CHARACTERIZING OSTEOSARCOPENIA IN SPINAL METASTASES PATIENTS UNDERGOING STEREOTACTIC BODY RADIOTHERAPY (SBRT): LEVERAGING DEEP LEARNING FOR IMPROVED OUTCOME PREDICTION

YESSICA CASTANO SAINZ

A THESIS

SUBMITTED TO THE FACULTY OF GRADUATE STUDIES

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS

FOR THE DEGREE OF

MASTER OF ARTS

GRADUATE PROGRAM IN

INFORMATION SYSTEMS & TECHNOLOGY

YORK UNIVERSITY

TORONTO, ONTARIO

JUNE 2025

Abstract

Stereotactic body radiotherapy (SBRT) is commonly used to treat spinal metastases, offering excellent local control and pain relief. However, it carries an average 14% risk of vertebral compression fractures (VCFs), and despite growing evidence linking osteosarcopenia to adverse clinical outcomes, musculoskeletal health is not routinely assessed during SBRT treatment planning. This thesis introduces a fully automated pipeline for extracting musculoskeletal biomarkers from CT, combining deep learning–based segmentation with vertebral landmark–guided cropping and volumetric analysis. Sarcopenia thresholds were derived for volumetric indices using height-based and vertebral-based normalization, guided by established literature cutoffs for the Psoas Muscle Index (PMI). Osteoporosis was defined using trabecular bone density. In this SBRT cohort, 58% of patients met criteria for osteoporosis, 45% for sarcopenia, and 31% for osteosarcopenia. In multivariable logistic regression analyses, significant associations between fracture risk and both osteoporosis and lower psoas muscle density were observed in specific models, warranting further investigation. Additionally, categorical definitions of sarcopenia and osteosarcopenia were significantly associated with reduced overall survival. The pipeline was extended to MRI using CT-based segmentations as weak labels for training nnU-Net models, achieving high segmentation accuracy and supporting future radiation-free musculoskeletal biomarker assessment.

To my mother, for being the strength behind all I do.

To my husband, for always being by my side.

Acknowledgements

I would like to sincerely thank those who made this thesis possible.

I'm deeply grateful to my supervisor, Dr. Stephen Chen, for his continuous guidance, encouragement, and insightful feedback throughout this journey. His thoughtful mentorship gave me the confidence and freedom to explore ideas independently, while always offering clear direction and steady support.

I also want to express my sincere appreciation to Dr. Michael Hardisty and Dr. Cari Whyne for their encouragement, guidance, and insightful discussions, and for welcoming me as part of their team at Sunnybrook.

I'm also grateful to the broader Sunnybrook team for their valuable collaboration. In particular, I would like to thank Raj, Saleh, and Declan for their helpful discussions, technical support, and contributions throughout the project.

A special thanks to my husband for his unwavering encouragement, constant support, and deep understanding throughout this journey.

Contents

List of Tables

List of Figures

Chapter 1

Introduction

Stereotactic body radiation therapy (SBRT) has become an important modality for the treatment of spinal metatasis offering precise targeting of cancerous lesions with minimal impact on surrounding tissues. This technique has shown great efficacy in controlling pain and slowing tumor progression [1, 2].

As patient survival improves, there is increasing attention to the long-term effects of SBRT on musculoskeletal health, particularly regarding bone and muscle integrity. Complications such as vertebral compression fractures (VCF) and muscle atrophy are common after treatment and can substantially affect patient quality of life and physical function [3, 4].

For patients with metastatic disease, especially those with spinal involvement, preserving musculoskeletal health is critical to reducing the burden of symptoms and optimizing outcomes. Musculoskeletal deterioration, influenced by declines in bone and muscle quality, can lead to reduced mobility, increased risk of fractures, and elevated mortality. Consequently, osteosarcopenia, the combined loss of bone mass and muscle mass, should be an important clinical focus in oncology, particularly in evaluating the long-term consequences of therapies

like SBRT [5, 6].

Osteosarcopenia has been consistently associated with adverse outcomes in cancer patients, including higher risk of fracture, decreased mobility, and increased mortality [7–9]. This underscores the importance of integrating musculoskeletal monitoring into cancer care pathways. Imaging biomarkers provide a noninvasive and quantitative approach to evaluating musculoskeletal health by measuring parameters such as muscle volume, fat infiltration, and bone density. These biomarkers offer clinicians valuable insight in predicting adverse events and forming treatment strategies [10]. In this thesis, we calculate quantitative imaging biomarkers such as trabecular bone density, vertebral body volume, psoas muscle volume, and cross-sectional area. Characterizing the musculoskeletal status of patients at the time of SBRT treatment planning could enhance risk stratification and allow early identification of individuals at higher risk for complications.

While computed tomography (CT) remains the clinical standard for bone assessment and is widely used in the extraction of musculoskeletal biomarkers, magnetic resonance imaging (MRI) provides superior soft tissue contrast, enabling the detailed visualization of muscle degeneration and fat infiltration. This makes MRI valuable for tracking musculoskeletal changes in patients undergoing SBRT, where both bone and muscle deterioration are concerns [11]. However, extracting musculoskeletal biomarkers from MRI presents technical challenges, including variations in image contrast, lower spatial resolution for bone structures, and the scarcity of large annotated datasets for training robust segmentation models. Extending biomarker pipelines to MRI is crucial for leveraging routinely acquired MRI scans without necessitating additional imaging, facilitating broader clinical applicability.

The work presented in this thesis is part of a broader research initiative conducted at Sunnybrook Research Institute, aimed at establishing the relationship between musculoskeletal

imaging biomarkers, derived from both CT and MRI at the time of SBRT treatment planning, and treatment outcomes. Specifically, this study focuses on analyzing baseline CT-derived biomarkers to characterize musculoskeletal health in patients with spinal metastases and to investigate their association with vertebral compression fracture risk. Concurrent efforts were made to extend the automated segmentation pipeline to MRI, enabling future extraction of musculoskeletal biomarkers from MRI data. By understanding these relationships, this work aims to contribute toward the development of an automated pipeline for musculoskeletal health assessment at the SBRT treatment planning stage, with the overarching goal of improving risk prediction, guiding therapy, and ultimately enhancing patient outcomes [6, 7, 12].

The remainder of this thesis is organized as follows. Chapter 2 provides background on SBRT, osteosarcopenia, and musculoskeletal imaging, including relevant anatomy and imaging modalities. Chapter 3 reviews current literature related to musculoskeletal biomarkers, sarcopenia and osteoporosis assessment, and deep learning-based segmentation approaches. Chapter 4 outlines the study design, ethical approvals, and datasets used. Chapter 5 details the development of the imaging pipeline, including segmentation, biomarker computation, and data integration, and presents the characterization of osteosarcopenia in spinal metastases patients using the extracted biomarkers. It also evaluates associations with vertebral fracture risk and overall survival. Chapter 6 describes the extension of the segmentation pipeline to MRI, including the training and evaluation of deep learning models for psoas and vertebral body segmentation using aligned CT-MRI data. Finally, Chapter 7 concludes the thesis and highlights potential directions for future research.

Chapter 2

Background

2.1 Stereotactic Body Radiotherapy (SBRT)

Stereotactic body radiation therapy (SBRT) is a highly focused radiation treatment that gives an intense dose of radiation concentrated on a tumor, while limiting the dose to the surrounding organs. It has become a treatment of choice for many people with limited volume tumors for whom surgery may not be an optimal treatment.

This technique enables the delivery of high doses of radiation in fewer treatment sessions, resulting in significant tumor control and pain relief. Clinical studies report local control rates exceeding 80% [13,14]. A key advantage of SBRT is its millimeter-level accuracy, which facilitates dose escalation, particularly for patients with radioresistant tumors [15, 16].

Despite its benefits, SBRT is associated with complications, notably vertebral compression fractures (VCFs). These fractures has been reported to occur in up to 40% of cases following SBRT, particularly with high-dose, single-fraction regimens [17, 18]. In low doses and fractionated SBRT, the risk of fracture is $< 5\%$ (close to the external beam radiother-

apy (EBRT) risk), while keeping a good local control of $> 80\%$. The risk of VCFs with conventional EBRT is approximately 3% [18, 19]. The higher incidence of VCFs in SBRT is attributed to increased radiation doses delivered to vertebral bodies, compromising bone structural integrity [17,19]. Additionally, muscle atrophy may arise as a consequence of radiation therapy, further affecting musculoskeletal health and potentially leading to functional impairments [20].

2.2 Osteosarcopenia in Cancer Patients

Osteosarcopenia is defined as the co-occurrence of osteoporosis, which involves low bone mass and microarchitectural deterioration of bone tissue and sarcopenia, the progressive and generalized loss of skeletal muscle mass and strength [5]. The combined presence of these conditions increases the likelihood of adverse outcomes such as falls, fractures, frailty, and mortality, imposing substantial personal and societal burdens [21].

This condition has gained growing attention in oncology due to its association with higher rates of functional decline, treatment-related complications, and reduced survival [22, 23]. In cancer populations, the prevalence of osteosarcopenia has been reported to range from 15% to 50%, with higher rates in patients with advanced disease [24, 25]. Among those with spinal metastases, osteosarcopenia has been specifically linked to an increased risk of vertebral compression fractures (VCFs), owing to compromised bone strength and structural integrity [26, 27].

Furthermore, in patients who sustain hip or spinal fractures, the presence of osteosarcopenia has been associated with significantly higher one-year mortality rates compared to those without the condition [9]. Recent studies have also shown that it serves as a strong

predictor of poor prognosis and treatment-related toxicity in oncology settings, emphasizing the need for early identification and integration into treatment planning [28].

2.3 Musculoskeletal Imaging

Imaging biomarkers are quantitative image-derived features that provide objective measures of anatomical or physiological characteristics. In musculoskeletal health assessment, these biomarkers are increasingly used to evaluate bone and muscle integrity through non-invasive imaging. They offer insights into disease presence, progression, and treatment response, including diagnosis of osteoporosis, sarcopenia, and osteosarcopenia [29].

In cancer care, musculoskeletal imaging biomarkers are particularly relevant for assessing treatment-related musculoskeletal decline. They can be used to measure parameters such as vertebral body density, cross sectional area or volume of the psoas muscle, and intramuscular fat infiltration—features that are closely linked to fracture risk, frailty, and overall survival.

Two imaging modalities are used in SBRT treatment planning: Computed Tomography (CT) and Magnetic Resonance Imaging (MRI). Each modality offers distinct advantages for musculoskeletal and plays a complementary role in comprehensive assessment.

2.3.1 CT and MRI Modalities

CT and MRI are commonly employed for musculoskeletal imaging due to their ability to provide detailed anatomical information. MRI is particularly favored for its superior soft tissue contrast (Figure 2.2), making it ideal for visualizing muscle morphology, fat infiltration, and soft-tissue injuries [30, 31]. It is frequently used to assess conditions involving tendons, ligaments, cartilage, and muscle degeneration—offering insights that are critical in sarcopenia

and related conditions [32, 33].

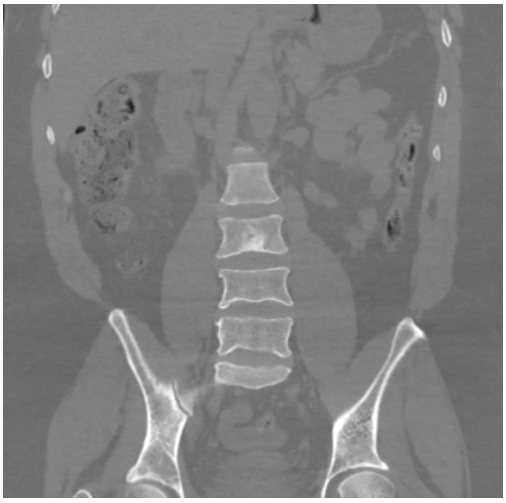


Figure 2.1: Computed Tomography (CT), as seen on a coronal scan.

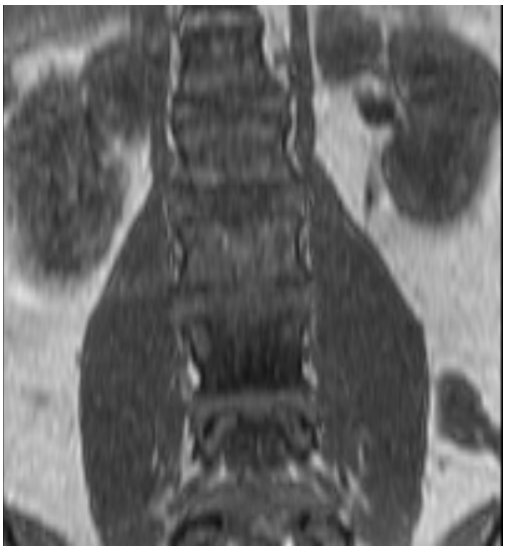


Figure 2.2: Magnetic Resonance Imaging (MRI), as seen on a coronal scan.

CT (Figure 2.1) provides high-resolution images of bone structures and is well-suited for evaluating bone density and detecting vertebral compression fractures. It is often used in acute settings due to its speed and accuracy. Advances in CT technology have improved its

spatial resolution and reduced radiation exposure, allowing detailed assessment of vertebral body morphology and structural integrity [34]. In combination with PET, CT also allows functional and anatomical imaging for certain musculoskeletal conditions [35].

The choice between CT and MRI depends on the clinical context: MRI is generally preferred for soft tissue assessment, while CT remains essential for evaluating bone pathology and trauma [30, 34].

2.3.2 The Psoas Muscle, Iliopsoas Complex, and Vertebral Bodies

The psoas major muscle, together with the iliacus, forms the iliopsoas complex—a major hip flexor with an important role in posture and locomotion. Anatomically, it originates from the transverse processes and vertebral bodies of T12 to L5, making it clearly visible on axial CT and MRI scans (Figure 2.3).

Due to its consistent shape and location, the psoas is commonly used in imaging studies as a proxy for overall muscle mass. The cross-sectional area (CSA) of the psoas at the L3 vertebral level is frequently used as a biomarker for sarcopenia. Similarly, the lumbar vertebral bodies—particularly L3 and L4—are used to assess bone health, often serving as reference points for vertebral body volume (VBV) and bone mineral density (BMD) calculations (Figure 2.4).. When height is unavailable, some indices normalize psoas muscle area at L3 to vertebral body area (e.g., L3-TPA/VB), enabling internal comparison of musculoskeletal status.

These anatomical structures form the foundation for musculoskeletal biomarker quantification in both clinical and research settings, particularly in patients with spinal metastases, where bone fragility and muscle wasting are common.

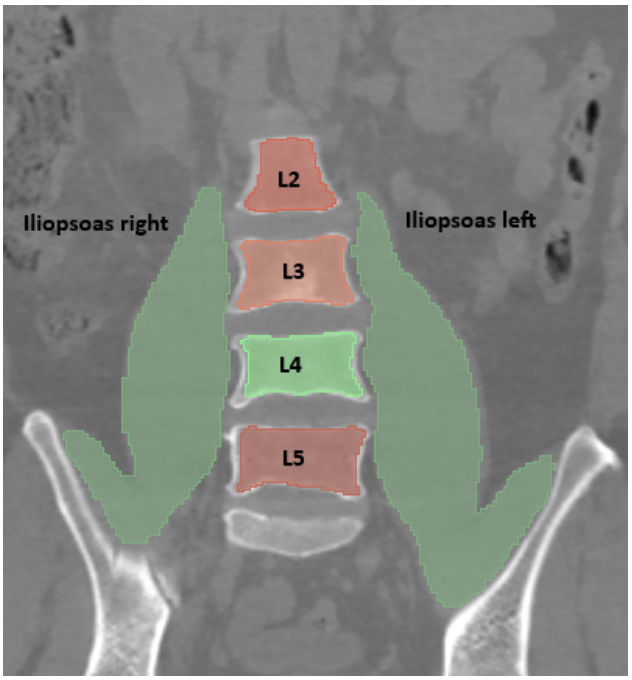


Figure 2.3: Anatomical position of the psoas major muscle (in green) in relation to the lumbar vertebrae and pelvis as seen on a coronal CT scan.

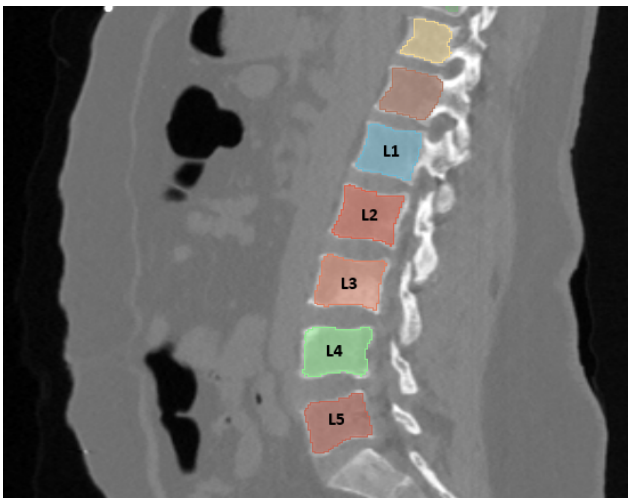


Figure 2.4: Lumbar vertebral bodies as seen on a sagittal CT scan.

2.3.3 Deep Learning in Musculoskeletal Image Analysis

Deep learning has revolutionized medical image analysis by enabling highly accurate and automated extraction of anatomical structures, which is particularly valuable in musculoskeletal

imaging. Among its applications, image segmentation—identifying and delineating specific anatomical regions such as the psoas muscle or vertebral bodies—plays a central role in the quantification of imaging biomarkers.

Accurate segmentation is a critical step in deriving biomarkers like muscle volume, cross-sectional area, and bone density from CT and MRI scans. Traditional manual segmentation is time-consuming and prone to variability, whereas deep learning approaches offer fast, consistent, and reproducible results across large datasets.

Architectures such as U-Net and its self-configuring variant nnU-Net have demonstrated exceptional performance in biomedical segmentation tasks, including the musculoskeletal system [36]. These networks operate by learning multi-scale representations of imaging data, allowing them to delineate both large and small anatomical structures with high precision.

In the context of osteosarcopenia assessment, segmentation models can be trained to isolate the psoas muscles and vertebral bodies—structures that are used for calculating muscle and bone biomarkers. Once segmented, post-processing algorithms can be applied to compute features such as muscle volume, density, and fat infiltration, as well as vertebral body volume and bone mineral metrics. These quantitative outputs are then used to assess musculoskeletal health and predict clinical outcomes.

Despite its advantages, implementing deep learning in clinical imaging workflows poses challenges. These include the need for large, annotated training datasets, robustness across imaging protocols, and interpretability of model predictions. Nevertheless, ongoing advances in transfer learning, data augmentation, and model generalization are addressing these limitations, making deep learning-based segmentation a cornerstone for reliable biomarker extraction in musculoskeletal imaging [37].

Chapter 3

Literature Review

3.1 Sarcopenia Assessment Using Imaging Biomarkers

Quantitative imaging biomarkers play a central role in the clinical and research assessment of sarcopenia. Among imaging modalities, computed tomography (CT) is most widely used due to its routine availability in oncologic care and its capacity to provide high-resolution cross-sectional views of musculoskeletal structures. The third lumbar vertebral level (L3) is the standard anatomical landmark for analysis, as the muscle cross-sectional area at this level correlates strongly with total body skeletal muscle mass and volume [38, 39].

Muscle quantification at L3 typically includes the paraspinal, psoas, and abdominal wall muscles. The area is normalized to the patient's height squared to derive the Skeletal Muscle Index (SMI), a well-established metric for identifying low muscle mass. When full muscle segmentation is not feasible, the Psoas Muscle Index (PMI), which considers only the bilateral psoas cross-sectional area, serves as a simplified alternative. However, PMI has been shown to correlate only moderately with SMI and may misclassify sarcopenia in some patients [39].

Normalization is critical for comparing individuals of different body sizes. While height-squared adjustment is standard, it poses challenges in retrospective studies where height is frequently missing or unreliable. To address this, alternative normalization strategies have been proposed—most notably, indexing to internal anatomical references such as vertebral body area (VBA) at L3 [6]. These adjustments allow for consistent biomarker calculation even in incomplete clinical datasets.

3.2 Thresholds for Defining Sarcopenia in the Literature

Sarcopenia is commonly defined by reduced skeletal muscle mass, often accompanied by diminished strength or physical performance. In imaging-based studies, skeletal muscle mass is typically assessed using computed tomography (CT), with cross-sectional areas measured at the third lumbar vertebra (L3). These areas are normalized to the patient's height squared, yielding either the Skeletal Muscle Index (SMI) or the Psoas Muscle Index (PMI), depending on whether all trunk muscles or only the psoas muscles are considered.

Thresholds for low SMI or PMI vary across studies and are generally derived using one of two approaches: statistical distributions from healthy young adults or outcome-based stratification in specific clinical populations. In general population studies, thresholds are often defined as values below the 5th percentile or two standard deviations below the mean of a healthy cohort. For example, in a large Chinese cohort, the 5th-percentile SMI thresholds were approximately 40.2 cm^2/m^2 for men and 31.6 cm^2/m^2 for women [40]. In Western cohorts, Prado et al. proposed higher thresholds—52.4 cm^2/m^2 for men and 38.5 cm^2/m^2 for women—based on survival analysis in obese cancer patients [41]. Martin et al. further adjusted SMI cutoffs based on BMI, proposing 43 cm^2/m^2 for men with BMI < 25, 53 cm^2/m^2

for men with BMI ≥ 25, and 41 cm^2/m^2 for women [42, 43].

The Psoas Muscle Index (PMI), calculated from the cross-sectional area of the bilateral psoas muscles at L3, offers a simpler alternative when full muscle segmentation is not feasible. While easier to compute, PMI may correlate only moderately with SMI and has been shown to misclassify sarcopenia in certain cases [39]. Nevertheless, PMI is widely used in clinical research, particularly in oncology and hepatology, where imaging is routinely available.

A comprehensive systematic review by Albano et al. [44] examined CT-based sarcopenia across 25 studies involving lymphoma patients. The review highlighted substantial variability in the thresholds used to define low muscle mass, with SMI values in men ranging from 43 to 56.8 cm^2/m^2 and in women from 31 to 47.4 cm^2/m^2. Some studies in the review also used PMI instead of SMI, with two distinct threshold sets reported: 6.36 cm^2/m^2 for men and 3.92 cm^2/m^2 for women, and 4.4/3.1 cm^2/m^2 in another study. Despite this heterogeneity, many studies demonstrated strong associations between low muscle indices—whether SMI or PMI—and adverse outcomes such as reduced progression-free survival, overall survival, and increased treatment toxicity.

In this thesis, the thresholds of 6.36 cm^2/m^2 for men and 3.92 cm^2/m^2 for women were adopted to define sarcopenia using PMI. These values were originally derived from healthy Japanese adults using the mean minus two standard deviations approach [45, 46]. Although these cut-offs were developed in an Asian population, they have demonstrated consistent prognostic value across multiple studies involving various cancer types and chronic diseases, particularly in surgical and oncology settings [44, 47–52].

To our knowledge, these specific thresholds have not yet been widely adopted in Western cancer cohorts, where population-specific cut-offs are more commonly derived. However, given the limited availability of validated Western PMI thresholds and the practicality of

applying a fixed, reproducible reference for muscle mass quantification, these well-established Japanese thresholds were used as a proxy to facilitate comparability and support initial classification efforts.

While most sarcopenia thresholds are based on CT imaging due to its wide availability and high reproducibility, MRI offers a radiation-free alternative with excellent soft-tissue contrast. Studies have shown a strong correlation between CT and MRI-derived muscle areas at L3, suggesting that existing CT-based cutoffs can be cautiously applied to MRI when acquisition protocols and segmentation methods are comparable [53]. However, standardized MRI-specific thresholds for sarcopenia have not yet been established.

In summary, sarcopenia thresholds derived from SMI or PMI are not universally standardized and depend on the population and outcome of interest. This thesis adopts PMI thresholds that have been validated in cancer populations and shown to reflect clinically meaningful differences in muscle mass and prognosis in our effort to derive thresholds for our volumetric measures.

3.3 Osteoporosis Assessment (BMD via DXA and CT Hounsfield Units)

Osteoporosis, defined as a systemic skeletal disease characterized by low bone mineral density (BMD) and microarchitectural deterioration of bone tissue, is conventionally diagnosed using dual-energy X-ray absorptiometry (DXA). The World Health Organization (WHO) defines osteoporosis based on a T-score of -2.5 or lower. However, the accessibility and reliability of DXA may be limited by factors such as improper patient positioning and scan interpretation. As an alternative, Hounsfield Unit (HU) measurements derived from CT

imaging have been explored as an opportunistic screening tool for reduced BMD. Thresholds of 135 HU have been proposed to provide a balance between sensitivity and specificity for diagnosing osteoporosis [54].

3.4 Deep Learning for Musculoskeletal Segmentation and Analysis

Deep learning techniques, particularly convolutional neural networks (CNNs), have become fundamental in the segmentation and analysis of musculoskeletal structures in medical imaging. Architectures like U-Net, introduced by Ronneberger et al., feature an encoder-decoder structure that enables accurate segmentation even with limited training data [55]. U-Net's design, which includes a contracting path to capture context and an expanding path for localization, has proven particularly well-suited for biomedical applications in CT and MRI.

In musculoskeletal imaging, several extensions of U-Net have been developed. For instance, V-Net introduced by Milletari et al. uses 3D convolutions to directly capture volumetric context, considerably improving segmentation accuracy for structures such as the vertebrae in MRI [56].

More recently, nnU-Net—an automated adaptation of U-Net—has demonstrated exceptional performance across diverse biomedical datasets. It configures itself based on dataset properties and has been recognized as a reliable standard for 3D medical image segmentation [57]. As highlighted by Xia et al. [58], nnU-Net remains highly competitive due to its adaptability and lack of manual tuning.

In their benchmarking study, Isensee et al. (2024) reaffirm nnU-Net's position as a state-of-the-art framework, consistently outperforming other contemporary architectures including

CNN-, Transformer-, and Mamba-based models [36].

In this thesis, nnU-Net was used to train deep learning models for psoas muscle and vertebral body segmentation in MRI volumes. CT-based segmentations from aligned CT-MRI pairs served as training references, enabling the development of MRI-specific models despite limited manually annotated MRI data.

Alternative deep learning strategies have also been explored. Swin-UNETR, a Transformer-based model, has demonstrated strong results in 3D medical segmentation tasks by modeling long-range dependencies [59]. Similarly, adversarial approaches such as those by Dong et al. combine U-Net and fully convolutional networks (FCNs) to improve segmentation consistency [60].

One challenge in training segmentation models in medical imaging is the limited availability of expert-annotated datasets—particularly for MRI. Strategies such as cross-modality learning, transfer learning, and synthetic data generation have been proposed to overcome this issue. Jiang et al. developed a cross-modality deep learning framework that combines CT and MRI data for lung tumor segmentation, improving performance by augmenting limited MRI data with CT [61].

Generative Adversarial Networks (GANs) have also been employed to synthesize CT from MRI, creating pseudo-CT volumes for tasks like radiation planning [62,63]. Transfer learning remains a valuable technique for adapting models trained on large public datasets to more specific medical tasks. For instance, Proietto et al. fine-tuned a 3D U-Net trained on CT data for MRI pancreas segmentation, improving performance in data-limited scenarios [64].

These strategies are particularly relevant to osteosarcopenia assessment, which often requires both bone and muscle imaging. The integration of CT and MRI, supported by

deep learning, offers a more complete and scalable approach to musculoskeletal biomarker extraction.

3.5 Gaps in Current Research

Despite advancements in deep learning and multimodal imaging, several gaps remain in the quantitative assessment of musculoskeletal health, particularly in oncology patients receiving spinal SBRT.

Most existing sarcopenia assessments involving the psoas muscle rely on manual or semi-automated 2D slice-based measurements, which are prone to inter-observer variability and sensitive to slice location [65]. These limitations reduce reproducibility and are insufficient for longitudinal monitoring or large-scale analysis.

While volumetric muscle analysis offers greater reliability, it is rarely implemented in clinical practice due to the time-intensive nature of manual segmentation. Deep learning–based approaches address this challenge but have not yet been widely validated for volumetric sarcopenia and osteosarcopenia biomarkers across diverse populations or imaging modalities.

Furthermore, most studies assess sarcopenia or osteoporosis using either CT or MRI in isolation. CT provides standardized attenuation-based muscle thresholds and is effective for bone assessment, whereas MRI offers superior soft-tissue contrast without radiation. However, few studies have investigated the harmonization or comparison of CT- and MRI-derived biomarkers, especially using aligned imaging or cross-modality learning approaches.

There is also limited application of imaging biomarkers in outcome prediction for high-risk groups, such as patients with spinal metastases treated with SBRT. Existing prognostic models—like the Spinal Instability Neoplastic Score (SINS)—do not incorporate quantitative

imaging features, missing an opportunity to improve risk stratification and guide treatment decisions.

This thesis addresses several of the identified gaps by developing a fully automated, pipeline for quantifying volumetric musculoskeletal biomarkers from CT, including psoas muscle volume and trabecular bone density. It establishes sarcopenia classification thresholds using both vertebral body–normalized indices and height-adjusted ratios, enabling application across cohorts with and without height data. These biomarkers are then applied to assess vertebral fracture risk and overall survival in a large cohort of patients treated with SBRT. Finally, the thesis initiates the development of MRI-based segmentation models using aligned CT-MRI datasets to support future multimodal biomarker extraction and analysis.

3.6 Objectives

To address the described omissions from the existing literature, the objectives of this thesis are:

1. **Develop a Fully Automated CT Biomarker Pipeline:** Enhance and automate an end-to-end pipeline for automated segmentation and quantification of bone and muscle biomarkers from CT, including vertebral body trabecular density, vertebral body volume, psoas muscle volume, and psoas cross-sectional area.

2. **Establish and Validate Sarcopenia Classification Thresholds:** Develop sarcopenia thresholds using both traditional height-based measures (PMI) and novel height-independent volumetric indices normalized to internal anatomical references (e.g., vertebral body volume).

3. **Apply Biomarkers to Clinical Outcome Analysis:** Characterize osteosarcopenia in spinal metastases patients at the time of SBRT treatment planning, and evaluate associations between CT-based biomarkers and clinical outcomes, including vertebral fracture risk and overall survival.

4. **Develop MRI Segmentation Models for Future Biomarker Use:** Train and evaluate deep learning models to segment psoas muscles and vertebral bodies bodies in MRI using aligned CT-based labels, laying the foundation for future MRI-based biomarker quantification.

Chapter 4

General Methodology

This thesis presents a retrospective analysis of imaging and clinical data from patients treated with spinal stereotactic body radiotherapy (SBRT) at Sunnybrook Health Sciences Centre. The work is divided into two main components.

The first component focused on CT-based biomarker development and clinical analysis. We developed and validated an automated pipeline for extracting quantitative imaging biomarkers related to sarcopenia and osteoporosis, including psoas muscle volume, psoas cross-sectional area, vertebral body volume, and trabecular bone density. To define sarcopenia, we adopted a volumetric approach using psoas muscle volume within a standardized anatomical region [L3-L4], normalized by vertebral body volume. This allowed for fully automated, height-independent classification. Height-dependent thresholds were also computed, as they may be preferred in contexts where patient height is available. Thresholds were derived from a subgroup of 80 patients (41 female, 39 male) out of the full SBRT-treated cohort of 143 patients (65 female, 78 male), validated in an external cohort, and then applied to the full cohort to define osteosarcopenia. These biomarkers were further

analyzed for associations with clinical outcomes, including vertebral fracture risk and overall survival.

The second component focused on extending the segmentation pipeline to MRI. Using aligned CT-MRI pairs, we transferred CT-based segmentations onto MRI volumes to generate labeled training data. We then developed and evaluated nnU-Net models for segmenting psoas muscles and vertebral bodies in MRI, enabling future MRI-based biomarker extraction and multimodal analysis.

4.1 Study Design and Ethics

This work is embedded within a retrospective cohort study conducted at the Sunnybrook Research Institute (SRI), in collaboration with the Holland Bone and Joint Bioengineering Program. The study focuses on patients treated with SBRT for spinal metastases and aims to evaluate musculoskeletal health through imaging-derived biomarkers.

Ethics approval was granted by the Sunnybrook Research Ethics Board (REB # 419-2015 [old], 2623 [new], Approval Date: 12-Nov-2015). Access to de-identified imaging and clinical data was granted following completion of institutional training and onboarding. All data analysis was conducted on secure, on-premise infrastructure.

4.2 Datasets and Cohort Description

This study incorporates three de-identified imaging datasets acquired at Sunnybrook Health Sciences Centre. These include:

- **Sunnybrook Spine SBRT Patient Cohort (SpineMets):** A longitudinal database

of patients with spinal metastases treated with SBRT between 2008 and 2022, containing CT imaging, tumor characteristics, and clinical outcomes.

- **Young Advanced Prostate Cancer Patient Cohort:** A cohort of 93 male patients under 55 years of age treated with systemic therapy for advanced prostate cancer, who underwent repeated lumbar spine CT imaging between 2009 and 2021.

- **Fused CT-MRI Subset of SBRT Patients:** A subset of 217 patients from the SpineMets cohort whose CT and MRI scans were aligned as part of SBRT treatment planning. These aligned pairs include both T1- and T2-weighted MRI scans and corresponding planning CTs.

Detailed descriptions of each dataset are provided in the subsections below.

4.2.1 Sunnybrook Spine SBRT Patient Cohort (SpineMets)

This cohort comprises patients with spinal metastases treated with SBRT at Sunnybrook Health Sciences Centre between January 1, 2008, and December 31, 2022. A total of 977 patients with 2364 spinal segments were treated. The database includes planning CT imaging, baseline clinical and tumor data, and longitudinal outcomes such as vertebral compression fractures.

A subset of 143 patients with CT scans covering the lumbar spine (L1–L5) was used for further analysis. Of these, 80 patients had recorded height information. Applications of this cohort in biomarker development and outcome analysis are detailed in Sections 5.2 and 5.3.5.

4.2.2 Young Advanced Prostate Cancer Patient Cohort

This dataset comprises 93 young patients ($<$ 55 years of age at diagnosis) diagnosed with advanced prostate cancer, who underwent a total of 644 CT scans at the Sunnybrook Odette Cancer Centre (Toronto, ON, Canada) between January 2009 and December 2021. All CT scans included imaging of the lumbar spine (L1–L5). These patients underwent regular imaging surveillance, typically ranging from one to three CT scans per year, as part of routine metastatic disease monitoring.

This cohort was used for the external validation of the sarcopenia thresholds developed within the SpineMets cohort. Although the prostate cancer cohort consisted exclusively of male patients, it provided a relatively younger and healthier reference group to assess the generalizability of the CT-based imaging biomarkers. A subset of 69 patients with available height measurements was also identified for this purpose. To minimize potential bias associated with variable imaging frequency, only baseline CT scans were included in the validation analysis.

These patients were selected due to their younger age and presumed absence of muscle wasting, making them a suitable reference population for evaluating sex-specific sarcopenia thresholds.

4.2.3 Fused CT-MRI Subset of SBRT Patients

The Fused Dataset comprises 217 CT scans that have been manually pre-aligned with their corresponding MRI scans (T1- and T2-weighted) and was used to support nnU-Net training for psoas muscle segmentation (Figure 4.1). This alignment ensures anatomical consistency across modalities, allowing the structural detail from CT-based segmentations to be used as

ground truth for training MRI segmentation models.

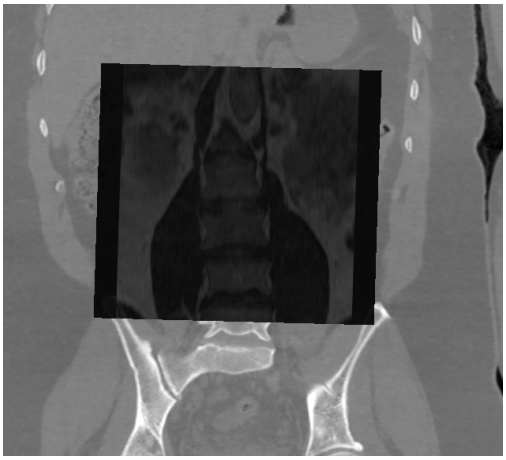


Figure 4.1: Registered CT-MRI T2-weighted image.

CT scans including the lumbar region (L1–L5) were selected for inclusion, as will be detailed in Chapter 6. Each CT scan was first segmented using the CT biomarker extraction pipeline described in Section 5.2, providing reference segmentations for training MRI-based models.

The following chapter presents the methodology and results of the osteosarcopenia analysis in our SBRT patient cohort, including biomarker extraction from CT, classification strategy, and associations with clinical outcomes.

Chapter 5

Characterization of Osteosarcopenia in Spinal Metastases Patients Using Quantitative Image-Based AI-Derived Biomarkers

This chapter describes the methods and analyses used to define and apply imaging-based musculoskeletal biomarkers for characterizing osteosarcopenia in a cohort of patients with spinal metastases undergoing SBRT.

We developed CT-derived metrics to assess sarcopenia and osteoporosis, enabling fully automated and height-independent quantification of muscle and bone health. Central to this work is the development of a fully automated biomarker extraction pipeline that leverages 3D segmentation of vertebral bodies and psoas muscles from CT images.

Our approach involved defining a volumetric index for sarcopenia and thresholds based

on validated relationships with traditional indices, and evaluating these thresholds using an external cohort of young advanced prostate cancer patients. We then applied these biomarkers to quantify osteosarcopenia prevalence, assess associations with vertebral fracture risk, and explore survival outcomes in patients treated with SBRT.

The analyses in this chapter provides a foundation for using artificial intelligence (AI)-derived image biomarkers in clinical risk stratification and treatment planning for spinal oncology.

5.1 Study Objectives

This study aimed to characterize osteosarcopenia in patients with spinal metastases using quantitative imaging biomarkers derived from routine computed tomography (CT) scans obtained for spine stereotactic body radiotherapy (SBRT) planning.

The primary objectives of the CT-based component were to:

- Develop a volumetric sarcopenia index using CT-based features and establish sex-specific thresholds through regression and ROC analysis.
- Validate the proposed sarcopenia thresholds in an independent cohort of young, advanced prostate cancer patients.
- Apply the derived thresholds to classify sarcopenia, osteoporosis, and osteosarcopenia in the SBRT-treated SpineMets cohort.
- Investigate associations between imaging-based classifications and vertebral fracture risk.
- Evaluate the prognostic value of musculoskeletal biomarkers for overall survival.

5.2 Biomarker Extraction Pipeline

A core component of our study was the extraction of quantitative musculoskeletal biomarkers from CT scans. This required accurate segmentation of anatomical structures and standardized computation of bone and muscle metrics. To support this, we developed and automated a dedicated processing pipeline tailored to compute imaging-based biomarkers relevant to osteosarcopenia.

As illustrated in Figure 5.1, the pipeline is a multi-step process designed to quantify musculoskeletal biomarkers from CT images, focusing on the lumbar spine and psoas muscles. It includes automated segmentation and quantification of both bone and muscle structures. Key biomarkers extracted include vertebral body trabecular bone density, vertebral body volume, psoas muscle volume, and the psoas cross-sectional area (CSA), typically computed at the L3 vertebral centroid. The entire pipeline was implemented in 3D Slicer, with custom Python scripts enabling consistent, automated execution across patient datasets.

While the pipeline implementation was started as part of earlier work, we extended and finalized its implementation for this thesis. Our contributions included automating the full pipeline for the patient cohort, integrating TotalSegmentator for psoas segmentation, integrating vertebra-guided cropping, adding a largest connected component filter to remove residual artifacts after cropping, and extending the CSA computation module to support measurements at the centroids of all segmented vertebrae—not only for the psoas but also for vertebral bodies.

1. **Segmentation of Vertebral Bodies.** Vertebrae were detected and labeled using VertDetect, a deep learning-based model designed at Sunnybrook Research Institute for automated vertebral detection and segmentation in CT scans [66]. VertDetect uses a

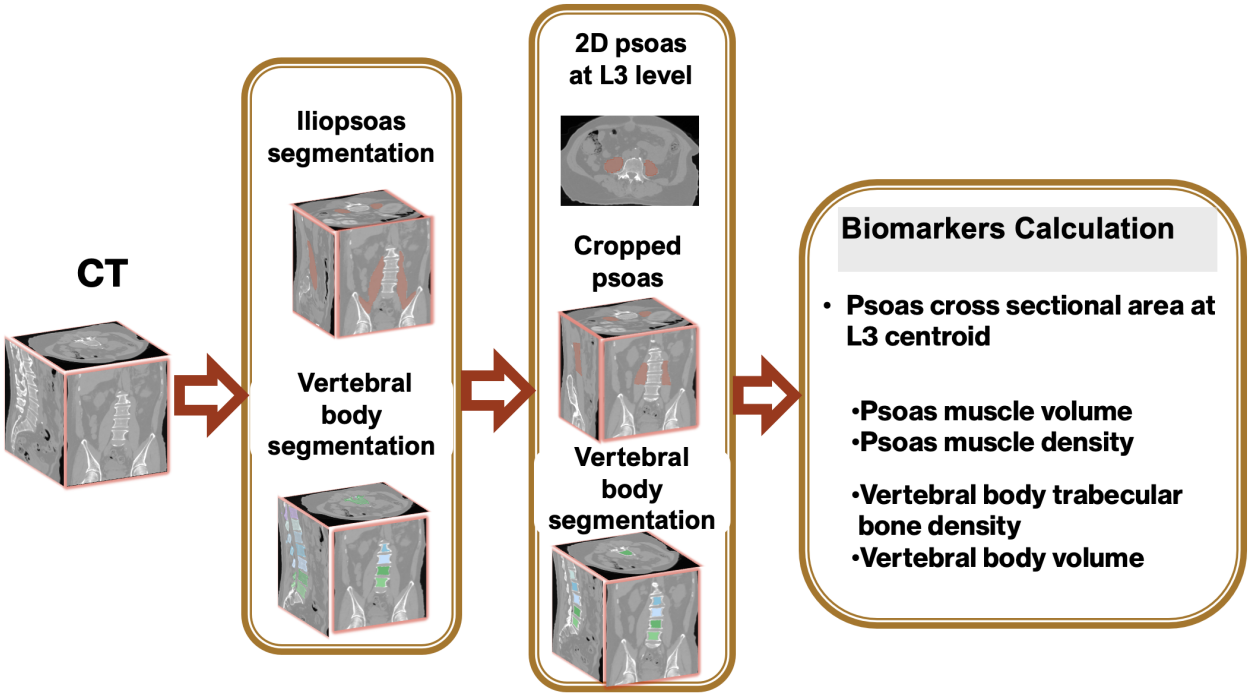


Figure 5.1: Biomarker extraction pipeline

Convolutional Neural Network (CNN) architecture with a shared ResNet backbone and a feature pyramid network (FPN) to extract both spinal and vertebral-level features. Additionally, a Graph Convolutional Network (GCN) layer enhances vertebral labeling by incorporating the known anatomical structure of the spine.

Vertebral body segmentation was then performed using a separate U-Net model, which takes the centroids provided by VertDetect as input to localize and segment individual vertebral bodies, which served as anatomical landmarks for subsequent muscle segmentation and biomarker quantification.

2. **Psoas Muscle Segmentation.** The psoas muscle was segmented using TotalSegmentator [67], a neural network-based (nnU-Net [68]) tool developed for segmentation of most major anatomical structures in abdominal CT volumes.

Validation and Adjustment of Psoas Volume Measurements. As part of this thesis, we integrated TotalSegmentator into the pipeline to automate psoas muscle segmentation. To validate its performance, we compared its output against ground truth segmentations manually delineated by a medical professional in a subset of four patients.

Table 5.1 shows the predicted psoas muscle volumes from TotalSegmentator and the corresponding manually segmented volumes for four CT cases. All volumes were cropped to a consistent anatomical region using the vertebra-guided cropping module described in step 3 .

Table 5.1: Comparison of TotalSegmentator and manually segmented psoas muscle volumes (in cm^3) across four CT cases.

Patient	**Manual Volume (cm^3)**	**TotalSegmentator Volume (cm^3)**	**Ratio (Manual / TS)**
Patient A	120.51	102.42	1.18
Patient B	138.26	118.13	1.17
Patient C	100.13	83.36	1.20
Patient D	137.50	118.85	1.16
Mean	—	—	**1.18**

To correct for systematic under-segmentation by TotalSegmentator, we calculated the ratio between manual and predicted psoas muscle volumes for the four manually annotated CT cases. The average ratio was approximately 1.18, indicating that TotalSegmentator tended to underestimate psoas volume. To ensure consistency across the dataset, we applied a correction factor of 1.19 to all predicted psoas volumes during

biomarker quantification.

3. **Standardized Psoas Muscle Cropping.** To ensure consistency in volume measurements across patients with varying orientations and scans with varying fields of view, the psoas muscle was cropped to the anatomical region spanning the L2/L3 to L4/L5 intervertebral disc levels. This region was selected based on its stable anatomical relationship with the vertebral column, minimizing variability due to patient positioning. Cropping was performed using automated vertebral segmentation to align muscle boundaries with vertebral endplates. The anatomical cropping planes were determined using the centroids of the vertebral body segmentations. Two planes were defined:

    - The first plane was defined midway between the centroids of the L4 and L5 vertebral bodies.
    - The second plane was defined midway between the centroids of the L2 and L3 vertebral bodies.

4. **Quantification of Biomarkers.** As part of the segmentation pipeline, key biomarkers were automatically extracted using 3D Slicer. These included the trabecular bone density (Hounsfield Units, HU) of the lumbar vertebral bodies (L1–L5), the volume of the lumbar vertebral bodies, and the volume and density (mean and standard deviation) of the segmented psoas muscle. The psoas cross-sectional area (CSA) was aslo determined at the level of the L3 vertebral body. Quantification was performed using the SegmentStatistics module for bone and muscle volume, and custom scripts to calculate the psoas CSA via the SegmentCrossSectionArea module.

5.3 Application of the Pipeline to the SpineMets Dataset

The full SpineMets dataset includes 977 patients and 2364 spinal segments treated with SBRT. For this study, we restricted our analysis to patients whose planning CT scans included full coverage of the lumbar vertebrae L1 through L5. To identify these cases, we used VertDetect [66] to extract vertebral body centroids and verify anatomical coverage.

Eligible CT volumes were then processed using the automated biomarker extraction pipeline described in Section 5.2. This generated muscle and bone biomarkers per scan, which were stored in structured CSV files and later merged with clinical variables for analysis.

Figure 5.2 summarizes the overall data processing workflow.

5.3.1 Biomarker Extraction from Imaging Data

Each CT volume covering L1–L5 was processed using the pipeline described in Section 5.2 to extract musculoskeletal biomarkers relevant to osteosarcopenia. These included trabecular bone density, vertebral body volume, psoas muscle volume, and psoas CSA.

The resulting biomarker values, along with the hashed patient ID and the full CT path, were saved in a structured CSV file, forming the imaging dataset to be linked with the clinical information.

The clinical component of the SpineMets dataset is highly detailed and includes a wide range of patient- and vertebra-level variables. Patient-level information encompasses demographic data (such as sex and age), cancer history, and survival status. Vertebra-level variables include the specific spinal segment treated, the SBRT treatment start date, radiation dose and number of fractions, performance status indicators (e.g., ECOG score, neurological

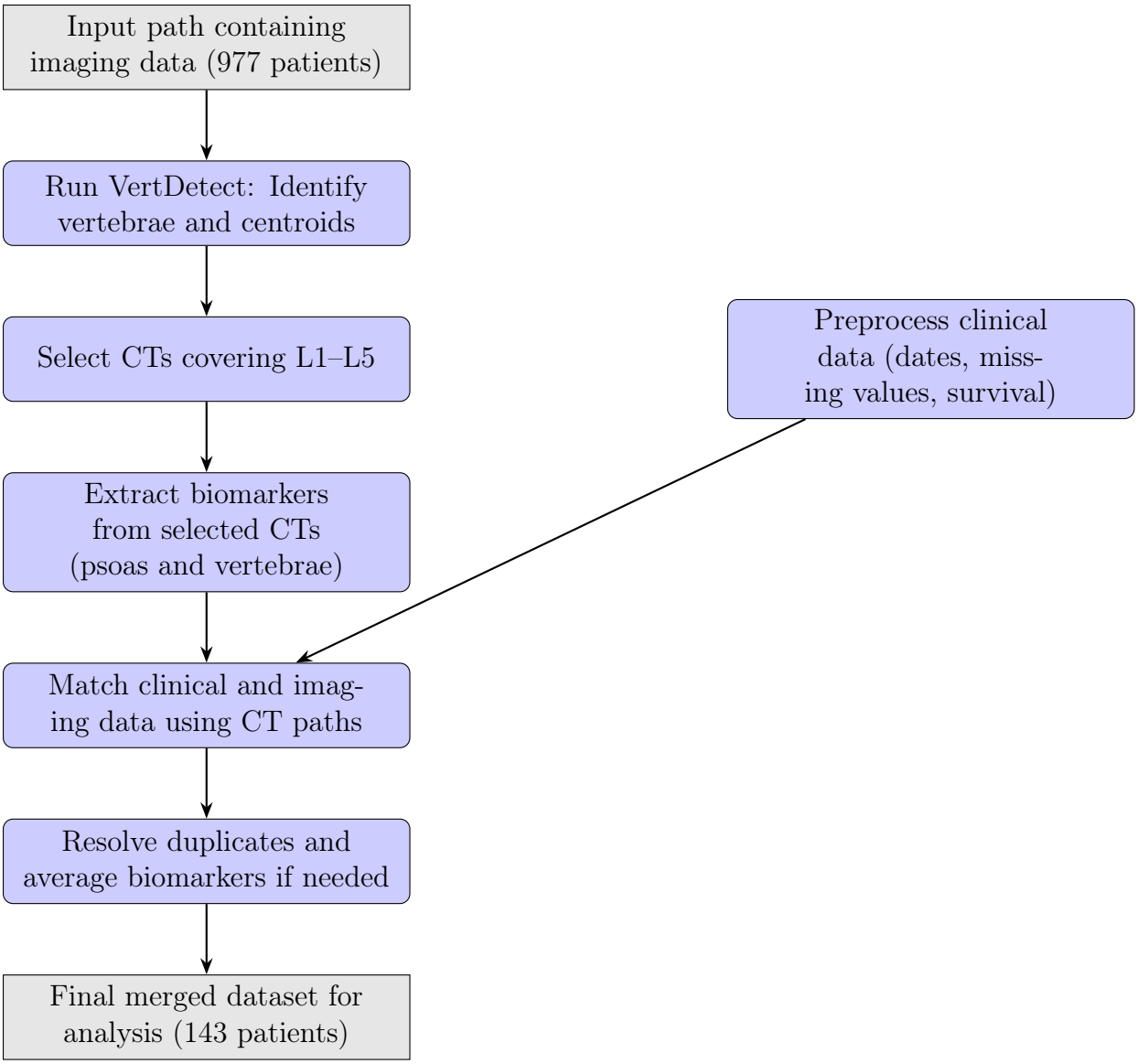


Figure 5.2: Flowchart summarizing the application of the biomarker extraction pipeline to the SpineMets dataset.

status), and disease-related characteristics including primary tumor site and the presence of spinal or extra-spinal metastases. Additionally, imaging metadata such as planning CT file paths is included. While the full set of clinical variables is retained for reference, Table 5.2 summarizes some of the most relevant features.

5.3.2 Clinical Data Preprocessing and Imaging Inclusion Criteria

Clinical data were preprocessed to standardize formats and address missing or inconsistent entries. All date-related columns including death date, treatment start and follow-up, were converted to datetime format to enable accurate time-based computations. Patient-level variables, such as sex, age at treatment, and prior cancer history, were filled within each patient's record to ensure consistency across multiple treatment entries. Similarly, visit-specific fields like ECOG score, neurological status, and SINS classification were imputed per patient and SBRT session. Demographic variables such as age and sex were extracted, alongside treatment-related information including vertebral levels targeted, prescribed radiation dose, and history of prior and new fractures. Patients often have multiple vertebrae treated with SBRT both at individual time points or over multiple time points.

To support survival analysis, we computed additional derived columns. The survival duration (in months) was calculated based on each patient's earliest SBRT start date and either the recorded death date or the most recent follow-up, whichever was available. We also created a `Death_Source` column to distinguish between patients with confirmed deaths and those censored at last follow-up.

5.3.3 Linking Clinical Records with Imaging Biomarkers

To merge imaging-derived biomarker data with clinical records, we used patient ID, scan date, and the vertebral levels imaged in the CT scan to match treated lesions with corresponding imaging. In some cases, a single treatment entry matched multiple CT scans; these entries were expanded by duplicating the treatment entry so that each CT path appeared in its own row.

Based on the imaging inclusion criteria described in Section 5.2 and the availability of relevant clinical data, 144 patients were initially identified from the SpineMets dataset (Section 4.2.1) as having planning CT scans covering the lumbar spine (L1–L5), where the psoas muscle could be consistently visualized and analyzed. One patient was excluded due to missing age data, resulting in a final analytic cohort of 143 patients. Height data, which is required for some sarcopenia measures, was only available for a subgroup of 80 patients.

5.3.4 Resolving Redundancies and Ambiguities in CT-to-Clinical Matching

After merging the imaging and clinical datasets, we identified two types of duplication issues that required resolution:

1. **Exact duplicates** — In 26 patients, multiple clinical rows linked to CT scans yielded identical biomarker values. These were likely due to the same CT scan being listed under multiple paths or plans. We removed these exact duplicates, retaining only the first occurrence of each unique patient-vertebra-treatment entry.

2. **Non-identical entries with overlapping clinical identifiers** — In 6 patients, a single vertebra was matched to multiple CT paths that produced slightly different biomarker values. For these cases, we averaged the biomarker values across scans to generate a consolidated, representative entry for each treated vertebra.

This ensured that the final merged dataset used for analysis contained one biomarker set per treated vertebra per patient.

To characterize the prevalence of osteosarcopenia, we used the first available CT scan per patient, ensuring consistent baseline assessment across the cohort. For fracture risk

analysis, we used the CT scan obtained during SBRT treatment planning, as it best reflects musculoskeletal status at the time of treatment.

5.3.5 SpineMets Patient Demographics and Clinical Data

Table 5.2 summarizes the demographic and clinical characteristics of the study population.

Table 5.2: Patient Demographics and Clinical Characteristics

Variable	**Value**
Sex (male)	78/143 (54.5%)
Age (years)	Median: 64.1 (Min: 28.4, Max: 90.5)
Mortality Rate	68/143 (47.6%)
Median Survival (months)	9.6 (IQR: 6.4–21.2)
ECOG Performance Status	
0 (Asymptomatic)	45/143 (31.5%)
1 (Symptomatic, ambulatory)	77/143 (53.8%)
2 (Symptomatic, <50% in bed)	17/143 (11.9%)
3 (Symptomatic, >50% in bed)	1/143 (0.7%)
4 (Bedbound)	2/143 (1.4%)
Neurological Status	
Normal (Motor and sensory intact)	120/143 (83.9%)
Abnormal, Incomplete	17/143 (11.9%)
Abnormal (Sensory only)	5/143 (3.5%)

Continued on next page

Table 5.2 – *Continued from previous page*

Variable	**Value**
Past Cancer History	
No history of past cancer	128/143 (89.5%)
Yes, past cancer history	14/143 (9.8%)
Primary Tumour Site	
Breast	33/143 (23.1%)
Non-Small Cell Lung Cancer (NSCLC)	28/143 (19.6%)
Renal Cell Carcinoma (RCC)	24/143 (16.8%)
Prostate	24/143 (16.8%)
Colon	11/143 (7.7%)
Melanoma	7/143 (4.9%)
Thyroid	4/143 (2.8%)
Hepatocellular Carcinoma (HCC)	3/143 (2.1%)
Head & Neck Cancer (H&N)	3/143 (2.1%)
Urothelial	2/143 (1.4%)
Endometrial	1/143 (0.7%)
Gastrointestinal Stromal Tumour (GIST)	1/143 (0.7%)
Esophagus	1/143 (0.7%)
Unknown primary	1/143 (0.7%)
Spine Metastases	
1 metastasis	39/143 (27.3%)

Continued on next page

Table 5.2 – *Continued from previous page*

Variable	**Value**
2 metastases	31/143 (21.7%)
3 metastases	17/143 (11.9%)
4 metastases	14/143 (9.8%)
≥5 metastases	31/143 (21.7%)
Other Metastases	
Bone Metastases	74/143 (51.7%)
Lung Metastases only	32/143 (22.4%)
Liver Metastases only	21/143 (14.7%)
Both Lung & Liver Metastases	16/143 (11.2%)
No Lung/Liver Mets	73/143 (51.0%)
Brain Metastases	26/143 (18.2%)
SBRT Site	
Lumbar (L1–L5)	96/143 (67.1%)
Thoracic (T1–T12)	42/143 (29.4%)
Cervical (C1–C7)	3/143 (2.1%)
Sacral (S1–S5)	2/143 (1.4%)
Spinal Disease Characteristics	
Baseline Vertebral Compression Fracture	29/143 (20.3%)
Epidural Disease	53/143 (37.1%)
Treatment History	

Continued on next page

Table 5.2 – *Continued from previous page*

Variable	**Value**
Previous Spine Surgery	20/143 (14.0%)
Previous Radiotherapy	41/143 (28.7%)
SINS Score (Spinal Instability)	
Stable (0–6)	73/143 (51.0%)
Potentially Unstable (7–12)	62/143 (43.4%)
Unstable (13–18)	7/143 (4.9%)

5.3.6 Statistical Methods

Statistical analyses were conducted to define sarcopenia thresholds, validate volumetric indices, and evaluate associations between musculoskeletal biomarkers and clinical outcomes including fracture risk and survival. All analyses were performed in Python using the `SciPy`, `lifelines`, and `statsmodels` libraries. A two-tailed p-value less than 0.05 was considered statistically significant.

To define sex-specific sarcopenia thresholds, linear regression was performed to model the relationship between traditional PMI (normalized by height squared) and CT-derived volumetric indices (e.g., PMV/VBV(L3), PMV/VBV(L3)$^{1/2}$). Regression parameters (slope, R^2, and MAE) were used to derive volumetric thresholds aligned with standard PMI cut-offs.

Thresholds were also derived using Receiver Operating Characteristic (ROC) curve analysis. The area under the ROC curve (AUC) was used to evaluate classification performance, and optimal cut-offs were selected by maximizing Youden's J statistic.

To validate thresholds across datasets, thresholds derived in the SpineMets cohort were

applied to an independent young advanced prostate cancer cohort, and agreement in sarcopenia classification was assessed descriptively.

To evaluate associations between categorical classifications (e.g., sarcopenia, osteoporosis, osteosarcopenia) and fracture risk, chi-square tests were used for the full cohort and sex-specific subgroups.

To compare continuous biomarkers between fractured and non-fractured vertebrae, Mann–Whitney U tests were used when data were not normally distributed, and Welch's t-tests were used when normality was assumed but variance equality was not met. Selection of test type was based on statistical assessment of data distribution.

Effect sizes for continuous comparisons were reported using Cohen's d. Post hoc power analyses were performed when findings approached significance to estimate the sample size needed to detect a true effect.

To assess the independent contributions of muscle and bone biomarkers to vertebral fracture risk, multivariate logistic regression models were fitted. Predictor variables included continuous imaging-derived features and binary classifications

Survival outcomes were analyzed using Kaplan–Meier estimators. Time-to-event was defined as time from SBRT start to death or last follow-up. Patients were stratified by biomarker-defined groups, and survival differences were assessed using the log-rank test. For analyses involving more than two groups (e.g., sarcopenic only, osteoporotic only, osteosarcopenic, or neither), a multivariate log-rank test was applied.

5.4 Osteoporosis Methodology

Osteoporosis was evaluated in vertebral bodies without metastatic involvement. Vertebral bodies with metastatic involvement were detected and excluded from bone density calculations. Metastatic involvement was detected within the vertebral bodies using a previously developed method by our group at Sunnybrook Research Institute. Tumour involvement causes variability in bone deposition, osteolytic tumours resorb bone and decrease bone density, whereas osteoblastic lesions cause excessive bone deposition and mineralization. Both of these types of tumour lead to a large distribution in bone density within the vertebral body. As such, standard deviation (STD > 75HU) of bone density within the trabecular centrum of the vertebral body was used to classify vertebrae as metastatically involved.

For each patient, the bone mineral density was then computed as the mean HU value across all remaining (non-metastatic) vertebral bodies. A threshold of 135 HU was used to classify patients into osteoporotic and non-osteoporotic groups. Specifically, patients with mean vertebral HU below 135 were classified as osteoporotic, while those with HU above this threshold were classified as non-osteoporotic. This approach enabled a consistent and objective stratification of patients based on quantitative imaging data.

5.5 Sarcopenia Methodology

5.5.1 Defining a CT-Based Volumetric Index for Sarcopenia

Sarcopenia is commonly assessed using imaging-based indices such as the Skeletal Muscle Index (SMI) or the Psoas Muscle Index (PMI), particularly in oncologic studies. Both indices are derived from the cross-sectional area (CSA) of skeletal muscle at the third lumbar

vertebra (L3) and normalized by the patient's height squared (H^2) to account for body size differences. While SMI includes multiple muscle groups, PMI focuses on the bilateral psoas muscles and is often used when full segmentation is not feasible. PMI is calculated as:

$$PMI = \frac{PCA}{H^2} \tag{5.1}$$

where PCA is the combined cross-sectional area of the left and right psoas muscles (cm^2), and H is the patient's height in meters.

PMI has shown clinical utility as a surrogate for muscle mass and has been associated with outcomes such as treatment toxicity, complications, and overall survival in various cancer populations [44, 69–71], also see section 3.2. However, both PMI and SMI present limitations. Their dependence on a single axial slice may lead to inaccurate representation of muscle mass, especially in patients with spinal deformities or tumors at L3. Moreover, both indices require accurate height data, which is frequently missing open datasets and in retrospective cohorts—as in our dataset, where only 56% of patients (80 of 143) had height recorded.

To overcome these limitations, we developed a CT-based volumetric alternative to PMI that removes the dependency on height and better reflects total psoas muscle mass. We proposed a Volumetric Psoas Muscle Index (vPMI), which normalizes the psoas muscle volume (PMV) by vertebral body volume (VBV) at the L3 level:

$$vPMI = \frac{PMV}{VBV_{L3}^{\text{power}}} \tag{5.2}$$

This approach removes the need for height and provides a more comprehensive representation of muscle mass. We tested several normalization strategies by varying the power term

(e.g., linear, square-root) and also evaluated multi-level VBV references (e.g., VBV_{L3+L4}).

Although the main motivation was to support analysis in patients without height data, we also derived height-normalized volumetric indices—vPMI(h), vPMI(h^2), and vPMI(h^3)—to enhance comparability with conventional methods such as PMI and BMI, which rely on height-based normalization. These indices may be preferred in studies or clinical settings where height is available and body-size adjustment is essential.

To establish sarcopenia thresholds, we analyzed a subset of 80 SpineMets patients with available height data. Volumetric indices were grouped into vertebral-based (e.g., PMV/VBV(L3), PMV/(VBV(L3)+VBV(L4)), PMV/$\text{VBV(L3)}^{1/2}$) and height-based (e.g., PMV/h^3, PMV/h^2, PMV/h) categories. We used linear regression and ROC analysis to align volumetric indices with sex-specific PMI cutoffs from literature [44], enabling classification consistent with conventional sarcopenia definitions.

Thresholds were validated in an independent cohort of 68 young prostate cancer patients (490 CTs). To avoid confounding from disease progression or treatment effects, baseline scans were used for threshold derivation.

5.5.2 Deriving and Validating Volumetric Sarcopenia Thresholds

Sex-specific thresholds for sarcopenia classification were derived using two complementary methods. First, linear regression was used to model the relationship between each volumetric index and the height-normalized Psoas Muscle Index (PMI), allowing standard PMI cutoffs (<6.36 cm^2/m^2 for men, <3.92 cm^2/m^2 for women) to be translated into volumetric thresholds. Second, receiver operating characteristic (ROC) curve analysis was performed to identify optimal thresholds by maximizing Youden's J statistic.

The strength of each volumetric index was evaluated using the coefficient of determination (R^2) and mean absolute error (MAE). Among the CT-only indices, PMV/VBV(L3)$^{1/2}$ showed the strongest agreement with traditional PMI, outperforming direct and quadratic normalization strategies. While height-based indices such as vPMI(h^3), vPMI(h^2), and vPMI(h) yielded even higher correlations ($R^2 > 0.85$) and AUCs, their reliance on patient height limits clinical utility in retrospective analyses.

Figures 5.3 and 5.4 illustrate the linear relationships for both CT-based and height-based indices in male and female patients, respectively. Table 5.3 summarizes the regression- and ROC-derived thresholds, along with AUC values for each index.

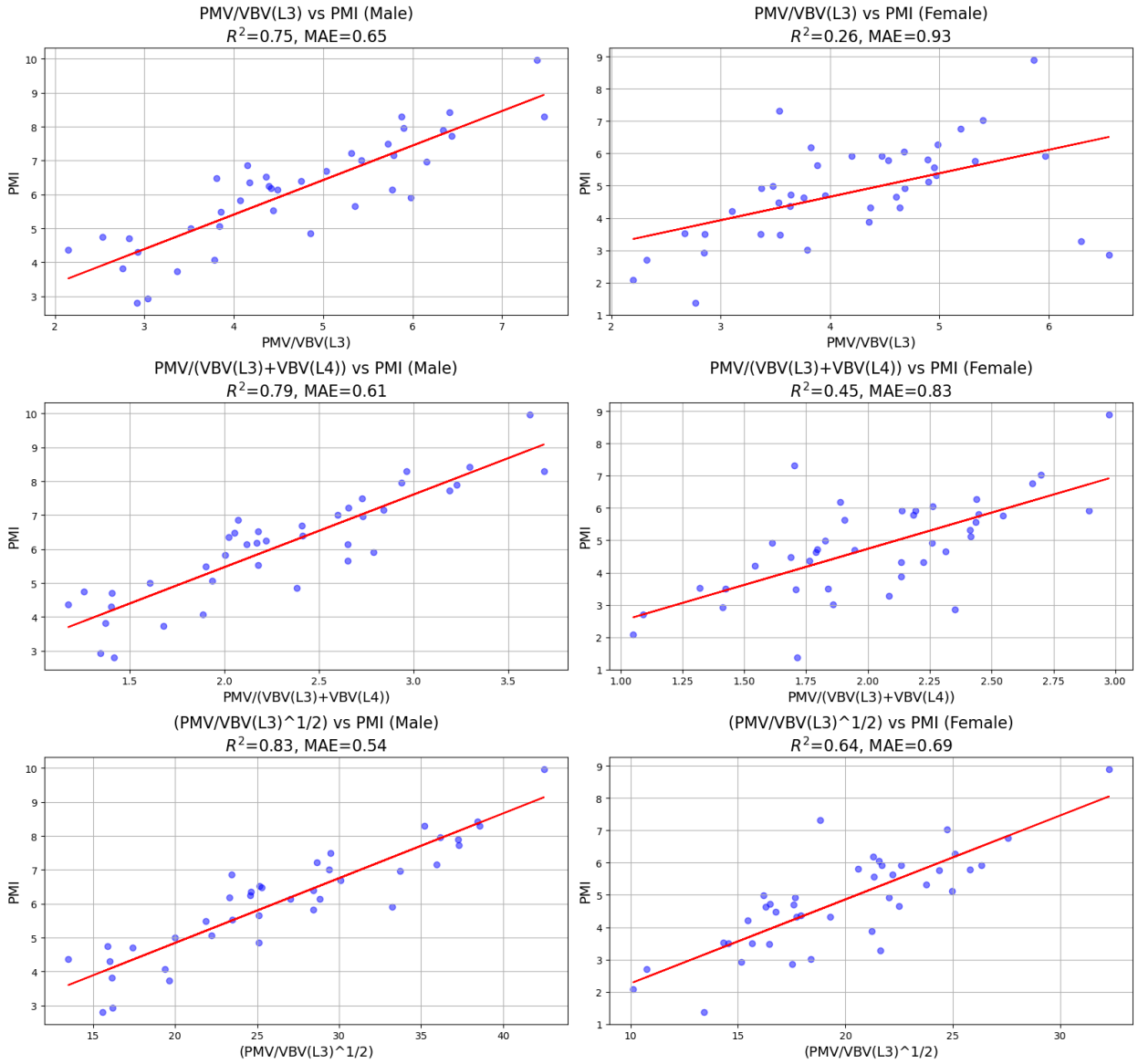


Figure 5.3: Linear regressions between PMI and CT-based volumetric indices (PMV/VBV(L3), PMV/(VBV(L3)+VBV(L4)), and PMV/VBV(L3)$^{1/2}$) for male and female SpineMets patients.

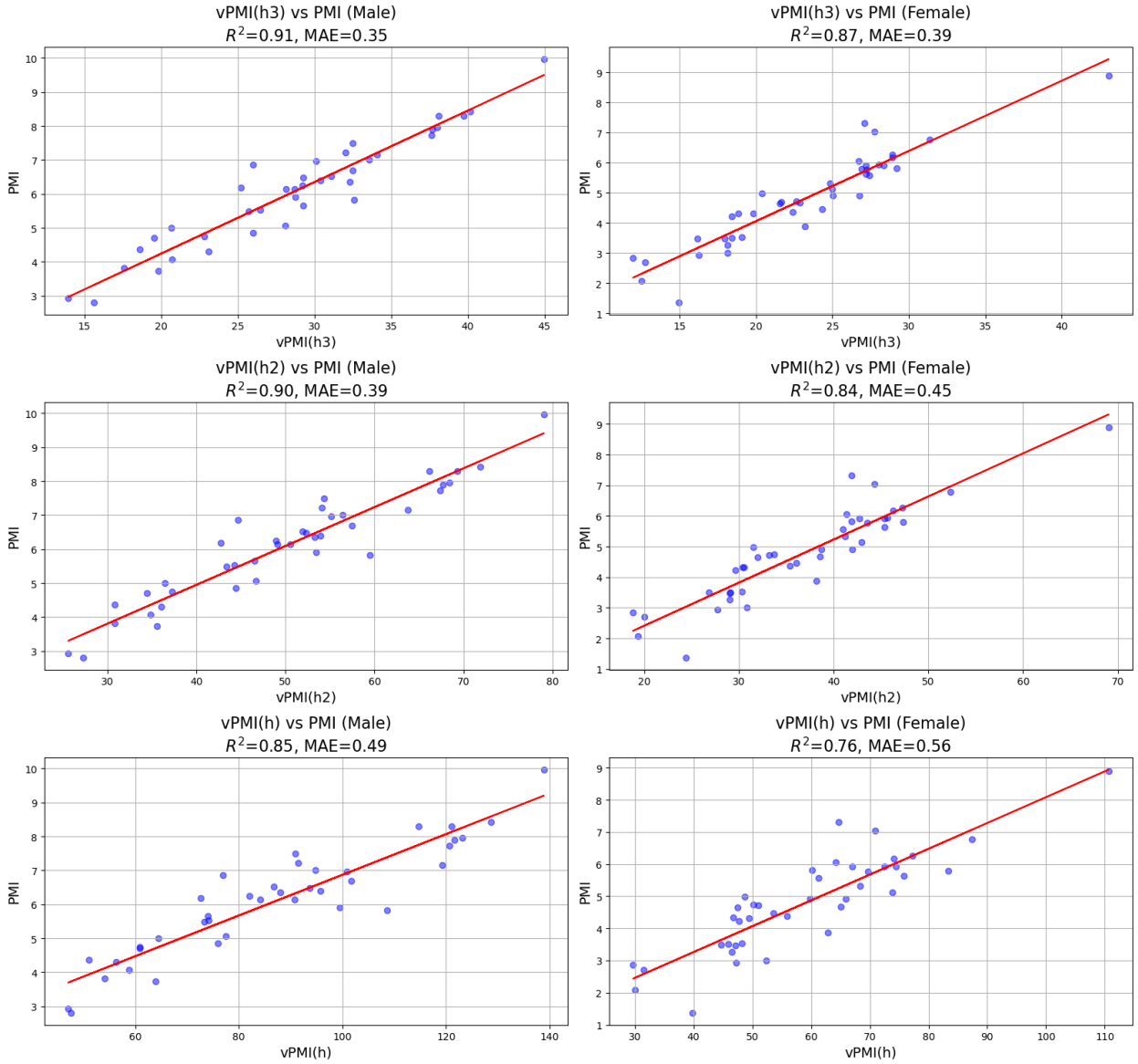


Figure 5.4: Linear regressions between PMI and height-based volumetric indices (vPMI(h^3), vPMI(h^2), and vPMI(h)) for male and female SpineMets patients.

Table 5.3: SpineMets Patients: Linear Regression and ROC Analysis for Volumetric Sarcopenia Indices

Volumetric Index	**Sex**	**Linear Threshold**	**ROC Threshold**	**ROC AUC**
vPMI(PMV/VBV(L3)$^{1/2}$)	Male	27.97	25.11	0.94
	Female	16.38	16.45	0.87
vPMI(PMV/VBV(L3))	Male	4.95	4.49	0.88
	Female	2.99	3.37	0.76
vPMI(PMV/(VBV(L3)+VBV(L4)))	Male	2.42	2.38	0.91
	Female	1.64	2.13	0.81
vPMI(h^3)	Male	30.05	29.20	0.94
	Female	19.38	19.05	0.97
vPMI(h^2)	Male	52.32	50.54	0.95
	Female	30.72	30.84	0.96
vPMI(h)	Male	91.42	90.72	0.94
	Female	48.18	48.20	0.93

Overall, PMV/VBV(L3)$^{1/2}$ was selected as the preferred CT-based surrogate for sarcopenia classification in the SpineMets cohort due to its strong performance across sexes (ROC AUC = 0.94 for males, 0.87 for females) and independence from height data. The derived regression-based and ROC-derived thresholds showed close agreement, supporting the robustness of this index for automated classification.

Validation of Volumetric Sarcopenia Thresholds in Young Advanced Prostate Cancer Patients

To evaluate sarcopenia classification robustness, we performed threshold derivation and validation using an independent cohort of young advanced prostate cancer patients (see Section 4.2.2).

Linear regression and receiver operating characteristic (ROC) analyses were applied to the young advanced prostate cancer cohort using the same methodology as in the SpineMets cohort. Thresholds were derived for all proposed volumetric sarcopenia indices, allowing direct comparison of their behavior across datasets.

Table 5.4 presents the comparison of thresholds between the SpineMets male cohort and the young advanced prostate cancer cohort. Thresholds were comparable across all volumetric indices. The PMV/VBV(L3)$^{1/2}$ index showed close agreement, with linear regression thresholds of 27.94 and 28.17 and ROC-derived thresholds of 25.11 and 28.41 for the SpineMets and young advanced prostate cancer cohorts, respectively. Height-normalized indices demonstrated similar patterns, with slightly greater variability.

To evaluate classification agreement, we applied the SpineMets-derived threshold of 27.94 for PMV/VBV(L3)$^{1/2}$ to all available scans in the prostate cancer cohort, including follow-up scans. Sarcopenia classification using this volumetric index was then compared to classification using the traditional PMI, and agreement was assessed using concordance plots (Figure 5.5) and Cohen's kappa statistic.

Table 5.4: Comparison of Sarcopenia Thresholds Between SpineMets Male and Young Advanced Prostate Cancer Cohorts

Volumetric Index	**SpineMets Linear**	**Young Linear**	**SpineMets ROC**	**Young ROC**
PMV/VBV(L3)$^{1/2}$	27.97	28.17	25.11	28.41
PMV/VBV(L3)	4.95	5.02	4.49	5.35
PMV/(VBV(L3)+VBV(L4))	2.42	2.50	2.38	2.70
vPMI(h^3)	30.05	30.12	29.20	30.23
vPMI(h^2)	52.32	52.59	50.54	54.34
vPMI(h)	91.42	90.73	90.72	89.18

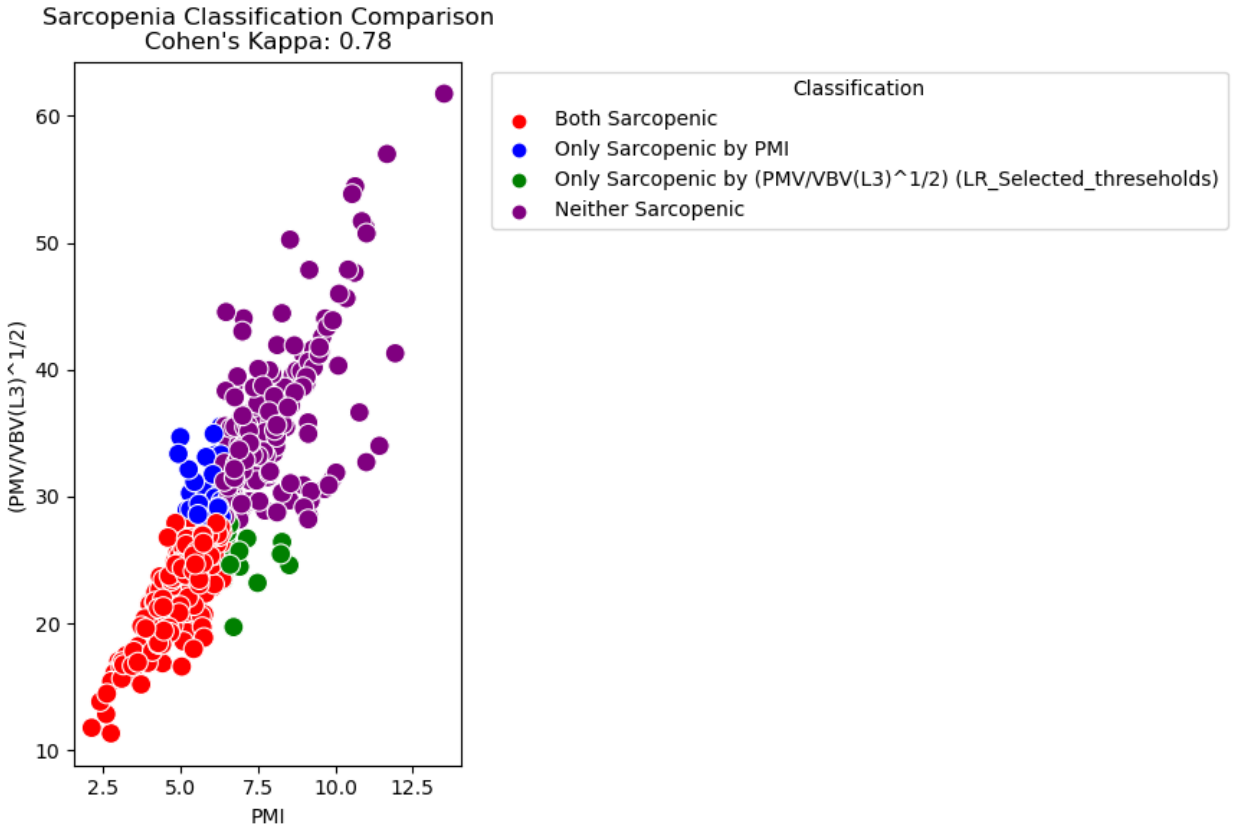


Figure 5.5: Sarcopenia classification comparison between PMI and PMV/VBV(L3)$^{1/2}$ using the SpineMets-derived linear regression threshold. Classification was performed on all available scans.

5.5.3 Sarcopenia classification methodology

To classify sarcopenia in our cohort, we adopted sex-specific thresholds derived from the volumetric index PMV/VBV(L3)$^{1/2}$, which was identified as the most consistent CT-based surrogate for the traditional Psoas Muscle Index (PMI). This approach enables sarcopenia classification without requiring external clinical information such as patient height, supporting a fully automated imaging-based pipeline.

Thresholds were obtained from linear regression analyses aligning PMV/VBV(L3)$^{1/2}$ with PMI, based on the subset of SpineMets patients with available height data. The thresholds used were:

$$\mathrm{PMV/VBV(L3)}^{1/2}_{\mathrm{men}} < 27.97, \quad \mathrm{PMV/VBV(L3)}^{1/2}_{\mathrm{women}} < 16.38 \tag{5.3}$$

These volumetric thresholds correspond to the conventional PMI cut-offs reported by Albano et al. [44], where sarcopenia is defined as:

$$\boldsymbol{PMI}_{\mathrm{men}} < 6.36\,\mathrm{cm}^2/\mathrm{m}^2, \quad \boldsymbol{PMI}_{\mathrm{women}} < 3.92\,\mathrm{cm}^2/\mathrm{m}^2 \tag{5.4}$$

5.6 Results

5.6.1 Psoas Cross-Sectional Area vs. Psoas Muscle Volume

We found a strong linear relationship between psoas cross-sectional area (CSA) and psoas muscle volume (PMV) in both the Patients with Spinal metastases and in the Young advanced prostate cancer patients. (Figure 5.6):

- SpineMets cohort (143 patients, 160 CT volumes): c = 0.1208, R^2 = 0.897, MAE = 1.28
- Young advanced prostate cancer cohort (69 patients, 490 CT volumes): c = 0.1184, R^2 = 0.906, MAE = 1.19

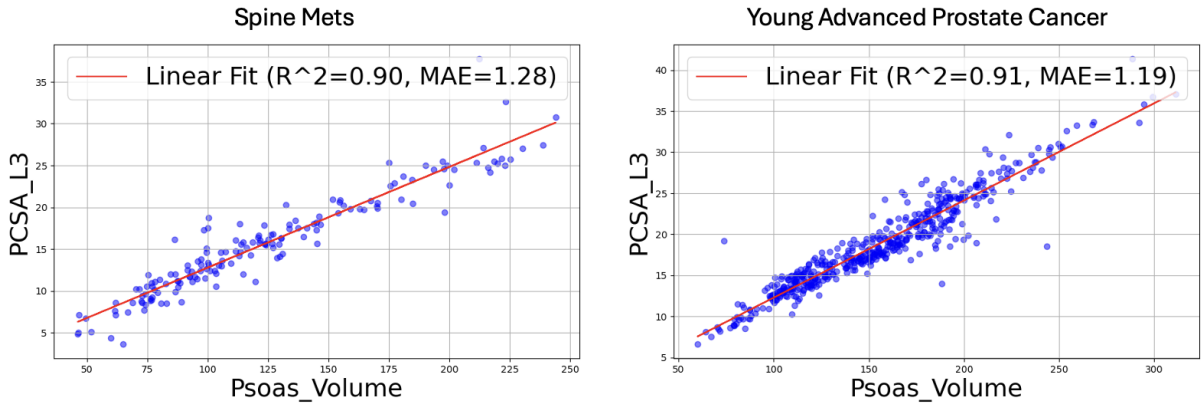


Figure 5.6: Linear regression of psoas cross-sectional area at L3 versus total psoas muscle volume across 143 SpineMets patients (left) and 69 young advanced prostate cancer patients (right).

5.6.2 Osteoporosis Prevalence

Figure 5.7 presents the distribution of osteoporotic and non-osteoporotic patients across different age groups. The data indicate a higher prevalence of osteoporosis in older patients, particularly those aged 60 and above, while younger age groups exhibit a lower incidence.

Figure 5.8 shows the distribution of the osteoporosis score, resulting in a total of 83 osteoporotic patients, 37/65 female and 46/78 male.

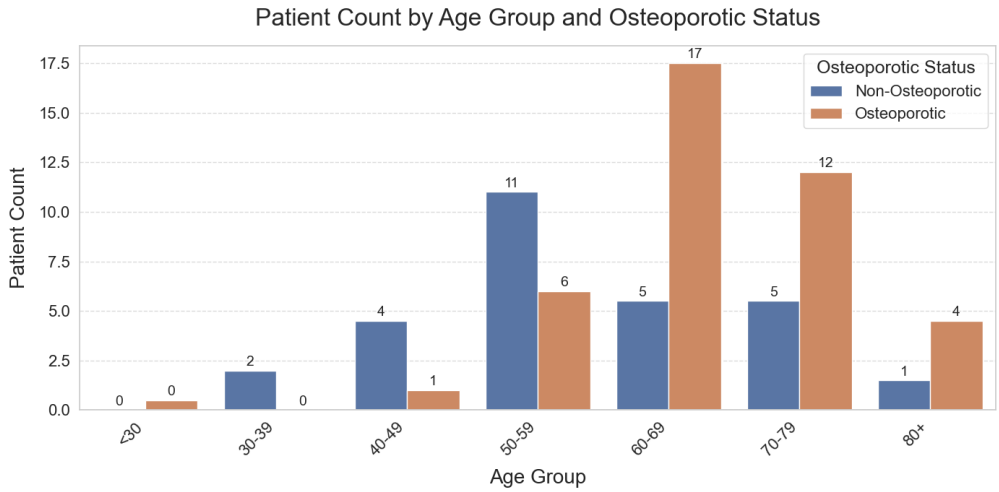


Figure 5.7: Patient Count by Age Group and Osteoporotic Status

5.6.3 Sarcopenia Prevalence

Based on the stablished thresholds, sarcopenia was observed in 64 out of 143 patients (45%) overall. When stratified by sex, 44 of 78 male patients (56%) and 20 of 65 female patients (31%) were classified as sarcopenic.

Figure 5.9 illustrates the distribution of PMV/VBV(L3)$^{1/2}$ values by sex, with dashed vertical lines indicating the sex-specific sarcopenia thresholds.

5.6.4 Osteosarcopenia Distribution

Using the previous classifications for osteoporosis and sarcopenia, patients were further classified into osteosarcopenia groups. Osteosarcopenia was the most prevalent condition in the cohort, observed in 44 out of 143 patients (31%). Additionally, 20 patients (14%) were classified as sarcopenic only, and 39 patients (27%) as osteoporotic only. The remaining 40 patients (28%) exhibited neither condition.

Table 5.5 presents the detailed distribution stratified by sex. Among female patients, 22

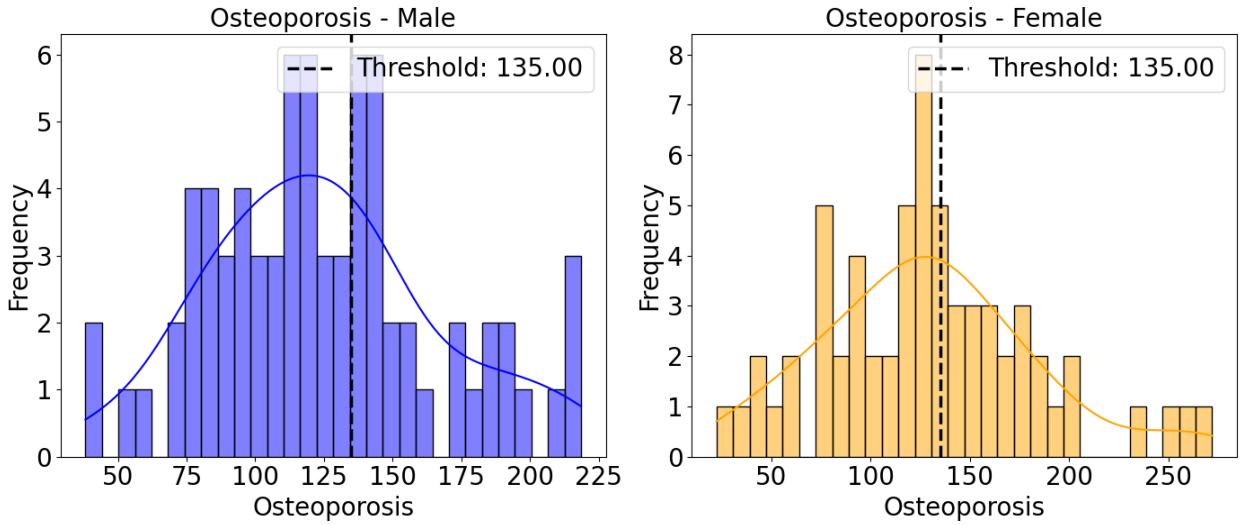


Figure 5.8: Osteoporosis distribution

were classified as having no condition, 23 as osteoporotic only, 14 as osteosarcopenic, and 6 as sarcopenic only. Among male patients, 18 had no condition, 16 were osteoporotic only, 30 were osteosarcopenic, and 14 were sarcopenic only.

These findings highlight the high prevalence of osteosarcopenia among spine SBRT patients and emphasize the importance of integrated muscle and bone health assessments in this population.

Table 5.5: Distribution of Osteosarcopenia Classification by Sex

Sex	Osteosarcopenia Classification	Patient Count
F	No condition	22
F	Osteoporotic Only	23
F	Osteosarcopenic	14
F	Sarcopenic Only	6
M	No condition	18
M	Osteoporotic Only	16
M	Osteosarcopenic	30
M	Sarcopenic Only	14

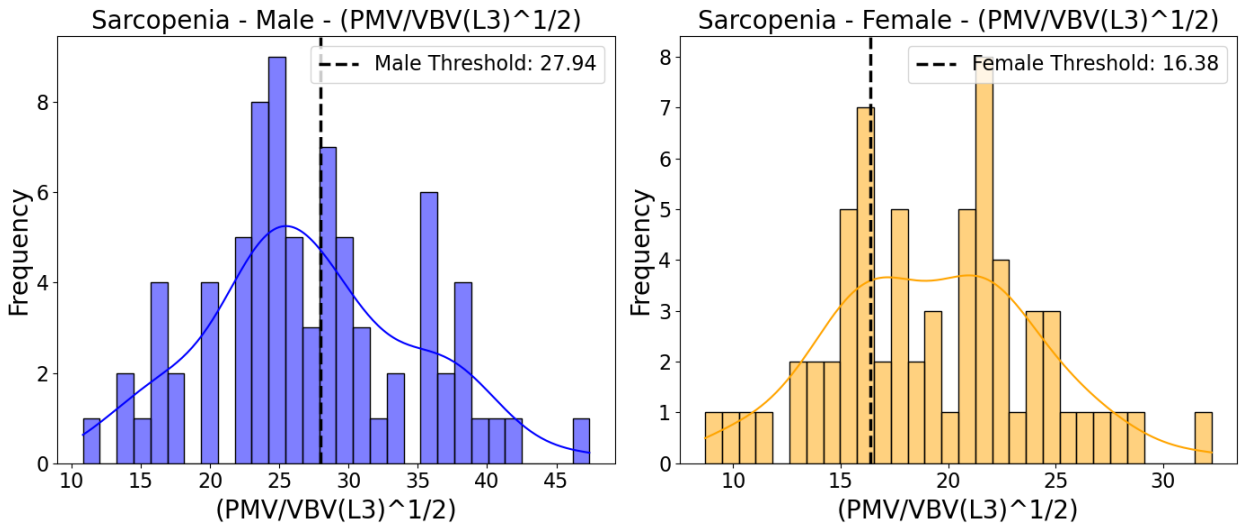


Figure 5.9: Distribution of PMV/VBV(L3)$^{1/2}$ values by sex. Dashed lines represent the sex-specific thresholds for sarcopenia classification.

5.6.5 Fracture Risk Analysis

This section evaluates the association between imaging biomarkers and vertebral fracture risk using both categorical and continuous variables. Analyses were performed separately for the full cohort and for sex-specific subgroups.

A total of 295 vertebrae from 143 patients were analyzed. Among these, 38 vertebrae (13%) experienced fractures. Thirty-two patients (22%) had at least one fractured vertebra. The fracture rate was higher among males (15%) than females (11%).

Chi-Squared tests were conducted to assess whether categorical classifications—including sarcopenia (based on PMI and volumetric indices), osteoporosis, and osteosarcopenia—were associated with vertebral fracture risk. Across the full cohort, none of the categorical variables demonstrated a statistically significant association with fracture occurrence ($p > 0.05$). Similarly, sex-specific subgroup analyses did not identify any significant associations in either males or females ($p > 0.05$ for all tests).

For continuous imaging biomarkers, Mann–Whitney U tests and Welch's t-tests were used depending on normality assumptions. Figure 5.10 visualizes the overall distribution of four continuous imaging biomarkers across the cohort: Psoas Density Mean, Psoas Volume, the normalized psoas volume (PMV/VBV(L3)$^{1/2}$), and Osteoporosis.

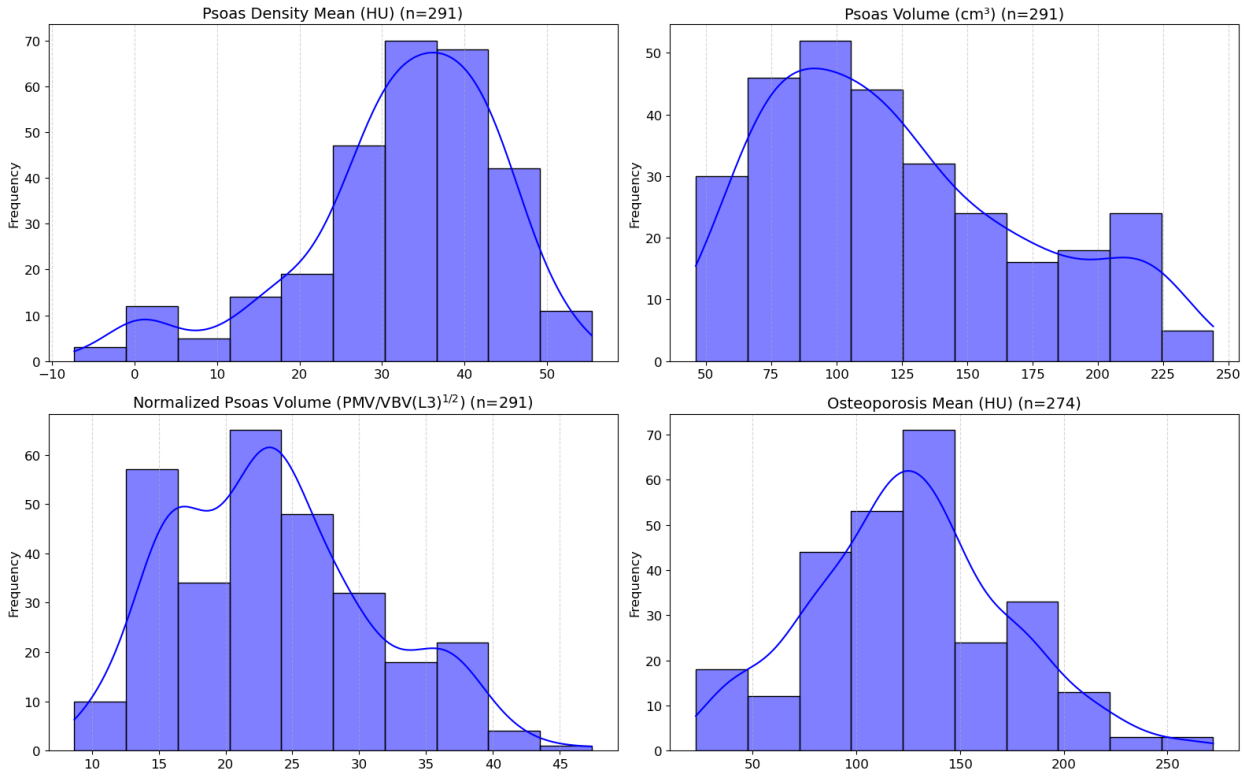


Figure 5.10: Distributions of selected imaging biomarkers across the full cohort.

Figure 5.11 shows the distribution of each biomarker by fracture status using boxplots. Each plot includes the statistical test used and the corresponding p-value.

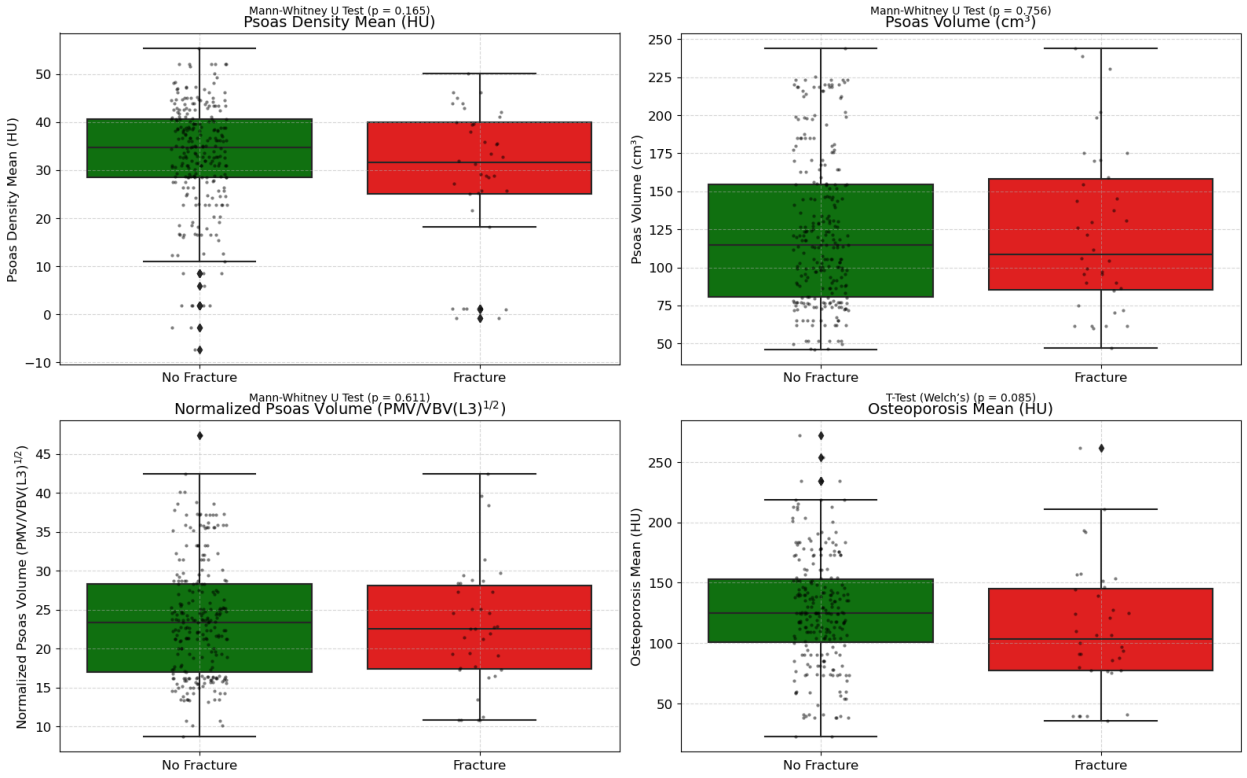


Figure 5.11: Boxplots comparing biomarker values between vertebrae with and without fracture.

Across all comparisons, no significant differences in biomarker values were found between fractured and non-fractured vertebrae in the cohort ($p > 0.05$ for all tests). Psoas Density had a median of 34.66 HU in the non-fracture group versus 31.68 HU in the fracture group ($p = 0.165$; Cohen's $d = -0.33$; power $= 0.47$), suggesting a possible difference with insufficient power to confirm it. Osteoporosis also trended lower in the fracture group (mean difference = –16.34 HU; $p = 0.085$; Cohen's $d = -0.33$), with post hoc analysis suggesting a required minimum sample size of 84 per group to detect this effect.

Other biomarkers such as Psoas Volume and PMV/VBV(L3)$^{1/2}$ showed smaller median differences (–5.9 cm^3 and –0.79 respectively), with very low power estimates (0.05 and 0.12), indicating that a substantially larger sample size would be required to detect moderate

effects.

In the male subgroup, the ratio PMV/VBV(L3) reached statistical significance (p = 0.0155) with a moderate effect size ($d = -0.41$). However, the power remained low (0.44), and a minimum sample size of 55 individuals per group would be required to confirm this finding reliably. No other biomarkers showed statistically significant differences in males or females. Among females, none of the continuous variables were significantly associated with fracture risk. The closest result was for osteoporosis, which had a non-significant p-value ($p = 0.351$) and very low power (0.06), limiting its interpretability.

We also explored fracture risk while accounting for potential interactions between biomarkers by fitting logistic regression models that incorporated both continuous and categorical imaging-derived features.

Osteoporosis, treated as a continuous variable, was significantly associated with fracture risk in models that included binary sarcopenia classification based on the volumetric PMV/VBV(L3)$^{1/2}$ ratio (Sarcopenic_PMV/VBV(L3)$^{1/2}$) and sex (p = 0.02). These findings suggest that trabecular bone density, is an independent predictor of fracture risk beyond the effects of binary muscle loss classification or sex.

Lower psoas muscle density was significantly associated with increased fracture risk in ($p = 0.022$; OR = 0.968, 95% CI: 0.943–0.996) in a model that adjusted for both Osteoporosis classification and Sarcopenic_PMV/VBV(L3)$^{1/2}$. In this model, neither osteoporosis (p = 0.244) nor the sarcopenia classification (p = 0.157) reached statistical significance. This result suggests that lower psoas muscle density is independently associated with greater fracture risk, beyond what is captured by binary osteoporosis or sarcopenia labels.

5.6.6 Survival Analysis

To evaluate the relationship between osteosarcopenia-related imaging biomarkers and patient survival, a survival analysis was conducted using the Kaplan-Meier method and log-rank tests. The Kaplan-Meier estimator is a non-parametric statistic widely used to estimate the probability of survival over time, accounting for both censored and uncensored data. This approach enables the visualization of survival probabilities across time intervals, even in the presence of incomplete follow-up data.

Censoring was applied to individuals for whom death had not been observed by the end of the follow-up period. The time-to-event variable was defined as the duration between the date of SBRT initiation and either the recorded date of death or the date of last clinical follow-up, whichever occurred first. Patients were considered deceased if a verified date of death was available and alive (i.e., censored) otherwise.

To compare survival distributions between groups (e.g., osteosarcopenic vs. non-osteosarcopenic), the log-rank test was used. This statistical test evaluates whether there is a significant difference in the survival curves of two or more groups, under the null hypothesis that the survival functions are identical across the groups. For analyses involving multiple risk groups (e.g., sarcopenic only, osteoporotic only, osteosarcopenic, and no condition), a multivariate log-rank test was employed to test for overall differences among the curves.

The median follow-up duration in our cohort of patients with spinal metastases treated with SBRT was 22.1 months (interquartile range [IQR]: 8.8–37.3 months). At the time of analysis, 48% of patients had died, while 52% were censored at their last follow-up.

Kaplan-Meier survival curves revealed that patients classified as osteosarcopenic exhibited slightly lower survival probabilities over time compared to those without osteosarcopenia.

The log-rank test indicated a statistically significant difference between groups ($p = 0.023$; see Figure 5.12), suggesting that osteosarcopenia may be associated with decreased overall survival.

When stratifying patients into four groups (no condition, sarcopenic only, osteoporotic only, and osteosarcopenic), the multivariate log-rank test also indicated a statistically significant difference in survival ($p = 0.0017$; Figure 5.13). These results support the relevance of osteosarcopenia classification in predicting patient survival.

Sarcopenia classification based on imaging-derived volumetric indices was also significantly associated with survival. When patients were stratified based on sarcopenia defined by PMV/VBV(L3)$^{1/2}$, the resulting survival curve showed statistically significant differences (log-rank $p = 0.0002$; see Figure 5.14). These findings suggest that sarcopenia, even when assessed using automated imaging biomarkers alone, may be predictive of poorer survival outcomes in patients with spinal metastases.

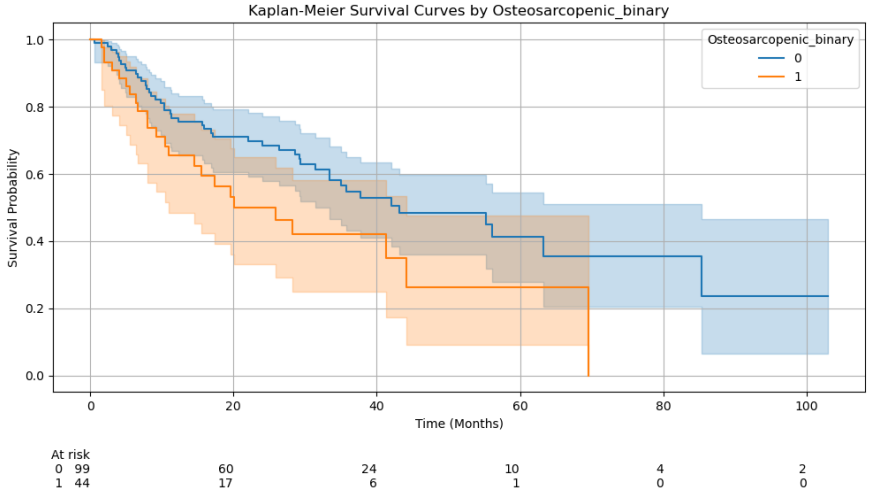


Figure 5.12: Kaplan-Meier survival curves comparing osteosarcopenic and non-osteosarcopenic patients. A statistically significant difference in survival was found (log-rank $p = 0.023$).

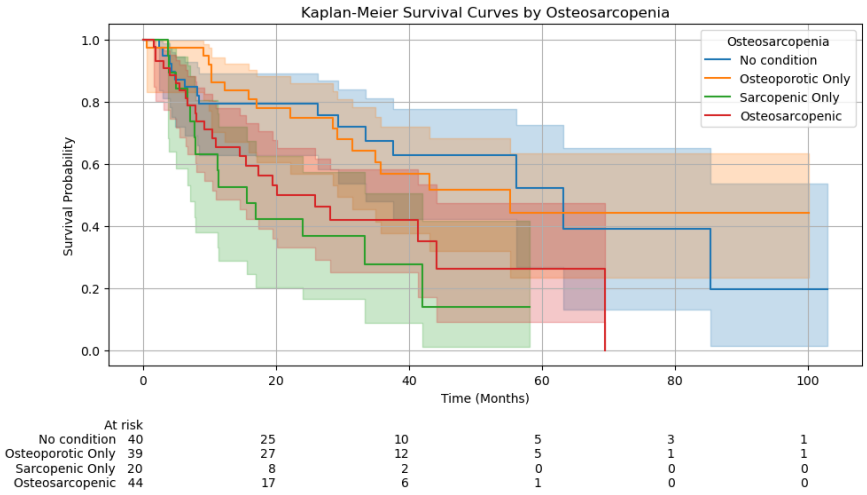


Figure 5.13: Kaplan-Meier survival curves stratified by: no condition, sarcopenic only, osteoporotic only, and osteosarcopenic. A statistically significant difference was detected (multivariate log-rank $p = 0.0017$).

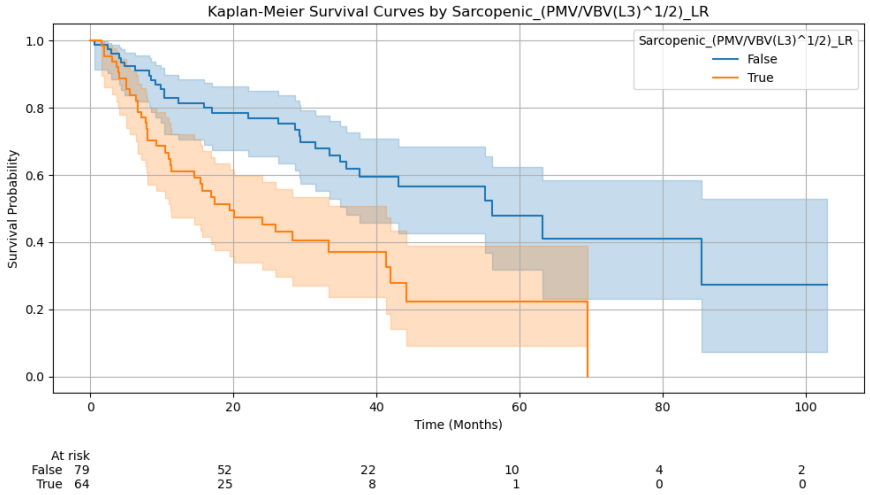


Figure 5.14: Kaplan-Meier survival curves comparing sarcopenic and non-sarcopenic patients. A statistically significant difference in survival was found (log-rank $p = 0.0002$).

5.7 Discussion

The strong relationship between the 3D volumetric and CSA based measures of sarcopenia support the use of volumetric metrics with the potential advantages of 3D based measures

for monitoring change between scans. Given these strong correlations and the need for height-independent metrics, volumetric indices such as PMV and vPMI provide a scalable and practical alternative for sarcopenia assessment, particularly in longitudinal studies or when height is unavailable.

Osteosarcopenia was the most prevalent condition in this cohort, affecting 31% of patients, consistent with prior studies highlighting the high burden of musculoskeletal deterioration in oncology populations. Given its established association with frailty and adverse outcomes, these findings reinforce the need for routine musculoskeletal assessments in patients undergoing SBRT.

Across the full cohort, categorical classifications of sarcopenia, osteoporosis, and osteosarcopenia were not significantly associated with fracture risk. Sex-stratified analyses also failed to show significant associations in either males or females, suggesting that categorical imaging-based indicators alone may be insufficiently sensitive for predicting vertebral fracture risk in this population.

Continuous imaging biomarkers offered additional insights. Although no comparisons reached statistical significance in the full cohort, psoas muscle density and bone density showed trends toward association with fracture risk (e.g., $p = 0.1364$ and $p = 0.0923$, respectively), but were underpowered to confirm such associations. In the male subgroup, the PMV/VBV(L3) ratio was significantly lower in patients with vertebral compression fractures ($p = 0.0155$), suggesting a possible relationship between reduced muscle volume relative to bone size and fracture risk. However, the low statistical power (0.44) and small fracture sample size limit interpretability.

Despite not reaching statistical significance in univariate comparisons, both psoas muscle density and osteoporosis were significantly associated with fracture risk in logistic regres-

sion models. Psoas density was a significant predictor of fracture risk both independently ($p = 0.035$) and after adjusting for osteoporosis and sarcopenia classification ($p = 0.022$). Osteoporosis also remained significant when adjusting for sarcopenia and sex ($p = 0.02$).

Regarding survival analysis, both sarcopenia and osteosarcopenia classifications were significantly associated with survival, supporting their relevance as prognostic indicators. These findings reinforce the potential of fully automated CT-based imaging biomarkers for integration into clinical risk stratification workflows in spinal metastasis patients undergoing SBRT.

Chapter 6

MRI-Based Segmentation of the Psoas Muscle and Vertebral Bodies

Musculoskeletal biomarkers derived from MRI are clinically valuable due to MRI's superior soft tissue contrast and absence of ionizing radiation. However, extracting these biomarkers requires accurate segmentation of muscle and bone structures, which is more challenging than in CT due to variability in image quality, contrast, and field of view. This chapter describes the development and evaluation of deep learning models for segmenting the psoas muscle and vertebral bodies in MRI using CT-derived labels from aligned datasets.

Segmenting the psoas muscle and vertebral bodies in MRI volumes presents unique challenges. Despite MRI's superior soft tissue contrast, it exhibits high variability across scanners and protocols, lacks standardized intensity scaling, and often has a limited field of view. These factors make it difficult to obtain training images with consistent anatomical coverage. While TotalSegmentator performs well on CT scans, it yields poor results on our MRI data. A key limitation is the scarcity of publicly available MRI datasets with consistent regions

and ground truth segmentations, which limits model training and evaluation.

To address these challenges, we leveraged a previously manually aligned dataset of CT and MRI scans to propagate labels from CT to MRI (see Section 6.2 and Section 6.1 for full details).

We extended the CT segmentation pipeline to MRI using nnU-Net [57], a self-configuring deep learning framework for biomedical image segmentation. We used the basic public implementation without manual modifications. nnU-Net automatically selects an appropriate model configuration (2D, 3D full resolution, or 3D cascade) and generates a "plan" tailored to the dataset. This plan defines spacing, patch size, architecture parameters, and batch size.

6.1 nnU-Net Training Pipeline

The following is a brief summary of the nnU-Net training process, as described by Isensee *et al.* [57], which we applied in this work:

1. **Preprocessing:** Images are resampled to near-isotropic resolution and intensity-normalized. Training patches are extracted with oversampling of foreground regions.

2. **Architecture Configuration:** Based on the dataset plan, nnU-Net adapts the architecture depth, number of feature maps, and kernel sizes to maximize GPU usage and context coverage.

3. **Data Augmentation:** Real-time augmentations such as rotations, scaling, flips, elastic deformations, and intensity shifts are applied.

4. **Model Optimization:** The model is trained with stochastic gradient descent and a combined Dice and cross-entropy loss. A polynomial learning rate decay is applied.

5. **Validation:** A 5-fold cross-validation is used. The best model per fold is saved based on Dice score.

6. **Ensembling:** Cross-validation predictions are averaged to improve final results.

7. **Postprocessing:** Small disconnected components are removed using connected component analysis.

Before training, nnU-Net requires a specific folder structure for dataset organization. Each dataset must include a folder for training images and a corresponding folder for segmentations. All files must be in NIfTI format and follow a strict naming convention (e.g., `image_0000.nii.gz` for images and `image.nii.gz` for corresponding labels). Additionally, a `dataset.json` file must be provided. This file specifies essential metadata such as the dataset name, label definitions, modality/channel mappings and the total number of images. This setup allows nnU-Net to automatically parse the dataset and generate an appropriate configuration plan for training.

Details for each trained model are presented in sections 6.6.2 and 6.6.3.

6.2 Vertebral Coverage Analysis

As part of the data preprocessing step, we examined the vertebral coverage in each scan. This was necessary because CT and MRI volumes often include different regions of the spine. Vertebral bodies were segmented for all CT images using VertDetect [66].

6.2.1 Implications for Model Training and future Longitudinal Analysis

The vertebral coverage analysis revealed a substantial disparity between CT and MRI scans. While CT volumes consistently encompassed the full lumbar spine (L1–L5), MRI scans tended to have more limited anatomical coverage. This discrepancy has implications for both training and downstream applications: in longitudinal studies or CT-MRI biomarker comparisons, it may be beneficial to define the psoas muscle region adaptively, based on the specific vertebral levels captured in each case.

These coverage limitations impacted which scan pairs could be used for training. Our inclusion criteria are described in detail in Sections 6.6.2 and 6.6.3.

6.3 nnU-Net Data Preparation Pipeline

We developed a data preparation pipeline that automates the steps of selecting appropriate scan pairs, resampling, labeling, and saving the volumes. This pipeline is modular and reusable, facilitating training for future models.

To prepare the data for nnU-Net training, the first step was to segment the psoas muscles and vertebral bodies in CT volumes. Because high-quality ground truth segmentations for MRI are scarce, particularly for musculoskeletal structures, we leveraged CT-MRI aligned pairs to propagate CT-based labels onto MRI volumes. This enabled weak supervision for training, eliminating the need for manual MRI annotations.

For psoas segmentation, we used TotalSegmentator to segment the CT volumes. The previously segmented vertebral bodies, obtained using VertDetect for the coverage analysis,

were also included as part of a systematic pipeline to match CT and MRI pairs that covered the region from L2 to the midpoint of L5. In addition to guiding region selection, these vertebral segmentations served as ground truth labels for training the nnU-Net models for vertebral body segmentation in MRI.

Figure 6.1 summarizes the process of transferring ground truth from CT segmentations to pre-registered T1 and T2 MRI volumes. The pipeline shown is for psoas segmentation; a similar process was used for vertebral bodies.

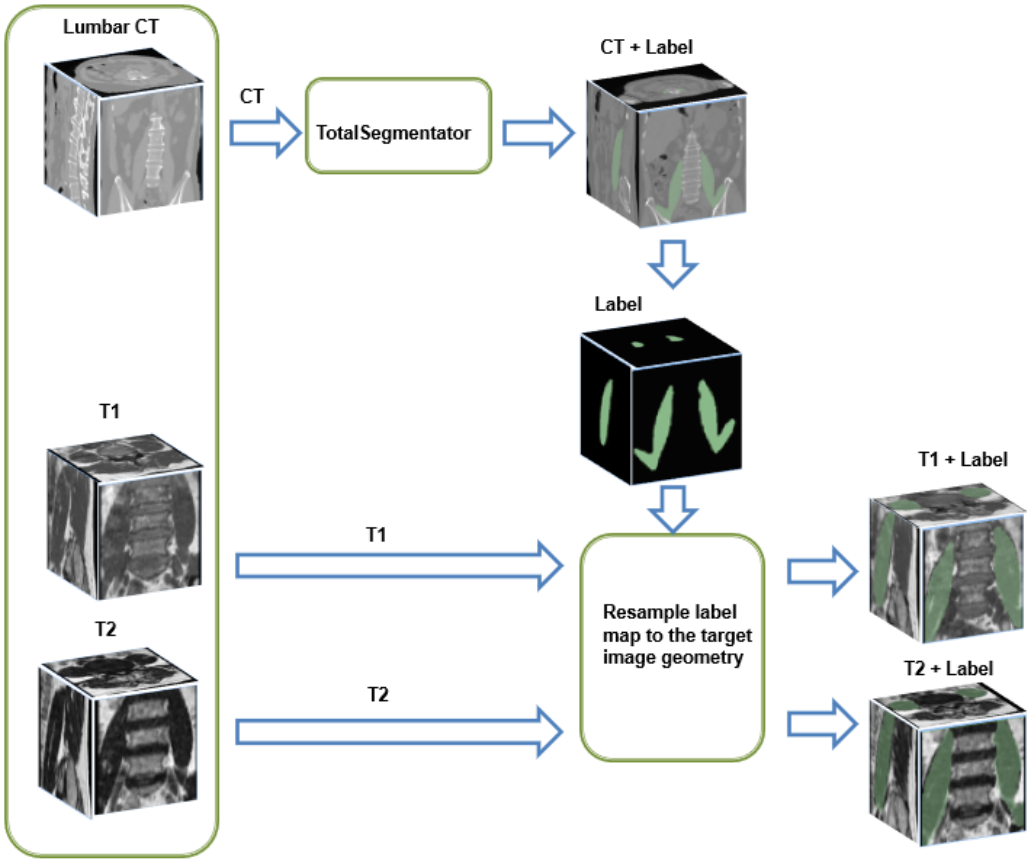


Figure 6.1: Data preparation pipeline for psoas segmentation, showing label propagation from CT to pre-registered T1 and T2 MRI volumes.

A single-channel model framework was implemented, meaning CT, T1, and T2 scans

were each used independently as inputs for training, despite originating from fused datasets. The fused CT-MRI dataset was used solely to propagate ground truth segmentations onto MRI volumes, as shown in Figure 6.1.

6.4 Psoas Segmentation Training Strategy

Based on the analysis of vertebral coverage on our CT and corresponding MRI, see section 6.2, and the vertebral body segmentations, we defined inclusion criteria for selecting training pairs. We retained all CT scans that included the region L2–L5 and continued to filter and retain CT-MRI scan pairs where the MRI included the region minimum of L2 vertebrae to midpoint of L5, ensuring that the relevant anatomy for psoas segmentation was visible.

Only pairs with exactly one T1-weighted and one T2-weighted MRI volume were included for model training. If multiple valid T1 or T2 scans were available, we processed them separately and saved them for potential use in validation. A total of 31 CT-MRI pairs (93 CT, T1 and T2) contained the required region, and 7 scans including 2 CT, 3 T1 and 2 T2 of images that included the region but were not paired (could be because more than one T1 or T2 existed for a CT) were used as a testing set.

The first model was trained with both left and right psoas labeled separately. A total of 31 CT-MRI pairs met all criteria and were used for training. An additional set of 7 scans that matched the anatomical region but were not perfectly paired (due to duplicate modalities) were used as a testing set.

6.5 Vertebral Body Segmentation Training Strategy

We trained three nnU-Net models using different combinations of CT and MRI training data, all using aligned volumes and a consistent label convention (C1 to L6 and T13, labels 1–26).

We evaluated the following three training configurations to assess the impact of anatomical coverage and data inclusion criteria::

- **Model 1 – CT (L2–L5) + MRI (L2 to mid-L5)**: Aligned MRI scans containing at least L2 to mid-L5 were selected. This matched the psoas training region and ensured consistent anatomical context.
- **Model 2 – CT (L2–L5) + any aligned MRI**: To increase the number of training samples, this configuration included all MRI scans aligned with CTs, regardless of MRI vertebral coverage.
- **Model 3 – All CTs + all MRIs**: In this setup, all available CT and MRI volumes were used, regardless of anatomical region. This maximized dataset size but introduced variability in vertebral coverage.

Each model was trained using 5-fold cross-validation.

6.6 Results

6.6.1 Vertebral Coverage Findings

Figure 6.2 shows the frequency of L1–L5 vertebrae across the CT and MRI scans in the fused dataset.

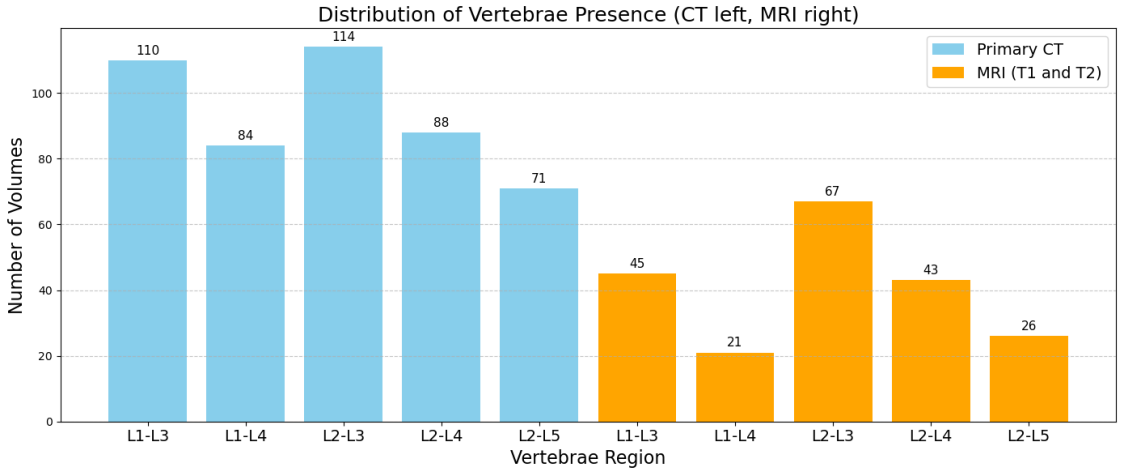


Figure 6.2: Presence of vertebral regions in CT (sky blue) and MRI (orange) scans across the fused dataset.

The CT scans covered a broader anatomical range, frequently including all five lumbar vertebrae. In contrast, MRI scans had limited field of view. The L2–L3 region was the most consistently present in both modalities, appearing in 114 CT and 67 MRI scans, respectively.

6.6.2 Psoas Segmentation Results

The nnU-Net model trained for separate segmentation of the left and right psoas muscles was evaluated using 5-fold cross-validation. Dice scores ranged from 0.84 to 0.90, and Jaccard indices from 0.75 to 0.83. The average Dice score across all folds was 0.88. A detailed breakdown is shown in Table 6.2.

Table 6.1: Per-class segmentation performance of nnU-Net model for left and right psoas muscles across 5-fold cross-validation. Metrics are reported separately for the left and right segments.

Fold	**Class**	**Dice Score**	**FNR**	**FPR**
Fold 0	Left	0.89	0.16	4.1×10^{-4}
Fold 0	Right	0.91	0.11	3.2×10^{-4}
Fold 1	Left	0.88	0.13	4.3×10^{-4}
Fold 1	Right	0.88	0.13	4.3×10^{-4}
Fold 2	Left	0.88	0.13	6.4×10^{-4}
Fold 2	Right	0.89	0.12	5.8×10^{-4}
Fold 3	Left	0.84	0.12	8.9×10^{-4}
Fold 3	Right	0.85	0.16	6.9×10^{-4}
Fold 4	Left	0.90	0.11	3.2×10^{-4}
Fold 4	Right	0.91	0.09	3.1×10^{-4}

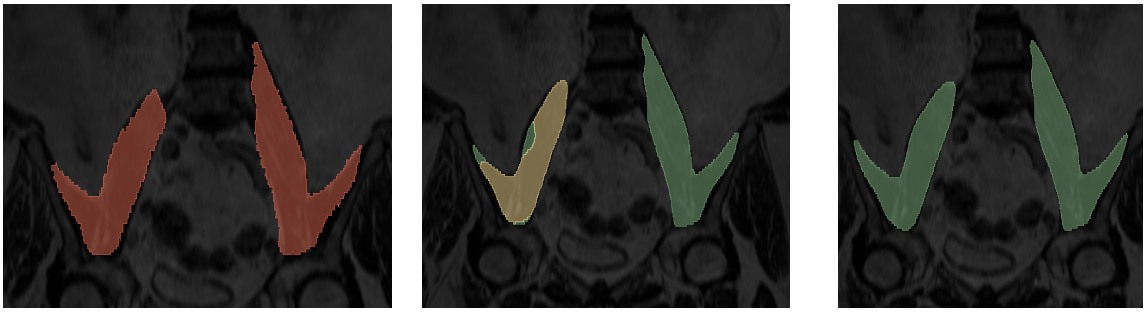
Table 6.2: Average performance of nnU-Net model for left and right psoas segmentation using 5-fold cross-validation.

Fold	**Dice Score**	**Jaccard Index**	**FNR**	**FPR**
Fold 0	0.90	0.83	0.17	3.6×10^{-4}
Fold 1	0.88	0.80	0.14	4.3×10^{-4}
Fold 2	0.88	0.80	0.13	5.1×10^{-4}
Fold 3	0.84	0.75	0.16	6.2×10^{-4}
Fold 4	0.90	0.83	0.11	3.3×10^{-4}
Mean	**0.88**	**0.80**	**0.14**	4.5×10^{-4}

While the results showed high overlap with the ground truth segmentations, the model

occasionally mislabeled left and right psoas muscles, more frequently in CT volumes, though some mislabeling was also observed in MRI. To mitigate this, we trained a version of the model where both psoas muscles were combined into a single segment. This approach improved the mean Dice score from 0.88 to 0.90 and reduced mislabeling, as detailed in Table 6.3.

Figure 6.3 illustrates a visual comparison between the separate and combined psoas segmentation models. While both models achieved similar quantitative metrics, the combined model yielded more anatomically consistent segmentations by avoiding left/right mislabeling. This difference is evident in the MRI slice, where the separate model introduces lateral asymmetry not present in the ground truth or the combined prediction.



(a) Ground Truth (b) Separate Prediction (c) Combined Prediction

Figure 6.3: Psoas Segmentation – Visual Comparison of Separate vs Combined Models. A coronal MRI slice is shown with: (a) ground truth segmentation, (b) prediction from the separate left/right psoas model, and (c) prediction from the combined model. The combined prediction exhibits improved anatomical consistency by reducing left/right mislabeling.

Table 6.3 shows the Dice and IoU metrics for each individual fold. The final row reports the cross-validation performance computed by nnU-Net by aggregating predictions across all test splits, yielding a Dice score of 0.90 and IoU of 0.82.

Table 6.3: Cross-validation performance of nnU-Net model for combined psoas segmentation.

Fold	**Dice Score**	**Jaccard Index**	**FNR**	**FPR**
Fold 0	0.91	0.83	0.12	7.1×10^{-4}
Fold 1	0.90	0.83	0.10	7.6×10^{-4}
Fold 2	0.90	0.83	0.10	9.8×10^{-4}
Fold 3	0.86	0.77	0.13	1.4×10^{-3}
Fold 4	0.93	0.87	0.07	5.2×10^{-4}
Crossval (0-4)	0.90	0.82	0.11	8.3×10^{-4}

Table 6.4 presents the performance of the combined psoas model on the 7-scan test set using CT-derived ground truth. CT volumes yielded the highest accuracy, while MRI volumes showed slightly lower performance. Most MRI errors were localized near the inferior lumbar boundary, likely due to the limited field of view in MRI and the reduced availability of ground truth labels beyond L5.

Table 6.4: Performance of the combined psoas segmentation model on 7 unseen test volumes using CT-derived ground truth. Volumes are grouped by subject and modality.

Subject	**Modality and Series**	**Dice**	**Jaccard**	**Vol. Sim.**	**FNR**	**FPR**
Patient 1	CT	0.95	0.90	-0.02	0.06	1.16×10^{-4}
	MRI – T1	0.90	0.82	-0.06	0.12	6.05×10^{-3}
	MRI – T2	0.91	0.83	-0.06	0.12	5.36×10^{-3}
Patient 2	CT	0.93	0.88	-0.02	0.07	2.19×10^{-4}
	MRI – T1 (pre-contrast)	0.87	0.76	-0.07	0.16	1.70×10^{-3}
	MRI – T1 (post-contrast)	0.87	0.77	-0.10	0.17	1.50×10^{-3}
	MRI – T2	0.86	0.76	-0.09	0.18	1.61×10^{-3}
Mean	—	**0.90**	**0.82**	**-0.06**	**0.13**	2.74×10^{-3}

Table 6.5 presents the performance of the combined psoas model evaluated against four manually segmented CT volumes, each preprocessed using the vertebra-guided cropping

module described in section 5.2. For comparison, Table 6.6 reports the performance of TotalSegmentator on the same cases. Both methods achieved Dice scores exceeding 0.85 across all cases, with the combined model slightly outperforming TotalSegmentator in Jaccard index and volume similarity.

Table 6.5: Performance of the combined psoas model against manual segmentations on 4 cropped CT volumes.

Patient	**Dice**	**Jaccard**	**Vol. Sim.**	**FNR**	**FPR**
Patient A	0.86	0.75	-0.27	0.24	1.01×10^{-5}
Patient B	0.92	0.85	-0.10	0.13	5.56×10^{-5}
Patient C	0.88	0.79	-0.20	0.19	1.55×10^{-5}
Patient D	0.91	0.83	-0.16	0.16	2.06×10^{-5}
Mean	**0.89**	**0.81**	**-0.17**	**0.18**	2.05×10^{-5}

Table 6.6: Performance of TotalSegmentator against manual segmentations on the same CT volumes.

Patient	**Dice**	**Jaccard**	**Vol. Sim.**	**FNR**	**FPR**
Patient A	0.88	0.79	-0.16	0.18	5.43×10^{-5}
Patient B	0.89	0.80	-0.16	0.18	5.54×10^{-5}
Patient C	0.86	0.75	-0.18	0.21	4.37×10^{-5}
Patient D	0.89	0.79	-0.14	0.17	6.31×10^{-5}
Mean	**0.88**	**0.78**	**-0.16**	**0.18**	5.17×10^{-5}

6.6.3 Vertebral Body Segmentation Results

Table 6.7 reports the cross-validation segmentation results for L1–L5 for each model.

Table 6.7: Cross-validation performance on L1–L5 for each vertebral body model.

Model	**Dice Score**	**Jaccard Index**	**FNR**	**FPR**
Model 1 (L2–mid-L5 MRI only)	0.81	0.74	0.15	1.2×10^{-4}
Model 2 (All MRI matched to CT L2–L5)	**0.82**	**0.75**	**0.14**	9.0×10^{-5}
Model 3 (All CT + all MRI, full spine)	0.73	0.67	0.13	6.0×10^{-5}

Best Fold per Model

The fold with the highest average Dice score across L1–L5 was identified for each model:

- **Model 1: Fold 2** – Dice Score: 0.86, Jaccard Index: 0.78
- **Model 2: Fold 2** – Dice Score: 0.86, Jaccard Index: 0.79
- **Model 3: Fold 2** – Dice Score: 0.76, Jaccard Index: 0.69

While Model 2 achieved the highest overall cross-validation performance, the improvement over Model 1 was marginal. Both models focused on the lumbar region and yielded similarly high accuracy, suggesting that the inclusion of additional MRI volumes in Model 2 may have provided a modest benefit. In contrast, Model 3, trained on a broader anatomical range, underperformed across all metrics.

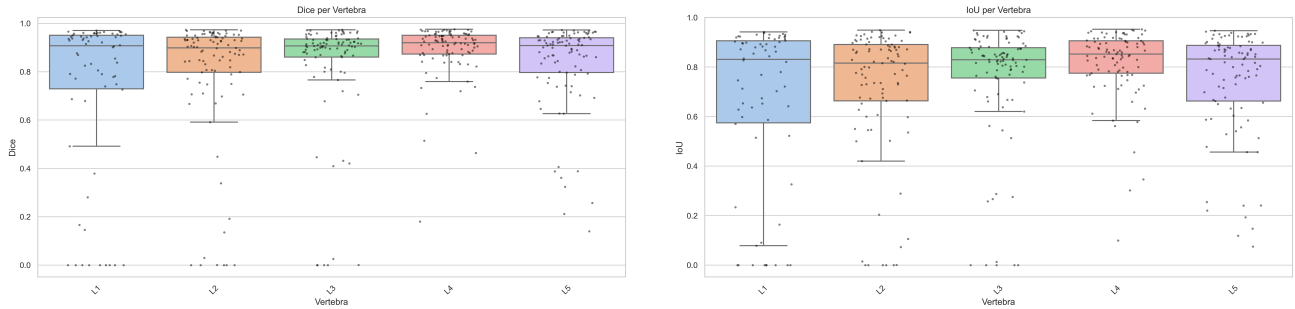
To further examine segmentation consistency, Table 6.8 presents Dice scores for each lumbar vertebra. L4 showed the best performance across all models, with Model 1 achieving the highest Dice score (0.89 ± 0.12). Model 3 produced lower and more variable results, consistent with its wider anatomical scope.

Table 6.8: Per-vertebra Dice score (mean ± std) for L1–L5 across all models.

Vertebra	**Model 1**	**Model 2**	**Model 3**
L1	0.73 ± 0.34	0.73 ± 0.34	0.67 ± 0.39
L2	0.79 ± 0.27	0.80 ± 0.26	0.73 ± 0.34
L3	0.82 ± 0.24	0.80 ± 0.27	0.73 ± 0.34
L4	0.89 ± 0.12	0.85 ± 0.20	0.77 ± 0.32
L5	0.83 ± 0.18	0.82 ± 0.22	0.75 ± 0.33

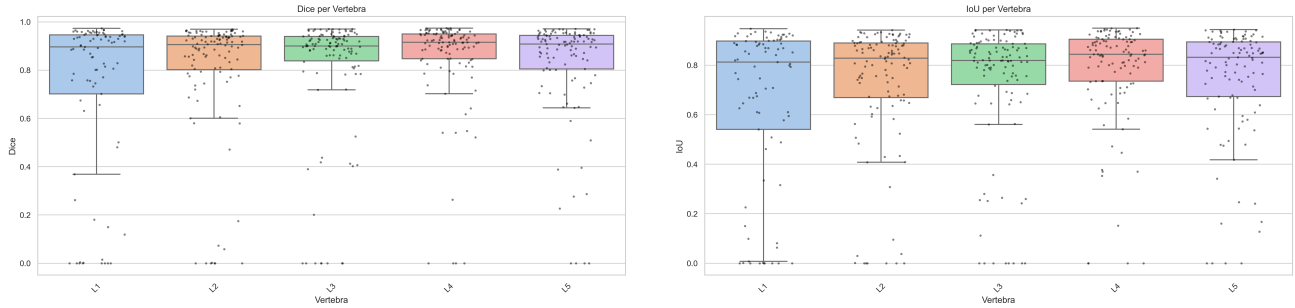
(a) Dice (M1)

(b) IoU (M1)



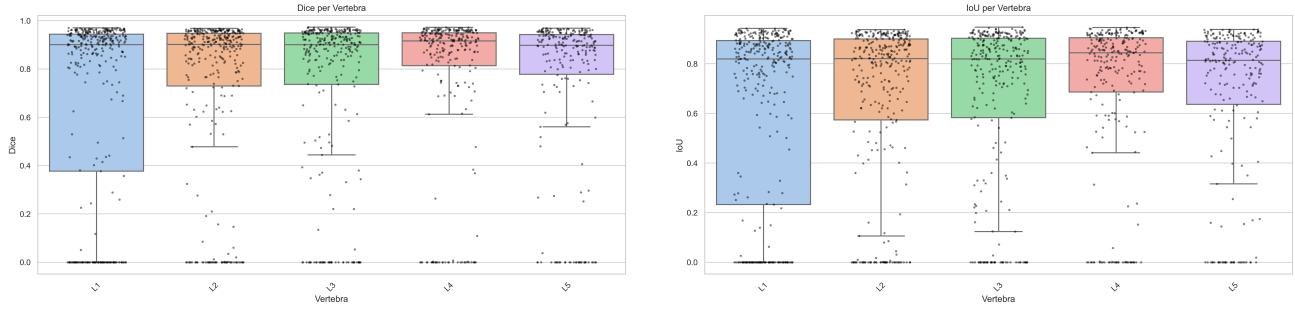
(c) Dice (M2)

(d) IoU (M2)



(e) Dice (M3)

(f) IoU (M3)



(g) Dice (M3) (C1–L6)

(h) IoU (M3) (C1–L6)

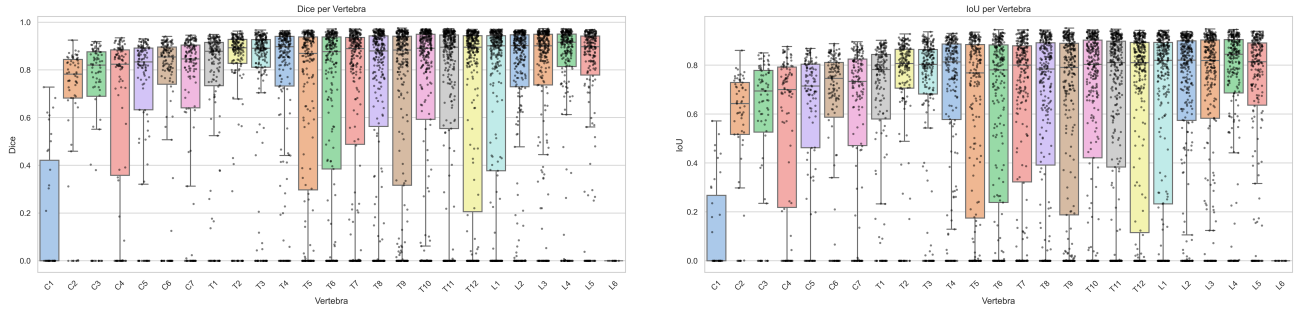
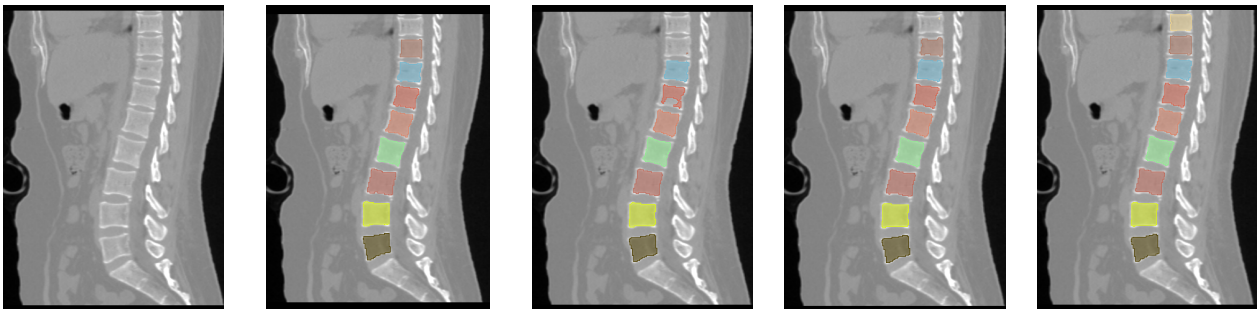


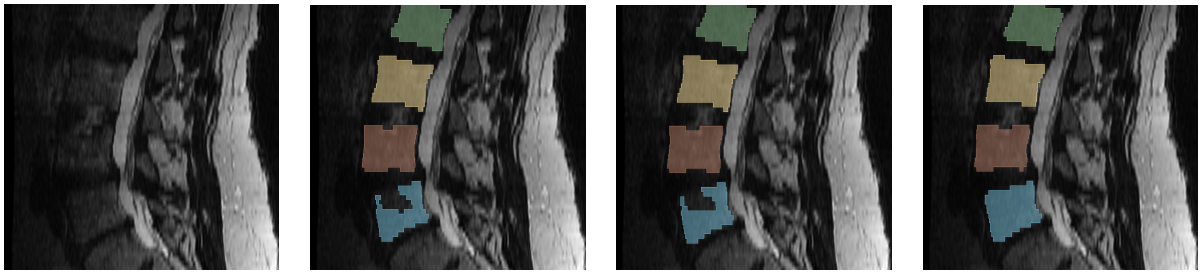
Figure 6.4: Boxplots showing segmentation performance (Dice and IoU) per vertebra for the three trained models. Each dot represents a single scan. Models 1 and 2 were trained and evaluated on the lumbar region (L1–L5), while Model 3 was trained on full-spine input (C1–L6).

Although Model 3 yielded slightly lower average cross-validation metrics compared to Models 1 and 2, its overall performance when running inference across all folds was promising when visually inspected. In particular, segmentation results on unseen images—both CT and MRI—appeared more anatomically consistent. While occasional mislabeling was observed in individual folds, especially for CT volumes, the full cross-validation setup mitigated these issues. As shown in Figure 6.5, Model 3 segmentations demonstrated greater coverage and alignment with vertebral structures, suggesting that its broader training set may have improved generalization. However, this qualitative observation warrants further validation using larger test sets and external datasets.



(a) CT Volume (b) Ground Truth (c) Model 1 (d) Model 2 (e) Model 3

Figure 6.5: Vertebral Body Segmentation – Visual comparison for a test CT volume not seen during training. The figure shows: (a) the original CT scan, (b) ground truth segmentation from VertDetect, and predictions from (c) Model 1, (d) Model 2, and (e) Model 3.



(a) MRI Volume (b) Model 1 (c) Model 2 (d) Model 3

Figure 6.6: Vertebral Body Segmentation on MRI – Visual comparison of model predictions on a previously unseen MRI scan. Panel (a) shows the input MRI, while (b)–(d) display the segmentation results from: (b) Model 1, (c) Model 2, and (d) Model 3. Ground truth is unavailable for this case.

6.7 Discussion

The psoas segmentation models achieved strong quantitative performance across folds, with Dice scores consistently above 0.88. Although the performance difference between the separate and combined psoas models was small, the combined model produced more anatomically consistent segmentations. This improvement is largely due to avoiding left–right mislabeling, which was occasionally observed in the separate-segment model. Visual inspection confirmed that the most accurate segmentation occurred within the L2–L5 region. Outside of this region, specifically the inferior edges of the MRI field of view, under-segmentation was more common. This supports the use of vertebral-based cropping as a necessary postprocessing step to ensure anatomical consistency for biomarker extraction.

For vertebral body segmentation, the nnU-Net models successfully delineated the lumbar vertebrae, with the highest cross-validation metrics observed when training was restricted to the lumbar region. However, when evaluating full inference across folds, the model trained on the entire spine (Model 3) appeared to produce more anatomically consistent segmentations

in unseen CT and MRI volumes. Despite this, some segmentations still exhibited incorrect vertebral label assignments. This mislabeling likely stems from the visual similarity between adjacent vertebrae and the absence of explicit spatial priors during training. Since nnU-Net performs semantic segmentation—predicting class labels at the voxel level without distinguishing between instances—it lacks the capability to enforce anatomical ordering across vertebral levels. As a result, adjacent vertebrae with similar appearance can be mislabeled, especially in datasets lacking full spine context. Furthermore, the reference segmentations used for training and evaluation were derived from VertDetect, which may itself introduce occasional label inaccuracies. Dice score variability can also arise when the predicted segmentation includes vertebrae partially visible at the image edges, while the corresponding ground truth does not—or vice versa—leading to artificially lowered overlap metrics. Prior studies have addressed these challenges by integrating postprocessing steps that incorporate anatomical knowledge, such as coordinate-based relabeling or spine-specific labeling frameworks using graph-based optimization or attention to vertebral centroids [72–74]. These approaches may be essential complements to improve label reliability in future models trained for vertebral labeling.

In summary, this chapter described the implementation of a fully automated MRI segmentation pipeline for the psoas muscle and vertebral bodies using CT-derived labels and nnU-Net. The psoas segmentation models demonstrated robust performance across T1 and T2 MRI sequences and are suitable for downstream biomarker analysis. The vertebral segmentation models were sensitive to anatomical coverage and label consistency but showed strong performance when focused on consistent lumbar regions. These results validate the feasibility of extending the CT-based biomarker pipeline to MRI, enabling future longitudinal or cross-modality analyses that avoid radiation exposure.

Chapter 7

Conclusions and Future Work

This thesis presented the development, validation, and preliminary clinical application of a fully automated pipeline for extracting musculoskeletal biomarkers from computed tomography (CT) scans in patients with spinal metastases undergoing stereotactic body radiotherapy (SBRT). The pipeline integrates deep learning–based segmentation with standardized anatomical cropping and post-processing, enabling consistent and reproducible quantification of clinically relevant bone and muscle features.

Specifically, it performs automated segmentation of vertebral bodies and psoas muscles, followed by the extraction of key biomarkers including trabecular bone density, vertebral body volume, psoas muscle volume and density, and cross-sectional area.

Osteoporosis was defined using a previously validated threshold based on vertebral bone density (in Hounsfield Units). For sarcopenia classification, volumetric indices were derived using CT-based muscle and bone measurements. These included both height-independent indices, by normalizing psoas muscle volume to an internal anatomical reference (vertebral body volume), and indices that incorporate patient height. Thresholds for each index

were statistically derived from a subset of patients with available height data and validated against an independent cohort of younger individuals. The inclusion of both index types ensures flexibility in application: height-based indices remain clinically valuable where height is available, while height-independent alternatives are particularly advantageous for retrospective or large-scale studies where such information may be missing.

The prevalence of osteosarcopenia was 31%, osteoporosis 58%, and sarcopenia 45% among the 143 patients with spinal metastases treated with SBRT, reflecting a considerable burden of musculoskeletal degeneration in this population.

While categorical definitions of sarcopenia, osteoporosis, and osteosarcopenia were not significantly associated with vertebral fracture incidence, some continuous imaging biomarkers showed statistically significant or near-significant associations that merit further evaluation. In univariate analyses, lower psoas muscle density and trabecular vertebral bone density showed trends toward association with fracture risk ($p = 0.165$ and $p = 0.085$, respectively), and among male patients, the volumetric sarcopenia index PMV/VBV(L3) was significantly lower in those who experienced fractures ($p = 0.016$). In multivariate logistic regression models, lower psoas muscle density was independently associated with increased fracture risk ($p = 0.022$), even after adjusting for osteoporosis and sarcopenia classification. Trabecular vertebral bone density also reached statistical significance in adjusted models ($p = 0.02$). These findings suggest that continuous imaging biomarkers may be more informative than binary classifications for identifying patients at elevated risk of skeletal complications following SBRT.

In terms of survival outcomes, sarcopenia defined using height-independent volumetric indices was significantly associated with reduced overall survival (log-rank $p = 0.0002$), and patients classified with osteosarcopenia also had significantly poorer survival (log-rank

$p = 0.023$). Together, these results highlight the prognostic relevance of imaging-derived biomarkers and support their integration into clinical workflows to enhance risk stratification and guide personalized treatment planning.

To extend the pipeline's applicability to MRI, this thesis developed a novel methodology to generate labeled MRI datasets by transferring segmentations from CT to aligned MRI volumes. These MRI volumes, labeled using CT-derived ground truth, enabled the training of nnU-Net segmentation models without requiring manually annotated MRI data. The psoas segmentation model achieved high performance, with mean Dice scores above 0.9 across validation folds, demonstrating the feasibility of accurate muscle segmentation directly from MRI. This results not only support radiation-free biomarker extraction but also enables future longitudinal studies using routinely acquired MRI scans.

In addition to model development, this thesis demonstrated how aligned multi-modality datasets can be leveraged to train MRI segmentation models in the absence of manual annotations. The use of CT-derived segmentations as weak supervision enabled a practical approach for building multi-channel models. Furthermore, the cropping module was enhanced with vertebra-guided logic, which is particularly valuable for ensuring consistency in longitudinal or multi-level analyses, especially in MRI, where the region encompassing the psoas muscle may be limited.

Although the nnU-Net models achieved high anatomical accuracy in vertebral body segmentation, occasional label inconsistencies were observed. However, when using full cross-validation and focusing on the lumbar region (L1–L5), vertebral labeling appeared more stable. It is also important to note that ground truth segmentations were derived from VertDetect, which may introduce labeling variability. Discrepancies between model predictions and ground truth could therefore reflect annotation noise rather than segmentation error.

Additionally, Dice variability may arise from partial vertebrae at the edges of the image, either present in the prediction or the ground truth, leading to artificially lower overlap scores.

Together, these contributions advance the field of automated musculoskeletal assessment and provide tools for noninvasive, reproducible, and clinically relevant characterization of osteosarcopenia in oncologic populations.

Summary of Contributions

This thesis contributes to the field of AI-based musculoskeletal imaging and cancer care through the following:

- **Automation of a CT-Based Biomarker Pipeline:** Improved and validated a 3D Slicer-based pipeline that segments vertebral bodies and psoas muscles, quantifies bone and muscle biomarkers, and integrates consistent cropping using vertebral landmarks.
- **Derivation and Validation of Sarcopenia Thresholds:** Statistically derived and externally validated thresholds for both height-based and height-independent volumetric sarcopenia indices. The height-independent indices, normalized to vertebral body volume, enable automated classification in datasets where height information is unavailable or unreliable.
- **Clinical Outcome Association Analysis:** Applied CT-derived biomarkers to an SBRT-treated cohort and identified statistically significant associations with survival and fracture risk in specific models, highlighting their potential utility and the need for validation in larger, independent cohorts.

- **MRI Segmentation Model Development:** Developed MRI-based segmentation models using nnU-Net, trained on CT-derived labels, to enable psoas and vertebral body segmentation despite the scarcity of manually annotated MRI data.
- **Pipeline Extension to MRI:** Implemented a reusable pipeline for aligning CT-MRI volumes, resampling segmentations, and preparing multi-modality training datasets for deep learning segmentation.

Future Work

This thesis established a robust framework for extracting musculoskeletal biomarkers from CT and MRI in patients with spinal metastases. However, several areas remain open for further development and investigation:

- **Validation and Refinement of Biomarker Thresholds.** This thesis proposed both height-independent and height-normalized CT-derived volumetric indices for sarcopenia classification. While all prevalence estimates and downstream analyses in this study were based on the height-independent index (normalized to the square root of vertebral body volume at L3), future studies may benefit from further evaluating the comparative performance of height-normalized indices, which could offer improved precision when accurate height data are available.
- **Expansion of Fracture Risk Analysis.** While preliminary associations between imaging biomarkers and fracture risk were observed, larger and more diverse cohorts — particularly with a higher number of post-SBRT fractures — may help strengthen statistical power and support the development of predictive models with clinical utility.

- **Improvement of MRI Vertebral Body Segmentation.** Although the psoas muscle segmentation model performed well, the vertebral body segmentation occasionally produced mislabeled vertebral levels. Label consistency improved when using cross-validation and focusing on the lumbar region. This behavior reflects nnU-Net's lack of anatomical priors for vertebral labeling. Future work could incorporate postprocessing strategies, such as coordinate-based or centroid-guided relabeling, to assign consistent vertebral labels and enable reliable vertebra-level biomarker analysis.

- **Completion of the Fully Automated MRI Pipeline.** With the current modules for segmentation, cropping, and biomarker extraction adapted to MRI, the foundation for a fully automated MRI-based pipeline is already in place. Finalizing this pipeline may require improvements in vertebral labeling and further automation of its component steps.

- **Comparison and Harmonization of CT and MRI Biomarkers.** With both CT- and MRI-based pipelines available, future studies could compare biomarker values derived from each modality to evaluate consistency and potential conversion factors. Assessing agreement and repeatability will be essential for longitudinal studies where patients may alternate between imaging modalities.

- **Assessment of Threshold Transferability from CT to MRI.** The sarcopenia thresholds established in this thesis were based on CT-derived metrics. Future work may explore whether these thresholds can be meaningfully transferred to MRI-based biomarkers or whether MRI-specific thresholds should be developed.

In summary, this thesis represents an important step toward integrating deep learning, medical imaging, and musculoskeletal biomarker analysis into a cohesive and clinically useful

framework for oncology patients. Continued research in these areas will further advance personalized risk prediction and therapeutic decision-making for individuals with spinal metastases undergoing SBRT.

Bibliography

[1] T. Sprave, V. Verma, R. Förster, I. Schlampp, T. Bruckner, T. Bostel, S. E. Welte, E. Tonndorf-Martini, N. H. Nicolay, J. Debus, *et al.*, "Randomized phase ii trial evaluating pain response in patients with spinal metastases following stereotactic body radiotherapy versus three-dimensional conformal radiotherapy," *Radiotherapy and Oncology*, vol. 128, no. 2, pp. 274–282, 2018.

[2] A. Sahgal, S. D. Myrehaug, S. Siva, G. L. Masucci, P. J. Maralani, M. Brundage, J. Butler, E. Chow, M. G. Fehlings, M. Foote, *et al.*, "Stereotactic body radiotherapy versus conventional external beam radiotherapy in patients with painful spinal metastases: an open-label, multicentre, randomised, controlled, phase 2/3 trial," *The Lancet Oncology*, vol. 22, no. 7, pp. 1023–1033, 2021.

[3] R. S. Guninski, F. Cuccia, F. Alongi, N. Andratschke, C. Belka, D. Bellut, M. Dahele, M. Josipovic, T. Kroese, P. Mancosu, *et al.*, "Efficacy and safety of sbrt for spine metastases: A systematic review and meta-analysis for preparation of an estro practice guideline," *Radiotherapy and Oncology*, p. 109969, 2023.

[4] A. Sahgal, E. G. Atenafu, S. Chao, A. Al-Omair, N. Boehling, E. H. Balagamwala, M. Cunha, I. Thibault, L. Angelov, P. Brown, *et al.*, "Vertebral compression fracture

after spine stereotactic body radiotherapy: a multi-institutional analysis with a focus on radiation dose and the spinal instability neoplastic score," *Journal of clinical oncology*, vol. 31, no. 27, pp. 3426–3431, 2013.

[5] J. Paintin, C. Cooper, and E. Dennison, "Osteosarcopenia," *British journal of hospital medicine*, vol. 79, no. 5, pp. 253–258, 2018.

[6] É. Bourassa-Moreau, A. Versteeg, E. Moskven, R. Charest-Morin, A. Flexman, T. Ailon, T. Dalkilic, C. Fisher, N. Dea, M. Boyd, *et al.*, "Sarcopenia, but not frailty, predicts early mortality and adverse events after emergent surgery for metastatic disease of the spine," *The Spine Journal*, vol. 20, no. 1, pp. 22–31, 2020.

[7] E. Moskven, E. Bourassa-Moreau, R. Charest-Morin, A. Flexman, and J. Street, "The impact of frailty and sarcopenia on postoperative outcomes in adult spine surgery. a systematic review of the literature," *The Spine Journal*, vol. 18, no. 12, pp. 2354–2369, 2018.

[8] A. M. Flexman, J. Street, and R. Charest-Morin, "The impact of frailty and sarcopenia on patient outcomes after complex spine surgery," *Current Opinion in Anesthesiology*, vol. 32, no. 5, pp. 609–615, 2019.

[9] J.-I. Yoo, H. Kim, Y.-C. Ha, H.-B. Kwon, and K.-H. Koo, "Osteosarcopenia in patients with hip fracture is related with high mortality," *Journal of Korean medical science*, vol. 33, no. 4, 2018.

[10] K. M. Sturgeon, K. M. Mathis, C. J. Rogers, K. H. Schmitz, and D. L. Waning, "Cancer- and chemotherapy-induced musculoskeletal degradation," *JBMR plus*, vol. 3, no. 3, p. e10187, 2019.

[11] E. Massaad, C. P. Bridge, A. Kiapour, M. S. Fourman, J. B. Duvall, I. D. Connolly, M. Hadzipasic, G. M. Shankar, K. P. Andriole, M. Rosenthal, *et al.*, "Evaluating frailty, mortality, and complications associated with metastatic spine tumor surgery using machine learning–derived body composition analysis," *Journal of Neurosurgery: Spine*, vol. 37, no. 2, pp. 263–273, 2022.

[12] A. M. Flexman, R. Charest-Morin, L. Stobart, J. Street, and C. J. Ryerson, "Frailty and postoperative outcomes in patients undergoing surgery for degenerative spine disease," *The spine journal*, vol. 16, no. 11, pp. 1315–1323, 2016.

[13] M. Guckenberger, "Dose-intensified stereotactic body radiotherapy for painful vertebral metastases: a randomized phase 3 trial," *Cancer*, vol. 130, no. 15, pp. 2713–2722, 2024.

[14] M. Huo, A. Sahgal, D. Pryor, K. Redmond, S. Lo, and M. Foote, "Stereotactic spine radiosurgery: review of safety and efficacy with respect to dose and fractionation," *Surgical Neurology International*, vol. 8, no. 1, p. 30, 2017.

[15] X. Wang, L. Rhines, A. Shiu, J. Yang, U. Selek, I. Gning, and E. Chang, "Stereotactic body radiation therapy for management of spinal metastases in patients without spinal cord compression: a phase 1-2 trial," *The Lancet Oncology*, vol. 13, no. 4, pp. 395–402, 2012.

[16] B. Bindels, "Stereotactic body and conventional radiotherapy for painful bone metastases," *JAMA Network Open*, vol. 7, no. 2, p. e2355409, 2024.

[17] S. Godin, A. Durham, L. Schiappacasse, E. Ozsahin, and F. Vilotte, "Vertebral compression fracture during stereotactic body radiotherapy for spinal metastasis: a rare case of tracking failure," *Cancer/Radiothérapie*, vol. 24, no. 8, pp. 866–869, 2020.

[18] M. Jawad, D. Fahim, P. Gerszten, J. Flíckinger, A. Sahgal, I. Grills, and M. Guckenberger, "Vertebral compression fractures after stereotactic body radiation therapy: a large, multi-institutional, multinational evaluation," *Journal of Neurosurgery Spine*, vol. 24, no. 6, pp. 928–936, 2016.

[19] N. Boehling, D. Grosshans, P. Allen, M. McAleer, A. Burton, S. Azeem, and E. Chang, "Vertebral compression fracture risk after stereotactic body radiotherapy for spinal metastases," *Journal of Neurosurgery Spine*, vol. 16, no. 4, pp. 379–386, 2012.

[20] D. Rivas, "Robotic stereotactic body radiotherapy for spine metastasis pain relief," *Applied Sciences*, vol. 14, no. 5, p. 1775, 2024.

[21] H. Hirschfeld, R. Kinsella, and G. Duque, "Osteosarcopenia: where bone, muscle, and fat collide," *Osteoporosis International*, vol. 28, pp. 2781–2790, 2017.

[22] S. S. Shachar, A. M. Deal, M. Weinberg, K. A. Nyrop, G. R. Williams, T. F. Nishijima, J. M. Benbow, and H. B. Muss, "Skeletal muscle measures as predictors of toxicity, hospitalization, and survival in patients with metastatic breast cancer receiving taxane-based chemotherapy," *Clinical Cancer Research*, vol. 23, no. 3, pp. 658–665, 2017.

[23] B. R. Nielsen, J. Abdulla, H. E. Andersen, P. Schwarz, and C. Suetta, "Sarcopenia and osteoporosis in older people: a systematic review and meta-analysis," *European geriatric medicine*, vol. 9, pp. 419–434, 2018.

[24] C. Wang, "Osteosarcopenia in patients with cancer: a systematic review and meta-analysis," *Medicine*, vol. 103, no. 45, p. e40476, 2024.

[25] C. Saeki, T. Kanai, K. Ueda, M. Nakano, T. Oikawa, Y. Torisu, M. Saruta, and A. Tsubota, "Osteosarcopenia predicts poor survival in patients with cirrhosis: a retrospective study," *BMC gastroenterology*, vol. 23, no. 1, p. 196, 2023.

[26] N. Boehling, D. Grosshans, P. Allen, M. McAleer, A. Burton, S. Azeem, and E. Chang, "Vertebral compression fracture risk after stereotactic body radiotherapy for spinal metastases," *Journal of Neurosurgery: Spine*, vol. 16, no. 4, pp. 379–386, 2012.

[27] M. S. Jawad, D. K. Fahim, P. C. Gerszten, J. C. Flickinger, A. Sahgal, I. S. Grills, J. Sheehan, R. Kersh, J. Shin, K. Oh, *et al.*, "Vertebral compression fractures after stereotactic body radiation therapy: a large, multi-institutional, multinational evaluation," *Journal of Neurosurgery: Spine*, vol. 24, no. 6, pp. 928–936, 2016.

[28] M. Mascarella, "Sarcopenia predicts short-term treatment-related toxicity in patients undergoing curative-intent therapy for head and neck cancer: a systematic review and meta-analysis," *Head & Neck*, vol. 46, no. 6, pp. 1500–1509, 2024.

[29] S. Chauvie, L. N. Mazzoni, and J. O'Doherty, "A review on the use of imaging biomarkers in oncology clinical trials: Quality assurance strategies for technical validation," *Tomography*, vol. 9, no. 5, pp. 1876–1902, 2023.

[30] H. Ibad *et al.*, "Computed tomography," *Investigative Radiology*, vol. 58, no. 1, pp. 99–110, 2022.

[31] B. Eck, M. Yang, J. Elias, *et al.*, "Quantitative mri for evaluation of musculoskeletal disease," *Investigative Radiology*, vol. 58, no. 1, pp. 60–75, 2022.

[32] N. Habibullah, "Use of point-of-care ultrasound in a low-resource setting to diagnose achilles tendon rupture and avulsion fracture of the calcaneal bone," *International Journal of Emergency Medicine*, vol. 16, no. 1, 2023.

[33] F. Wuennemann *et al.*, "Diagnostic performance of 3d-multi-echo-data-image-combination (medic) for evaluating slap lesions of the shoulder," *BMC Musculoskeletal Disorders*, vol. 20, no. 1, 2019.

[34] M. Wong *et al.*, "Four-dimensional computed tomography: musculoskeletal applications," *Canadian Journal of Surgery*, vol. 65, no. 3, pp. E388–E393, 2022.

[35] J. Yoder *et al.*, "Applications of pet–computed tomography–magnetic resonance in the management of benign musculoskeletal disorders," *Pet Clinics*, vol. 14, no. 1, pp. 1–15, 2019.

[36] F. Isensee, T. Wald, C. Ulrich, M. Baumgartner, S. Roy, K. Maier-Hein, and P. F. Jaeger, "nnu-net revisited: A call for rigorous validation in 3d medical image segmentation," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pp. 488–498, Springer, 2024.

[37] X. Liu, K. Gao, B. Liu, C. Pan, K. Liang, L. Yan, J. Ma, F. He, S. Zhang, S. Pan, *et al.*, "Advances in deep learning-based medical image analysis," *Health Data Science*, vol. 2021, 2021.

[38] V. Binay Safer and U. Safer, "Usefulness and limitations of single-slice computed tomography analysis at the third lumbar region in the assessment of sarcopenia," *Critical Care*, vol. 17, pp. 1–2, 2013.

[39] F. Pigneur, M. Di Palma, B. Raynard, A. Guibal, F. Cohen, N. Daidj, R. Aziza, M. El Hajjam, G. Louis, F. Goldwasser, *et al.*, "Psoas muscle index is not representative of skeletal muscle index for evaluating cancer sarcopenia," *Journal of Cachexia, Sarcopenia and Muscle*, vol. 14, no. 4, pp. 1613–1620, 2023.

[40] M. Kong, N. Geng, Y. Zhou, N. Lin, W. Song, M. Xu, S. Li, Y. Piao, Z. Han, R. Guo, *et al.*, "Defining reference values for low skeletal muscle index at the l3 vertebra level based on computed tomography in healthy adults: a multicentre study," *Clinical Nutrition*, vol. 41, no. 2, pp. 396–404, 2022.

[41] C. M. Prado, J. R. Lieffers, L. J. McCargar, T. Reiman, M. B. Sawyer, L. Martin, and V. E. Baracos, "Prevalence and clinical implications of sarcopenic obesity in patients with solid tumours of the respiratory and gastrointestinal tracts: a population-based study," *The lancet oncology*, vol. 9, no. 7, pp. 629–635, 2008.

[42] L. Martin, L. Birdsell, N. MacDonald, T. Reiman, M. T. Clandinin, L. J. McCargar, R. Murphy, S. Ghosh, M. B. Sawyer, and V. E. Baracos, "Cancer cachexia in the age of obesity: skeletal muscle depletion is a powerful prognostic factor, independent of body mass index," *Journal of clinical oncology*, vol. 31, no. 12, pp. 1539–1547, 2013.

[43] R. D. Dolan, A. S. Almasaudi, L. B. Dieu, P. G. Horgan, S. T. McSorley, and D. C. McMillan, "The relationship between computed tomography-derived body composition, systemic inflammatory response, and survival in patients undergoing surgery for colorectal cancer," *Journal of cachexia, sarcopenia and muscle*, vol. 10, no. 1, pp. 111–122, 2019.

[44] D. Albano, F. Dondi, M. Ravanelli, A. Tucci, D. Farina, R. Giubbini, G. Treglia, and F. Bertagna, "Prognostic role of" radiological" sarcopenia in lymphoma: A systematic review," *Clinical Lymphoma Myeloma and Leukemia*, vol. 22, no. 5, pp. e340–e349, 2022.

[45] Y. Hamaguchi, T. Kaido, S. Okumura, A. Kobayashi, A. Hammad, Y. Tamai, N. Inagaki, and S. Uemoto, "Proposal for new diagnostic criteria for low skeletal muscle mass based on computed tomography imaging in asian adults," *Nutrition*, vol. 32, no. 11-12, pp. 1200–1205, 2016.

[46] H. Nishikawa and S. Nishiguchi, "Sarcopenia in liver disease," *The Evolving Landscape of Liver Cirrhosis Management*, pp. 119–130, 2019.

[47] Y. Yuri, H. Nishikawa, H. Enomoto, A. Ishii, Y. Iwata, Y. Miyamoto, N. Ishii, K. Hasegawa, C. Nakano, T. Nishimura, *et al.*, "Implication of psoas muscle index on survival for hepatocellular carcinoma undergoing radiofrequency ablation therapy," *Journal of Cancer*, vol. 8, no. 9, p. 1507, 2017.

[48] M. Zhang, Y. Xiong, M. Chen, D. Xu, K. Xu, and W. Tian, "Psoas muscle mass index and peak expiratory flow as measures of sarcopenia: relation to outcomes of elderly patients with resectable esophageal cancer," *Frontiers in Oncology*, vol. 13, p. 1303877, 2023.

[49] T. Ishida, T. Makino, M. Yamasaki, K. Tanaka, Y. Miyazaki, T. Takahashi, Y. Kurokawa, M. Motoori, Y. Kimura, K. Nakajima, *et al.*, "Impact of measurement of skeletal muscle mass on clinical outcomes in patients with esophageal cancer undergoing esophagectomy after neoadjuvant chemotherapy," *Surgery*, vol. 166, no. 6, pp. 1041–1047, 2019.

[50] S. Mayanagi, A. Ishikawa, K. Matsui, S. Matsuda, T. Irino, R. Nakamura, K. Fukuda, N. Wada, H. Kawakubo, N. Hijikata, *et al.*, "Association of preoperative sarcopenia with postoperative dysphagia in patients with thoracic esophageal cancer," *Diseases of the Esophagus*, vol. 34, no. 9, p. doaa121, 2021.

[51] Y. Sato, Y. Tanaka, T. Imai, H. Ojio, J. Mase, Y. Hatanaka, T. Suetsugu, H. Fujii, N. Okumura, N. Matsuhashi, *et al.*, "Effect of biweekly-dcf neoadjuvant chemotherapy on skeletal muscle mass of esophageal cancer patients," *Gan to Kagaku ryoho. Cancer & Chemotherapy*, vol. 49, no. 1, pp. 47–52, 2022.

[52] R. Nakamura, Y. Inage, R. Tobita, S. Yoneyama, T. Numata, K. Ota, H. Yanai, T. Endo, Y. Inadome, S. Sakashita, *et al.*, "Sarcopenia in resected nsclc: effect on postoperative outcomes," *Journal of Thoracic Oncology*, vol. 13, no. 7, pp. 895–903, 2018.

[53] A. S. Tagliafico, B. Bignotti, L. Torri, and F. Rossi, "Sarcopenia: how to measure, when and why," *La radiologia medica*, vol. 127, no. 3, pp. 228–237, 2022.

[54] D. P. Ahern, J. M. McDonnell, M. Riffault, *et al.*, "A meta-analysis of the diagnostic accuracy of hounsfield units on computed topography relative to dual-energy x-ray absorptiometry for the diagnosis of osteoporosis in the spine surgery population," *The Spine Journal*, vol. 21, no. 10, pp. 1738–1749, 2021.

[55] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *Medical image computing and computer-assisted intervention–MICCAI 2015: 18th international conference, Munich, Germany, October 5-9, 2015, proceedings, part III 18*, pp. 234–241, Springer, 2015.

[56] F. Milletari, N. Navab, and S.-A. Ahmadi, "V-net: Fully convolutional neural networks for volumetric medical image segmentation," in *2016 Fourth International Conference on 3D Vision (3DV)*, pp. 565–571, IEEE, 2016.

[57] F. Isensee, P. F. Jaeger, S. A. Kohl, J. Petersen, and K. H. Maier-Hein, "nnu-net: a self-configuring method for deep learning-based biomedical image segmentation," *Nature methods*, vol. 18, no. 2, pp. 203–211, 2021.

[58] Q. Xia, H. Zheng, H. Zou, D. Luo, H. Tang, L. Li, and B. Jiang, "A comprehensive review of deep learning for medical image segmentation," *Neurocomputing*, vol. 613, p. 128740, 2025.

[59] A. Hatamizadeh, V. Nath, Y. Tang, D. Yang, H. R. Roth, and D. Xu, "Swin unetr: Swin transformers for semantic segmentation of brain tumors in mri images," in *Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries*, pp. 272–284, Springer, 2022.

[60] Y. Dong, K. Zhang, S. Ying, *et al.*, "Unsupervised domain adaptation for automatic estimation of cardiothoracic ratio on chest x-ray," *IEEE Access*, vol. 7, pp. 100704–100712, 2019.

[61] J. Jiang, Y.-C. Hu, N. Tyagi, P. Zhang, A. Rimner, J. O. Deasy, and H. Veeraraghavan, "Cross-modality (ct-mri) prior augmented deep learning for robust lung tumor segmentation from small mr datasets," *Medical physics*, vol. 46, no. 10, pp. 4392–4404, 2019.

[62] A. Bahrami, A. Karimian, E. Fatemizadeh, H. Arabi, and H. Zaidi, "A new deep convolutional neural network design with efficient learning capability: Application to ct image synthesis from mri," *Medical physics*, vol. 47, no. 10, pp. 5158–5171, 2020.

[63] Y. Lei, J. Harms, T. Wang, Y. Liu, H.-K. Shu, A. B. Jani, W. J. Curran, H. Mao, T. Liu, and X. Yang, "Mri-only based synthetic ct generation using dense cycle consistent generative adversarial networks," *Medical physics*, vol. 46, no. 8, pp. 3565–3581, 2019.

[64] F. Proietto Salanitri, G. Bellitto, I. Irmakci, S. Palazzo, U. Bagci, and C. Spampinato, "Hierarchical 3d feature learning forpancreas segmentation," in *Machine Learning in Medical Imaging: 12th International Workshop, MLMI 2021, Held in Conjunction with MICCAI 2021, Strasbourg, France, September 27, 2021, Proceedings 12*, pp. 238–247, Springer, 2021.

[65] R. D. Boutin, L. Yao, R. J. Canter, and L. Lenchik, "Sarcopenia: current concepts and imaging implications," *American Journal of Roentgenology*, vol. 205, no. 3, pp. W255–W266, 2015.

[66] G. Klein, M. Hardisty, C. Whyne, and A. L. Martel, "Vertdetect: Fully end-to-end 3d vertebral instance segmentation model," 2023.

[67] J. Wasserthal, H.-C. Breit, M. T. Meyer, M. Pradella, D. Hinck, A. W. Sauter, T. Heye, D. T. Boll, J. Cyriac, S. Yang, *et al.*, "Totalsegmentator: robust segmentation of 104 anatomic structures in ct images," *Radiology: Artificial Intelligence*, vol. 5, no. 5, 2023.

[68] F. Isensee, P. F. Jaeger, S. A. Kohl, J. Petersen, and K. H. Maier-Hein, "nnu-net: a self-configuring method for deep learning-based biomedical image segmentation," *Nature methods*, vol. 18, no. 2, pp. 203–211, 2021.

[69] K. Imai, K. Takai, M. Aiba, S. Unome, T. Miwa, T. Hanai, H. Sakai, Y. Shirakami, A. Suetsugu, and M. Shimizu, "Psoas muscle index as an independent predictor of

survival in patients with hepatocellular carcinoma receiving systemic targeted therapy," *Cancers*, vol. 17, no. 2, p. 209, 2025.

[70] S. U. Engelmann, C. Pickl, M. Haas, S. Kaelble, V. Hartmann, M. Firsching, L. Lehmann, M. Guzvic, B. W. van Rhijn, J. Breyer, *et al.*, "Body composition of patients undergoing radical cystectomy for bladder cancer: sarcopenia, low psoas muscle index, and myosteatosis are independent risk factors for mortality," *Cancers*, vol. 15, no. 6, p. 1778, 2023.

[71] Z. Benedek, S. Todor-Boér, L. Kocsis, O. Bauer, N. Suciu, and M. F. Coroș, "Psoas muscle index defined by computer tomography predicts the presence of postoperative complications in colorectal cancer surgery," *Medicina*, vol. 57, no. 5, p. 472, 2021.

[72] A. e. a. Sekuboyina, "Labeling vertebrae with two-stage 3d cnns: integrating spatial configuration into deep learning," *Medical Image Analysis*, vol. 73, p. 102146, 2020.

[73] N. e. a. Lessmann, "Iterative fully convolutional neural networks for automatic vertebra segmentation and identification," *Medical Image Analysis*, vol. 53, pp. 142–155, 2019.

[74] A. e. a. Mukhopadhyay, "Deep image-to-image network for vertebra segmentation and identification," in *ISBI*, pp. 1599–1603, 2020.