PRISM-3D: Periskeletal Region–aware Imaging with Segmentation-guided Modeling Using 3D Deep Learning for NSCLC Survival Prediction

Tianxi Liang
Wake Technical Community College
Raleigh, North Carolina, USA
tliang2@my.waketech.edu

Sai Maruvada
Wake Technical Community College
Raleigh, North Carolina, USA
shmaruvada@my.waketech.edu

Nishanth Sathisha
Wake Technical Community College
Raleigh, North Carolina, USA
nsathisha@my.waketech.edu

***Abstract*—Non–small cell lung cancer (NSCLC) exhibits substantial survival heterogeneity even among patients with similar clinical staging, limiting the effectiveness of tumor-centric prognostic models. We propose PRISM-3D, a periskeletal region–aware, segmentation-guided 3D deep learning framework for NSCLC survival prediction that explicitly incorporates patient-level anatomical context. Using 422 chest CT scans from the Lung1 cohort of The Cancer Imaging Archive, periskeletal anatomical regions were manually annotated and used to train an nnU-Net segmentation model achieving a Dice score of 0.946. The resulting masks were integrated as an explicit spatial prior for a 3D ResNet-18 classifier trained to predict two-year survival. On a held-out internal validation cohort of 122 patients, PRISM-3D achieved an AUC of approximately 0.72, matching or exceeding reported performance of prior radiomics- and deep learning–based LUNG1 benchmarks. Grad-CAM analysis demonstrates that model predictions are driven by periskeletal anatomical regions rather than intrapulmonary features alone, indicating that periskeletal context provides complementary and interpretable prognostic information beyond tumor-centric imaging representations.**

***Index Terms*—Non–small cell lung cancer, survival prediction, periskeletal anatomy, segmentation-guided learning, 3D convolutional neural networks, nnU-Net, model interpretability**

I. Introduction

Non–small cell lung cancer (NSCLC) accounts for approximately 85% of lung cancer diagnoses and remains one of the leading causes of cancer-related mortality worldwide [1], [2]. In the United States alone, more than 230,000 new cases are diagnosed annually, with five-year survival rates varying by over 50% across patients with similar clinical stage and treatment regimens [2], [3]. This pronounced heterogeneity highlights the limitations of current prognostic stratification strategies and motivates the development of improved imaging-based survival prediction models.

Computed tomography (CT) imaging plays a central role in NSCLC diagnosis, staging, and treatment planning [4]. Beyond its clinical utility, CT has increasingly been leveraged for outcome prediction through radiomics and deep learning approaches. Radiomics-based methods extract handcrafted intensity, texture, and shape features from segmented tumor regions and apply classical machine learning or survival models [5], [6]. More recent studies employ end-to-end convolutional neural networks (CNNs) to directly learn prognostic representations from volumetric CT data [7]. Despite these advances, most existing approaches remain tumor-centric, restricting model input to intratumoral or peritumoral regions and implicitly assuming that tumor appearance alone dominates patient prognosis [7], [8].

However, growing clinical evidence indicates that patient-level systemic health plays a critical role in cancer outcomes. In particular, musculoskeletal condition—including muscle mass, composition, and quality—has been shown to correlate strongly with frailty, treatment tolerance, and overall survival in NSCLC patients [9]. Measures of sarcopenia and body composition derived from CT imaging provide prognostic information independent of tumor characteristics [10], [11]. These indicators are visually encoded within periskeletal anatomical regions surrounding the lung, such as paraspinal and thoracic musculature, yet remain underexplored in modern deep learning–based survival modeling.

We posit that this omission represents a critical gap in existing imaging-based prognostic pipelines. While attention mechanisms and implicit feature learning may capture contextual information, they lack explicit anatomical grounding and offer limited interpretability [12]. In contrast, explicitly modeling periskeletal regions introduces a biologically motivated spatial prior that enables deep networks to integrate patient-level physiological context alongside tumor-related cues in a transparent and anatomically meaningful manner [13].

In this work, we propose PRISM-3D, a periskeletal region–aware, segmentation-guided 3D deep learning framework for NSCLC survival prediction. Using 422 chest CT scans from the Lung1 cohort of The Cancer Imaging Archive, we demonstrate that PRISM-3D achieves competitive survival prediction performance while providing interpretable evidence—via Grad-CAM visualization—that model attention aligns with periskeletal anatomical structures. By moving beyond tumor-centric analysis, PRISM-3D highlights the prognostic value of patient-level anatomical context in imaging-based oncology models.

II. Related Work

Imaging-based survival prediction for NSCLC has primarily followed two methodological paradigms: radiomics-driven modeling and end-to-end deep learning. Radiomics approaches extract handcrafted intensity, texture, and shape features from segmented tumor regions and apply classical machine learning or survival analysis techniques, such as Cox proportional hazards models [5], [6]. While radiomics methods offer some interpretability, they are highly sensitive to segmentation variability, imaging protocols, and feature selection, often limiting generalization across cohorts [14].

More recent work has shifted toward deep learning–based prognostic modeling, leveraging 3D convolutional neural networks (CNNs) to learn features directly from volumetric CT data [7]. These approaches reduce reliance on handcrafted features and have demonstrated improved predictive performance; however, they largely remain tumor-centric and depend on implicit feature learning or attention mechanisms, making anatomical interpretation challenging.

Segmentation-guided learning has been explored as a means of introducing anatomical priors into deep networks, particularly for lesion classification and outcome prediction [13]. By constraining model input to predefined regions of interest, such approaches can reduce background noise and improve robustness. Nevertheless, existing methods predominantly focus on tumor or organ-level regions and do not explicitly model patient-level anatomical context related to systemic health.

In parallel, body composition and sarcopenia studies have established strong associations between skeletal muscle characteristics and cancer outcomes, including treatment tolerance and overall survival [9], [10]. These analyses typically rely on manual or semi-automated measurements and are not integrated into modern 3D deep learning pipelines.

In contrast to prior work, PRISM-3D explicitly incorporates periskeletal anatomical regions—which encode patient-level physiological condition—into a segmentation-guided 3D deep learning framework for NSCLC survival prediction. To our knowledge, this represents the first integration of high-fidelity periskeletal segmentation with interpretable 3D CNN–based survival modeling in this context.

On the public LUNG1 cohort, Braghetto *et al.* benchmarked radiomics and deep learning pipelines for 2-year overall survival prediction and reported best test AUC values of 0.63–0.67, with a clinical-only baseline of 0.59 [8]. Beyond LUNG1-specific studies, deep learning prognosis models trained on multi-institutional CT cohorts have achieved AUCs around 0.70–0.71 [15], and more recent approaches have reported AUC up to 0.78 under external-test evaluations [16]. These results motivate anatomically informed priors that can improve robustness and interpretability without relying solely on tumor appearance.

III. Materials and Methods

We summarize the end-to-end workflow in Fig. 1. The following subsections describe the dataset, periskeletal segmentation, segmentation-guided input construction, survival classification, and interpretability procedure.

TABLE I
Patient demographics and clinical characteristics for the Lung1 cohort.

Characteristic	**All**
Patients, n	422
Age, mean $\pm$ SD	68.0 ± 10.1
Age, median (IQR)	69 (61–76)
Age missing, n	22
Sex, n (%)	
Male	290 (68.7)
Female	132 (31.3)
Sex missing, n	0
Overall stage, n (%)	
I	93 (22.1)
II	40 (9.5)
IIIa	112 (26.6)
IIIb	176 (41.8)
Stage missing, n	1
Survival time (days), median (IQR)	546 (261–1397)
Events, n (%)	373 (88.4)
Censored, n (%)	49 (11.6)

A. Dataset

This study utilizes the NSCLC-Radiomics (Lung1) dataset obtained from The Cancer Imaging Archive (TCIA) [5], [17], which comprises chest computed tomography (CT) scans with associated clinical and survival metadata. The cohort includes 422 patients diagnosed with non–small cell lung cancer and represents a diverse distribution of disease stages and histological subtypes. Survival outcomes were formulated as a binary classification task using a two-year survival threshold, consistent with prior imaging-based prognostic studies [7], [8]. A summary of patient demographics and clinical characteristics is provided in Table I.

To ensure unbiased evaluation and prevent information leakage, the dataset was partitioned into a training cohort of 300 cases and a held-out internal validation cohort of 122 cases. All preprocessing steps, model training procedures, and hyperparameter selection were conducted exclusively on the training cohort.

B. Periskeletal Region Segmentation

To capture patient-level physiological information beyond intratumoral appearance, periskeletal anatomical regions surrounding the lung were defined and manually curated as ground-truth annotations. These regions encompass the chest wall musculature and adjacent skeletal structures hypothesized to reflect systemic health, frailty, and treatment tolerance relevant to cancer prognosis.

Manual annotations were performed using ITK-SNAP, an interactive medical image segmentation and visualization tool widely adopted in clinical research [18]. A representative example of the annotation process is shown in Fig. 2. In cases

where anatomical boundaries were ambiguous, annotators consulted board-certified clinicians to ensure anatomical consistency. These annotations served as supervision for training an automated segmentation model.

Segmentation was performed using the nnU-Net framework, a self-configuring deep learning pipeline designed for biomedical image segmentation [19]. nnU-Net automatically adapts network architecture, preprocessing strategies, data augmentation policies, and training schedules to dataset-specific characteristics, thereby reducing manual design bias and improving reproducibility.

Model optimization employed the Dice similarity coefficient (DSC), defined as

$$\mathrm{DSC} = \frac{2|P \cap G|}{|P| + |G|} \tag{1}$$

where P and G denote the predicted and ground-truth segmentation masks, respectively. Quantitative segmentation performance and qualitative visualizations are reported in the Results section.

C. Segmentation-guided Input Construction

Following segmentation, periskeletal masks were incorporated directly into the survival prediction pipeline to provide explicit anatomical guidance. For each patient, the CT volume was paired with its corresponding binary periskeletal segmentation mask, forming a multi-channel volumetric input to the classification network.

This segmentation-guided formulation introduces a spatial inductive bias that constrains the model to attend to anatomically meaningful regions while suppressing irrelevant background information. Unlike implicit attention mechanisms, this explicit guidance enforces anatomically grounded feature learning and facilitates downstream interpretability.

D. 3D ResNet-18 Survival Classification Network

Survival prediction was performed using a three-dimensional ResNet-18 architecture adapted for volumetric medical imaging [20]. Residual networks leverage identity skip connections to stabilize gradient propagation and enable efficient optimization of deep architectures. Each residual block learns a residual mapping of the form

$$\mathbf{y} = \mathcal{F}(\mathbf{x}, \mathbf{W}) + \mathbf{x} \tag{2}$$

where $\mathbf{x}$ denotes the input feature map and $\mathcal{F}$ represents the residual function parameterized by weights $\mathbf{W}$.

The network was configured for binary survival classification using a sigmoid-activated output layer and trained with binary cross-entropy loss.

E. Model Interpretability via Grad-CAM

To assess whether the proposed model learns anatomically meaningful representations, interpretability was evaluated using Gradient-weighted Class Activation Mapping (Grad-CAM), a post hoc visualization technique for convolutional neural networks [21]. Grad-CAM computes importance weights for feature maps in a target convolutional layer as

$$\alpha_k = \frac{1}{Z} \sum_{i,j,l} \frac{\partial y}{\partial A_k^{ijl}} \tag{3}$$

where y denotes the model output and Z is a normalization constant.

The resulting class activation map is computed as

$$L_{\text{Grad-CAM}} = \mathrm{ReLU}\left(\sum_k \alpha_k A_k \right) \tag{4}$$

Grad-CAM visualizations are used to qualitatively evaluate whether model attention aligns with periskeletal anatomical regions and are reported in the Results section.

IV. Results

A. Periskeletal Segmentation Performance

The nnU-Net model achieved high segmentation performance on the Lung1 cohort. Across all 422 cases, the model obtained a Dice similarity coefficient (DSC) of 0.946 and a mean Intersection-over-Union (mIoU) of 0.897. Precision and recall were 0.944 and 0.948, respectively. The false positive rate was 0.0008, and overall voxel-wise accuracy was 0.998.

Figure 3 summarizes the optimization behavior of the nnU-Net model. Training and validation losses decreased rapidly during the initial epochs and reached a stable plateau thereafter. Dice scores increased consistently throughout training and converged above 0.94, with reduced variance in later epochs, reflecting stable segmentation behavior across training iterations.

Figure 4 provides a representative qualitative comparison between manually annotated ground-truth periskeletal masks and model predictions. The predicted segmentation closely aligns with the ground truth across the chest wall musculature and adjacent skeletal regions, with minor discrepancies observed primarily along anatomical boundaries.

B. Survival Classification Performance

The segmentation-guided 3D ResNet-18 model was evaluated on a held-out internal validation cohort of 122 patients. The model achieved an area under the receiver operating characteristic curve (AUC) of 0.72 at epoch 72. At a decision threshold of 0.5, the model yielded a recall of 0.918 and a precision of 0.421.

Training performance at epoch 72 included a training AUC of 0.72 and an accuracy of 0.647. Figure 5 illustrates training and internal validation performance across epochs, including loss trajectories and AUC trends.

C. Model Interpretability and Attention Analysis

Gradient-weighted Class Activation Mapping (Grad-CAM) was applied to the trained segmentation-guided 3D ResNet-18 model to visualize spatial regions contributing to two-year survival predictions. Grad-CAM attribution maps were

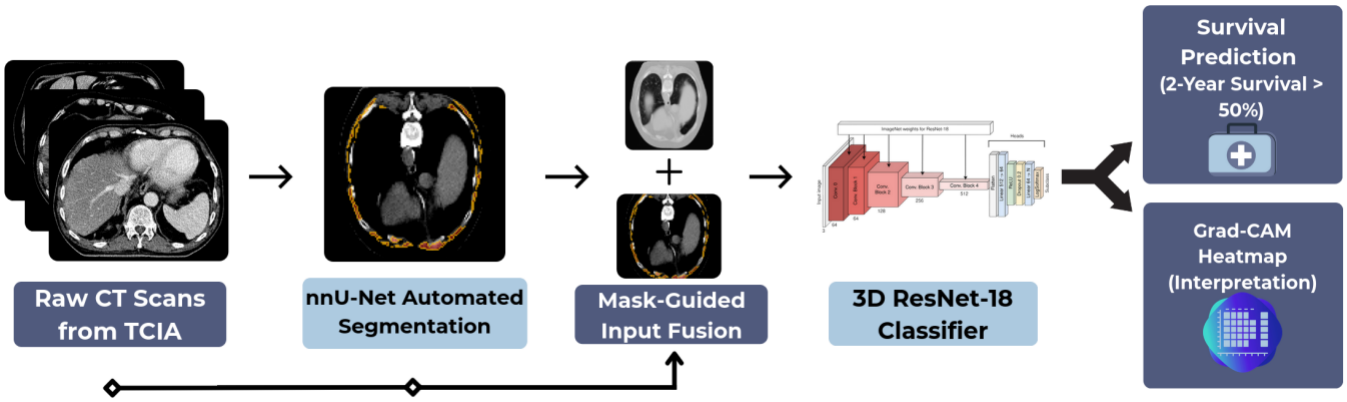

Fig. 1. Overview of the PRISM-3D pipeline. Raw chest CT volumes from TCIA are first processed by an nnU-Net model to obtain periskeletal segmentations. The CT volume and the corresponding binary mask are concatenated as a multi-channel input to a 3D ResNet-18 classifier for two-year survival prediction. Grad-CAM is applied to visualize anatomically grounded model attention.

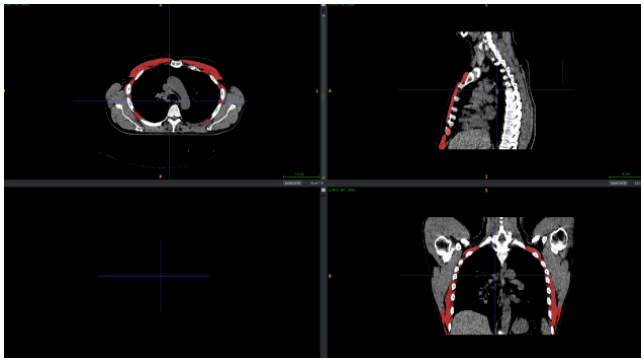

Fig. 2. Manual annotation of periskeletal anatomical regions using ITK-SNAP for Case 47 (slice 226). Shown are synchronized axial (top left), sagittal (top right), and coronal (bottom right) CT views with manually delineated periskeletal muscle regions overlaid in red. Annotations encompass the chest wall musculature and adjacent skeletal structures surrounding the lung fields, excluding intrapulmonary tissue. These manually curated masks serve as ground-truth labels for training the nnU-Net segmentation model.

Fig. 3. Training dynamics of the nnU-Net periskeletal segmentation model. The plot shows training loss (blue) and validation loss (red) on the left axis, along with Dice scores (green, right axis) computed during training. The dotted curve represents per-iteration Dice, while the solid curve shows its moving average.

generated from the final convolutional layer and overlaid on corresponding CT slices for qualitative inspection.

Figure 6 shows a representative Grad-CAM heatmap for a case from the internal validation cohort. High-activation regions are primarily localized to periskeletal anatomical structures, including chest wall musculature and adjacent skeletal boundaries. In contrast, comparatively lower activation is observed within intrapulmonary regions and surrounding background tissue.

Across examined validation cases, Grad-CAM visualizations exhibited consistent spatial localization patterns, with model attention concentrated within anatomically defined periskeletal regions provided by the segmentation-guided input. No systematic concentration of activation within intrapulmonary regions was observed in these qualitative assessments.

These results indicate that the segmentation-guided formulation directs model attention toward periskeletal anatomical regions during survival prediction, as reflected by the spatial

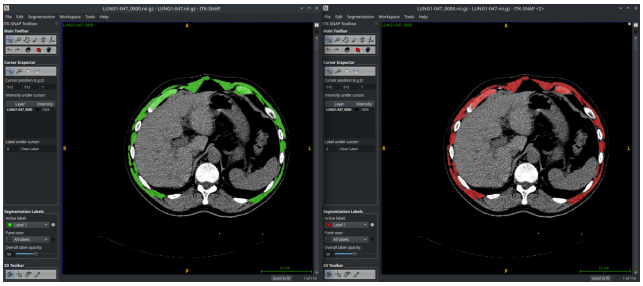

Fig. 4. Qualitative comparison of nnU-Net periskeletal segmentation results for Case 47 (slice 1). The left panel shows the manually annotated ground-truth periskeletal mask (green), while the right panel shows the nnU-Net prediction (red), overlaid on the corresponding axial CT slice.

distribution of Grad-CAM activations.

Table II summarizes the performance of PRISM-3D relative to representative NSCLC survival prediction approaches. Under direct internal evaluation on the LUNG1 cohort, PRISM-

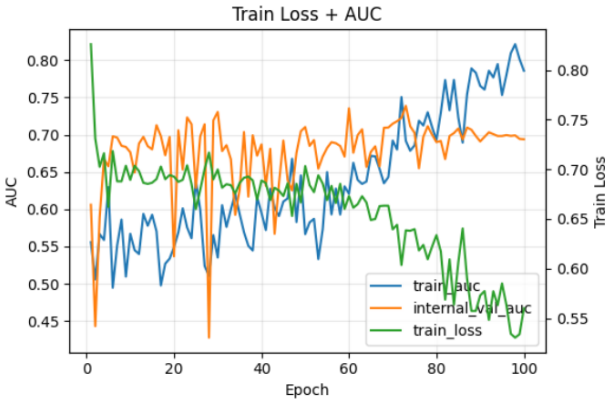

Fig. 5. Training dynamics of the segmentation-guided 3D ResNet-18 survival classification model. The plot shows training loss (green, left axis), training AUC (blue, right axis), and internal validation AUC (orange, right axis) across epochs.

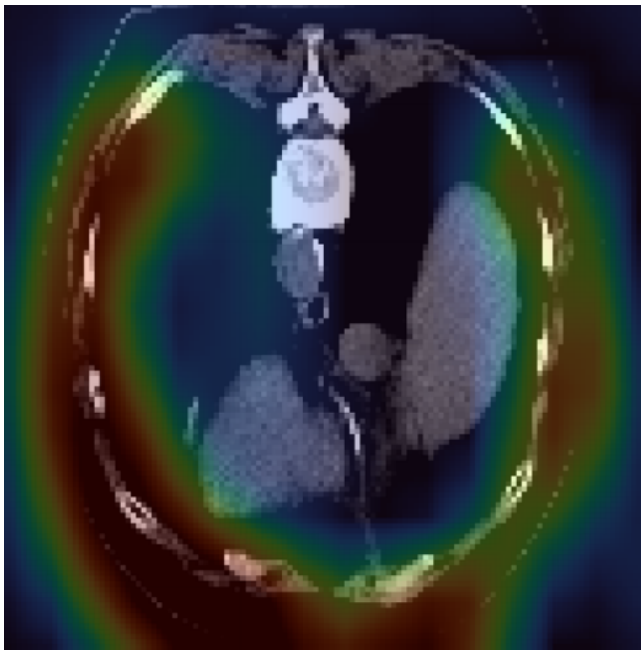

Fig. 6. Grad-CAM visualization for the segmentation-guided 3D ResNet-18 model on Case 5 (slice 25). Warmer colors indicate regions with higher contribution to the predicted two-year survival outcome.

3D achieved an AUC of 0.72 for two-year overall survival prediction. This performance exceeds the reported clinical-only baseline (AUC = 0.59) and is higher than the best-performing tumor-centric deep learning model (AUC = 0.63 $\pm$ 0.03) evaluated under the same internal validation protocol. PRISM-3D also outperformed the strongest radiomics-based and radiomics–deep learning hybrid approaches reported on LUNG1, which achieved AUC values of approximately 0.67.

For contextual reference, Table II additionally reports results from prior deep learning–based prognosis models evaluated on external or multi-institutional cohorts. These studies report AUC values in the range of approximately 0.70–0.71, with some approaches achieving higher performance under external-test settings. While these results are not directly comparable due to differences in cohort composition, data harmonization, and evaluation protocols, they provide a reference range for state-of-the-art performance in NSCLC survival prediction.

TABLE II
PERFORMANCE COMPARISON FOR 2-YEAR SURVIVAL PREDICTION.

Method	**Evaluation**	**AUC**
Direct internal evaluation		
PRISM-3D (ours)	internal	0.72
Clinical-only baseline [8]	internal	0.59
Radiomics best [8]	internal	0.67 $\pm$ 0.03
Deep learning best [8]	internal	0.63 $\pm$ 0.03
Radiomics + deep best [8]	internal	0.67 $\pm$ 0.02
External Context		
Hosny *et al.* [15]	external cohort	$\sim$0.70–0.71
Khodabakhshi *et al.* [16]	external test	up to 0.78

V. DISCUSSION

This work presents PRISM-3D, a periskeletal region–aware, segmentation-guided 3D deep learning framework for two-year overall survival prediction in NSCLC from routine chest CT. In contrast to tumor-centric pipelines that primarily encode intratumoral and peritumoral appearance, PRISM-3D introduces an explicit anatomical prior derived from periskeletal structures surrounding the thoracic cavity. On the public Lung1 cohort, PRISM-3D achieved competitive discriminative performance while providing anatomically grounded attribution patterns via Grad-CAM. Together, these findings support the feasibility of leveraging structured, non-tumor anatomical context for imaging-based prognosis and motivate broader investigation into anatomy-aware survival modeling.

A. Periskeletal Context as a Prognostic Signal

A central contribution of PRISM-3D is the operationalization of "periskeletal context" as a well-defined, learnable imaging representation. The nnU-Net segmentation model achieved high agreement with manual annotations (Dice = 0.946), indicating that the proposed periskeletal region definition is both consistent and automatable at cohort scale. This is important because segmentation-guided frameworks are only as reliable as the anatomical masks that shape model input; in PRISM-3D, the strong segmentation fidelity reduces the likelihood that downstream performance is driven by unstable or noisy region definitions.

The survival prediction results suggest that restricting and guiding model input toward periskeletal anatomy can yield meaningful prognostic discrimination. Notably, PRISM-3D reached an internal AUC of 0.72 on Lung1, exceeding the reported clinical-only baseline and improving upon tumor-centric deep learning results reported under comparable internal evaluation protocols. These results are consistent with the broader clinical literature linking skeletal muscle condition and body composition to cancer outcomes, but PRISM-3D differs

in how this information is incorporated: rather than measuring a small set of handcrafted sarcopenia markers, the model learns volumetric representations from a spatially constrained, anatomically motivated field-of-view. This design allows the network to utilize multiple periskeletal cues (e.g., musculature distribution, chest wall composition, skeletal boundaries) without requiring manual feature engineering.

B. Interpretability, Anatomical Grounding, and Grad-CAM Evidence

A practical limitation of many prognosis models is that high-level performance does not guarantee clinically plausible decision mechanisms. PRISM-3D partially addresses this issue by combining an explicit segmentation prior with post hoc attribution. In qualitative Grad-CAM analyses, activation consistently localized to periskeletal structures—including chest wall musculature and adjacent skeletal boundaries—rather than concentrating within intrapulmonary tissue. This attention pattern is aligned with the intended modeling objective of PRISM-3D: to prioritize non-tumor anatomical context that is typically excluded or underutilized in tumor-centric pipelines.

Importantly, the interpretability signal is not solely derived from Grad-CAM. The model is explicitly conditioned on periskeletal masks through a multi-channel input formulation, which constrains the learned representation space and provides a direct anatomical handle on what regions are emphasized. In this sense, Grad-CAM serves as complementary evidence that the segmentation-guided formulation is functioning as designed: the regions receiving high attribution correspond to the anatomical prior supplied during training. This coupling between input design (explicit masks) and attribution behavior (localized activations) strengthens transparency relative to approaches that rely exclusively on implicit attention mechanisms or unconstrained full-volume inputs.

C. Operating Point and Sensitivity–Precision Tradeoff

PRISM-3D exhibited a high-recall operating profile on the internal validation cohort, with comparatively lower precision at the default threshold. This behavior may reflect a conservative decision boundary that prioritizes identifying patients at higher risk of not surviving two years. While threshold selection is application-dependent, these results suggest that PRISM-3D may be more suitable for screening-style risk stratification (where missed high-risk cases are costly) than for settings requiring high positive predictive value without further clinical context. In future deployments, calibration procedures and threshold tuning based on clinical utility (e.g., decision-curve analysis or cost-sensitive optimization) could better align model outputs with specific workflow needs.

D. Relativity to Prior Prognostic Modeling

PRISM-3D contributes a distinct perspective to CT-based NSCLC prognosis in three ways. First, it shifts the primary modeling focus from tumor appearance to an explicitly defined periskeletal anatomical envelope, enabling prognosis modeling that is not intrinsically dependent on tumor segmentation quality or intratumoral feature dominance. Second, it introduces a segmentation-guided inductive bias that is anatomically interpretable by construction; the mask channel provides a structured prior that is consistent across patients and can be audited independently of the classifier. Third, PRISM-3D pairs this anatomical prior with qualitative attribution analysis, providing a coherent narrative that connects region definition, model input construction, and observed attention patterns. This combination distinguishes PRISM-3D from prior work that either (i) uses tumor-centric ROIs, (ii) relies on implicit contextual learning without anatomical constraints, or (iii) performs separate body-composition analyses without integration into an end-to-end 3D deep learning pipeline.

E. Comparison with Prior Models and Implications

When placed alongside representative benchmarks, PRISM-3D matches or exceeds tumor-centric radiomics and deep learning baselines reported on Lung1 while remaining comparatively lightweight in architecture (3D ResNet-18) and explicit in anatomical targeting. While some external-test studies report higher AUC values under different cohort conditions, those results are typically obtained with differing data distributions, preprocessing, and evaluation setups, limiting direct comparability. Within the Lung1 internal evaluation context, PRISM-3D narrows the gap between anatomy-aware interpretability and competitive predictive performance. These findings suggest that periskeletal anatomical context may provide complementary information to conventional tumor-centric representations and may be particularly valuable when tumor signals alone are insufficient to explain outcome heterogeneity.

F. Limitations and Future Directions

This study has several limitations that should be considered when interpreting the results. First, evaluation was limited to internal validation on a single public cohort; the generalizability of periskeletal priors across institutions, scanners, and protocols remains to be established. Second, survival was formulated as a binary two-year endpoint rather than a time-to-event outcome; while this aligns with prior Lung1 benchmarks, it does not fully utilize censored survival information. Third, the periskeletal region definition was manually curated for this study, and although segmentation performance was high, annotation guidelines and inter-annotator variability were not formally quantified. Fourth, Grad-CAM provides qualitative attribution but does not by itself prove causal dependence; quantitative overlap metrics between attributions and masks, along with controlled ablations (e.g., mask-only vs CT-only vs combined inputs), would further strengthen mechanistic claims.

Future work will focus on (i) external validation on multi-institutional datasets to assess robustness under domain shift, (ii) extension to time-to-event modeling (e.g., deep Cox or discrete-time survival losses), (iii) systematic ablation studies to isolate the contribution of the periskeletal mask channel and the CT channel, and (iv) integration of clinical covariates to

evaluate whether periskeletal representations provide additive value beyond standard prognostic variables. Despite these limitations, PRISM-3D demonstrates that periskeletal region–aware, segmentation-guided modeling is technically feasible, yields competitive performance on a widely used benchmark, and produces anatomically grounded attribution patterns that support transparency in imaging-based prognosis.

VI. Conclusion

By explicitly incorporating anatomically defined periskeletal regions into a 3D CNN pipeline, PRISM-3D extends imaging-based prognostic modeling beyond tumor-centric analysis. Results on a large public CT cohort demonstrate that periskeletal segmentation can be reliably learned and leveraged for survival prediction while maintaining interpretability through anatomically grounded attention. These findings support the integration of patient-level anatomical context as a complementary source of prognostic information in imaging-based oncology models.

Acknowledgements

The authors disclose the use of an AI-based language model (ChatGPT, OpenAI) for limited assistance with language editing of the manuscript. All scientific content, experimental design, data analysis, and conclusions are solely those of the authors [22].

References

[1] F. Bray, J. Ferlay, I. Soerjomataram, R. L. Siegel, L. A. Torre, and A. Jemal, "Global cancer statistics 2018: Globocan estimates of incidence and mortality worldwide for 36 cancers in 185 countries," *CA: A Cancer Journal for Clinicians*, vol. 68, no. 6, pp. 394–424, 2018.

[2] R. L. Siegel, K. D. Miller, N. S. Wagle, and A. Jemal, "Cancer statistics, 2024," *CA: A Cancer Journal for Clinicians*, vol. 74, no. 1, pp. 12–49, 2024.

[3] G. A. Woodard, K. D. Jones, and D. M. Jablons, "Heterogeneity of survival in stage i non–small cell lung cancer," *Journal of Thoracic Oncology*, vol. 11, no. 5, pp. 692–700, 2016.

[4] F. C. Detterbeck, D. J. Boffa, A. W. Kim, and L. T. Tanoue, "The eighth edition lung cancer stage classification," *Chest*, vol. 151, no. 1, pp. 193–203, 2017.

[5] H. J. Aerts, E. R. Velazquez, R. T. Leijenaar *et al.*, "Decoding tumour phenotype by noninvasive imaging using a quantitative radiomics approach," *Nature Communications*, vol. 5, no. 1, p. 4006, 2014.

[6] C. Parmar, R. T. Leijenaar, P. Grossmann *et al.*, "Robust radiomics feature quantification using semiautomatic volumetric segmentation," *PLOS ONE*, vol. 9, no. 7, p. e102107, 2014.

[7] A. Hosny, C. Parmar, J. Quackenbush, L. H. Schwartz, and H. J. Aerts, "Deep learning for lung cancer prognostication: A retrospective multi-cohort radiomics study," *PLOS Medicine*, vol. 15, no. 11, p. e1002711, 2018.

[8] A. Braghetto, F. Marturano, M. Paiusco *et al.*, "Radiomics and deep learning methods for the prediction of 2-year overall survival in lung1 dataset," *Scientific Reports*, vol. 12, p. 14132, 2022.

[9] M. Mourtzakis and C. M. Prado, "Sarcopenia in cancer: Definitions, etiology, and clinical implications," *The Lancet Oncology*, vol. 19, no. 12, pp. e657–e666, 2018.

[10] S. S. Shachar, G. R. Williams, H. B. Muss, and T. F. Nishijima, "Sarcopenia in patients with cancer: A systematic review and meta-analysis," *The Oncologist*, vol. 21, no. 5, pp. 627–635, 2017.

[11] C. M. Prado, V. E. Baracos, L. J. McCargar *et al.*, "Body composition as an independent determinant of 5-fluorouracil–based chemotherapy toxicity," *Clinical Cancer Research*, vol. 14, no. 9, pp. 3263–3268, 2008.

[12] A. Holzinger, G. Langs, H. Denk, K. Zatloukal, and H. Müller, "What do we need to build explainable ai systems for the medical domain?" *arXiv preprint arXiv:1712.09923*, 2017.

[13] H. R. Roth *et al.*, "Improving computer-aided detection using convolutional neural networks and random view aggregation," *IEEE Transactions on Medical Imaging*, vol. 35, no. 5, pp. 1170–1181, 2016.

[14] A. Zwanenburg, M. Vallières, M. A. Abdalah *et al.*, "The image biomarker standardization initiative: standardized quantitative radiomics for high-throughput image-based phenotyping," *Radiology*, vol. 295, no. 2, pp. 328–338, 2020.

[15] A. Hosny, H. J. W. L. Aerts *et al.*, "Deep learning for lung cancer prognostication: A retrospective multi-cohort radiomics study," *PLOS Medicine*, 2018.

[16] Z. Khodabakhshi *et al.*, "A deep learning approach to predict 2-year overall survival in non-small cell lung cancer patients using ct imaging," in *Proceedings of Machine Learning Research: Medical Imaging with Deep Learning (MIDL)*, 2022.

[17] H. J. W. L. Aerts, L. Wee, E. Rios Velazquez, R. T. H. Leijenaar, C. Parmar, P. Grossmann, S. Carvalho, J. Bussink, R. Monshouwer, B. Haibe-Kains *et al.*, "Data from nsclc-radiomics," https://www.cancerimagingarchive.net, 2019.

[18] P. A. Yushkevich, J. Piven, H. C. Hazlett, R. G. Smith, S. Ho, J. C. Gee, and G. Gerig, "User-guided 3d active contour segmentation of anatomical structures: Significantly improved efficiency and reliability," *NeuroImage*, vol. 31, no. 3, pp. 1116–1128, 2006.

[19] F. Isensee, P. F. Jaeger, S. A. 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.

[20] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016, pp. 770–778.

[21] R. R. Selvaraju, M. Cogswell, A. Das, R. Vedantam, D. Parikh, and D. Batra, "Grad-cam: Visual explanations from deep networks via gradient-based localization," in *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, 2017, pp. 618–626.

[22] OpenAI, "ChatGPT," Large language model, 2026, accessed: 2026-01-28.