Hounsfield Unit Ranges as Inductive Bias for Intra-Clinical Learning of Data-Efficient CT Segmentation Models

Benjamin Meyer[*†‖], Pascal J. Sager[*†‖],
Ahmed Abdulkadir[*], Benjamin F. Grewe[†], Philipp Schuetz[‡], Thilo Stadelmann[*¶], Felice Burn[§]
{mebr,sage,abdk,stdm}@zhaw.ch, bgrewe@ethz.ch, {philipp.schuetz,felice.burn}@ksa.ch
[*]Centre for Artificial Intelligence, Zurich University of Applied Sciences, Winterthur, Switzerland
[†]Institute of Neuroinformatics, ETH Zurich & University of Zurich, Zurich, Switzerland
[‡]University Clinic of Medicine, Aarau Cantonal Hospital, Aarau, Switzerland
[§]AI & Data Science CoE, Aarau Cantonal Hospital, Aarau, Switzerland
[¶]Fellow, ECLT (European Centre for Living Technology, Venice, Italy) and Senior Member, IEEE
[‖]The first two authors contributed equally to this work.

Abstract—Automated tissue segmentation in medical imaging plays a critical role in clinical AI-assisted decision-making, and particularly in the assessment of body composition from CT scans. However, acquiring data and annotations of sufficient quality to train deep-learning models is expensive and time-consuming. In this work, we propose a novel approach to improve data efficiency and model accuracy by leveraging domain knowledge about biologically relevant tissue-specific Hounsfield unit (HU) ranges as an inductive bias for learning. Specifically, we extend the input representation of deep learning-based segmentation models with binary masks indicating potential tissue types, where each binary mask is created from thresholds derived from medical literature. Our method not only enhances segmentation performance by up to 5 % for intramuscular adipose tissue but surpasses the performance of the baseline model with 50 % of the training data. Our easy-to-apply method thus improves data efficiency and facilitates the development and use of segmentation models in resource-constrained clinical settings.

Index Terms—Tissue segmentation, body composition analysis, medical imaging, data efficiency, physiology-informed machine learning, deep learning
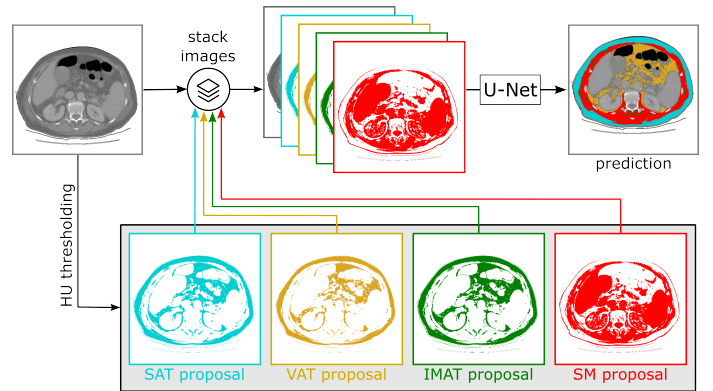
Fig. 1: Overview of our approach: Binary masks that indicate potential tissue-specific locations are extracted using HU ranges as thresholds and concatenated with the original CT image as additional input channels for a machine learning model.

I. INTRODUCTION

Deep neural networks have shown remarkable performance across various domains, including health and medical imaging [1]–[4]. They have huge potential in clinical decision support systems, assisting physicians in making more informed diagnoses and better treatment choices [5]–[7]. An important task in medical imaging is the assessment of a patient's body composition [8], such as quantifying muscle mass and different types of adipose tissue [9]. A widely adopted approach for estimating body composition is applying tissue segmentation to a specific two-dimensional (2D) slice extracted from a three-dimensional computed tomography (CT) scan. Tissue segmentation labels the tissue type at the pixel level, allowing for the calculation of tissue areas and subsequent body composition analysis (BCA).

Despite the importance of automated BCA, only a few publicly ready-to-use systems with limited scope exist [10]–[12]. Furthermore, annotated datasets are rarely shared due to strong data privacy constraints and strict regulations regarding medical data usage. Consequently, institutions such as hospitals and software manufacturers must train (or fine-tune) custom segmentation models on internal annotated data to adapt to in-house imaging equipment and procedures [13], [14]. The necessary annotations must be provided by medical experts, such as radiologists, making the labeling process both expensive and time-intensive to the point of infeasibility given ordinary clinical operations as a main task. This raises our research question: *How can the number of required training samples be reduced for machine learning-based medical imaging models while maintaining the performance necessary for clinical application?*

Existing approaches to mitigate data scarcity in medical segmentation use techniques like data augmentation, transfer learning, domain adaptation, or self-supervised techniques (for a detailed review, see Section II-B). While these techniques can reduce the dependency on annotated data, they often require careful optimization of hyperparameters, model architectures, and learning strategies, making them complex to implement effectively.

In this paper, we introduce an easy-to-apply approach to reduce the dependence on labeled training data. Specifically, taking the data-centric perspective [15], [16] into account, we leverage established tissue-specific ranges of so-called Hounsfield unit (HU) values from the medical literature, defined on the Hounsfield scale (a standardized measure of radiodensity in CT imaging), as an inductive bias to learn segmentation models: Each tissue type exhibits a well-defined HU range [17], providing natural prior knowledge that can guide the learning process. For instance, cortical bone typically falls within the range of $+500$ to $+1900$ HU, whereas different adipose tissue is found within -190 to -30 HU [18]. By incorporating HU ranges as prior knowledge, the model is informed about plausible tissue types for each pixel and can use this information as an inductive bias to guide learning, thereby narrowing the search space and improving data efficiency as well as segmentation performance.

We make the following contributions: (1) *Utilization of HU ranges:* We propose incorporating HU ranges as a novel and effective inductive bias to learn tissue segmentation economically and enhance the final neural network's performance. (2) *Comprehensive experiments:* We conduct extensive experiments across various model and dataset sizes, demonstrating robust improvements. (3) *Tissue-level analysis:* We highlight that our method is especially helpful for identifying intramuscular adipose tissue (IMAT), a clinically relevant tissue type representing muscle health conditions. Our contributions have direct relevance for clinical practice, enabling the efficient training and clinical application of tailored segmentation models in-house.

II. RELATED WORK

A. Abdominal CT segmentation for BCA

CT, alongside magnetic resonance imaging (MRI), is considered the most accurate modality for BCA [17], where different tissue types are classified at the pixel level. It has been demonstrated that 2D slices acquired at the level of the first (L1) [9] or third lumbar vertebra (L3) [19] as well as the tenth (T10) [20] or twelfth thoracic vertebra (T12) [21] serve as a reliable surrogate by segmenting for the assessment of abdominal body compositions, enabling an efficient estimation of body composition without requiring full-volume segmentation [22]. For a comprehensive discussion on alternative anatomical levels, refer to [9]. Recent advancements in automated BCA predominantly utilize deep learning-based segmentation models, particularly U-Net architectures [23]–[25] or variations of it [26]. These models are trained on CT slices where each pixel's value corresponds to a Hounsfield Unit (HU), enabling

the model to learn the relationship between HU values and tissue types.

In this study, we build upon these existing methodologies and enhance the well-established U-Net framework [27] by incorporating *HU ranges*, derived from the medical literature as an inductive bias during pre-processing. By leveraging medical textbook knowledge in this manner, we employ an effective inductive bias that facilitates more data-efficient learning and enhances final segmentation performance.

B. Label scarcity in medical image segmentation

Various methodologies have been developed to address the challenge of missing or partially labeled segmentation data [28]. *Data augmentation* techniques artificially expand the dataset by applying transformations to the original data, thereby generating multiple variations [29]–[32]. *Domain adaptation and transfer learning* facilitate knowledge transfer from related domains or tasks, enhancing model generalization [13], [33]–[35]. *Weakly-supervised learning* leverages unlabeled data in conjunction with limited annotations to improve model performance [14], [36].

Similar to data augmentation, we use a pre-processing step (creation of HU range-based binary masks) that is applied before model training. However, rather than expanding the dataset by new examples, our method enriches each example with additional domain knowledge (fed into the network as additional channels of the original input example), akin to approaches in other domains such as physics-informed neural networks [37]. This creates strong physiology-informed inductive biases in neural network training.

Prior research has demonstrated that incorporating domain knowledge into medical imaging is a highly effective strategy for addressing label scarcity [14]. For example, a related approach is *HU window* pre-processing [38], [39] that normalizes and clips HU values rather than deriving simple binary masks (for details, see Appendix A). Importantly, our approach is complementary to existing methodologies, offering a versatile addition to current techniques for reducing data requirements in medical segmentation.

III. HOUNSFIELD UNIT RANGES AS INDUCTIVE BIAS

The goal of tissue segmentation for BCA is to classify each pixel of a 2D CT slice into a tissue class or the background class. We denote the CT slice, a 2D image of HU values, as matrix $\mathbf{X}_0 \in \mathbb{R}^{h,w}$, where h and w denote the height and width of the slice. The HU pixel values in $\mathbf{X}_0$ typically range from -1000 (air) to approximately $+3000$ (dense bone) [40], while foreign objects such as metal implants can exhibit HU values up to $+30\,000$. The pixel-wise classification outputs a segmentation mask $\hat{\mathbf{Y}} \in (\mathcal{C} \cup \{BG\})^{h,w}$ of the same shape as the input image, where $\mathcal{C}$ represents the set of tissue classes and BG the background class.

Fig. 1 visualizes how we integrate the prior knowledge about tissue-specific HU ranges into the model. On top of the input $\mathbf{X}_0$ we stack $|\mathcal{C}|$ binary masks $\{\mathbf{X}_c \mid c \in \mathcal{C}\}$, yielding $\mathbf{X} \in$

TABLE I: HU ranges for each tissue type.

Tissue type	HU range	
	From	To
Skeletal muscle (SM)	-29	150
Intramuscular adipose tissue (IMAT)	-190	-30
Visceral adipose tissue (VAT)	-150	-50
Subcutaneous adipose tissue (SAT)	-190	-30

$\mathbb{R}^{(1+|\mathcal{C}|),h,w}$. Given $\mathbf{X}_0(i,j)$ be the pixel value at row i and column j in matrix $\mathbf{X}_0$, $\mathbf{X}_c$ for class c is defined as:

$$\mathbf{X}_c(i,j) = \begin{cases} 1, & \text{if } \mathbf{X}_0(i,j) \text{ within the HU range of class } c, \\ 0, & \text{otherwise.} \end{cases}$$

Thus, the first channel of the input $\mathbf{X}$ contains the "original" CT image $\mathbf{X}_0$, while the additional channels contain the binary masks $\mathbf{X}_c$, each indicating if a class c is plausible at specific pixel locations according to medical textbook knowledge on HU ranges. Table I provides an overview of the tissue types and corresponding HU ranges used in this study. Equivalent HU ranges (IMAT, SAT) lead to identical binary masks and could also be represented as a single mask. Overlapping HU ranges (e.g., VAT, SAT) lead to overlapping masks, indicating multiple possible tissue types at some locations. Adding the respective information as additional input channels effectively constrains the model's decision space, reducing ambiguity and improving segmentation accuracy by leveraging well-established radiodensity properties.

IV. EXPERIMENTAL SETUP

We evaluate our proposed approach using widely adopted deep learning architectures for medical image segmentation, training them with default hyperparameter settings. To ensure a rigorous assessment, we employ standard validation procedures in accordance with best practices.

A. Model

For the segmentation task, we build upon a 2D U-Net model [27], a widely recognized state-of-the-art architecture in this domain, known for its effectiveness even in data-limited scenarios. The U-Net model follows an encoder-decoder structure with skip connections to preserve spatial information. To test our approach on different model sizes, we train three model variants with different initial feature map sizes (16, 32, or 64 channels). All configurations are based on a standard U-Net architecture and maintain a consistent depth of five encoding and decoding stages. They follow the standard practice of doubling the number of feature maps at each downsampling stage and ending with a 1×1 kernel for pixel-wise prediction [27]. The models contain around 2M (*U-Net-small*), 8M (*U-Net-base*), and 31M (*U-Net-large*) parameters, respectively.

Unlike the original U-Net architecture, we employ padded convolutions and add batch normalization after each convolutional layer in the encoder and decoder.

B. Data

We conduct our experiments using CT data obtained from the Cantonal Hospital Aarau, Switzerland. Our dataset contains 176 CT slices in 512×512 resolution from 44 patients with malnutritional conditions. We segment four tissue types: The skeleton muscle (SM), intramuscular adipose tissue (IMAT), visceral adipose tissue (VAT), and subcutaneous adipose tissue (SAT). For each patient, we label four CT slices acquired at four vertebral levels: L1, L3, T10, and T12. Due to data privacy regulations and national legislation (HRA), we are unable to publicly share imaging data, associated annotations, or model predictions.

To demonstrate the effectiveness of our approach on different dataset sizes, we sample subsets only containing fractions (10 %, 20 %, ..., 100 %) of the available data. The subsets are *nested*, meaning a bigger subset includes smaller subsets, simulating the growth of an increasingly large dataset.

C. Annotation process

Following a thorough review, cases of unsuitable quality for manual analysis were excluded due to skeletal muscle tissue extending beyond the field of view. No severe anatomical or technical abnormalities—such as body habitus variations, metal artifacts, or massive fluid accumulation—necessitated exclusion. The remaining slices were classified based on characteristics commonly encountered in routine CT scans, including subcutaneous adipose tissue extending beyond the image field and mild image quality impairments (e.g., streaking or graininess). After dataset review, all eligible slices underwent semi-assisted manual segmentation using Slicomatic™ software from TomoVision at the specified vertebral levels [41].

D. Training parameters

Each model is trained using the standard Adam optimizer [42] with its default settings: learning rate $\alpha = 0.001$, $\beta_1 = 0.9$, $\beta_2 = 0.999$, and $\epsilon = 10^{-8}$. The optimization process minimizes the cross-entropy loss function using a batch size of 4. Regardless of the dataset fraction used, we maintain a fixed number of 1800 optimization steps across all settings to ensure convergence and fair comparison of model performance.

E. Validation

All experiments are conducted using a 3-fold cross-validation across patients. Folds are generated using the same random seed, enabling consistent comparisons across all experiments.

Performance metrics are reported as the mean value computed across the three folds. For each fold, all evaluation metrics are computed based on the best model checkpoint, as determined by the lowest validation loss. We report the Dice score, which quantifies the overlap between the predicted segmentation (P) and the ground truth segmentation (G) and is defined as:
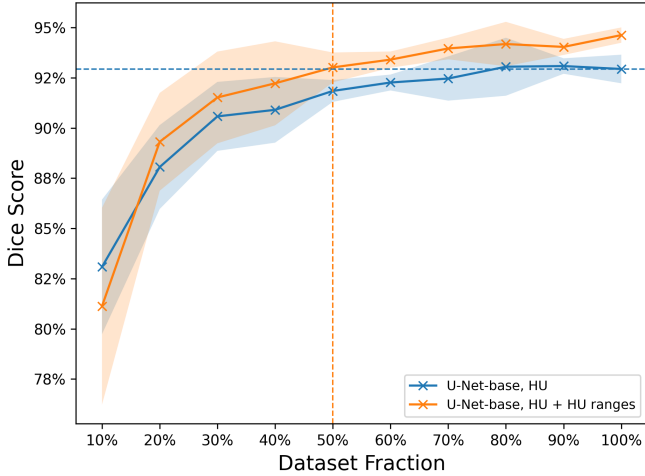
$$\frac{2 \cdot |P \cap G|}{|P| + |G|}$$

Fig. 2: Our model using HU ranges as additional features shows consistent improvement *across different dataset sizes* over the baseline without them. The horizontal blue line indicates the *U-Net base, HU* baseline performance using the entire dataset. The vertical orange line indicates that our approach surpasses this performance using 50 % of the dataset.

Additionally, we report the precision, the recall, and the F_1 score. All metrics are macro-averaged (calculated per tissue class and then averaged across classes), which is common practice to account for class imbalance. To quantify the uncertainty in the estimated means, we report the 95 % confidence interval, estimated using the standard deviation across folds.

V. RESULTS

A. Reduced data needs

Fig. 2 presents the Dice score for the U-Net-base model across varying dataset fractions, comparing performance with and without the proposed HU ranges as additional features. The results show our approach consistently improves performance by approximately 1 % to 2 % across dataset fractions except for the 10 % fraction. This smallest dataset fraction only contains 4 patients, which increases variability in cross-validation and results in wider confidence intervals.

More importantly, as highlighted in Fig. 2, the model utilizing HU ranges surpasses the performance of the baseline when only using 50 % of the training data. This finding underscores the *data efficiency* of our approach, suggesting that incorporating HU ranges not only enhances segmentation accuracy but also significantly reduces the amount of labeled data needed to achieve a given performance level. Consequently, our method offers substantial practical benefits by mitigating annotation costs and improving data efficiency in training deep learning models in and for clinical practice.

B. Performance improvements

Table II summarizes the experimental results, comparing the performance of U-Net models of varying sizes with and without our proposed approach of incorporating HU ranges
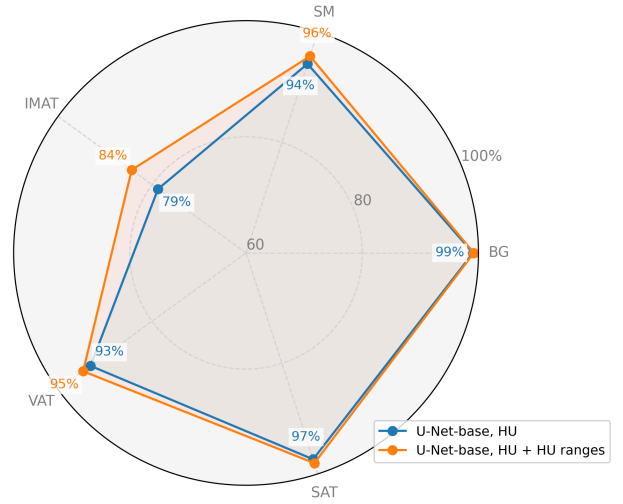

Fig. 3: Dice score *per tissue type* and background (BG). Giving the model a physiology-informed inductive bias via HU ranges-based masks keeps (BG, SAT) or improves performance, with IMAT improvements being of high clinical value.

as additional features to learn from. Across all model sizes, the inclusion of HU ranges consistently improves segmentation performance across multiple metrics, with the Dice score improving by approximately 1 % to 2 %, reducing the remaining error by 20 %. The improvements hold consistently across different U-Net model sizes and dataset fractions, demonstrating the applicability across diverse training settings.

Our baseline results without our proposed additional inputs align with previously reported findings on similar datasets. For example, [43] achieve a Dice score of 97 % using their internal dataset of 1496 CT scans of the L3 vertebra from 922 patients, segmenting the tissue types SM, SAT, and VAT. Our *U-Net-base, HU* model achieves a comparable Dice score of 95 % on the same tissue types (i.e., when ignoring IMAT) despite using a smaller dataset of only 176 CT scans from 44 patients for training and evaluation.

By expanding the number of input channels with engineered masks, the first convolution has few additional parameters, leading to increased computational demand and memory needs. Empirically, we found that the training time was at most $1.17\times$ higher than the network without the HU ranges. This efficiency makes our approach both computationally feasible and practically advantageous.

C. Performance per tissue type

Fig. 3 shows the dice score per tissue type for the U-Net-base model, comparing performance with and without the inclusion of HU ranges. While the segmentation performance remains nearly identical for background and SAT, a modest improvement is observed for SM and VAT. However, a substantial performance increase of approximately 5 % is achieved for IMAT.

IMAT is recognized as an important biomarker for various medical diagnoses [44], [45], yet it remains challenging to

TABLE II: Experimental results. Score ranges represent the 95 % confidence interval. Dice score, Precision, Recall, and F_1-Score are all macro averages across tissue classes.

Model	Number of parameters	Dice score		Precision		Recall		F_1-Score	
U-Net-small, HU	1 943 829	93.0 %	±0.3 %	92.5 %	±1.5 %	92.7 %	±0.9 %	92.5 %	±0.3 %
U-Net-small, HU + HU ranges	1 944 549	**94.7 %**	±0.3 %	**93.6 %**	±0.4 %	**95.3 %**	±0.4 %	**94.4 %**	±0.3 %
U-Net-base, HU	7 765 541	93.6 %	±0.3 %	92.7 %	±0.4 %	93.8 %	±0.7 %	93.2 %	±0.4 %
U-Net-base, HU + HU ranges	7 766 981	**94.8 %**	±0.3 %	**94.0 %**	±0.3 %	**95.0 %**	±0.9 %	**94.5 %**	±0.3 %
U-Net-large, HU	31 042 629	93.1 %	±0.6 %	92.6 %	±1.1 %	92.7 %	±2.2 %	92.6 %	±0.7 %
U-Net-large, HU + HU ranges	31 045 509	**94.6 %**	±0.1 %	**93.2 %**	±0.3 %	**95.3 %**	±0.6 %	**94.2 %**	±0.1 %

TABLE III: Segmentation performance when trained using only binary masks (excluding raw HU values from the CT images).

Model	Dice score		F_1-Score	
U-Net-small, (only) HU ranges	94.2 %	±0.3 %	93.8 %	±0.4 %
U-Net-base, (only) HU ranges	94.4 %	±0.3 %	94.0 %	±0.2 %
U-Net-large, (only) HU ranges	94.0 %	±0.4 %	93.6 %	±0.4 %

segment it accurately [23], [24] because its regions are smaller and more spread out than other tissues. Thus, the incorporation of HU ranges not only enhances overall segmentation performance but also significantly improves the detection of IMAT, a clinically relevant yet difficult-to-identify tissue type.

VI. ABLATION STUDIES

A. Removing raw HU values

To evaluate the sufficiency of the derived features $\mathbf{X}_c$ alone, we conduct experiments excluding the original HU values $\mathbf{X}_0$ (the original CT image) from the input. This modification reduces the available input information, as the binary masks are *derived from* the HU values but are *not invertible*. However, as shown in Table III, the performance remains high across all model sizes, even surpassing the results obtained with raw HU values alone (Table II). This confirms that the inductive bias introduced by our "feature engineering" provides a more effective representation of the information needed for semantic segmentation of CT images with respect to BCA.

B. Removing individual binary masks during inference

To further evaluate the contribution of individual binary masks, we measure the performance drop when artificially removing a single binary mask *during inference* by setting all locations to zero. Table IV presents the results using the final checkpoint of our *U-Net-base, HU + HU ranges* model.

The middle column shows the drop in performance for a specific tissue type if its binary mask is removed. The segmentation performance for SM and VAT drops substantially to 0 % when we remove their corresponding binary mask at inference, suggesting that the trained model relies heavily on the binary mask to segment these tissue types. Since IMAT and SAT share identical thresholds (Table I), we remove their (identical) binary masks together. Removing those masks

results in a notable decrease in segmentation performance, particularly for IMAT, where the score drops by over 70 %. Overall, the results in the middle column suggest that the model utilizes a tissue's binary mask to segment the specific tissue type.

The right column shows the drop in performance for the *remaining* tissue classes (all tissue classes whose masks we did not remove). Notably, this score is also affected, dropping 24 % if we remove the SM binary mask and around 8 % to 9 % if removing other masks. While we suspect that the binary masks help the model to better detect other tissue types as well, this overall performance drop on other tissue types could also be due to having unrealistic input, as the binary masks don't match CT input image, and thus are out of distribution.

TABLE IV: Effect of removing individual binary masks during inference. Numbers (left → right) show the Dice score drop from unchanged binary mask (left) to removed binary mask (right). The first column shows which tissue's binary mask we remove. The second column shows the Dice score drop for the tissue classes whose mask we remove. The third column shows the Dice score drop for the *remaining* tissue classes (all tissue classes whose masks we did not remove). Last row: As IMAT and SAT share the same HU range, they are removed together, with the first tissue Dice score for IMAT and the second for SAT.

Removed mask	Tissue Dice score	Other Dice score
SM	95.1 % → 0.0 %	93.4 % → 69.3 %
VAT	94.1 % → 0.0 %	93.4 % → 84.8 %
IMAT and SAT	83.2 % → 10.6 % 98.0 % → 68.5 %	94.8 % → 86.5 %

VII. DISCUSSION AND CONCLUSIONS

We introduced a new method for semantic segmentation of CT images for BCA by adding binary masks derived from medical textbook knowledge of HU ranges per tissue class to each input image as additional channels. Considering this physiology-informed prior knowledge as an inductive bias, our model achieves enhanced segmentation performance with drastically reduced labeled training data requirements: We show a consistent improvement across dataset and model sizes

and, in our case, that our model surpasses the performance of the baseline model using only 50 % of the data.

Compared to a related method that uses multiple HU windows [38], our method achieves superior results, especially in the data-limited setting (see Appendix A for details).

The standardized nature of HU ranges ensures broad applicability of our approach across diverse clinical settings, facilitating CT segmentation in resource-limited environments. This has direct implications for clinical practice in tasks like BCA, aiding in identifying conditions such as sarcopenia and obesity.

Despite these advantages, HU-based binary masks provide a coarse tissue approximation and may introduce ambiguities where HU ranges overlap, such as between muscle and different types of adipose tissue. Addressing these limitations in future work with adaptive thresholding or probabilistic priors could refine performance. Additionally, incorporating additional anatomical constraints or shape priors may further enhance segmentation robustness. Furthermore, the proposed approach should be evaluated on different datasets and tasks to demonstrate its broad generalizability.

In conclusion, our results highlight the efficacy of leveraging physiology-informed inductive bias in medical image segmentation, demonstrating a broadly applicable approach that optimizes data efficiency and facilitates clinical applicability.

REFERENCES

[1] J. Schmidhuber, "Deep learning in neural networks: An overview," *Neural Networks*, vol. 61, pp. 85–117, Jan. 2015.

[2] T. Stadelmann, V. Tolkachev, B. Sick, J. Stampfli, and O. Dürr, "Beyond imagenet: Deep learning in industrial practice," in *Applied Data Science*, M. Braschler, T. Stadelmann, and K. Stockinger, Eds. Cham: Springer International Publishing, 2019, pp. 205–232.

[3] R. Emberger, J. M. Boss, D. Baumann, M. Seric, S. Huo *et al.*, "Video object detection for privacy-preserving patient monitoring in intensive care," in *Proceedings of the 10th Swiss Conference on Data Science (SDS)*. Zurich, Switzerland: IEEE, Jun. 2023, pp. 85–88.

[4] M. Amirian, J. A. Montoya-Zegarra, I. Herzig, P. Eggenberger Hotz, L. Lichtensteiger *et al.*, "Mitigation of motion-induced artifacts in cone beam computed tomography using deep convolutional neural networks," *Medical Physics*, vol. 50, no. 10, pp. 6228–6242, Oct. 2023.

[5] U. Kose, O. Deperlioglu, J. Alzubi, and B. Patrut, *Deep Learning for Medical Decision Support Systems*, ser. Studies in Computational Intelligence. Singapore: Springer, 2021, vol. 909.

[6] R. Kabir, H. Z. Syed, D. Vinnakota, M. Sivasubramanian, G. Hitch *et al.*, "Deep learning for clinical decision-making and improved healthcare outcome," in *Deep Learning in Personalized Healthcare and Decision Support*. Elsevier, 2023, pp. 187–201.

[7] P. R. Jermain, M. Oswald, T. Langdun, S. Wright, A. Khan *et al.*, "Deep learning-based cell segmentation for rapid optical cytopathology of thyroid cancer," *Scientific Reports*, vol. 14, no. 1, p. 16389, Jul. 2024.

[8] S. Ali and J. M. Garcia, "Sarcopenia, cachexia and aging: Diagnosis, mechanisms and therapeutic options - a mini-review," *Gerontology*, vol. 60, no. 4, pp. 294–305, 2014.

[9] A. Tolonen, T. Pakarinen, A. Sassi, J. Kyttä, W. Cancino *et al.*, "Methodology, clinical applications, and future directions of body composition analysis using computed tomography (ct) images: A review," *European Journal of Radiology*, vol. 145, p. 109943, Dec. 2021.

[10] J. Wasserthal, H.-C. Breit, M. T. Meyer, M. Pradella, D. Hinck *et al.*, "Totalsegmentator: Robust segmentation of 104 anatomic structures in ct images," *Radiology: Artificial Intelligence*, vol. 5, no. 5, p. e230024, 2023.

[11] D. Makrakis, K. Rounis, A.-P. Tsigkas, A. Georgiou, N. Galanakis *et al.*, "Effect of body tissue composition on the outcome of patients with metastatic non-small cell lung cancer treated with pd-1/pd-l1 inhibitors," *PLOS ONE*, vol. 18, no. 2, p. e0277708, 2023.

[12] M. V. Dietz, K. Popuri, L. Janssen, M. Salehin, D. Ma *et al.*, "Evaluation of a fully automated computed tomography image segmentation method for fast and accurate body composition measurements," *Nutrition*, vol. 129, p. 112592, 2025.

[13] M. Amirian, J. A. Montoya-Zegarra, J. Gruss, Y. D. Stebler, A. S. Bozkir *et al.*, "Prepnet: A convolutional auto-encoder to homogenize ct scans for cross-dataset medical image analysis," in *Proceedings of the 14th International Congress on Image and Signal Processing, BioMedical Engineering and Informatics (CISP-BMEI)*. Shanghai, China: IEEE, Oct. 2021, pp. 1–7.

[14] P. Sager, S. Salzmann, F. Burn, and T. Stadelmann, "Unsupervised domain adaptation for vertebrae detection and identification in 3d ct volumes using a domain sanity loss," *Journal of Imaging*, vol. 8, no. 8, p. 222, Aug. 2022.

[15] T. Stadelmann, T. Klamt, and P. H. Merkt, "Data centrism and the core of data science as a scientific discipline," *Archives of Data Science*, vol. 8, no. 2, 2022.

[16] P.-P. Luley, J. M. Deriu, P. Yan, G. A. Schatte, and T. Stadelmann, "From concept to implementation: The data-centric development process for ai in industry," in *Proceedings of the 10th Swiss Conference on Data Science (SDS)*. Zurich, Switzerland: IEEE, Jun. 2023, pp. 73–76.

[17] C. Beaudart, E. McCloskey, O. Bruyère, M. Cesari, Y. Rolland *et al.*, "Sarcopenia in daily practice: assessment and management," *BMC Geriatrics*, vol. 16, no. 1, p. 170, Dec. 2016.

[18] L. Mueller, N. Mentil, N. Staub, S. Griot, T. Olpe *et al.*, "Association of thoracic skeletal muscle index with clinical outcome and response to nutritional interventions in patients at risk of malnutrition–secondary analysis of a randomized trial," *Nutrients*, vol. 15, no. 4, p. 817, 2023.

[19] B. Amini, S. P. Boyle, R. D. Boutin, and L. Lenchik, "Approaches to assessment of muscle mass and myosteatosis on computed tomography: A systematic review," *The Journals of Gerontology: Series A*, vol. 74, no. 10, pp. 1671–1678, Sep. 2019.

[20] B. A. Derstine, S. A. Holcombe, B. E. Ross, N. C. Wang, G. L. Su *et al.*, "Skeletal muscle cutoff values for sarcopenia diagnosis using t10 to l5 measurements in a healthy us population," *Scientific Reports*, vol. 8, no. 1, p. 11369, Jul. 2018.

[21] Y. H. Cho, K.-H. Do, E. J. Chae, S. H. Choi, K.-W. Jo *et al.*, "Association of chest ct-based quantitative measures of muscle and fat with post-lung transplant survival and morbidity: a single institutional retrospective cohort study in korean population," *Korean journal of radiology*, vol. 20, no. 3, pp. 522–530, 2019.

[22] W. Shen, M. Punyanitya, Z. Wang, D. Gallagher, M.-P. St.-Onge *et al.*, "Total body skeletal muscle and adipose tissue volumes: estimation from a single abdominal cross-sectional image," *Journal of Applied Physiology*, vol. 97, no. 6, pp. 2333–2338, Dec. 2004.

[23] M. T. Paris, P. Tandon, D. K. Heyland, H. Furberg, T. Premji *et al.*, "Automated body composition analysis of clinically acquired computed tomography scans using neural networks," *Clinical Nutrition*, vol. 39, no. 10, pp. 3049–3055, Oct. 2020.

[24] R. Hemke, C. G. Buckless, A. Tsao, B. Wang, and M. Torriani, "Deep learning for automated segmentation of pelvic muscles, fat, and bone from ct studies for body composition assessment," *Skeletal Radiology*, vol. 49, no. 3, pp. 387–395, Mar. 2020.

[25] S. Koitka, L. Kroll, E. Malamutmann, A. Oezcelik, and F. Nensa, "Fully automated body composition analysis in routine ct imaging using 3d semantic segmentation convolutional neural networks," *European Radiology*, vol. 31, no. 4, pp. 1795–1804, Apr. 2021.

[26] S. Dabiri, K. Popuri, E. M. Cespedes Feliciano, B. J. Caan, V. E. Baracos *et al.*, "Muscle segmentation in axial computed tomography (ct) images at the lumbar (l3) and thoracic (t4) levels for body composition analysis," *Computerized Medical Imaging and Graphics*, vol. 75, pp. 47–55, Jul. 2019.

[27] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *Medical Image Computing and Computer-Assisted Intervention*, ser. MICCAI, N. Navab, J. Hornegger, W. M. Wells, and A. F. Frangi, Eds. Cham: Springer, 2015, vol. 9351, pp. 234–241.

[28] N. Simmler, P. Sager, P. Andermatt, R. Chavarriaga, F.-P. Schilling *et al.*, "A survey of un-, weakly-, and semi-supervised learning methods for noisy, missing and partial labels in industrial vision applications," in *Proceedings of the 8th Swiss Conference on Data Science (SDS)*. Lucerne, Switzerland: IEEE, Jun. 2021, pp. 26–31.

[29] M. Gardner, Y. B. Bouchta, A. Mylonas, M. Mueller, C. Cheng *et al.*, "Realistic ct data augmentation for accurate deep-learning based

segmentation of head and neck tumors in kv images acquired during radiation therapy," *Medical Physics*, vol. 50, no. 7, pp. 4206–4219, Jul. 2023.

[30] E. Goceri, "Medical image data augmentation: techniques, comparisons and interpretations," *Artificial Intelligence Review*, vol. 56, no. 11, pp. 12 561–12 605, Nov. 2023.

[31] L. Tuggener, R. Emberger, A. Ghosh, P. Sager, Y. P. Satyawan *et al.*, "Real world music object recognition," *Transactions of the International Society for Music Information Retrieval*, vol. 7, no. 1, pp. 1–14, Jan. 2024.

[32] K. Y. Lim, J. E. Ko, Y. N. Hwang, S. G. Lee, and S. M. Kim, "Transraunet: A deep neural network with reverse attention module using hu windowing augmentation for robust liver vessel segmentation in full resolution of ct images," *Diagnostics*, vol. 15, no. 2, p. 118, 2025.

[33] H. Guan and M. Liu, "Domain adaptation for medical image analysis: A survey," *IEEE Transactions on Biomedical Engineering*, vol. 69, no. 3, pp. 1173–1185, Mar. 2022.

[34] D. Karimi, S. K. Warfield, and A. Gholipour, "Transfer learning in medical image segmentation: New insights from analysis of the dynamics of model parameters and learned representations," *Artificial Intelligence in Medicine*, vol. 116, p. 102078, Jun. 2021.

[35] P. Yan, A. Abdulkadir, P.-P. Luley, M. Rosenthal, G. A. Schatte *et al.*, "A comprehensive survey of deep transfer learning for anomaly detection in industrial time series: Methods, applications, and directions," *IEEE Access*, vol. 12, pp. 3768–3789, 2024.

[36] B. Felfeliyan, N. D. Forkert, A. Hareendranathan, D. Cornel, Y. Zhou *et al.*, "Self-supervised-rcnn for medical image segmentation with limited data annotation," *Computerized Medical Imaging and Graphics*, vol. 109, p. 102297, Oct. 2023.

[37] S. Cuomo, V. S. Di Cola, F. Giampaolo, G. Rozza, M. Raissi *et al.*, "Scientific machine learning through physics–informed neural networks: Where we are and what's next," *Journal of Scientific Computing*, vol. 92, no. 3, p. 88, Sep. 2022.

[38] M. Kloenne, S. Niehaus, L. Lampe, A. Merola, J. Reinelt *et al.*, "Domain-specific cues improve robustness of deep learning-based segmentation of ct volumes," *Scientific Reports*, vol. 10, no. 1, p. 10712, Jul. 2020.

[39] D. Steybe, P. Poxleitner, M. C. Metzger, L. S. Brandenburg, R. Schmelzeisen *et al.*, "Automated segmentation of head ct scans for computer-assisted craniomaxillofacial surgery applying a hierarchical patch-based stack of convolutional neural networks," *International journal of computer assisted radiology and surgery*, vol. 17, no. 11, pp. 2093–2101, 2022.

[40] R. Bibb, D. Eggbeer, and A. Paterson, *Medical Modelling: The Application of Advanced Design and Rapid Prototyping Techniques in Medicine*. Woodhead Publishing, 2014, ch. 2, pp. 7–34.

[41] A. Baumgartner, T. Olpe, S. Griot, N. Mentil, N. Staub *et al.*, "Association of ct-based diagnosis of sarcopenia with prognosis and treatment response in patients at risk of malnutrition–a secondary analysis of the effect of early nutritional support on frailty, functional outcomes, and recovery of malnourished medical inpatients trial (effort) trial," *Clinical Nutrition*, vol. 42, no. 2, pp. 199–207, 2023.

[42] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," in *Proceedings of the 3rd International Conference on Learning Representations*, ser. ICLR, May 2015.

[43] J. Ha, T. Park, H.-K. Kim, Y. Shin, Y. Ko *et al.*, "Development of a fully automatic deep learning system for l3 selection and body composition assessment on computed tomography," *Scientific Reports*, vol. 11, no. 1, p. 21656, Nov. 2021.

[44] J. P. Lim, M. S. Chong, L. Tay, Y. X. Yang, B. P. Leung *et al.*, "Intermuscular adipose tissue is associated with adipose tissue inflammation and poorer functional performance in central adiposity," *Archives of Gerontology and Geriatrics*, vol. 81, pp. 1–7, Mar. 2019.

[45] I. G. P. Suka Aryana, I. B. Paulus, , S. Kalra, , D. Daniella, , R. A. T. Kuswardhani *et al.*, "The important role of intermuscular adipose tissue on metabolic changes interconnecting obesity, ageing and exercise: A systematic review," *European Endocrinology*, vol. 19, no. 1, p. 54, 2023.

APPENDIX

HU windowing is an alternative pre-processing technique proposed by [38], which, like our approach, incorporates the HU ranges as an inductive bias. Instead of utilizing binary masks, HU windowing creates multiple HU windows by clipping the HU values to each HU range and normalizes the values between 0 and 1. For a given tissue class c, let L_c and U_c denote the lower and upper bounds of the HU range, respectively. The transformed input representation is then defined as:

$$
X_c(i,j) = \begin{cases} \frac{L_c}{|U_c - L_c|}, & \text{if } X_0(i,j) < L_c, \\ \frac{U_c}{|U_c - L_c|}, & \text{if } X_0(i,j) > U_c, \\ \frac{X_0(i,j)}{|U_c - L_c|}, & \text{otherwise.} \end{cases} \tag{1}
$$

The results using HU windowing in Table V indicate that while our results are slightly superior, both inductive bias strategies yield comparable performance across different model sizes. However, binary masks have a stronger inductive bias as HU values are *reduced* to binary values, leading to fewer possible inputs. Consequently, models trained with binary masks should learn simpler yet more robust patterns independent of the specific HU values. Fig. 4 indicates that our approach works better in *data-limited* circumstances, but further investigation is required to determine whether specific segmentation tasks or dataset characteristics favor one encoding method over the other.

TABLE V: Results of HU windowing on our dataset.

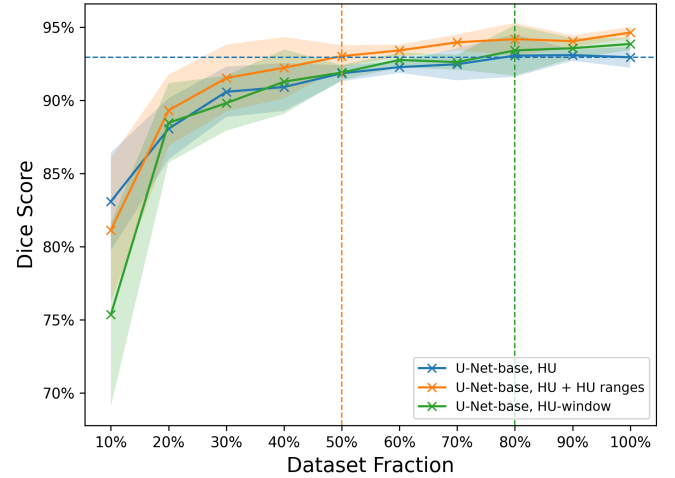| Model	Dice score
U-Net-small, HU windows	93.7 % ±0.5 %
U-Net-base, HU windows	94.1 % ±0.4 %
U-Net-large, HU windows	94.0 % ±0.2 %

Fig. 4: Fig. 2 with HU windowing experiments added. The vertical green line indicates that HU windowing surpasses the baseline performance with 80 % of the dataset.