a comprehensive assessment of the quality of community radiotherapy treatments, critically informing efforts to improve radiotherapeutic care nationwide.

However, human review of retrospective treatment plans is labor-intensive and nearly

impossible for large datasets on the order of thousands of patients. Hence, our goal in this study is to demonstrate a methodology for automated plan quality evaluation that can be used to assess community-based treatments against the standards of our academic institution. The proposed methodology is based on an automated analysis of specific quality indicators to identify potentially low-quality plans for further human review. Ideally, a small subset of treatments, whose quality indicators are outside a sufficiently broad range of standards of care, will be selected for human review. In this chapter, we focus on intact prostate and prostate bed radiotherapy plans, with or without the inclusion of seminal vesicle (SV) and lymph node (LN) targets.

Existing work on large-scale plan quality assessment has focused on the optimality of dose distributions relative to a Pareto front using knowledge-based planning (KBP) [39, 105, 2, 96]. Studies have demonstrated that KBP can identify treatment plans with suboptimal dosimetric quality within multi-institutional databases. Also, KBP has been proposed as a tool to regularize plan quality in prospective clinical trials [55, 88, 68]. However, other important aspects of plan quality, upstream of those covered by KBP, remain significantly unaddressed in plan quality analyses of large-scale databases.

In this chapter, we introduce novel algorithms for quality control that can flag potential low-quality plans based on the following five indicators:

- number of days between the computed tomography (CT) simulation and the treatment plan approval;
- planning target volume (PTV) margins about the clinical target volume (CTV);
- the appropriate use of fiducial markers in intact prostate treatments;
- prescription dose fractionations (℞) relative to widely accepted guidelines;
- PTV geometric shape descriptors used to flag potentially erroneous target volume contours.

The automated methods that we develop for computing these quality indicators will be applied to a dataset of over 1800 prostate and prostate bed patients to establish a reference of standards of care. Then the methods will be applied to a set of radiotherapy plans from external cancer centers to demonstrate the utility of the methodology.

More details on the quality indicators are found below.

Days to plan approval

Biological responses to radiation are time-dependent. Refer to [41] for a comprehensive text on radiobiology. Studies have shown that excessive delays in starting radiation treatment after receiving a diagnosis and CT simulation are correlated with reduced tumor control [19, 48]. Failure mode and effects analysis (FMEA) of treatment delays are highlighted in [103].

Concurrent with these existing studies on the importance of reducing time delays, we will retrospectively evaluate the number of days between CT simulation and plan approval. We will compare the number of days on a cumulative distribution plot between the internal and external datapoints.

PTV margins

A wide range of PTV margin recommendations exist in clinical research. Studies have found a correlation between radiotherapy delivery technique and average reported PTV margins [104]. For instance, positioning with external markers and in-room lasers corresponds to an average margin of 10 mm, whereas positioning with implanted fiducial markers corresponds to an average margin of less than 6 mm. Another study concluded that for image-guided radiotherapy (IGRT), at least a 4.5 mm margin is required to ensure full dose coverage to an intact prostate CTV [97]. Inclusion of seminal vesicles and pelvic lymph nodes in target delineations results in larger reported margins to accommodate for greater translational displacements. Notably, smaller margins have been correlated with decreased radiation toxicity effects in patients [94].

To our knowledge, however, no study has attempted to quantify the range of PTV margins that actually exists in clinical practice retrospectively. Our work elucidates an automatic cal-

culation of PTV margins and compares margins between treatment plans from our academic institution and from neighboring community hospitals.

Fiducial markers

The use of fiducial markers has been shown to improve the margins that morphologically expand a prostate CTV into a prostate PTV. One study showed that image guidance with implanted fiducial markers and daily electronic portal imaging (EPI) allows for narrower prostate PTV margins without compromising target coverage [91]. A related study found that electronic portal images of fiducial marker positions can accurately track intra-fraction motion, thereby significantly reducing PTV margins [20]. For prostate brachytherapy treatments (e.g., with iodine-125), gold seed implants were shown to provide an effective means of prostate localization and to improve dose conformality [100].

Our work includes an automated analysis to detect whether seeds are implanted in a prostate. Moreover, if seeds are detected, we automatically count the number of such seeds. Seed detection can be combined with the intended PTV margin calculation to assess the existing range of clinical practice.

Prescription doses

In 2023, a blue-ribbon panel of prostate cancer experts organized by the US Department of Veterans Affairs (VA) National Radiation Oncology Program developed updated measures and metrics for prostate radiotherapy [93]. The recommended prescription dose fractionations will be used to compare against the values that our automated algorithms calculate. Additionally, frequently used prescriptions from our academic institution that were absent in the VA consensus guidelines will be incorporated.

Our results will fill an invaluable knowledge gap regarding the capacity to which dose prescriptions in external-beam radiotherapy follow such guidelines.

PTV geometric descriptors

In medical imaging, geometric descriptors have been introduced for binary classification and segmentation tasks [44, 45, 58, 57]. Geometric descriptors of a binary mask are typically

invariant to translations and rotations, depending solely on the curvature of its boundary [73, 79]. A feature set of geometric descriptors may also be optimized to identify a weighted linear combination that maximally distinguishes a training set of contours for a particular anatomical site.

Concurrently, we may use deep learning approaches to perform target volume auto-contouring routines. The predicted contours can be compared to the approved physician-delineated contours as a measurement of plan quality. Auto-contouring intact prostate and prostate bed CTVs is explored in Chapter 4.

3.2 Data collection

The data underlying our investigation consisted of intact prostate and prostate bed plans extracted from clinical treatment planning databases at our institution. Note that by *prostate bed* plans we are referring to adjuvant or salvage postoperative radiotherapy plans for patients who underwent radical prostatectomy (a surgical removal of the prostate gland, seminal vesicles, and vas deferens). Data collection proceeded as follows.

3.2.1 DICOM query and retrieval

To create our dataset of treatment planning data for prostate radiotherapy patients, a software application was developed to query, retrieve, organize, and store DICOM files. Clinical data from our institution is managed by the ARIA Radiation Oncology Information System (Varian Medical Systems, Inc., Palo Alto, CA) and by MIM (MIM Software, Inc., Cleveland, OH). Our software was developed using the open-source *pynetdicom* package that implements the DICOM protocol [33]. The program allows a local workstation at our institution to access our clinical database and store copies of all relevant DICOM radiotherapy planning files. A general workflow of the program is described below. Recall the DICOM Standard outlined in §2.2. Also recall that patients are identified by their medical record number (MRN), and

that DICOM objects are identified by a unique identifier (UID).

Patient MRNs and study UIDs

In our dataset, we searched for study dates between the years 2012 and 2021. Quality assurance (QA) phantoms with an MRN containing the substring "QA" were ignored. The relevant series modalities that we queried for this investigation are:

- *RTPlan* for treatment planning information [30];
- *CT* for the planning CT medical image [23];
- *RTStruct* for the structure set containing anatomical contours [27];
- *RTDose* for the planned radiation dose distribution [26].

All of these modalities were required to be present within a study in order for the study to be queried.

RTPlan series UIDs and frame of reference UIDs

Given a list of patient MRNs and study UIDs for a particular day being queried, all associated RTPlan series UIDs and their frame of reference UIDs were recorded. Observed that this step cannot be combined with the first step on account of the DICOM query and retrieval information model hierarchy [32].

External-beam prostate plans

External-beam prostate plans were identified through a sequence of conditional checks:

- QA plans with the substring "QA" in the RTPlan label were ignored.
- Intact prostate plans were found by matching RTPlan labels containing any of the substrings "prst" or "prost". Prostate bed plans were found by matching RTPlan labels containing the substring "pbed".
- High dose rate (HDR) brachytherapy plans were ignored through the conditions:
  - The RTPlan label did not contain the substring "HDR";

  - The manufacturer was not Nucletron (Elekta, Stockholm, Sweden);
  - The brachytherapy treatment type tag was absent in both the RTPlan and the RTPlan fraction scheme module [29].

- Rejected plans were ignored. Observe that it was not sufficient to select only approved plans, since boost plans were often unapproved despite containing valid treatment planning data and contributing to the overall planned dose.

At our institution, initial plans typically included a substring "i1" in the RTPlan label, while sequential boosts typically included substrings "b1", "b2", etc. Simultaneous integrated boosts (SIBs) typically included the substring "s1". However, such nomenclature was not reliable for external community plans.

For every relevant plan, the object and frame of reference UIDs were recorded.

Associated planning CTs

There are two strategies for determining the relevant planning computed tomography (CT) as opposed to daily cone-beam computed tomography (CBCT) scans. One strategy would be to retrieve RTReg modalities [21] and take the most occurring referenced image series UID. The strategy that we chose was to query CTs and retrieve the series with a frame of reference UID matching that of a given RTPlan.

Associated RTStructs

Another query was initiated to retrieve structure sets. Oftentimes, radiation oncology residents perform multiple delineation revisions until a final structure set is approved by the attending physician. Thus, multiple RTStructs may exist with the same frame of reference as that of a retrieved RTPlan. One strategy would be to retrieve the singular RTStruct with an approved status. Note that in many cases, the most recent RTStruct belonged to a chart round session and was typically unapproved. A more reliable strategy that we chose instead is to record the referenced RTStruct object UID within a given RTPlan and then to retrieve the RTStruct with matching object UID.

Associated RTDoses

Finally, the prostate planning doses were queried by setting the referenced RTPlan object UID to match a given RTPlan object UID. We also ensured that the dose summation type was *plan*, indicating that the dose distribution represented the delivery of all planned fraction groups.

3.3 Data selection

3.3.1 Data-driven approach to remove duplicates

Occasionally, a single radiotherapy plan appeared twice in our dataset under two different patient IDs. Such duplicates could occur if the data was anonymized for testing purposes, or if the data originated from an external clinic and was imported into our clinical databases under a different generated patient ID. In order to remove any duplicate treatment planning datapoints and ensure that each datapoint was unique, a data-driven approach to automatically identify duplicates was created. Within the RTPlan object, the following items were recorded:

- target names and prescription doses in the radiotherapy prescription module [31, 98];
- number of fractions and beams planned in the fraction scheme module.

Additionally, within the associated RTStruct object, the following items were recorded:

- all region-of-interest (ROI) names in the structure set module [27];
- the numbers of contour points for all ROIs in the ROI contour module [28].

Duplicates were found if there was a perfect agreement in these recorded DICOM values between any two treatment plans. The plan with the most recent structure set date was maintained in our dataset, while the other duplicates were removed. Observe that this process effectively deanonymizes the dataset.

3.3.2 Identification of plans from external clinics

A subset of treatment plans in our dataset originated from community hospitals not affiliated with our institution. These cases were used to test our quality review algorithms and were set aside during algorithm development. The purpose of automated quality review on such an external subset is to provide a proof-of-concept for handling larger amounts of community data. In order to determine which treatment plans were internal (i.e., originated from our institution), two conditions were checked. First, the patient MRN was required to be seven digits with no alphabetical characters. Second, the following DICOM information was matched with known parameters from our institution:

- institution name and address in the CT object;
- CT station name in the CT object;
- physician of record (alternatively, reviewer name) in the RTStruct object.

All four of the proposed quality metrics incorporate the CTV and PTV as inputs. There was considerable variation in target volume naming conventions. Meanwhile, plans frequently contained multiple target volumes, such as lymph node target volumes and target optimization volumes where the target is cropped out of a neighboring organ at risk (OAR). We developed the following algorithms to automatically identify unique prostate CTV and PTV structures with high confidence and without human intervention.

3.3.3 Miscellaneous selection criteria

One treatment plan in the dataset was anonymized and did not have sufficient information in its CT object to confirm whether the datapoint originated from our institution. An expert medical physicist in our research group reviewed this case and determined that it should be included as an internal datapoint.

We recognized that one external datapoint belonged to a proton radiotherapy plan, on account of its target structure being labeled as an optimization target volume (OTV) instead of a PTV. This plan was excluded from further analysis.

A breakdown of the number of patients and plans acquired throughout the data selection process is shown in Table 3.1.

Selection step	Number of patients	Number of plans
External-beam prostate radiotherapy	1416	1818
Remove duplicates	1395	1671
Internal cases	1328	1577
External cases	67	94

Table 3.1: Number of prostate radiotherapy patients and treatment plans acquired throughout the quality review data selection process. Our investigation focused on external-beam prostate treatments that were planned on CT scans.

3.4 Identification of target volume structures

Many of the proposed quality metrics require an automated means of identifying the prostate CTV and PTV. We observed considerable variability in target volume nomenclature. Moreover, treatment plans frequently contained multiple targets, such as lymph node targets and optimization ROIs. We developed the following procedure to uniquely identify prostate and lymph node target volumes with high confidence and without human intervention.

3.4.1 Prostate PTV

Every structure set module contained a collection of closed planar ROIs. A subset of potential PTV contours was acquired by systematically filtering out ROIs based on their names. Contours were omitted if their names started with the characters "O_", "R_", "T_", or "x" or "z", indicating OARs, rings, tuning structures, and optimization structures, respectively. Contours whose names contained common normal tissue substrings such as "bladder",

"body", "bowel", "rectum", or abbreviations therein, were omitted. Finally, contours whose names did not either start with the characters "P_" or contain the characters "PTV" were omitted.

The resulting list of potential PTVs was then sorted into two groups: one containing the prostate exclusively, and one containing lymph nodes. Our algorithm first checked for the existence of a combined prostate and lymph node structure. The potential PTV with the largest volume was compared against unions of disjoint pairs of the remaining potential PTVs. If an overlap exceeded 95%, then the largest structure was ignored, unless it was included in the list of dose targets referenced in the RTPlan.

Subsequently, structures were removed from the list of potential prostate PTVs if their volume exceeded 400 cm^3 or if their centroid was outside the contour boundary. By inspection, both of these features were found to indicate lymph node targets with high accuracy. Finally, structures were removed from the list of potential prostate PTVs if the mean dose to the contour was less than 50% of the maximum dose $d_{\max}$ in the three-dimensional dose distribution. If more than one candidate PTV remained after the above filters were applied, then the structure with the largest volume was selected as the unique prostate PTV.

Note that the list of dose targets referenced in the RTPlan was not always reliable, since targets were sometimes either absent or mislabeled as OARs. We observed that in our treatment planning system Eclipse (Varian Medical Systems, Inc., Palo Alto, CA), dose target ROI names were monitored by the private DICOM tag $(3267, 1000)$ created by Varian [98]. The values of the tag are represented in Unicode and can be easily converted into readable text.

3.4.2 Prostate CTV

Similarly, candidate CTVs were obtained from the complete list of available ROIs in the structure set. Contours whose names began with the characters "R_", "T_", "x", or "z" were omitted. Unlike the methodology for finding PTVs, contours whose names began with the

characters "O_" were included as potential CTVs, since SV targets were sometimes labeled as OARs. Contours whose names contained substrings such as "body", "seed", "opt", and abbreviations therein, were omitted.

Contours were removed from the list of candidate CTVs if more than 10% of their volume was outside the identified PTV. To ensure that PTV optimization structures were ignored, a Euclidean distance transform with respect to the CTV candidate boundary was computed, and the distance transform values on the PTV boundary were recorded. We required that the 25^{th} percentile surface distance was at least 1 mm. Also, if the binary mask of a contour contained 5 or more disconnected components, then the contour was declared to be an optimization structure (e.g., an isodose contour).

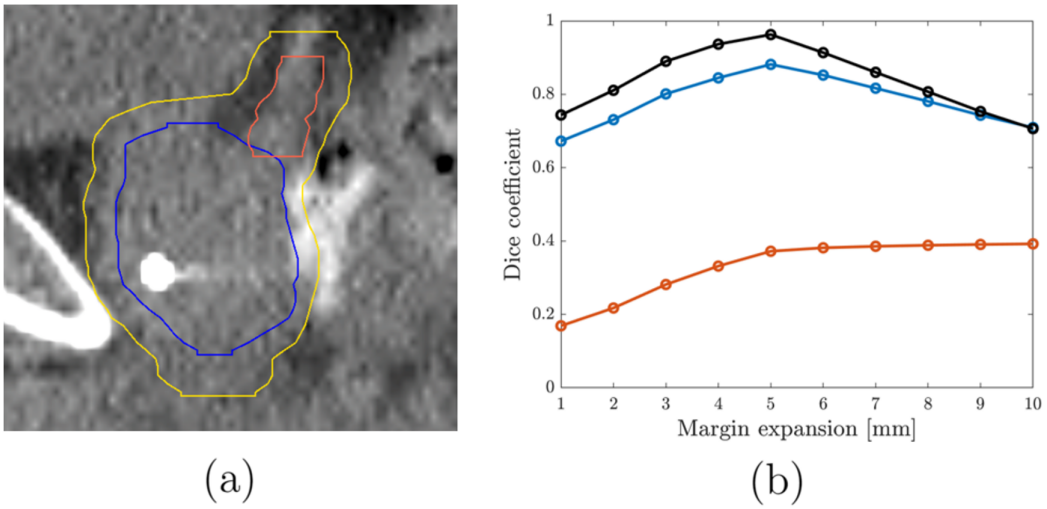
A unique CTV was then identified from the list of candidate CTVs by performing successive isotropic morphological expansions of 1 mm and maximizing a similarity metric with respect to the PTV. Specifically, the similarity metric that we used was the Sørensen-Dice coefficient $D(X, Y)$ between two sets (or in our context, binary masks) X and Y, defined by the formula

$$D(X, Y) = 2 \cdot \frac{|X \cap Y|}{|X| + |Y|}. \tag{3.1}$$

Since it was frequently the case the prostates and seminal vesicles were contoured as separate CTVs, either one ROI or the union of two ROIs in the list of potential CTVs was considered. Figure 3.1 illustrates this process.

3.5 Deep learning model to classify plan type

We designed a deep convolutional neural network (CNN) to classify whether a given plan in the dataset corresponded to either an intact prostate or a prostate bed case. The data analyzed for model development was composed of a random sample of internal plans only: 290 training, 37 validation, and 37 testing, with approximately equal amounts of intact and bed cases. Ground truth labels were assigned based on our internal plan naming conventions,



(a) (b)

Figure 3.1: Example of calculating approximate isotropic margins by maximizing Sørensen-Dice coefficients between the PTV and morphological expansions of the CTV. Here we conclude that the correct CTV combines the prostate and SVs, and that the approximate margin is 5 mm. (a) Sagittal view of a prostate PTV (yellow contour) enclosing a prostate CTV (blue contour) and SV CTV (orange contour). (b) Sørensen-Dice coefficients as a function of the radius of the spherical morphological structuring element. Blue line: prostate CTV only. Orange line: SV CTV only. Black line: combined prostate and SV CTVs.

where "bed" or "bd" substrings indicated prostate bed plans and "pros" or "prst" substrings indicated intact prostate plans. For every plan, the three-dimensional CT scan and CTV binary masks were combined into a two-channel image. The CT voxel intensities were pre-processed by setting the mean to zero and standard deviation to one. We chose a standard CNN architecture with four layers, followed by a global average pooling layer and fully connected layers to reduce the network down to two classifier neurons.

Plan type determination of the external cases was a hybrid of plan naming conventions and the CNN results. If an external case contained either of the substrings "bed" or "bd" in its plan name or target volume names, then the plan was immediately classified as a prostate bed plan. Otherwise, our CNN classifier was applied to the plan to decide its proper category.

A breakdown of the dataset categorized by intact prostate and prostate bed cases is shown in Table 3.2. Note that in this table, we manually verified the correct plan categories for all external cases. The accuracy of the CNN model is highlighted in the results below.

Plan type classification is an important step in the automated quality analysis. Intact prostate target volume delineations are more streamlined due to easily-recognizable anatomical borders, whereas prostate bed target volume delineations generally have more inter-physician variability [102]. Moreover, the prescription doses to intact prostates and prostate beds differ, as we will also encounter in the results below.

Category	Number of patients	Number of plans
Internal intact	1040	1156
Internal bed	288	421
External intact	35	52
External bed	32	42

Table 3.2: Number of prostate radiotherapy patients and treatment plans categorized according to plan type (intact or bed). Prostate bed cases refer to adjuvant or salvage postoperative radiotherapy plans for patients who underwent radical prostatectomy.

A random sample of 30 internal intact and 30 internal bed was created to evaluate the accuracy of our quality indicator algorithms. All available internal and external plans were analyzed in order to assess and compare treatment planning quality between our institution and clinics in the greater community.

3.6 Computation of quality metrics

Having identified the prostate target volumes, we proceeded to compute treatment planning quality indicators. Volumetric CT images were resampled onto a uniform and isotropic grid with 1 mm spacing per voxel. Contour coordinates and dose distributions were resampled onto their associated CT scan, and contours were converted into volumetric binary masks.

3.6.1 Days to plan approval

The days between CT simulation and plan approval require the following two DICOM tags:

- content date in the CT object;

- review date in the RTPlan object.

These tags are stored as strings in year-month-day (*YYYYMMDD*) format. Converting to chronological date-time variables in a programming environment and taking the difference in dates yields the desired quality indicator.

3.6.2 PTV margins

The PTV margins used for planning were retrospectively estimated by computing distance transforms. Let

$$\begin{aligned} T_X : \mathbb{R}^3 &\to \mathbb{R} \\ x &\mapsto T_X(x) \end{aligned} \tag{3.2}$$

denote a Euclidean distance transform with respect to a subset $X \subset \mathbb{R}^3$. We devised a formula to find the boundary (i.e., the surface) $\partial X \subset \mathbb{R}^3$ of X,

$$\partial X(x) = \begin{cases} 1 & T_{X^{\mathrm{c}}}(x) = 1 \\ 0 & \text{otherwise} \end{cases}, \tag{3.3}$$

where X^{c} is the set-theoretic complement of X. The mean μ of the distances of the PTV boundary points with respect to the CTV boundary was recorded as the estimated margin, and the standard deviation σ was recorded as a measurement of margin anisotropy. In other words,

$$\text{margin} = \mu\Big(\big\{T_{\partial \mathrm{CTV}}(x) \mid x \in \partial \mathrm{PTV}\big\}\Big), \tag{3.4}$$

$$\text{anisotropy} = \sigma\Big(\big\{T_{\partial \mathrm{CTV}}(x) \mid x \in \partial \mathrm{PTV}\big\}\Big). \tag{3.5}$$

3.6.3 Fiducial markers

A fiducial marker counting algorithm was developed and trained on a subset of internal prostate cases. Voxels corresponding to fiducial markers implanted in the prostate were

located by thresholding within the binary mask of the PTV. We fixed a threshold CT number intensity of 2000 Hounsfield units (HU), as it exceeds the upper limit of cancellous and cortical bone yet does not exceed the CT number of metal implants. If the RTStruct included fiducial marker contours, indicated by the substrings "beacon", "fid", "mark", or "seed", then the search space also included the union of all such contours with a slightly lower threshold of 1500 HU. We lowered the threshold within fiducial marker contours to account for microcalcifications that may have been used by the physician during treatment planning.

Voxels within the PTV or union of fiducial marker contours exceeding their respective intensity thresholds were selected as seed coordinates. In order to cluster the coordinates and count the number of fiducial markers, Euclidean distances between all pairs of coordinates were computed. If any pair of voxels was less than 5 mm apart, then the pair was regarded as belonging to the same fiducial marker. The coordinate with lower intensity was removed from the list of representative points. This process repeated until each marker had one representative point and all such points were at least 5 mm apart. See Figure 3.2 for an illustration of the counting algorithm. The number of representative points detected was therefore used as the automated number of fiducial markers counted.

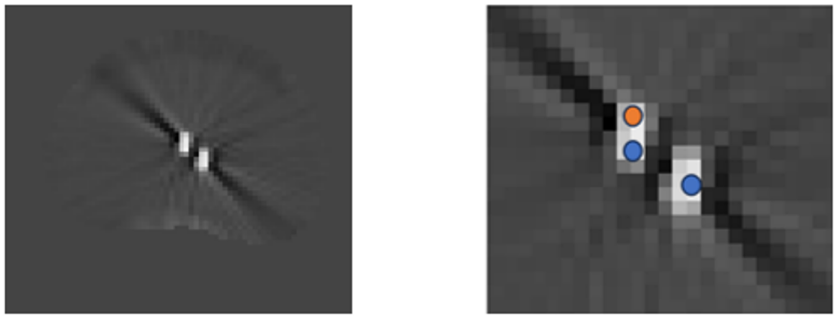


Figure 3.2: Example of calculating the number of gold seeds surgically implanted in a prostate. This CT slice was correctly found to have two seeds. (a) A particular axial slice of a CT scan, masked by its PTV. (b) Blue: unique representative points for the seeds. Orange: a duplicate point within 5 mm of a representative point.

3.6.4 Prescription doses

Next, the prescription dose (℞) to the prostate PTV was recorded. The dose distribution within the PTV was used to reproduce a dose-volume histogram (DVH) curve. Dose values from zero to the maximum dose $d_{\max}$ in the dose distribution were sampled in increments of 10 cGy. A linear interpolation was formulated to extract the planning dose prescribed to 95% of the PTV (which we will refer to as the $D95$ point on the DVH curve). The number of planned treatment fractions was easily extracted from the RTPlan object.

If a patient was treated with initial and boost plans, prescription doses from all plans were analyzed. For each plan, the prostate PTV was identified and the $D95$ dose was computed. Subsets of up to three separate doses and fractionations, if available, were accumulated and compared to standard recommended prescriptions from our institution and from the VA consensus guidelines.

Plans were flagged if no match was found for any subset of treatment plans up to a dose tolerance of 5%.

3.6.5 PTV geometric descriptors

Three-dimensional geometric shape descriptors of the PTVs were used as the final quality indicator. Such descriptors allowed us to identify PTV contours with unusual shapes compared to the set of all PTVs within our institutional database of prostate plans, presumed to encompass the range of acceptable standards of care. The four dimensionless shape descriptors that we computed are:

- volume V, normalized by our cutoff parameter of 400 cm^3;
- sphericity G_{sph}, defined by

$$G_{\text{sph}} = \frac{(36\pi V^2)^{1/3}}{A}, \tag{3.6}$$

  where V is volume and A is surface area;

- solidity G_{sol}, defined by

$$G_{\text{sol}} = \frac{V}{V_{\text{conv}}}, \tag{3.7}$$

  where V_{conv} is the convex hull of the target;

- extent G_{ext}, defined by

$$G_{\text{ext}} = \frac{V}{V_{\text{box}}}, \tag{3.8}$$

  where V_{box} is the bounding box of the target.

A correlation clearly exists between these variables, so it was helpful to project down to a two-dimensional subspace. Traditional dimensionality reduction algorithms include the widely-studied principal component analysis (PCA) [54], where datapoints are fit to an ellipsoid. However, we had greater success in preserving the inherent relationship between datapoints using locally linear embedding (LLE), a manifold learning technique introduced in [86]. See Appendix C for a derivation of the LLE algorithm. Embedding vectors and weights were trained on internal datapoints only with a fixed number of nearest neighbors. External datapoints were embedded using the pre-trained embedding vectors and weights with the same number of nearest neighbors.

3.7 Results

3.7.1 Target volume identification

The accuracy of PTV identification on the internal random sample was 100%. On the external plans, PTV identification accuracy was $64/67 \approx 96\%$. Among the three failures, two were due to the presence of an ROI that expands the PTV by 1 mm, which we presumed to be an optimization structure. Since our algorithm defaulted to choosing the largest available candidate PTV, and since we fixed a condition that the 25^{th} percentile surface distance ought to be at least 1 mm, such optimization structures were selected. For practical purposes,

optimization structures of this type perform equivalently to the correct PTV for measuring quality indicators. The remaining failure was a prostate bed case where the patient was re-simulated for a boost, yet the initial prostate bed PTV contour was also present in the structure set registered to the boost CT scan.

The accuracy of CTV identification on the internal random sample was 100%. On the external plans, CTV identification accuracy was $65/67 \approx 97\%$. Among the two failures, one was due to the presence of an ROI that expands the CTV by 3 mm. By the same reasoning as the PTV failures mentioned above, our algorithm selected the larger expanded ROI. The second failure occurred because our algorithm did not select any CTV. Upon review by an expert medical physicist, the correct CTV was composed of three distinct ROIs, whereas our code only accommodated at most two ROIs for computational efficiency. The medical physicist noted that the dosimetry on this treatment plan was particularly poor.

3.7.2 Plan type classification

Our CNN plan type classifier obtained a 97.5% validation accuracy and 100% testing accuracy. When applied to the external cases and compared with manually-determined ground truth category labels, the CNN accuracy was $62/67 \approx 93\%$. Three failures likely occurred due to significant bladder contrast, and two failures occurred for unapparent reasons. We noticed that two of the failures (both with bladder contrast) had seed implants detected in the prostate. Under the condition that seed implants are present exclusively for intact cases, the automated accuracy improves to $64/67 \approx 96\%$. It is reasonable to believe that an increased amount of datapoints and data augmentations would enable the deep learning model to recognize seed implants as a feature of intact cases. When evaluating quality indicators on the external plans, we relied on the ground truth labels for more accurate results.

3.7.3 Days to plan approval

Figure 3.3 shows cumulative distribution functions on the number of days between CT simulation and plan approval for internal and external plans. A point (x, y) on the cumulative curves indicates that y percent of plans were approved within x days.

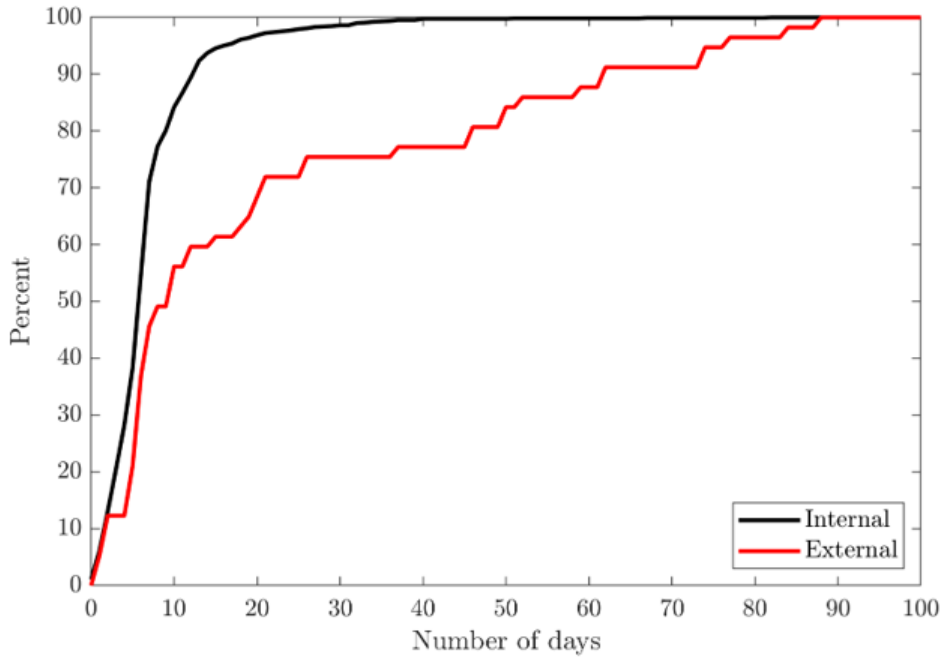


Figure 3.3: Number of days between CT simulation and plan approval for external-beam prostate radiotherapy treatment plans, expressed as cumulative distributions.

3.7.4 PTV margins

To validate our surface distance transform algorithm for computing margins, we created a custom workflow in MIM that returns the average Hausdorff distance between a pair of CTV and PTV contours. Figure 3.4 depicts a plot of both the automated and manual method results on the internal random sample. Error bars of one standard deviation are included for the automated algorithm. Interestingly, it appears that there is a slight bias in our methodology that underestimates the margins compared to the result in MIM, likely due to some inherent algorithmic discrepancy in contour voxelization or surface distance definitions.

Regardless, the average difference between algorithms is 0.24 ± 0.13 mm, a negligible amount that indicates a strong validation of our automated margin calculation methodology.

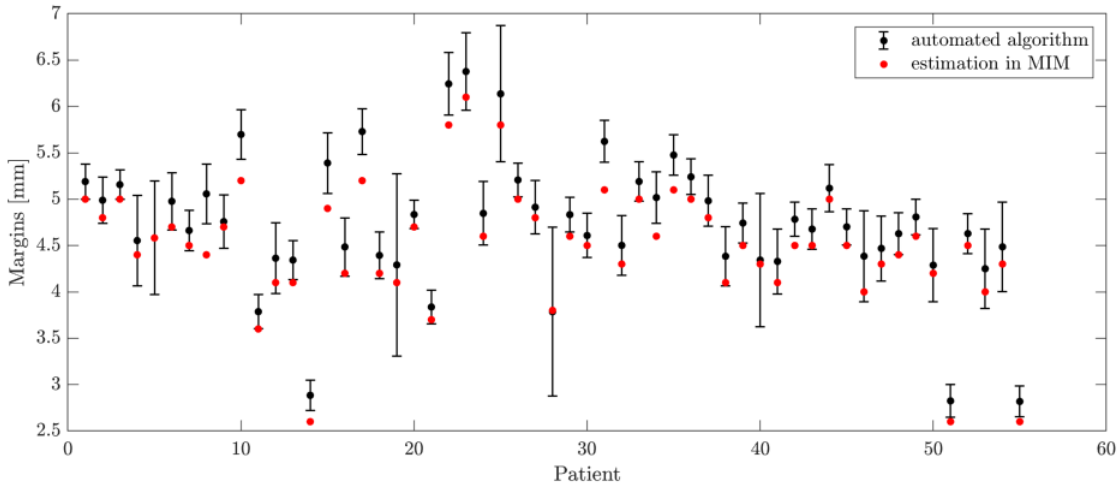


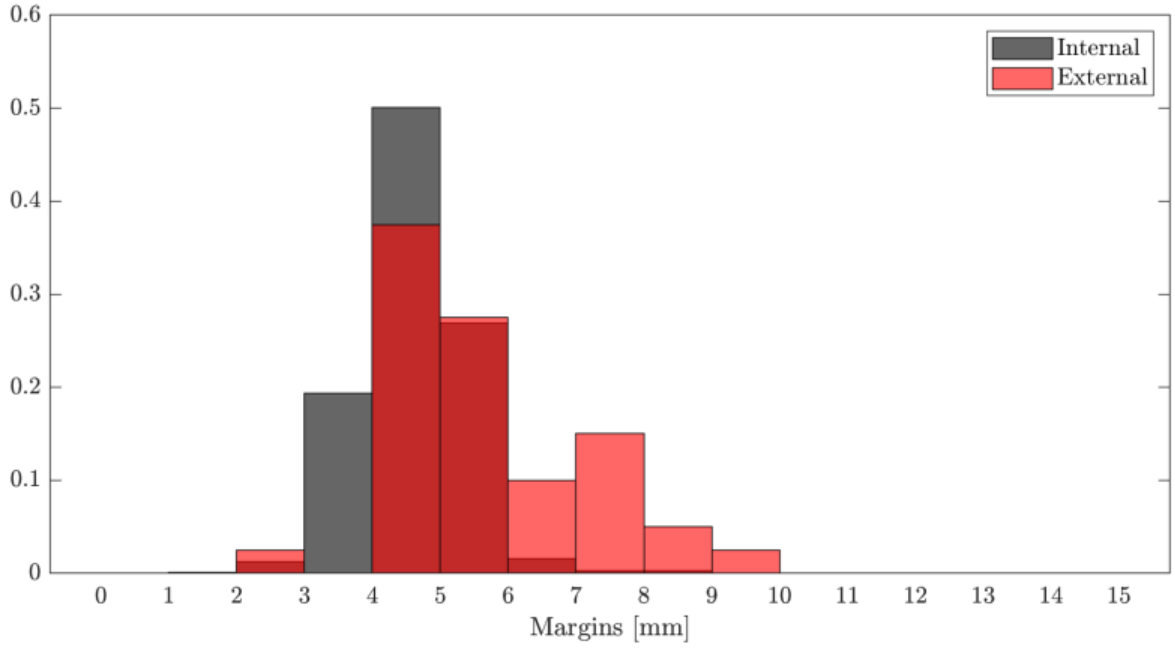
Figure 3.4: Comparison between our automated algorithm and our manual custom MIM workflow to compute PTV margins.

Figure 3.5 illustrates normalized histograms of the range of margins practiced at our institution and in the greater community. To evaluate the statistical significance of the differences in mean margins, we performed two-tailed independent t-tests [60]. Equal variance between the internal and external distributions was not assumed. The t-tests reveal that at a significance level of 95% ($\alpha = 0.05$), we have $p = 8.6 \times 10^{-5}$ for intact prostate margins and $p = 1.1 \times 10^{-2}$ for prostate bed margins. Therefore, the differences in mean margins are statistically significant.

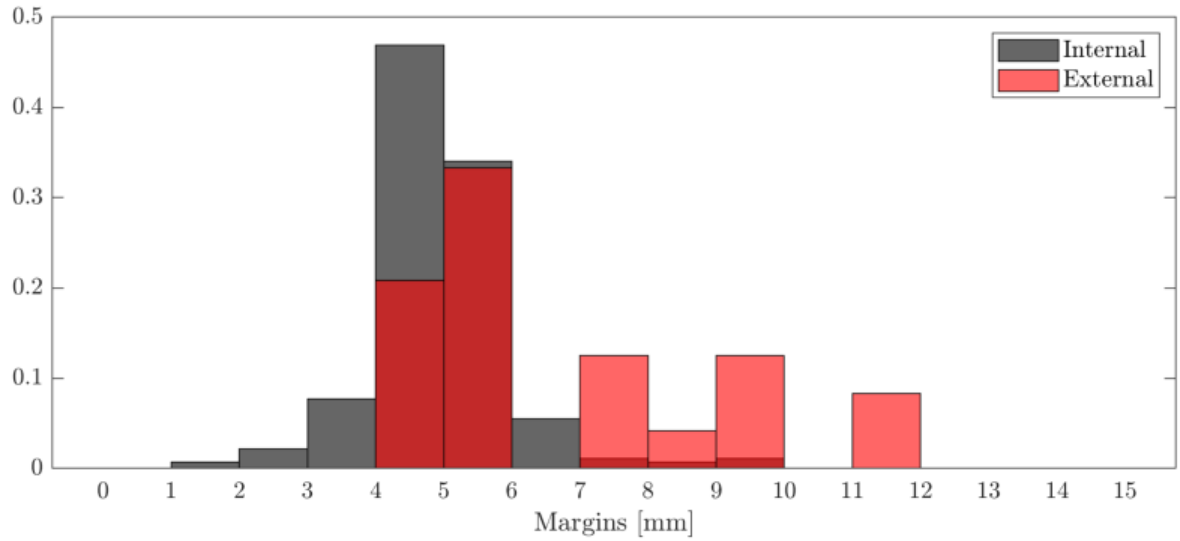
3.7.5 Fiducial markers

In Figure 3.6, we run our seed counting algorithm on half of the internal random sample, comprised of mostly intact prostate plans and two prostate bed plans for illustrative purposes. We obtained 100% seed counting accuracy on this sample.

Figure 3.7 shows normalized histograms of our seed counting algorithm applied to all intact prostate plans.



(a)



(b)

Figure 3.5: Normalized histograms of prostate PTV margins implemented in radiotherapy practice. (a) Intact prostate targets. (b) Prostate bed targets.

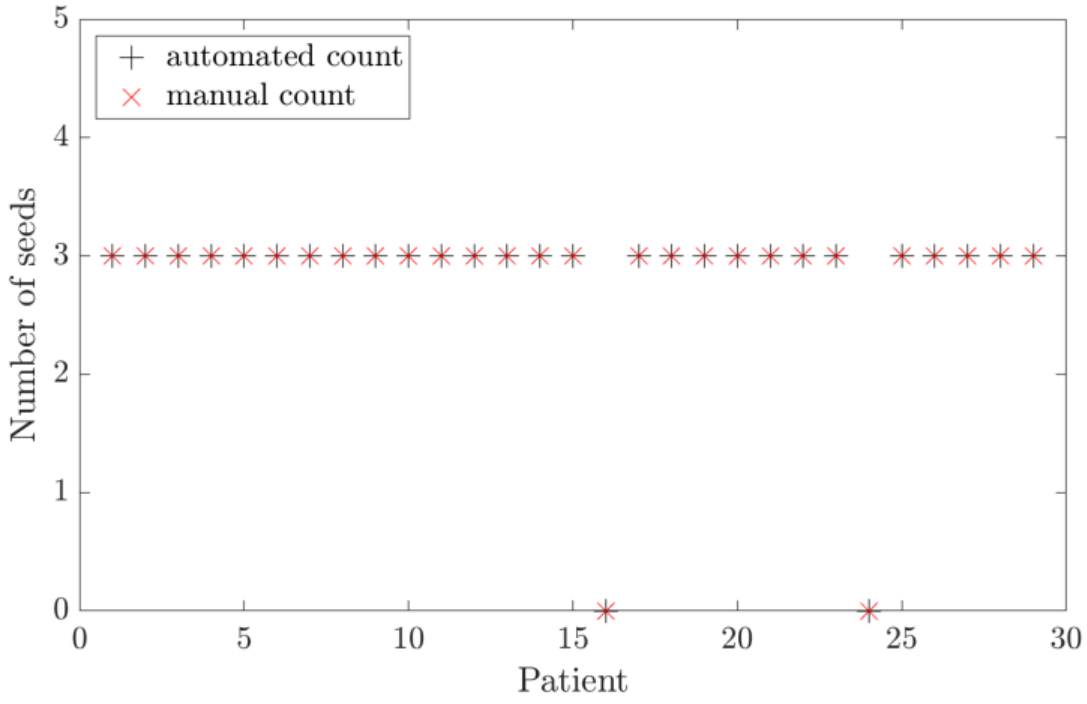


Figure 3.6: Comparison between our automated algorithm and a manual MIM workflow to compute PTV margins.

Internal cases with no seeds detected were reviewed. If the tumor was not on the prostate itself, such as for nodule and prostate-specific membrane antigen (PSMA) biomarker targets, then the case was removed from the plot. One case was also removed because the patient received a transurethral resection of the prostate (TURP).

Internal cases with more than five seeds detected corresponded to re-treatments of patients previously treated with low dose rate (LDR) brachytherapy.

3.7.6 Prescription doses

To validate our automated prescription dose ($\mathbb{R} = D95$) algorithm, we created a custom workflow in MIM that takes a DVH curve generated in MIM and returns its $D95$ point. For each internal sample patient, we also looked up the treatment intent prescription dose in the clinical treatment plan (CTP) document. Figure 3.8 depicts a plot of both the automated and manually recorded CTP treatment intent prescription results on the internal sample.

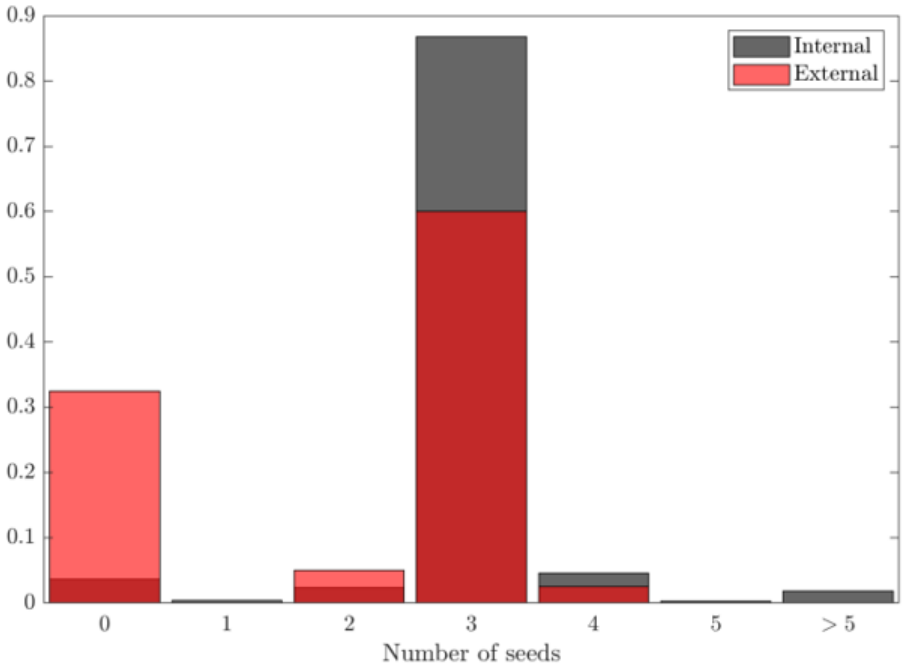


Figure 3.7: Normalized histograms of automated seed counting within intact prostate PTVs.

We found a perfect match between our automated methods and both manual methods of using the CTP document and the MIM workflow (the latter is not plotted). The average difference between the automated prescription and CTP treatment intent prescription was 0.050 ± 0.22 Gy.

In Figure 3.9, we depict normalized histograms of intact prostate prescription doses. Prescriptions in bold are listed in the VA consensus guidelines, whereas the rest are common prescriptions practiced at our institution.

Observe that the vast majority (approximately 80%) of internal intact prostate prescriptions were hypofractionated, indicating stereotactic body radiotherapy (SBRT) procedures delivering 40 Gy in 5 fractions. Meanwhile, the vast majority of external intact prostate prescriptions were conventionally fractionated.

Internal cases that did not match any of the prescription dose categories (up to a 5% tolerance) were reviewed. Some cases were removed from the plot because a boost was missing from our dataset. Recall that treatment plans involving magnetic resonance imaging

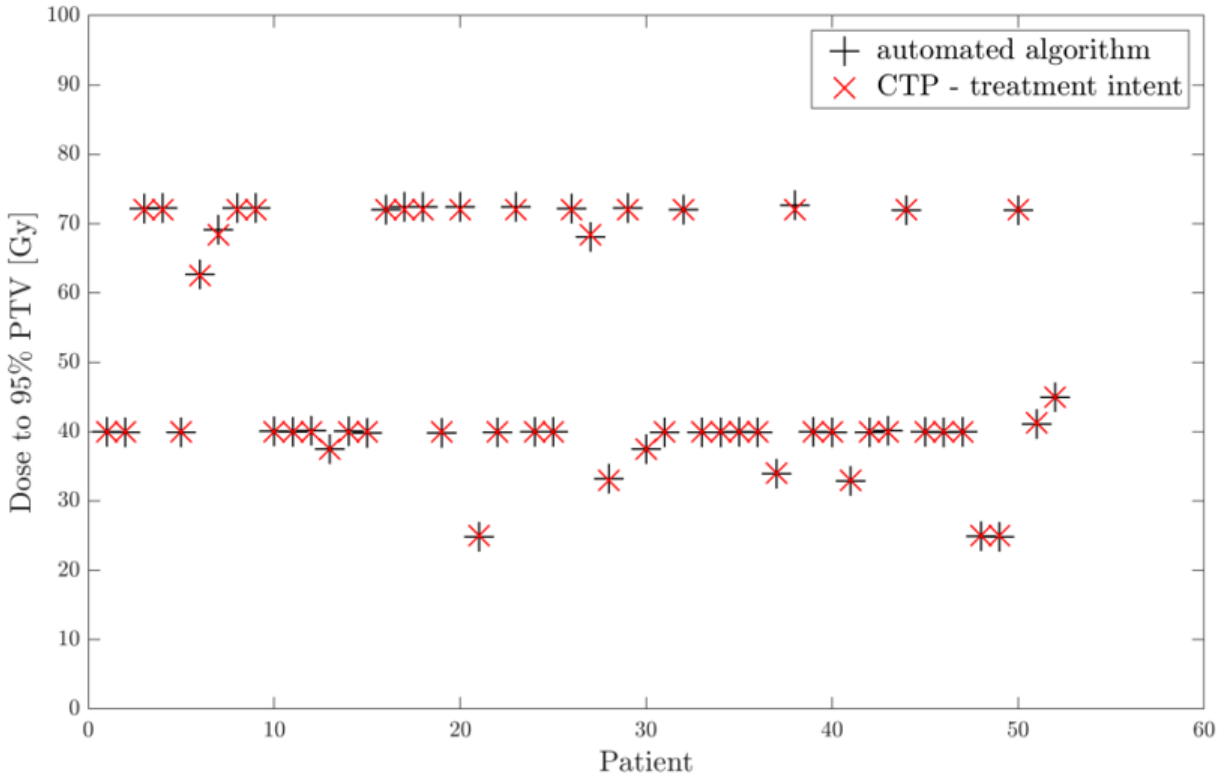


Figure 3.8: Comparison between our automated algorithm to compute prescription doses and a manual lookup of CTP treatment intents.

(MRI) were not considered in this study. Occasionally, a prostate patient may be transferred from our Varian TrueBeam STx and Novalis Tx radiotherapy linacs to our ViewRay MRIdian magnetic resonance guided linac (ViewRay, Inc., Denver, CO) and vice versa. Therefore, these patients were mistakenly flagged because dose distributions created on the ViewRay treatment planning system were never queried. Other reasons for false positive cases include urothelial cancers, recurrent cancers, wrong PTV selection (e.g., PSMA biomarker targets), and clinical trials with experimental doses (e.g., 24 Gy in 3 fractions).

Of particular emphasis, an expert medical physicist determined that one flagged internal cases corresponded to an unnoticed treatment planning error. The plan was intended to be an SBRT plan with 40 Gy delivered in 5 fractions. However, the plan was improperly normalized with a modified treatment percentage of 95% in Eclipse, resulting in a PTV overexposure by 2 Gy. This case was reported to our institution's incident learning system.

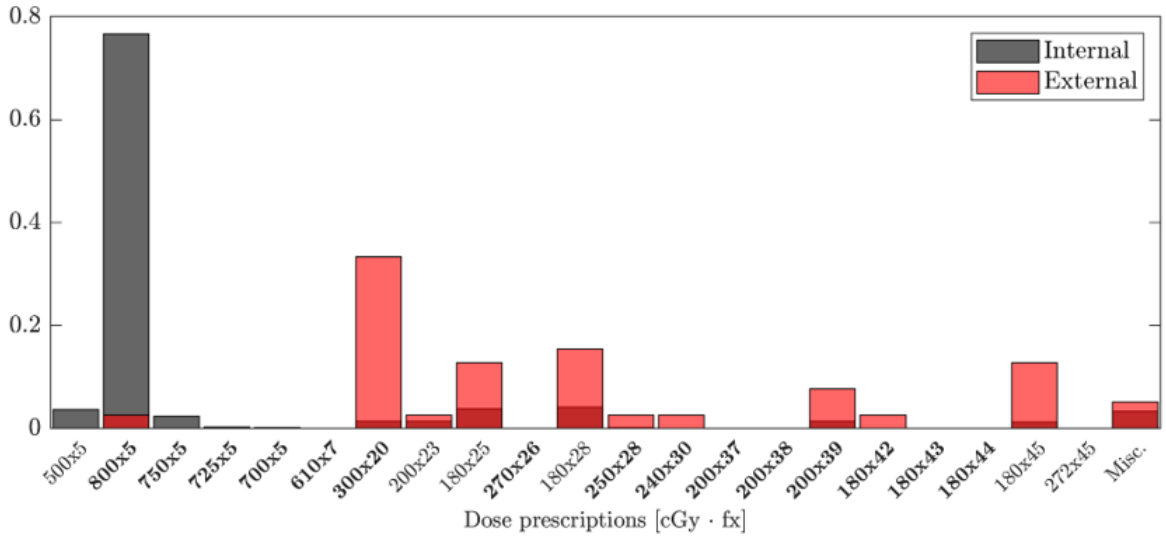


Figure 3.9: Normalized histograms of intact prostate prescription doses in units of cGy · fx. Prescriptions in bold are listed in the VA consensus guidelines.

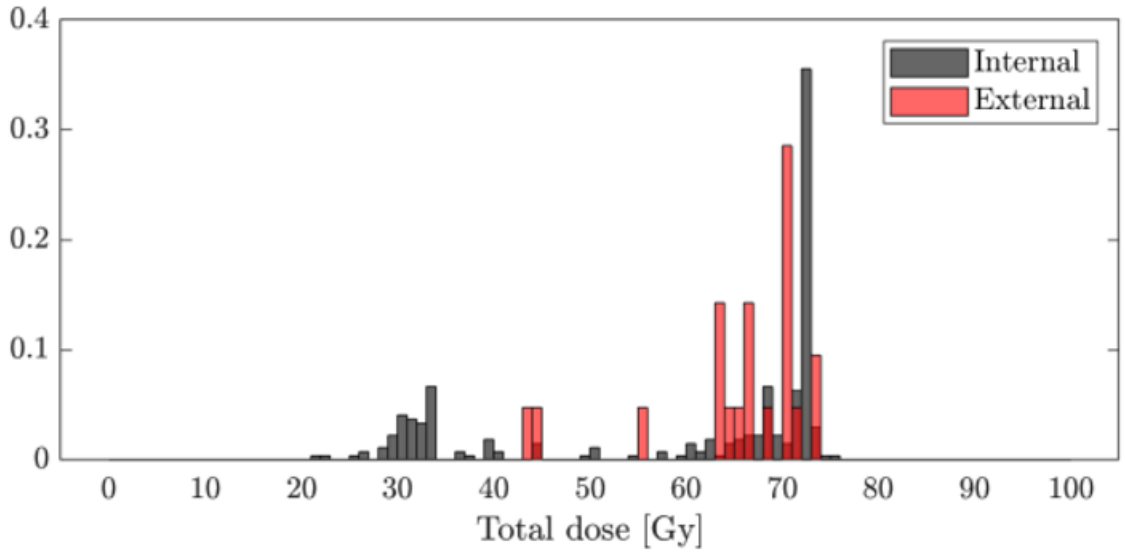
In Figure 3.10, we depict normalized histograms of prostate bed prescription doses. The VA publication recommends the following prescription ranges for prostatectomy external-beam radiotherapy patients:

- Total dose of $6400 \leq D \leq 7200$ cGy;
- Dose per fraction of $180 \leq D/F \leq 200$ cGy/fx.

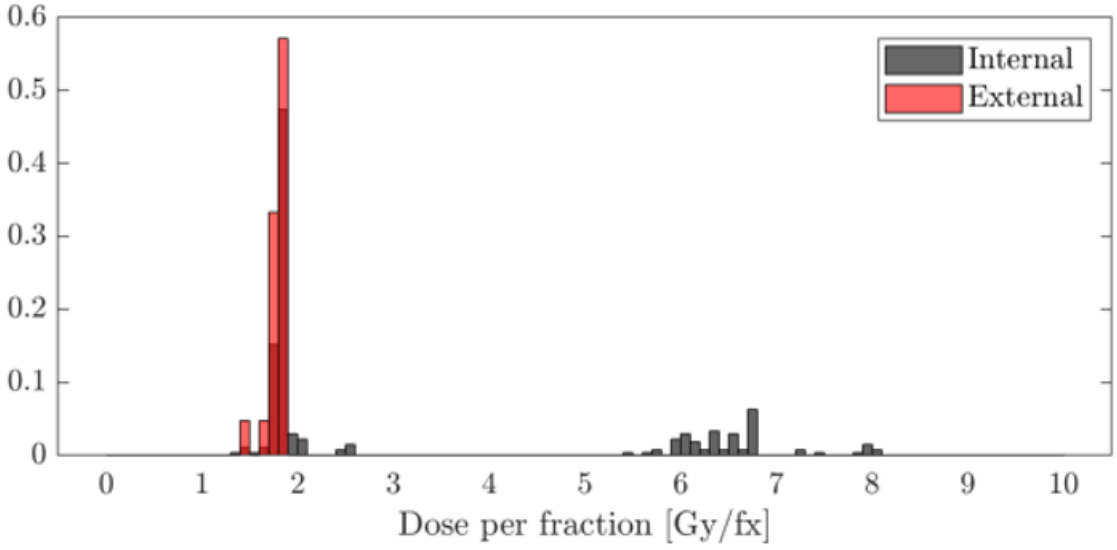
Observe that there is actually a broader range of standards of care among the internal cases than among the external cases, especially for the dose per fraction metric. A few cases with total doses between 40 Gy and 60 Gy were due to factors such as PTV under-coverage and unusual prescriptions (e.g., to prostate nodules).

3.7.7 PTV geometric descriptors

Finally, in Figure 3.11 we present the LLE manifold learning shape descriptors trained on internal cases and tested on external cases. We used a fixed hyperparameter of $K = 10$ nearest neighbors for intact prostate shape descriptor embeddings and $K = 5$ nearest neighbors for prostate bed shape descriptor embeddings. In total, we found one external intact outlier



(a)



(b)

Figure 3.10: Normalized histograms of prostate bed prescription doses. (a) Total dose in units of cGy. (b) Dose per fraction in units of cGy/fx.

and one external bed outlier, each taken with respect to a circle about the centroid enclosing 99% of internal datapoints.

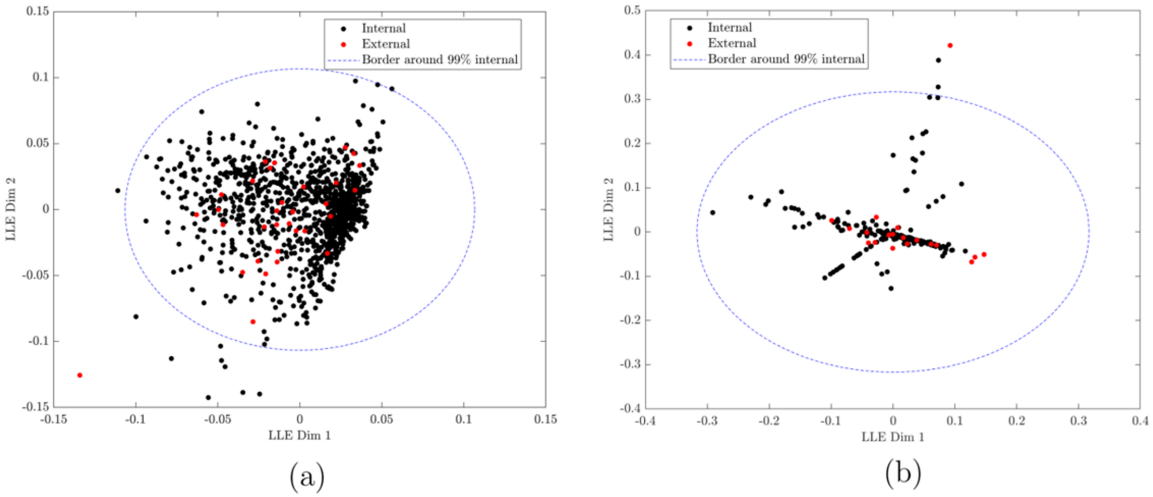


Figure 3.11: LLE dimensionality reduction of PTV shape descriptors onto a two-dimensional manifold Ω. Each blue dashed line represents a circle about the centroid enclosing 99% of internal datapoints. (a) Intact prostate plans. (b) Prostate bed plans.

3.8 Discussion

Our goal in this work was to develop a set of standardized, quantitative, and automatic measurements of radiotherapy plan quality. We performed quality measurements on our internal database of prostate treatment plans and a dataset of external community plans available at our institution. The results elucidated an observable range of standards of care in the external dataset. Observe that our quality metrics deliberately exclude those using volumetric dose distributions (including the quality of achieved dose to normal tissues and target volumes), as these measurements have been explored in KBP applications.

In order to flag a subset of external prostate plans for human review, we needed to fix certain thresholds of the specified quality indicators. The set of cutoff metrics that we applied in this study is summarized in the following list:

- Intact prostate PTV margins exceed 8 mm and seeds are detected;

- Intact prostate PTV margins less than 4 mm and no seeds are detected;
- Prostate bed PTV margins exceed 10 mm;
- Intact prostate PTV dose is not recognized as a standard prescription;
- Prostate bed PTV total dose or dose per fraction is outside the range specified by the VA consensus guidelines;
- Intact prostate PTV embedded shape descriptor vector is an outlier;
- Prostate bed PTV embedded shape descriptor vector is an outlier.

In our envisioned framework, plans flagged by the automated methods would be manually reviewed. Thus, the automated methods should have a high true-positive rate of detecting *low quality* plans while maintaining a sufficiently small false-positive rate such that the number of manually reviewed plans would not be unduly burdensome.

Our results demonstrate overall high accuracy of our automated methodology to compute plan quality measures. Failure to compute any of the measures would typically result in flagging the plan for human review, thus we believe that false-negatives would be rarer than false-positives.

Limitations of our investigation include a few items. First, for at least some quality indicators, no objective quality standard exists, which means that indicator cutoff values are currently exploratory. Second, we often do not have access to physician decision-making or additional clinical variables, especially for cases that are several years old. Hence, we cannot necessarily exclude the fact that deviations from standard practice were justified. Future improvements of our study could include addressing lymph node target volumes in prostate plans. We also foresee extending our methods to other anatomical treatment sites such as the head and neck.

3.9 Conclusion

Automated algorithms were developed to evaluate treatment planning quality for prostate radiotherapy patients. The algorithms were systematically applied on treatment plans from our academic institution and from hospitals in the greater community, which allowed us to compare and assess the existing range of standards of care in clinical practice. Precise cutoff metrics enabled treatment plan outliers to be flagged for retrospective human review. Our methods have future applications to a dedicated quality improvement project at the West Los Angeles VA Medical Center.

Chapter 4

Intact prostate and prostate bed target volume auto-contouring

4.1 Introduction

Our manifold learning approach to quantifying the range of target volume shape descriptors has led us to investigate target volume auto-segmentation with deep learning. Intact prostate delineation is more straightforward than prostate bed delineation, as there are clear anatomical boundaries of intact prostates visible in computed tomography (CT) images. However, prostate bed delineation for post-prostatectomy patients is more subject to inter-physician variability [102]. Seeing as there is no clinically-reliable software available to delineate prostate beds, our study develops a deep neural network for auto-segmentation trained on presumed high-quality clinical target volumes (CTVs) from our institution. The results of the network applied to contours from external community plans may be used as an automated prostate treatment planning quality evaluation tool.

For insightful presentations on existing prostate bed delineation guidelines for physicians, see [85, 76]. Depending on the tumor pathology, variations in anterior and posterior contour convexity exist throughout both the consensus guidelines and retrospective clinical data from

our institution. Therefore, the primary novelty of this study is to compare an overall model trained on prostate bed CTVs with multiple smaller models depending on the convexity of the anterior and posterior contours. We hypothesize that subdividing the dataset into categories of similar geometric descriptors will lead to improved performance of deep learning models. A related study investigated physician-based model subdivision and achieved improvements in auto-segmentation accuracy [5], whereas our method is purely data-driven and achieves similar state-of-the-art performance. Before proceeding to the methodology on prostate bed auto-segmentation, let us first analyze intact prostate segmentation with clinically available software.

4.2 MIM auto-contouring routine

Our institution has access to the clinically-approved auto-contouring software in MIM called Contour ProtégéAI+ (MIM Software, Inc., Cleveland, OH). Given a particular prostate radiotherapy (RT) study, the CT image may be exported to a designated server that generates an RTStruct with artificial intelligence (AI). We opened the physician-approved RTStruct alongside the AI RTStruct and measured the Sørensen-Dice similarity coefficients between CTVs using a custom MIM workflow. Oftentimes, the CTV automatically identified in §3.4.2 included proximal seminal vesicles (SVs). In these cases, we defined AI generated proximal SVs as the intersection between the AI prostate contour expanded by 1 cm and the AI SV contour.

A random sample of 30 intact prostate CTVs and all available 35 external intact prostate CTVs (see Table 3.2) were evaluated using Contour ProtégéAI+.

4.3 Deep learning dataset

4.3.1 Data selection

The dataset in this study consisted of prostate bed CTVs. Initial data collection and selection steps were equivalent to §3.2 and §3.3, respectively. To further regulate the data for training and inference purposes, we performed a handful of additional filters. First, we required that structure sets associated with boost plans were ignored in favor of initial plans and simultaneous integrated boosts (SIBs). In accordance with our institution's RTPlan naming convention, substrings of "i1" and "s1" in the RT plan label were used to determine the plan type of initial or SIB, respectively. Then, the Sørensen-Dice similarity coefficient between the expanded CTV and the planning target volume (PTV) computed in §3.4.2 was required to be at least 90%, which would indicate reasonably isotropic margins. Finally, we regulated the CTV datapoints by their geometric descriptors of volume, sphericity (Equation 3.6), and solidity (Equation 3.7) in order to avoid outlier targets. These criteria are summarized in Table 4.1.

Selection step	Number of patients	Number of plans
Internal bed	288	421
Initial or SIB	255	271
$D > 0.90$	236	248
$100 \leq V \leq 350$ [cm^3]	210	222
$0.7 \leq G_{\text{sph}} \leq 0.9$	194	205
$0.7 \leq G_{\text{sol}} \leq 0.9$	186	197

Table 4.1: Additional selection criteria that were applied prior to generating a dataset of prostate bed CTVs for deep learning auto-segmentation.

Our subdivision of the selected post-prostatectomy patients into training, validation, and testing groups is summarized in Table 4.2. The subdivision was achieved by permuting the list of patient medical record numbers (MRNs) using a pseudo-random number generator. We ensured that each datapoint belonged to a unique patient, hence the number of distinct CTV datapoints was equivalent to the number of patients.

Data subset	Number of datapoints
Training	149
Validation	19
Testing	18

Table 4.2: Number of prostate bed CTV datapoints belonging to unique post-prostatectomy radiotherapy patients placed in training, validation, and testing subsets.

Finally, we subdivided the training, validation, and testing data based on geometric shape descriptors. Anterior and posterior halves of the binary CTV masks with respect to their centroids were acquired. For both regions, two-dimensional axial measurements of G_{sol} were obtained, and the minimum values over all slices were recorded. Among all the minimum two-dimensional anterior and posterior solidities for the training and validation data, the 50^{th} percentile solidities were calculated. This procedure enabled us to devise four subsets for more refined model development, which we colloquially referred to as *anterior concave*, *anterior convex*, *posterior concave*, and *posterior convex* V-Net models.

4.3.2 Datapoint generation

Since the number of trainable parameters in a deep convolutional neural network (CNN) increases dramatically with the size of the input tensor, it was helpful in our methodology to use images with a small slice thickness and a small amount of in-plane pixels. For every datapoint, we opened the volumetric CT, converted to Hounsfield units (HU), and resampled to a uniform isotropic grid spacing of 1 mm per voxel. Then, we opened the CTV that was identified automatically in §3.4.2. Subsequently, we semi-automatically identified bladder and rectum organs-at-risk (OARs) by searching for substrings "O_bladder", "O_rectum", and abbreviations therein, in the region-of-interest (ROI) names. These substrings were typically sufficient to find unique OARs in the structure set, although sometimes manual assistance was necessary. The ROI contour points for the OARs were converted into binary masks on the resampled CT grid.

In order to assist in reducing input tensor sizes, we first resized the CT, CTV, and bladder and rectum OARs to a lower-resolution of 2 mm per voxel. The CT image was resized using cubic interpolation, whereas the binary masks were resized using nearest neighbor interpolation. All images were padded or cropped such that the in-plane grid sizes were momentarily set to 512×512. The CT image and binary masks were then cropped to an in-plane size of 256×256 about the centroid of the CTV. The number of overlapping k-sliced image stacks N was determined according to the formula

$$N = \left\lfloor \frac{S_{\text{axial}} - k}{S_{\text{stride}}} \right\rfloor + 1,$$

where S_{axial} is the number of axial slices intended to be covered and S_{stride} is the stride of the overlapping stacks. In this study, we chose to cover 64 axial slices about the centroid of the CTV with a stack size of 32 slices and a stride of 1 for both training and inference data. Therefore, we recorded $N = 33$ overlapping image stacks for every unique patient.

4.3.3 V-Net data loading

At this stage in the methodology, our images had dimensions $256 \times 256 \times 32$. During the data loading pre-processing stage, we reduced the in-plane dimensions to 128×128 pixels and apply data augmentations. For each datapoint, the CT image was normalized to a mean of zero and a standard deviation of one. We randomly applied reflections in the horizontal direction to the set of CT, CTV, and OARs. Next, we randomly applied small rotations to the set of images of degree $|\theta| \leq 2^\circ$ about the image centers. Then, we randomly applied translations $|t_{x,y}| \leq 10$ pixels in the horizontal and vertical directions independently. Finally, the set of images was cropped to 128×128 pixels, as desired.

4.4 V-Net architecture

The V-Net models were developed using the open-source library TensorFlow (Google AI, Mountain View, CA) with the open-source frontend interface Keras. Training was implemented on an NVIDIA RTX A5000 graphics processing unit (GPU) (Nvidia Corporation, Santa Clara, CA). Refer to Figure 4.1 for details on the deep learning architecture. There are two phases of a V-Net neural network, namely an encoder phase that transforms the input into a latent representation, and a decoder phase that produces the desired segmentation prediction. Skip connections (involving tensor addition and channel concatenation) allowed for higher-level features in the encoder layers to be retained in the decoder layers. Besides the final layer, all convolution layers were succeeded by rectified linear unit (ReLU) activation functions and batch normalization (BN). In order to arrive at probabilities for a binary mask prediction, the final convolution consisted of a single filter and a sigmoid activation function. In general, binary classification tasks in Keras can be achieved either through one-hot encoding with a softmax activation and categorical cross-entropy loss, or through vector encoding with a sigmoid activation and binary cross-entropy loss. Down-sampling (i.e, pooling) and up-sampling were carried out entirely via convolutional operations.

The tensor sent to the V-Net consisted of two channels, the first being the CT and the second being the sum of the bladder and rectum binary masks. The ground truth was the CTV binary mask. A probability density function (PDF) of prostate bed CTV binary mask predictions was output.

Training parameters included an Adam optimizer [61] with a learning rate of 10^{-4}, a batch size of 16, a step of 20 iterations per epoch, and a maximum number of 200 epochs.

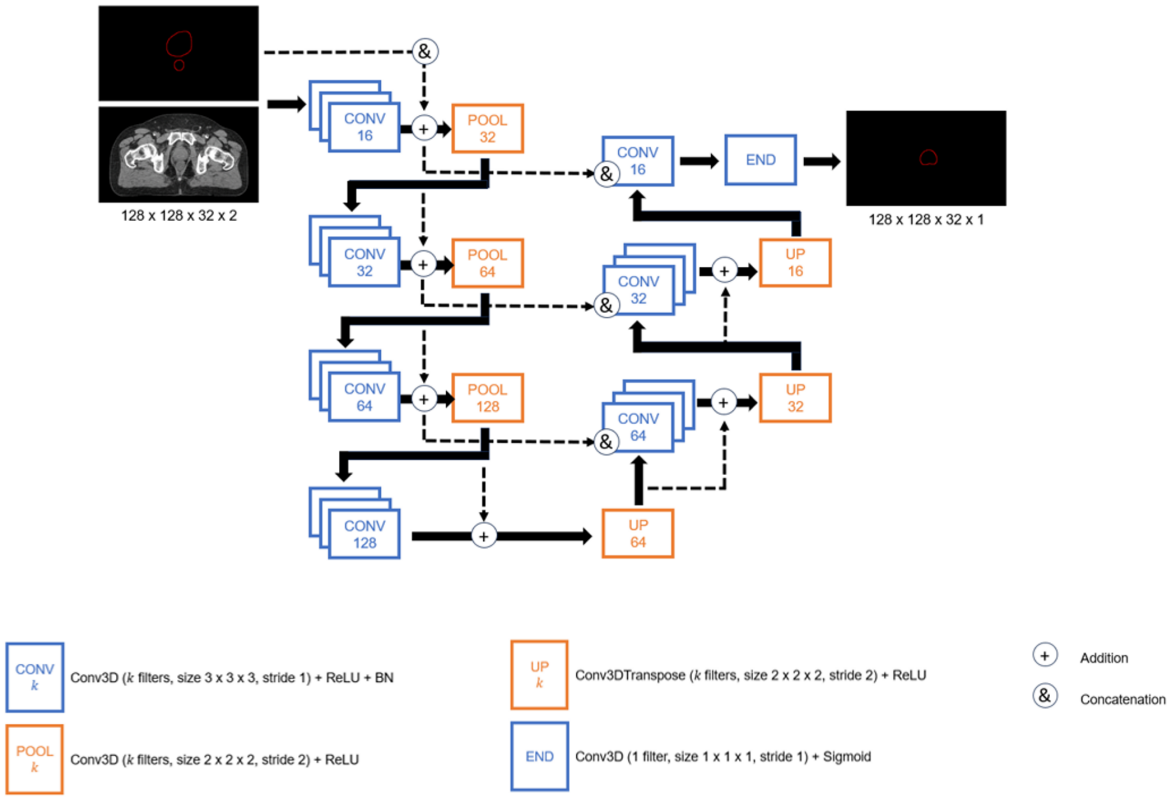


Figure 4.1: Schematic of the V-Net deep learning architecture that we implemented for training prostate bed auto-segmentation models.

4.5 Objective function

The objective function (i.e., loss function) that we used to train the V-Net models is composed of two terms. The first term is a standard Sørensen-Dice loss

$$L_{\text{Dice}}(X, Y) = 1 - 2 \cdot \frac{|X \cap Y|}{|X| + |Y|} \tag{4.1}$$

(cf., Equation 3.1) that was successfully learned by the original V-Net introduced in [77].

The second term is a boundary loss that computes Euclidean surface distance transforms between the ground truth and predicted contours. The authors in [56] argue for a boundary loss function that involves a signed *level set* representation of the surface distances. Given a

subset $X \subset \mathbb{R}^3$, define a binary indicator function $\iota_X : \mathbb{R}^3 \to \{0, 1\}$ for which

$$\iota_X(x) = \begin{cases} 1 & x \in X \\ 0 & \text{otherwise.} \end{cases} \tag{4.2}$$

Recall the distance transform map T_X (Equation 3.2) and boundary subset ∂X (Equation 3.3) applied in §3.6.2 for computing contour margins. Define a signed boundary distance map ϕ_X for X according to

$$\begin{aligned} \phi_X &: \mathbb{R}^3 \to \mathbb{R} \\ x \mapsto \phi_X(x) &= \begin{cases} -T_{\partial X}(x) & x \in X \\ T_{\partial X}(x) & \text{otherwise.} \end{cases} \end{aligned} \tag{4.3}$$

In other words, the signed distance between a point $x \in \mathbb{R}^3$ and a boundary ∂X is negative if $x \in X$ and positive if $x \notin X$. Interpreting X as the ground truth binary mask and Y as the predicted binary mask, the desired boundary loss term is

$$\begin{aligned} L_{\text{boundary}}(X, Y) &= \int_{\mathbb{R}^3} \mathrm{d}x \, \phi_X(x) \iota_Y(x) \\ &= \int_Y \mathrm{d}x \, \phi_X(x). \end{aligned} \tag{4.4}$$

The Dice loss is intended to be the dominant component of the total objective function L, whereas the boundary loss is a regularizer. Hence, we should multiply the boundary loss term by a small weighting parameter α, possibly scheduled as a function of the current training epoch number. All together, we have

$$\begin{aligned} L(X, Y) &= 1 - 2 \cdot \frac{|X \cap Y|}{|X| + |Y|} && + \alpha \cdot \int_Y \mathrm{d}x \, \phi_X(x) \\ &= L_{\text{Dice}}(X, Y) && + \alpha \cdot L_{\text{boundary}}(X, Y). \end{aligned} \tag{4.5}$$

4.6 Inference post-processing

For the inference (i.e., testing) datapoints, post-processing operations were applied prior to evaluating the Sørensen-Dice similarity coefficients with the ground truth CTVs. First, we devised a *majority vote* algorithm among the overlapping stack predictions. For each stack, we ran the V-Net model and output the tensor of probabilities. To yield a binary mask prediction, any voxel with probability exceeding 0.5 was mapped to 1, otherwise the voxel was mapped to 0. Another tensor accumulated the predictions over all stacks.

By inspection, the number of stacks overlapping a given slice (i.e., the number of model predictions) across the original 64 axial slices goes from 1 to 32, then from 32 back down to 1, in unit increments. Alternatively, the number of overlapping stacks may be formulated by a convolution whose kernel may be adjusted (e.g., to favor the predictions from middle slices). Dividing the number of model predictions in half yields a majority vote number for each slice. Given a particular slice, if the sum of all predictions for a voxel exceeds the majority vote number of the slice, then the final prediction for that voxel was mapped to 1, otherwise the voxel was mapped to 0. The final prediction is given by the middle 32 slices of this tensor.

The second post-processing operation that we applied was to remove *islands* from the V-Net prediction. We fixed an area parameter of 100 pix, so that any connected component with area less than 100 pixels was removed from the predicted mask.

Finally, we used a morphological closing operation to smooth the predicted contour. For computational efficiency, we closed each axial slice separately. We fixed a circle of radius 5 pixels for the morphological structuring element.

4.7 Results

4.7.1 Intact prostate contours

Figure 4.2 depicts normalized histograms of the MIM Contour ProtégéAI+ accuracy when applied to the internal sample and external intact prostate plans. Mean accuracy (as quantified by the Sørensen-Dice similarity coefficient) of the AI generated intact prostate contours was 0.84 on the internal sample and 0.82 on the external data. To evaluate the discrepancy in mean accuracy, we performed a two-tailed independent t-test [60]. Equal variance between the internal and external distributions was not assumed. The t-test revealed that at a significance level of 95% ($\alpha = 0.05$), we have $p = 0.11$, which implies there is no statistically significant difference. This result reflects the fact that intact prostate delineation is reasonably straightforward on CT images. Based on the figure, one could surmise a Dice coefficient cutoff of 0.7, below which a plan would be automatically flagged for human review.

4.7.2 Prostate bed contours

In the experimental results shown here, we used a fixed boundary loss weighting of $\alpha = 10^{-2}$. An initial model using all training and validation data from Table 4.2 was trained initially. Then we trained the anterior concave, anterior convex, posterior concave, and posterior convex models, each of which had about half as much data by construction. The refined model was defined to encapsulate the range of predictions from all sub-models by taking the prediction with maximal Dice coefficient. As we hypothesized, the Dice coefficients improved between the initial and refined models, indicating that the sub-models frequently outperformed the initial model.

Figures 4.3 and 4.4 show boxplots of these results. For the internal data, the mean Dice coefficient was 0.79 for the initial model and 0.83 for the refined aggregate model. A two-tailed paired t-test at a significance level of 95% was performed to compare the mean accuracies. We found that $p = 9.3 \times 10^{-2}$, which indicates the improvement in mean

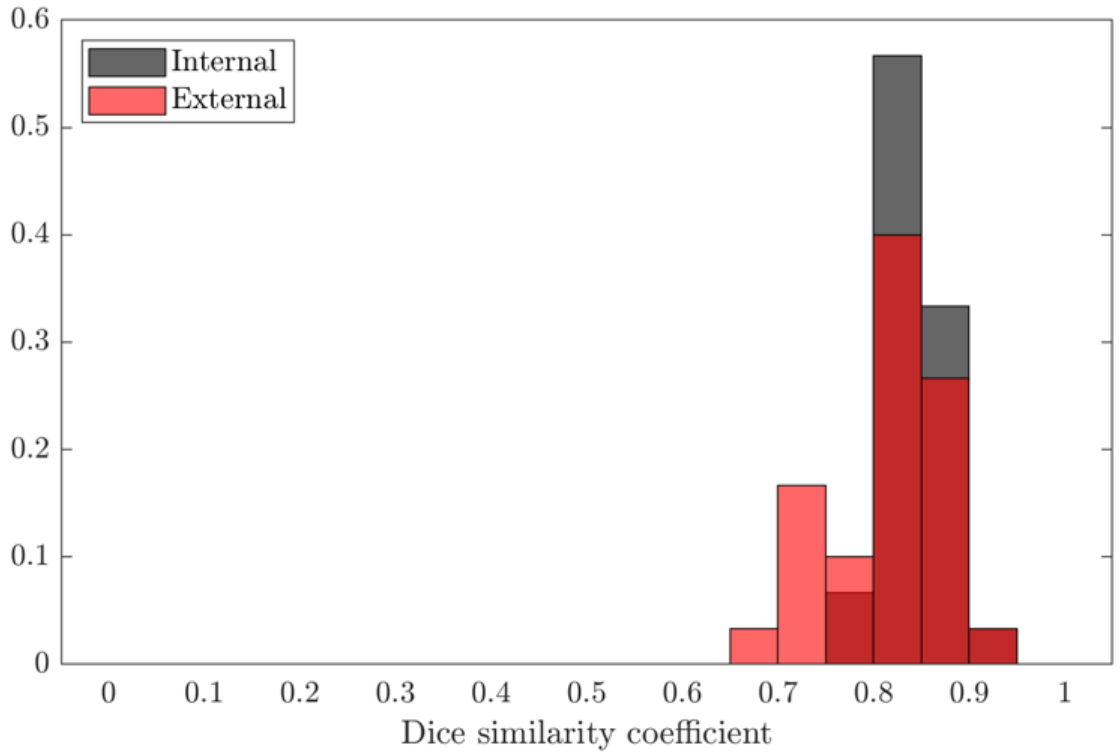


Figure 4.2: Normalized histograms of MIM Contour ProtégéAI+ accuracy on intact prostates, as quantified by Sørensen-Dice similarity coefficients. The difference in mean accuracy between the internal sample data and the external data was not statistically significant, which is evidence of the fact that intact prostate delineation is reasonably straightforward on CT images.

accuracy is statistically significant. For the external data, the mean Dice coefficient was 0.60 for the initial model and 0.63 for the refined aggregate model. Another t-test revealed that $p = 6.4 \times 10^{-2}$, which indicates the improvement in mean accuracy is statistically significant. It is interesting to point out that the overall accuracies are much lower on the external data, which highlights the vast range in standards of care between our institution and community clinics.

Finally, Figure 4.5 shows normalized histograms of the refined V-Net model accuracies on the internal and external testing data.

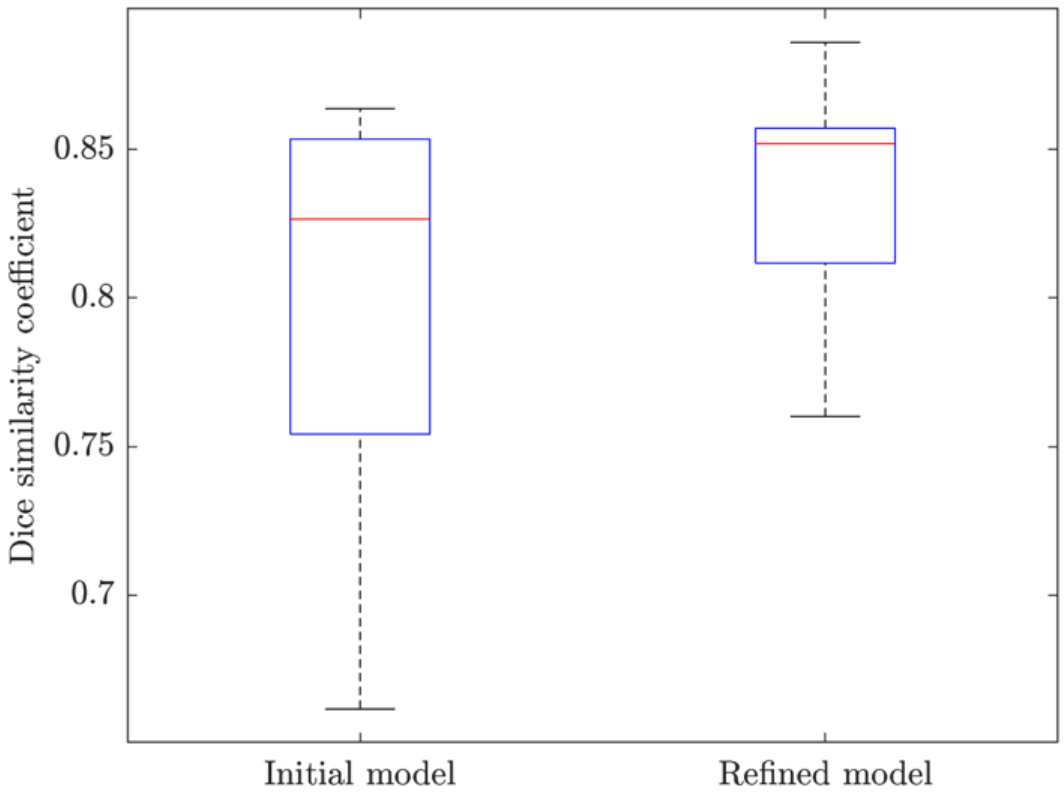


Figure 4.3: Boxplot of the initial and refined V-Net models run on the internal testing data. The initial model was trained and validated on all internal data selected for this study. On the other hand, the refined model took into account the *anterior concave*, *anterior convex*, *posterior concave*, and *posterior convex* V-Net sub-models, which were trained on subsets separated by 50^{th} percentile solidities.

4.8 Conclusion

In this chapter, we developed state-of-the art prostate bed auto-segmentation models for quality review applications. The models were trained on data-driven subsets of the queried prostate bed plans, specifically subsets based on the anterior and posterior convexity of the CTV. On the internal and external testing data, we observed statistically significant improvements in auto-segmentation accuracy using the refined sub-models compared with the initial model trained on all prostate bed plans. Potential low-quality CTV delineations among external community plans were identified for manual review.

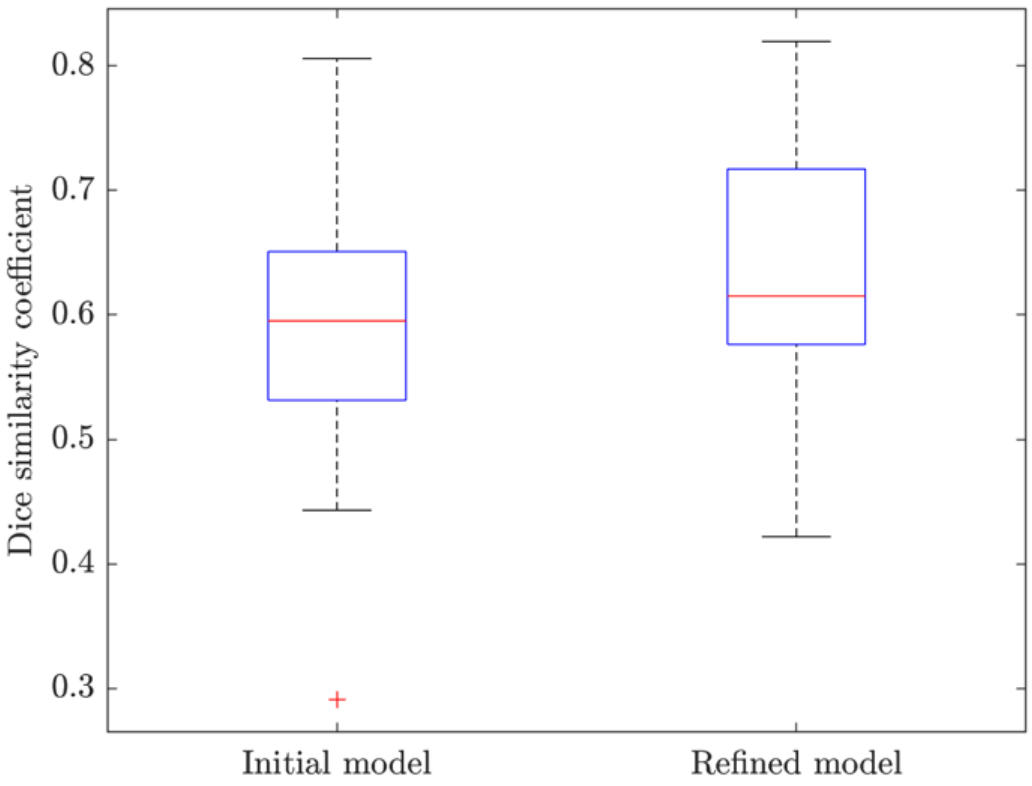


Figure 4.4: Boxplot of the initial and refined V-Net models run on the external testing data. The initial model was trained and validated on all internal data selected for this study. On the other hand, the refined model took into account the *anterior concave*, *anterior convex*, *posterior concave*, and *posterior convex* V-Net sub-models, which were trained on subsets separated by 50^{th} percentile solidities.

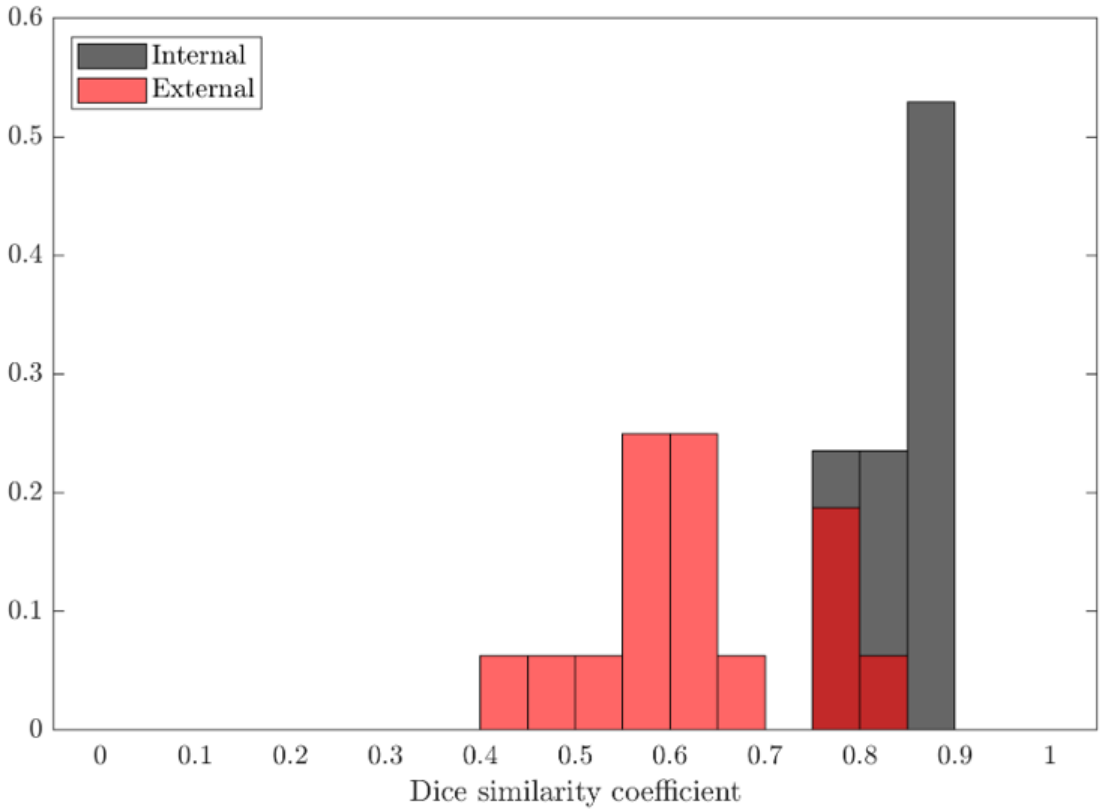


Figure 4.5: Normalized histograms of the refined V-Net model accuracies on the internal and external testing data.

Appendix A

Siddon-Jacobs ray tracing

Our implementation of Siddon-Jacobs ray tracing for DRR rendering closely follows the methodology in [42]. Here we provide details on the algorithm.

Suppose we have a rectangular cuboid with dimensions $N_x, N_y, N_z \in \mathbb{N}$ divided into regularly-spaced grid points with voxel spacing $\Delta_x, \Delta_y, \Delta_z \in \mathbb{R}^+$. Then the coordinates of the grid points are given by

$$(i\Delta_x, j\Delta_y, k\Delta_z), \quad i,j,k \in \mathbb{N}, \quad 0 \leq i < N_x, \quad 0 \leq j < N_y, \quad 0 \leq k < N_z. \tag{A.1}$$

Here we assume for simplicity that the origin coincides with a corner of the volume.

A line segment of interest, which may or may not pass through the volume, is defined by distinct endpoints

$$\mathbf{p}_a = (p_{a,x}, p_{a,y}, p_{a,z}), \quad \mathbf{p}_b = (p_{b,x}, p_{b,y}, p_{b,z}), \quad \mathbf{p}_a, \mathbf{p}_b \in \mathbb{R}^3.$$

The line segment may be parameterized by path length according to

$$\mathbf{p}(\gamma) = \mathbf{p}_a + \frac{\gamma}{L}(\mathbf{p}_b - \mathbf{p}_a), \quad \gamma \in \mathbb{R}, \quad 0 \leq \gamma \leq \ell, \quad \ell = \|\mathbf{p}_b - \mathbf{p}_a\|. \tag{A.2}$$

Hence, we may assign parametric values $\boldsymbol{\gamma} = (\gamma_x, \gamma_y, \gamma_z) \in \mathbb{R}^3$ at the grid points, where

$$\gamma_x(i) = \ell \cdot \frac{i\Delta_x - p_{a,x}}{p_{b,x} - p_{a,x}}, \quad \gamma_y(j) = \ell \cdot \frac{j\Delta_y - p_{a,y}}{p_{b,y} - p_{a,y}}, \quad \gamma_z(k) = \ell \cdot \frac{k\Delta_z - p_{a,z}}{p_{b,z} - p_{a,z}}.$$

The minimum and maximum parametric values where the line segment intersects with the surface of the rectangular cuboid are then given by

$$\begin{aligned} \gamma_{\min} = \max\Big\{ & 0, \\ & \min\{\gamma_x(0), \gamma_x(N_x - 1)\}, \\ & \min\{\gamma_y(0), \gamma_y(N_y - 1)\}, \\ & \min\{\gamma_z(0), \gamma_z(N_z - 1)\} \Big\} \end{aligned}$$

and

$$\begin{aligned} \gamma_{\max} = \min\Big\{ & \ell, \\ & \max\{\gamma_x(0), \gamma_x(N_x - 1)\}, \\ & \max\{\gamma_y(0), \gamma_y(N_y - 1)\}, \\ & \max\{\gamma_z(0), \gamma_z(N_z - 1)\} \Big\}. \end{aligned}$$

Care must be taken if any of the denominators in $\boldsymbol{\gamma}$ vanishes. In particular, undefined values ought to be ignored while computing $\gamma_{\min}$ and $\gamma_{\max}$.

To find the indices of the first voxel intersected by the line segment, we introduce

$$\mathbf{p}_{\min} = \mathbf{p}(\gamma_{\min})$$

and then take

$$i_{\min} = \frac{p_{\min,x}}{\Delta_x}, \quad j_{\min} = \frac{p_{\min,y}}{\Delta_y}, \quad k_{\min} = \frac{p_{\min,z}}{\Delta_z}.$$

By inspection, there are two cases to consider for each component. Let us illustrate with the x-component:

- If $i_{\min} \in \mathbb{N}$ and $p_{b,x} < p_{a,x}$, then subtract one

$$i_{\min} \mapsto i_{\min} - 1;$$

- Otherwise, round down

$$i_{\min} \mapsto \lfloor i_{\min} \rfloor.$$

Repeat for the y- and z-components. This process ensures that $i_{\min}, j_{\min}, k_{\min} \in \mathbb{N}$ are the intended voxel indices. Let $\mathbf{v}_{\min} = (i_{\min}, j_{\min}, k_{\min})$.

Next, we find the parameters where the line segment first intersects parallel planes inside the volume. Let i', j', k' be the indices of the grid coordinates nearest to $\mathbf{p}_{\min}$, such that

$$p_{\min,x} \leq i'\Delta_x, \quad p_{\min,y} \leq j'\Delta_y, \quad p_{\min,z} \leq k'\Delta_z.$$

For each component, subtract one if the ray travels backward. In other words, if $p_{b,x} < p_{a,x}$, then $i' \mapsto i' - 1$, and likewise for the y- and z-components. The associated parametric vector is $\boldsymbol{\gamma}' = (\gamma_x(i'), \gamma_y(j'), \gamma_z(k'))$. Any undefined parameters ought to be replaced with $\gamma_{\max} + 1$.

The entire ray tracing algorithm may now be summarized according to the following steps:

1. Initialize $\mathbf{v} = \mathbf{v}_{\min}$ and $\boldsymbol{\alpha} = \boldsymbol{\gamma}'$.

2. Let $\mu = \operatorname{argmin}\{\alpha_x, \alpha_y, \alpha_z\}$ be the axis of the smallest parametric component of $\boldsymbol{\alpha}$.

3. If $\alpha_\mu > \gamma_{\max}$, then conclude ray tracing. Otherwise, increment α_μ by

$$\alpha_\mu \mapsto \alpha_\mu + \left| \frac{\ell}{p_{b,\mu} - p_{a,\mu}} \right|. \tag{A.3}$$

4. Increment to the next adjacent voxel,

$$v_\mu \mapsto \begin{cases} v_\mu + 1 & p_{a,\mu} < p_{b,\mu} \\ v_\mu - 1 & p_{a,\mu} \geq p_{b,\mu} \end{cases}. \tag{A.4}$$

Suppose the ray intersects a particular voxel $\mathbf{v}(M)$ at iteration $M \in \mathbb{N}$ of the algorithm. Let $\alpha_\mu(M)$ be the value of α_μ at iteration M *prior* to incrementing. Then the distance that the ray travels inside $\mathbf{v}(M)$ is simply given by the parametric difference $\alpha_\mu(M) - \alpha_\mu(M-1)$. Note that to find the distance inside the first voxel, we need to declare $\alpha_\mu(-1) = \gamma_{\min}$. For the purpose of DRR rendering, these distances may be weighted by linear attenuation coefficients to obtain a radiological path length.

Appendix B

RQ decomposition

Applied mathematics students are undoubtedly familiar with QR decompositions. Here we provide a brief overview of QR decompositions, and then we highlight a slight variant that is more pertinent to projectional geometry, namely RQ decompositions.

Consider a standard inner product space $\mathbb{C}^N$ with $\langle \mathbf{u}, \mathbf{v} \rangle = \mathbf{u}^\dagger \mathbf{v}$. Suppose we are given an $N \times N$ matrix

$$\mathbf{A} = \begin{pmatrix} \mathbf{a}_1 & \cdots & \mathbf{a}_N \end{pmatrix} \in \mathcal{M}_N(\mathbb{C})$$

acting on $\mathbb{C}^N$. Using the Gram-Schmidt process with the familiar projection map

$$\text{proj}_{\mathbf{u}} \mathbf{v} = \frac{\langle \mathbf{u}, \mathbf{v} \rangle}{\langle \mathbf{u}, \mathbf{u} \rangle} \mathbf{u}, \tag{B.1}$$

we can recursively define an orthonormal basis

$$\mathbf{e}_j = \frac{\mathbf{u}_j}{\|\mathbf{u}_j\|}, \quad \mathbf{u}_j = \mathbf{a}_j - \sum_{i=1}^{j-1} \text{proj}_{\mathbf{u}_i} \mathbf{a}_j, \quad 1 \leq j \leq N. \tag{B.2}$$

The column vectors over this basis are

$$\mathbf{a}_j = \sum_{i=1}^{j} \langle \mathbf{e}_i, \mathbf{a}_j \rangle \mathbf{e}_i, \quad \langle \mathbf{e}_j, \mathbf{a}_j \rangle = \|\mathbf{u}_j\|, \quad 1 \leq j \leq N. \tag{B.3}$$

Subsequently, by defining a unitary matrix

$$\mathbf{Q} = \begin{pmatrix} \mathbf{e}_1 & \cdots & \mathbf{e}_N \end{pmatrix} \in \mathcal{U}_N$$

and an upper-triangular matrix

$$\mathbf{R} = \begin{pmatrix} \langle \mathbf{e}_1, \mathbf{a}_1 \rangle & \langle \mathbf{e}_1, \mathbf{a}_2 \rangle & \langle \mathbf{e}_1, \mathbf{a}_3 \rangle & \cdots & \langle \mathbf{e}_1, \mathbf{a}_N \rangle \\ 0 & \langle \mathbf{e}_2, \mathbf{a}_2 \rangle & \langle \mathbf{e}_2, \mathbf{a}_3 \rangle & \cdots & \langle \mathbf{e}_2, \mathbf{a}_N \rangle \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & \langle \mathbf{e}_N, \mathbf{a}_N \rangle \end{pmatrix},$$

we obtain a QR decomposition $\mathbf{A} = \mathbf{QR}$.

To arrive at an RQ decomposition $\mathbf{A} = \mathbf{RQ}$, it suffices to apply the same Gram-Schmidt process on the row vectors of $\mathbf{A}$,

$$\mathbf{A} = \begin{pmatrix} \mathbf{a}_N \\ \vdots \\ \mathbf{a}_1 \end{pmatrix} \in \mathcal{M}_N(\mathbb{C}).$$

Instead of iterating forwards over the columns, we iterate backwards over the rows. The unitary matrix needs to be modified accordingly,

$$\mathbf{Q} = \begin{pmatrix} \mathbf{e}_N \\ \vdots \\ \mathbf{e}_1 \end{pmatrix} \in \mathcal{U}_N.$$

Finally, $\mathbf{R}$ is easily found by computing $\mathbf{R} = \mathbf{AQ}^\dagger$.

Appendix C

LLE manifold learning

Here we derive the locally linear embedding (LLE) algorithm introduced in [86]. Consider N real-valued vectors $\mathbf{X}_i \in \mathbb{R}^D$, $i = 1, \ldots, N$. Our goal is to project these D-dimensional vectors onto a manifold $\Omega \subset \mathbb{R}^d$ of dimension $d < D$ that preserves their local relationship. Assume that any vector can be reconstructed by taking a linear combination of its nearest neighbors. Let $\mathbf{W} \in \mathcal{M}_N(\mathbb{R})$ be the desired matrix comprising the reconstruction weights, with entry W_{ij} indicating the contribution of the j^{th} vector to the i^{th} reconstruction. We constrain the weights such that $\sum_j W_{ij} = 1$ for all $i = 1, \ldots, N$. The reconstruction error is given by

$$E_D(\mathbf{W}) = \sum_{i=1}^{N} \left\| \mathbf{X}_i - \sum_{j=1}^{N} W_{ij} \mathbf{X}_j \right\|^2 . \tag{C.1}$$

A hyperparameter of the LLE algorithm is the number of nearest neighbors $K < D$. If K is sufficiently small, then $\mathbf{W}$ will be a sparse matrix.

In the LLE formalism, we assume that the weights of the embedded vectors $\mathbf{Y}_i \in \Omega$ are governed by the same matrix $\mathbf{W}$. Then the manifold embedding reconstruction error is given by a similar expression,

$$E_d(\mathbf{Y}) = \sum_{i=1}^{N} \left\| \mathbf{Y}_i - \sum_{j=1}^{N} W_{ij} \mathbf{Y}_j \right\|^2 , \tag{C.2}$$

where $\mathbf{Y} = \begin{pmatrix} \mathbf{Y}_1 & \cdots & \mathbf{Y}_N \end{pmatrix} \in \mathcal{M}_{d,N}(\mathbb{R})$.

Consider the i^{th} term of the objective function in Equation C.1, for any $i = 1, \ldots, N$. Let $\boldsymbol{\eta}_1, \ldots, \boldsymbol{\eta}_K$ be the K nearest neighbors of $\mathbf{X}_i$. For ease of notation, let $\mathbf{W}_i$ refer not only to the complete sparse dual-vector in the i^{th} row of $\mathbf{W}$, but also to the K nonzero values in the i^{th} row of $\mathbf{W}$. Then we have

$$\begin{aligned} E_D^{(i)}(\mathbf{W}_i) &= \left\| \mathbf{X}_i - \sum_{j=1}^{K} W_{ij} \boldsymbol{\eta}_j \right\|^2 \\ &= \left\| \sum_{j=1}^{K} W_{ij} (\mathbf{X}_i - \boldsymbol{\eta}_j) \right\|^2 \\ &= \sum_{j=1}^{K} \sum_{k=1}^{K} W_{ij} W_{ik} (\mathbf{X}_i - \boldsymbol{\eta}_j)^T (\mathbf{X}_i - \boldsymbol{\eta}_k) \\ &= \sum_{j=1}^{K} \sum_{k=1}^{K} W_{ij} W_{ik} C_{jk}, \end{aligned}$$

where

$$C_{jk} = (\mathbf{X}_i - \boldsymbol{\eta}_j)^T (\mathbf{X}_i - \boldsymbol{\eta}_k) \tag{C.3}$$

are the entries of a local covariance matrix $\mathbf{C}^{(i)} \in \mathcal{M}_K(\mathbb{R})$. Thus, the i^{th} term of the objective function has a Lagrangian

$$\mathcal{L}_D^{(i)}(\mathbf{W}_i) = \sum_{j=1}^{K} \sum_{k=1}^{K} W_{ij} W_{ik} C_{jk} + \lambda \cdot \left(1 - \sum_{j=1}^{K} W_{ij} \right). \tag{C.4}$$

The dual-vector weights $\mathbf{W}_i$ that minimize $E_D^{(i)}$ are given by the Euler-Lagrange equations for C.4 with respect to W_{ij},

$$2 \sum_{k=1}^{K} W_{ik} C_{jk} - \lambda = 0.$$

Therefore, it suffices to solve

$$\mathbf{C}^{(i)} \cdot \mathbf{W}_i^T = \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix} \in \mathbb{R}^K \tag{C.5}$$

for $\mathbf{W}_i$ initially, and then normalize $\mathbf{W}_i$ afterward.

Moving on to the low-dimensional space, let us first expand Equation C.2 into

$$\begin{aligned} E_d(\mathbf{Y}) &= \sum_{i=1}^{N} \left\| \mathbf{Y}_i - \sum_{j=1}^{K} W_{ij} \mathbf{Y}_j \right\|^2 \\ &= \sum_{i=1}^{N} \left(\mathbf{Y}_i - \sum_{j=1}^{K} W_{ij} \mathbf{Y}_j \right) \left(\mathbf{Y}_i^T - \sum_{k=1}^{K} W_{ik} \mathbf{Y}_k^T \right). \end{aligned}$$

This expression motivates a promotion from a scalar-valued objective function into a matrix-valued objective function

$$\mathbf{E}_d(\mathbf{Y}) = \mathbf{Y}(\mathbf{I}_N - \mathbf{W})(\mathbf{I}_N - \mathbf{W})^T \mathbf{Y}^T \in \mathcal{M}_d(\mathbb{R}), \tag{C.6}$$

where $\mathbf{I}_N \in \mathcal{M}_N(\mathbb{R})$ is an identity matrix. Accordingly, define the matrix

$$\mathbf{M} = (\mathbf{I}_N - \mathbf{W})(\mathbf{I}_N - \mathbf{W})^T \in \mathcal{M}_N(\mathbb{R}). \tag{C.7}$$

Suppose the embedding vectors constructing the manifold are constrained to have zero mean and unit covariance. Then the matrix-valued Lagrangian is given by

$$\mathcal{L}_d(\mathbf{Y}) = \mathbf{Y}\mathbf{M}\mathbf{Y}^T + \mu \cdot (\mathbf{Y}\mathbf{A}) + \nu \cdot \left(\mathbf{I}_d - \frac{1}{N} \mathbf{Y}\mathbf{Y}^T \right) \in \mathcal{M}_d(\mathbb{R}), \tag{C.8}$$

where $\mathbf{I}_d \in \mathcal{M}_d(\mathbb{R})$ is an identity matrix and

$$\mathbf{A} = \frac{1}{N} \cdot \begin{pmatrix} 1 & \cdots & 1 \\ \vdots & \ddots & \vdots \\ 1 & \cdots & 1 \end{pmatrix} \in \mathcal{M}_{N,d}(\mathbb{R}).$$

The Euler-Lagrange equations with respect to $\mathbf{Y}$ are

$$2\mathbf{Y}\mathbf{M} + \mu \cdot \mathbf{A}^T - \nu \cdot \frac{1}{N}\mathbf{Y} = 0. \tag{C.9}$$

Observe that Equation C.9 holds for any μ, a property that reflects translational invariance of the manifold Ω. Hence, it suffices to set $\mu = 0$, in which case $\mathbf{Y}^T$ is an eigenmatrix of $\mathbf{M}^T$:

$$\mathbf{M}^T\mathbf{Y}^T = \frac{\nu}{2N}\mathbf{Y}^T. \tag{C.10}$$

The d eigenvectors with the smallest eigenvalues are retained, not including the lowest eigenvector with vanishing eigenvalue.

The entire LLE algorithm may now be summarized according to the following steps:

1. For each high-dimensional point $\mathbf{X}_i \in \mathbb{R}^D$:
   (a) find the K nearest neighbors $\boldsymbol{\eta}_1, \ldots, \boldsymbol{\eta}_K$;
   (b) compute the local covariance matrix $\mathbf{C}^{(i)} \in \mathcal{M}_K(\mathbb{R})$ using Equation C.3;
   (c) invert $\mathbf{C}^{(i)}$ using Equation C.5 and normalize to obtain the weights $\mathbf{W}_i$.
2. Use Equation C.7 to obtain $\mathbf{M}$.
3. Find the d-smallest nontrivial eigenvectors of $\mathbf{M}^T$ according to Equation C.10.

Bibliography

[1] Edward Angel. *OpenGL: A Primer*. 3rd ed. Pearson, 2007.

[2] Lindsey M. Appenzoller et al. "Predicting dose-volume histograms for organs-at-risk in IMRT planning". In: *Medical Physics* 39.12 (2012), pp. 7446–7461. DOI: 10.1118/1.4761864.

[3] National Electrical Manufacturers Association. *DICOM*. Managed by the Medical Imaging and Technology Alliance, a division of the National Electrical Manufacturers Association. URL: https://www.dicomstandard.org/.

[4] French Nuclear Safety Authority. "Patient repositioning imaging: vertebra identification error". In: *Patient safety - Paving the way for progress* 12 (June 2018). URL: https://www.french-nuclear-safety.fr/asn-informs/publications#publications-for-the-professionals.

[5] Anjali Balagopal et al. "PSA-Net: Deep learning-based physician style-aware segmentation network for postoperative prostate cancer clinical target volumes". In: *Artificial Intelligence in Medicine* 121.102195 (2021). DOI: 10.1016/j.artmed.2021.102195.

[6] Brainlab. *ExacTrac Version 6.2: Clinical User Guide, Volume 1/2*. Revision 1.0. Brainlab AG Germany. 2016.

[7] Alexis Bujold et al. "Image-guided radiotherapy: has it influenced patient outcomes?" In: *Seminars in Radiation Oncology* 22.1 (2012), pp. 50–61. DOI: 10.1016/j.semradonc.2011.09.001.

[8] Richard L. Burden and J. Douglas Faires. *Numerical Analysis*. 9th ed. Cengage Learning, 2010.

[9] Lei Cai, Jingyang Gao, and Di Zhao. "A review of the application of deep learning in medical image classification and segmentation". In: *Annals of Translational Medicine* 8.11 (2020). DOI: 10.21037/atm.2020.02.44.

[10] George Casella and Roger L. Berger. *Statistical Inference*. 2nd ed. Cengage Learning, 2001.

[11] John A. Charters, Pascal Bertram, and James M. Lamb. "Offline generator for digitally reconstructed radiographs of a commercial stereoscopic radiotherapy image-guidance system". In: *Journal of Applied Clinical Medical Physics* 23.3 (2022). DOI: 10.1002/acm2.13492.

[12] John A. Charters, Amar Kishan, and James M. Lamb. "Application of super-resolution reconstruction to prostate 0.35 T MRI scans". In: *AAPM 63rd Annual Meeting & Exhibition*. Virtual: American Association of Physicists in Medicine, July 2021.

[13] John A. Charters et al. "Automated detection of vertebral body misalignments in orthogonal kV and MV guided radiotherapy: application to a comprehensive retrospective dataset". In: *Biomedical Physics and Engineering Express* 10.2 (2024). DOI: 10.1088/2057-1976/ad2baa.

[14] John A. Charters et al. "Automated patient positioning error detection with orthogonal kV and unpaired MV treatment radiographs". In: *AAPM 65th Annual Meeting & Exhibition*. Houston, TX: American Association of Physicists in Medicine, July 2023.

[15] John A. Charters et al. "Dosimetric evaluation of respiratory gating on a 0.35-T magnetic resonance guided radiotherapy linac". In: *AAPM 64th Annual Meeting & Exhibition*. Washington, DC: American Association of Physicists in Medicine, July 2022.

[16] John A. Charters et al. “Dosimetric evaluation of respiratory gating on a 0.35-T magnetic resonance-guided radiotherapy linac”. In: *Journal of Applied Clinical Medical Physics* 23.9 (2022). DOI: 10.1002/acm2.13666.

[17] John A. Charters et al. “Improving DWI spatial resolution with respiratory motion modeling on an interleaved bSSFP-DWI sequence”. In: *AAPM 63rd Annual Meeting & Exhibition*. Virtual: American Association of Physicists in Medicine, July 2021.

[18] Xuxin Chen et al. “Recent advances and clinical applications of deep learning in medical image analysis”. In: *Medical Image Analysis* 79.102444 (2022). DOI: 10.1016/j.media.2022.102444.

[19] Zheng Chen et al. “The relationship between waiting time for radiotherapy and clinical outcomes: A systematic review of the literature”. In: *Radiotherapy and Oncology* 87.1 (2008), pp. 3–16. DOI: 10.1016/j.radonc.2007.11.016.

[20] Patrick Cheung et al. “Individualized planning target volumes for intrafraction motion during hypofractionated intensity-modulated radiotherapy boost for prostate cancer”. In: *International Journal of Radiation Oncology, Biology, Physics* 62.2 (2005), pp. 418–425. DOI: 10.1016/j.ijrobp.2004.09.051.

[21] DICOM Standards Committee. *DICOM PS3.3 2013 - Information Object Definitions: C.20 Spatial Registration*. NEMA. 2013. URL: https://dicom.nema.org/dicom/2013/output/chtml/part03/sect_C.20.html.

[22] DICOM Standards Committee. *DICOM PS3.3 2016e - Information Object Definitions: C.7.6.2 Image Plane Module*. NEMA. 2016. URL: https://dicom.nema.org/medical/Dicom/2016e/output/chtml/part03/sect_C.7.6.2.html.

[23] DICOM Standards Committee. *DICOM PS3.3 2016e - Information Object Definitions: C.8.2.1 CT Image Module*. NEMA. 2016. URL: https://dicom.nema.org/medical/dicom/2016e/output/chtml/part03/sect_C.8.2.html.

[24] DICOM Standards Committee. *DICOM PS3.3 2016e - Information Object Definitions: C.8.8.14 RT Beams Module*. NEMA. 2016. URL: https://dicom.nema.org/medical/Dicom/2016e/output/chtml/part03/sect_C.8.8.14.html.

[25] DICOM Standards Committee. *DICOM PS3.3 2016e - Information Object Definitions: C.8.8.2 RT Image Module*. NEMA. 2016. URL: https://dicom.nema.org/medical/dicom/2016e/output/chtml/part03/sect_C.8.8.2.html.

[26] DICOM Standards Committee. *DICOM PS3.3 2016e - Information Object Definitions: C.8.8.3 RT Dose Module*. NEMA. 2016. URL: https://dicom.nema.org/medical/dicom/2016e/output/chtml/part03/sect_C.8.8.3.html.

[27] DICOM Standards Committee. *DICOM PS3.3 2016e - Information Object Definitions: C.8.8.5 Structure Set Module*. NEMA. 2016. URL: https://dicom.nema.org/medical/Dicom/2016e/output/chtml/part03/sect_C.8.8.5.html.

[28] DICOM Standards Committee. *DICOM PS3.3 2016e - Information Object Definitions: C.8.8.6 ROI Contour Module*. NEMA. 2016. URL: https://dicom.nema.org/medical/dicom/2016e/output/chtml/part03/sect_C.8.8.6.html.

[29] DICOM Standards Committee. *DICOM PS3.3 2016e - Information Object Definitions: C.8.8.9 RT Fraction Scheme Module*. NEMA. 2016. URL: https://dicom.nema.org/medical/Dicom/2016e/output/chtml/part03/sect_C.8.8.13.html.

[30] DICOM Standards Committee. *DICOM PS3.3 2016e - Information Object Definitions: C.8.8.9 RT General Plan Module*. NEMA. 2016. URL: https://dicom.nema.org/medical/Dicom/2016e/output/chtml/part03/sect_C.8.8.9.html.

[31] DICOM Standards Committee. *DICOM PS3.3 2016e - Information Object Definitions: C.8.8.9 RT Prescription Module*. NEMA. 2016. URL: https://dicom.nema.org/medical/dicom/2016e/output/chtml/part03/sect_C.8.8.10.html.

[32] DICOM Standards Committee. *DICOM PS3.4 2013 - Service Class Specifications: C.3 Standard Query/Retrieve Information Models*. NEMA. 2013. URL: https://dicom.nema.org/dicom/2013/Output/chtml/part04/sect_C.3.html.

[33] pynetdicom contributors. *pynetdicom*. A Python implementation of the DICOM networking protocol. URL: https://github.com/pydicom/pynetdicom.

[34] Michael M. Dominello et al. "Ten-year trends in safe radiation therapy delivery and results of a radiation therapy quality assurance intervention". In: *Practical Radiation Oncology* 5.6 (2015), pp. 665–671. DOI: 10.1016/j.prro.2015.08.004.

[35] Birte Domnik and Christian Dax. "Renderer Matrices from X-ray Camera Calibration". Jan. 2021.

[36] Nancy Epstein. "A perspective on wrong level, wrong side, and wrong site spine surgery". In: *Surgical Neurology International* 12.286 (2021). DOI: 10.25259/SNI_402_2021.

[37] Gary Ezzell et al. "Common error pathways seen in the RO-ILS data that demonstrate opportunities for improving treatment safety". In: *Practical Radiation Oncology* 8.2 (2018), pp. 123–132. DOI: 10.1016/j.prro.2017.10.007.

[38] Seth D. Frey et al. "Developing clinical indicators to effectively monitor a large network of radiation oncology practices". In: *JCO Oncology Practice* 37.27 (2019). DOI: 10.1200/JCO.2019.37.27_suppl.23.

[39] Yaorong Ge and Q. J. Wu. "Knowledge-based planning for intensity-modulated radiation therapy: A review of data-driven approaches". In: *Medical Physics* 46.6 (2019), pp. 2760–2775. DOI: 10.1002/mp.13526.

[40] Benyamin Ghojogh and Mark Crowley. "Linear and quadratic discriminant analysis: tutorial". In: *arXiv* (2019). DOI: 10.48550/arXiv.1906.02590.

[41] Eric J. Hall and Amato J. Giaccia. *Radiobiology for the Radiologist*. 8th ed. Wolters Kluwer, 2018.

[42] G. Han, Z. Liang, and J. You. "A fast ray-tracing technique for TCT and ECT studies". In: *1999 IEEE Nuclear Science Symposium. Conference Record. 1999 Nuclear Science Symposium and Medical Imaging Conference (Cat. No. 99CH37019)*. Vol. 3. Seattle, WA, USA: Institute of Electrical and Electronics Engineers, 1999, pp. 1515–1518. DOI: 10.1109/NSSMIC.1999.842846.

[43] Kaiming He et al. "Deep residual learning for image recognition". In: *2016 IEEE Conference on Computer Vision and Pattern Recognition*. Microsoft Research. Las Vegas, NV, USA: Institute of Electrical and Electronics Engineers, 2016, pp. 770–778. DOI: 10.1109/CVPR.2016.90.

[44] Tobias Heimann and Hans-Peter Meinzer. "Statistical shape models for 3D medical image segmentation: a review". In: *Medical Image Analysis* 13.4 (2009), pp. 543–563. DOI: 10.1016/j.media.2009.05.004.

[45] Byung-Woo Hong et al. "Shape representation based on integral kernels: Application to image matching and segmentation". In: *2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*. New York, NY, USA: Institute of Electrical and Electronics Engineers, 2006. DOI: 10.1109/CVPR.2006.277.

[46] Gao Huang et al. "Convolutional networks with dense connectivity". In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* 44.12 (2022), pp. 8704–8716. DOI: 10.1109/TPAMI.2019.2918284.

[47] Grace Huang et al. "Error in the delivery of radiation therapy: results of a quality assurance review". In: *International Journal of Radiation Oncology, Biology, Physics* 61.5 (2005), pp. 1590–1595. DOI: 10.1016/j.ijrobp.2004.10.017.

[48] Jenny Huang et al. "Does delay in starting treatment affect the outcomes of radiotherapy? A systematic review". In: *Journal of Clinical Oncology* 21.3 (2003), pp. 555–563. DOI: 10.1200/JCO.2003.04.171.

[49] J.H. Hubbell and S.M. Seltzer. *X-Ray Mass Attenuation Coefficients: NIST Standard Reference Database 126*. Tables of X-Ray Mass Attenuation Coefficients and Mass Energy-Absorption Coefficients from 1 keV to 20 MeV for Elements Z = 1 to 92 and 48 Additional Substances of Dosimetric Interest. NIST. July 2004. DOI: 10.18434/T4D01F.

[50] M. Saiful Huq et al. "The report of Task Group 100 of the AAPM: Application of risk analysis methods to radiation therapy quality management". In: *Medical Physics* 43.7 (2016), p. 4209. DOI: 10.1118/1.4947547.

[51] Shyam S. Jani, Daniel A. Low, and James M. Lamb. "Automatic detection of patient identification and positioning errors in radiation therapy treatment using 3-dimensional setup images". In: *Practical Radiation Oncology* 5.5 (2015), pp. 304–311. DOI: 10.1016/j.prro.2015.06.004.

[52] Jian-Yue Jin et al. "2D/3D image fusion for accurate target localization and evaluation of a mask based stereotactic system in fractionated stereotactic radiotherapy of cranial lesions". In: *Medical Physics* 33.12 (2006), pp. 4557–4566. DOI: 10.1118/1.2392605.

[53] Jian-Yue Jin et al. "Use of the BrainLAB ExacTrac X-Ray 6D system in image-guided radiotherapy". In: *Medical Dosimetry* 33.2 (2008), pp. 124–134. DOI: 10.1016/j.meddos.2008.02.005.

[54] I.T. Jolliffe. *Principal Component Analysis*. 2nd ed. Springer Series in Statistics. Springer, 2002.

[55] James A. Kavanaugh et al. "Multi-institutional validation of a knowledge-based planning model for patients enrolled in RTOG 0617: Implications for plan quality controls in cooperative group trials". In: *Practical Radiation Oncology* 9.2 (2019), pp. 218–227. DOI: 10.1016/j.prro.2018.11.007.

[56] Hoel Kervadec et al. “Boundary loss for highly unbalanced segmentation”. In: *Medical Image Analysis* 67.101851 (2021). DOI: 10.1016/j.media.2020.101851.

[57] J. Kilday, F. Palmieri, and M. D. Fox. “Classifying mammographic lesions using computerized image analysis”. In: *IEEE Transactions on Medical Imaging* 12.4 (1993), pp. 664–669. DOI: 10.1109/42.251116.

[58] Se H. Kim et al. “Computer-aided image analysis of focal hepatic lesions in ultrasonography: preliminary results”. In: *Abdominal Imaging* 34.2 (2009), pp. 183–191. DOI: 10.1007/s00261-008-9383-9.

[59] Yuto Kimura et al. “Error detection model developed using a multi-task convolutional neural network in patient-specific quality assurance for volumetric-modulated arc therapy”. In: *Medical Physics* 48.9 (2021), pp. 4769–4783. DOI: 10.1002/mp.15031.

[60] Andrew P. King and Robert J. Eckersley. *Statistics for Biomedical Engineers and Scientists: How to Visualize and Analyze Data*. Elsevier Ltd., 2019. DOI: 10.1016/C2018-0-02241-0.

[61] Diederik P. Kingma and Jimmy Ba. “Adam: A method for stochastic optimization”. In: 3rd International Conference for Learning Representations. San Diego, CA, USA: ICLR, 2015. DOI: 10.48550/arXiv.1412.6980.

[62] Stefan Klein et al. “elastix: a toolbox for intensity-based medical image registration”. In: *IEEE Transactions on Medical Imaging* 29.1 (2010), pp. 196–205. DOI: 10.1109/TMI.2009.2035616.

[63] Simon L. Krogh et al. “A national repository of complete radiotherapy plans: design, results, and experiences”. In: *Acta Oncologica (Stockholm, Sweden)* 62.10 (2023), pp. 1161–1168. DOI: 10.1080/0284186X.2023.2270143.

[64] James M. Lamb, Nzhde Agazaryan, and Daniel A. Low. “Automated patient identification and localization error detection using 2-dimensional to 3-dimensional registration of kilovoltage x-ray setup images”. In: *International Journal of Radiation*

Oncology, Biology, Physics 87.2 (2013), pp. 390–393. DOI: 10.1016/j.ijrobp.2013.05.021.

[65] Michele Larobina. "Thirty years of the DICOM Standard". In: *Tomography* 9.5 (2023), pp. 1829–1838. DOI: 10.3390/tomography9050145.

[66] Hugo de Las Heras Gala et al. "Quality control in cone-beam computed tomography (CBCT) EFOMP-ESTRO-IAEA protocol (summary report)". In: *Physica Medica* 39 (2017), pp. 67–72. DOI: 10.1016/j.ejmp.2017.05.069.

[67] L. Lemieux et al. "A patient-to-computed-tomography image registration method based on digitally reconstructed radiographs". In: *Medical Physics* 21.11 (1994), pp. 1749–1760. DOI: 10.1118/1.597276.

[68] Nan Li et al. "Highly efficient training, refinement, and validation of a knowledge-based planning quality-control system for radiation therapy clinical trials". In: *International Journal of Radiation Oncology, Biology, Physics* 97.1 (2017), pp. 164–172. DOI: 10.1016/j.ijrobp.2016.10.005.

[69] Geert Litjens et al. "A survey on deep learning in medical image analysis". In: *Medical Image Analysis* 42 (2017), pp. 60–88. DOI: 10.1016/j.media.2017.07.005.

[70] Ethan B. Ludmir et al. "Implementation and efficacy of a large-scale radiation oncology case-based peer-review quality program across a multinational cancer network". In: *JCO Oncology Practice* 37.27 (2019). DOI: 10.1200/JCO.2019.37.27_suppl.1.

[71] S.M.H. Luk et al. "Improving the quality of care in radiation oncology using artificial intelligence". In: *Clinical Oncology (Royal College of Radiologists (Great Britain))* 34.2 (2022), pp. 89–98. DOI: 10.1016/j.clon.2021.11.011.

[72] Dishane C. Luximon et al. "Development and interinstitutional validation of an automatic vertebral-body misalignment error detector for cone-beam CT-guided radiotherapy". In: *Medical Physics* 49.10 (2022), pp. 6410–6423. DOI: 10.1002/mp.15927.

[73] Siddharth Manay et al. “Integral invariants for shape matching”. In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* 28.10 (2006), pp. 1602–1618. DOI: 10.1109/TPAMI.2006.208.

[74] Danielle N. Margalit et al. “Technological advancements and error rates in radiation therapy delivery”. In: *International Journal of Radiation Oncology, Biology, Physics* 81.4 (2011), pp. 673–679. DOI: 10.1016/j.ijrobp.2011.04.036.

[75] Matthew McCormick et al. “ITK: enabling reproducible research and open science”. In: *Frontiers in Neuroinformatics* 8.13 (2014). DOI: 10.3389/fninf.2014.00013.

[76] Jeff M. Michalski et al. “Development of RTOG consensus guidelines for the definition of the clinical target volume for postoperative conformal radiation therapy for prostate cancer”. In: *International Journal of Radiation Oncology, Biology, Physics* 76.2 (2010), pp. 361–368. DOI: 10.1016/j.ijrobp.2009.02.006.

[77] Fausto Milletari, Nassir Navab, and Seyed-Ahmad Ahmadi. “V-Net: Fully convolutional neural networks for volumetric medical image segmentation”. In: *2016 Fourth International Conference on 3D Vision (3DV)*. Stanford, CA, USA: Institute of Electrical and Electronics Engineers, 2016. DOI: 10.1109/3DV.2016.79.

[78] N. Nabavizadeh et al. “Image guided radiation therapy (IGRT) practice patterns and IGRT’s impact on workflow and treatment planning: results from a national survey of American Society for Radiation Oncology members”. In: *International Journal of Radiation Oncology, Biology, Physics* 94.4 (2016), pp. 850–857. DOI: 10.1016/j.ijrobp.2015.09.035.

[79] Sandy A. Napel et al. “Automated retrieval of CT images of liver lesions on the basis of image similarity: method and preliminary results”. In: *Radiology* 256.1 (2010), pp. 243–252. DOI: 10.1148/radiol.10091694.

[80] Benjamin E. Nelms et al. "Variation in external beam treatment plan quality: An inter-institutional study of planners and planning systems". In: *Practical Radiation Oncology* 2.4 (2012), pp. 296–305. DOI: 10.1016/j.prro.2011.11.012.

[81] Tucker J. Netherton et al. "An automated treatment planning framework for spinal radiation therapy and vertebral-level second check". In: *International Journal of Radiation Oncology, Biology, Physics* 114.3 (2022), pp. 516–528. DOI: 10.1016/j.ijrobp.2022.06.083.

[82] Adam C. Olson et al. "Quality assurance analysis of a large multicenter practice: does increased complexity of intensity-modulated radiotherapy lead to increased error frequency?" In: *International Journal of Radiation Oncology, Biology, Physics* 82.1 (2012), pp. 77–82. DOI: 10.1016/j.ijrobp.2011.01.033.

[83] Rachel Petragallo, John A. Charters, and James M. Lamb. "A convolutional neural network-based retrospective search for previously unreported treatment errors in a stereoscopic planar X-ray setup image database". In: *AAPM 65th Annual Meeting & Exhibition*. Houston, TX: American Association of Physicists in Medicine, July 2023.

[84] Biao Qu et al. "Current development and prospects of deep learning in spine image analysis: a literature review". In: *Quantitative Imaging in Medicine and Surgery* 12.6 (2022), pp. 3454–3479. DOI: 10.21037/qims-21-939.

[85] Sophie Robin et al. "Prostate bed delineation guidelines for postoperative radiation therapy: On behalf of the Francophone Group of Urological Radiation Therapy". In: *International Journal of Radiation Oncology, Biology, Physics* 109.5 (2021), pp. 1243–1253. DOI: 10.1016/j.ijrobp.2020.11.010.

[86] S.T. Roweis and L.K. Saul. "Nonlinear dimensionality reduction by locally linear embedding". In: *Science (New York, NY)* 290.5500 (2000), pp. 2323–2326. DOI: 10.1126/science.290.5500.2323.

[87] Gregory A. Russo et al. "Daily orthogonal kilovoltage imaging using a gantry-mounted on-board imaging system results in a reduction in radiation therapy delivery errors". In: *International Journal of Radiation Oncology, Biology, Physics* 84.3 (2012), pp. 596–601. DOI: 10.1016/j.ijrobp.2012.01.033.

[88] Carolin Schubert et al. "Intercenter validation of a knowledge based model for automated planning of volumetric modulated arc therapy for prostate cancer. The experience of the German RapidPlan Consortium". In: *PLOS One* 12.5 (2017). DOI: 10.1371/journal.pone.0178034.

[89] Satomi Shiraishi and Kevin L. Moore. "Knowledge-based prediction of three-dimensional dose distributions for external beam radiotherapy". In: *Medical Physics* 43.1 (2015), pp. 378–387. DOI: 10.1118/1.4938583.

[90] R.L. Siddon. "Fast calculation of the exact radiological path for a three-dimensional CT array". In: *Medical Physics* 12.2 (1985), pp. 252–255. DOI: 10.1118/1.595715.

[91] David Skarsgard et al. "Planning target volume margins for prostate radiotherapy using daily electronic portal imaging and implanted fiducial markers". In: *Radiation Oncology (London, England)* 5.52 (2010). DOI: 10.1186/1748-717X-5-52.

[92] Sandie Smith et al. "Quality management in radiation therapy: A 15 year review of incident reporting in two integrated cancer centres". In: *Technical Innovations and Patient Support in Radiation Oncology* 6 (2020), pp. 15–20. DOI: 10.1016/j.tipsro.2020.02.001.

[93] Abhishek A. Solanki et al. "Consensus quality measures and dose constraints for prostate cancer from the Veterans Affairs Radiation Oncology Quality Surveillance Program and American Society for Radiation Oncology Expert Panel". In: *Practical Radiation Oncology* 13.2 (2023), pp. 149–165. DOI: 10.1016/j.prro.2022.08.018.

[94] Joen Sveistrup et al. "Improvement in toxicity in high risk prostate cancer patients treated with image-guided intensity-modulated radiotherapy compared to 3D confor-

mal radiotherapy without daily image guidance". In: *Radiation Oncology (London, England)* 9.44 (2014). DOI: 10.1186/1748-717X-9-44.

[95] Nikhil G. Thaker et al. "Assessing the quality of a radiation oncology case-based, peer-review program in an integrated academic and community cancer center network". In: *JCO Oncology Practice* 12.4 (2016). DOI: 10.1200/JOP.2015.005983.

[96] Jim P. Tol et al. "Evaluation of a knowledge-based planning solution for head and neck cancer". In: *International Journal of Radiation Oncology, Biology, Physics* 91.3 (2015), pp. 612–620. DOI: 10.1016/j.ijrobp.2014.11.014.

[97] Jen-San Tsai, Bizhan Micaily, and Curtis Miyamoto. "Optimization and quality assurance of an image-guided radiation therapy system for intensity-modulated radiation therapy radiotherapy". In: *Medical Dosimetry* 37.3 (2012), pp. 321–333. DOI: 10.1016/j.meddos.2011.11.006.

[98] Varian. *TrueBeam Treatment Delivery Console V4.0 DICOM Conformance Statement*. Version 7.0. Varian Medical Systems, Inc. Apr. 2023.

[99] Sheri M. Weintraub et al. "Human factor associations with safety events in radiation therapy". In: *International Journal of Radiation Oncology, Biology, Physics* 22.10 (2021), pp. 288–294. DOI: 10.1002/acm2.13420.

[100] James S. Welsh et al. "Fiducial markers implanted during prostate brachytherapy for guiding conformal external beam radiation therapy". In: *Technology in Cancer Research and Treatment* 3.4 (2004), pp. 359–364. DOI: 10.1177/153303460400300405.

[101] J. Westberg et al. "A DICOM based radiotherapy plan database for research collaboration and reporting". In: *Journal of Physics: Conference Series* 489.012100 (2014). DOI: 10.1088/1742-6596/489/1/012100.

[102] Xuanang Xu et al. "Asymmetric multi-task attention network for prostate bed segmentation in computed tomography images". In: *Medical Image Analysis* 72.102116 (2021). DOI: 10.1016/j.media.2021.102116.

[103] Zhengzheng Xu et al. “Evaluating radiotherapy treatment delay using Failure Mode and Effects Analysis (FMEA)”. In: *Radiotherapy and Oncology* 137 (2019), pp. 102–109. DOI: 10.1016/j.radonc.2019.04.016.

[104] Slav Yartsev and Glenn Bauman. “Target margins in radiotherapy of prostate cancer”. In: *The British Journal of Radiology* 89.1067 (2016). DOI: 10.1259/bjr.20160312.

[105] Xiaofeng Zhu et al. “A planning quality evaluation tool for prostate adaptive IMRT based on machine learning”. In: *Medical Physics* 38.2 (2011), pp. 719–726. DOI: 10.1118/1.3539749.