

VR-DiagNet: Medical Volumetric and Radiomic Diagnosis Networks with Interpretable Clinician-like Optimizing Visual Inspection - Supplementary Materials

Anonymous Authors

This supplementary material is organized as follows: §1 introduces the advantage of VR-DiagNet in terms of convergence as discussed in the “Ablation Study” of the main text. §2 provides additional visualization of the normalized feature strength in radiomics across six datasets, complementing “Visualization” in the main text. §3 completes the remaining ablation experiments as elaborated in “Ablation Study” of the main text. §4 describes the zero-centered position encoding used in Equation 6 of the main text. §5 provides detailed information about the model hyperparameter settings across all six datasets, augmenting the content of “Settings” in the main text. §6 supplements the second paragraph of “Visualization” by explaining the planning process over ten rounds across all six datasets. §7 elaborates on the different performances achieved by a traditional machine learning classifier on radiomic features extracted in two different ways, as discussed in the second paragraph of “Ablation Study” in the main text. Finally, in §8, we provide a detailed discussion of the limitations of this study and potential future research directions.

1 CONVERGENCE ANALYSIS

Figure S1 illustrates the convergence comparison between VR-DiagNet and the ACS method [3] on the validation sets of six datasets. Our approach outperforms the ACS method on five of six datasets, with the remaining exhibiting comparable performance. Using the same backbone showcases the advantage of our clinician-like planning, whereas the dense encoding employed in ACS fails to achieve this. However, it is noteworthy to mention that a higher degree of oscillation is observed in our approach, indicating potential room for improvement in stability in future iterations of our work.

2 RADIOMIC FEATURES VISUALIZATION

The figures below, denoted as Figure S2 (Organ3D), Figure S3 (Nodule3D), Figure S4 (Fracture3D), Figure S5 (Adrenal3D), Figure S6 (Vessel3D), and Figure S7 (Synapse3D), illustrate the normalized feature strength of raw radiomic features juxtaposed with multimodal-refined radiomic features. These visualizations eloquently illustrate the discernible alterations in discriminability across various features. Notably, the refined features manifest significant macroscopic changes in inter-class discriminability, particularly evident in cases such as Organ3D, Nodule3D, and Adrenal3D.

However, having clinical experts in the loop is essential for a more comprehensive and profound interpretation. This clinicians in the loop could entail delving into the intricate relationship between specific radiomic features and corresponding organs or lesions before and after refinement. Such interdisciplinary efforts are indispensable for unlocking deeper insights and maximizing our findings’ clinical relevance and applicability.

3 ABLATION STUDY

Does volumetric input matter? We commence our investigation by examining the significance of volumetric input in our task. Given the intricate three-dimensional nature of the tasks, we hypothesize that incorporating neighboring knowledge could potentially enhance task performance due to the lack of three-dimensional spatial information in 2D slices. As summarized in Table S1, the results substantiate this conjecture, demonstrating a consistent performance enhancement as the neighboring window size increases within the tested range of values. We refrained from exploring larger window sizes due to the limited number of slices (28) in the task and the need to balance accuracy and computational cost.

Table S1: The study of neighboring windows. Averaged AUROC (↑) and ACC (↑) across six datasets are reported.

Neighboring window	AUROC	ACC
$\mathcal{N}(1)$	0.851	0.799
$\mathcal{N}(3)$	0.865	0.817
$\mathcal{N}(5)$	0.873	0.831

What depth of experience tree do the tasks need? As illustrated in Table S2, each dataset achieves optimal performance under distinct configurations, challenging the notion that deeper layers inherently lead to superior results. We attribute this variance to differences in task complexity influenced by various factors: 1) target size in three dimensions and 2) the emphasis on texture detection or morphology.

Removing black slices constrains the maximum tree depth for Fracture3D, Adrenal3D, and Vessel3D. This constraint leads to a uniform tree depth adoption across all volumes within a given dataset, resulting in “N/A” entries within the table. These findings emphasize the importance of selecting dataset-specific tree depths for optimal performance.

Table S2: The comparison of tree layers. ACC (↑) is reported.

L	Org.	Nod.	Fra.	Adr.	Ves.	Syn.	Avg.
2	0.938	0.872	0.523	0.793	0.917	0.800	0.807
3	0.947	0.884	0.561	0.818	0.948	0.830	0.831
4	0.947	0.865	0.548	0.803	N/A	0.823	N/A
5	0.952	0.875	0.545	N/A	N/A	0.828	N/A
6	0.963	0.861	N/A	N/A	N/A	0.822	N/A
7	0.958	0.874	N/A	N/A	N/A	0.840	N/A

Is the default balancing factor in UCB1 compatible? We examine the influence of the Monte Carlo Tree Search balancing factor c as presented in Table S3. It is observed that the default value $\sqrt{2}$ utilized in UCB1 yields the optimal performance.

Table S3: The comparison of balancing factors. Averaged AUROC (\uparrow) and ACC (\uparrow) across six datasets are presented.

c	AUROC	ACC
1	0.865	0.828
$\sqrt{2}$	0.875	0.836
2	0.858	0.828

How many MCTS iterations do we need? Table S4 shows the impact of varying numbers of MCTS iterations, with the number of tree layers derived from Table S2 for each dataset. Notably, high sensitivity is also observed regarding hyper-parameter L across different datasets. One might wonder how the model's performance would fare if we select the top- L "SoIs" rather than those identified through the search process for classifier training. We adopt a straightforward approach to investigate this: determining the maximum value in each slice-level prediction distribution based on hierarchical priori. Here, the calculations between slices are treated as independent of each other. Subsequently, the "SoIs" with the top- L predicted values are identified. Notably, this process remains **class-agnostic**. As illustrated by the results in the last three rows of Table S4, in the absence of a planner, the classifier can still achieve improved performance as the neighboring window expands, albeit inferior to the MCTS scheme. The result highlights 1) the significance of the conditional clinician-like visual inspection process in our VR-DiagNet and 2) the efficacy of the hierarchical priori design in providing a robust initial starting point for our model.

Table S4: The impact of MCTS iterations on ACC (\uparrow).

N_{Mc}	Org.	Nod.	Fra.	Adr.	Ves.	Syn.	Avg.
140	0.956	0.863	0.545	0.820	0.941	0.841	0.828
112	0.959	0.863	0.530	0.806	0.935	0.841	0.822
84	0.963	0.884	0.561	0.818	0.948	0.840	0.836
56	0.955	0.875	0.538	0.820	0.942	0.837	0.828
28	0.958	0.872	0.532	0.809	0.953	0.840	0.827
14	0.953	0.870	0.542	0.809	0.945	0.832	0.825
top- L w/ $N(1)$	0.900	0.811	0.532	0.802	0.869	0.770	0.781
top- L w/ $N(3)$	0.929	0.805	0.528	0.819	0.905	0.781	0.795
top- L w/ $N(5)$	0.941	0.813	0.538	0.815	0.911	0.780	0.800

How to improve the stability of the search process? We approach enhancing search process stability through the lens of inter-round momentum as delineated in Table S5. Our findings reveal that while the non-momentum approach yields commendable results, the momentum scheme consistently outperforms the non-momentum design across all six datasets when endowed with a judiciously chosen parameter. The results underscore the effectiveness of the momentum scheme integrated into our proposed planner.

Table S5: Influence of inter-round momentum in the proposed planner. ACC (\uparrow) is reported.

m	Org.	Nod.	Fra.	Adr.	Ves.	Syn.	Avg.
N/A	0.959	0.870	0.549	0.801	0.924	0.838	0.824
0.2	0.955	0.860	0.554	0.817	0.934	0.843	0.827
0.4	0.945	0.877	0.545	0.804	0.942	0.843	0.826
0.6	0.955	0.868	0.553	0.808	0.949	0.846	0.830
0.8	0.963	0.884	0.561	0.820	0.953	0.841	0.837
1.0	0.958	0.877	0.547	0.795	0.942	0.846	0.828

4 ZERO-CENTERED POSITION ENCODING

The approach adopted for computing one-dimensional position encoding is outlined as Equation 1:

$$\begin{aligned} \text{zpe}(d, 2i) &= \sin\left(\frac{d}{10000^{\frac{2i}{N_{Emb}}}}\right) \\ \text{zpe}(d, 2i+1) &= \cos\left(\frac{d}{10000^{\frac{2i}{N_{Emb}}}}\right) \end{aligned} \quad (1)$$

In our scenario, $d \in [-\lfloor \frac{D-1}{2} \rfloor, \dots, \lfloor \frac{D-1}{2} \rfloor]$ represents the zero-centered slice index in a volume, where D denotes the volume's slice count. We establish the coordinate system with its center positioned at the 0-index. The variable i takes integer values within the interval $[0, 1024]$.

5 HYPER-PARAMETERS

We present an overview of the configuration of all hyper-parameters pertinent to our methodology in Table S6. It is worth noting that the accurate selection of these hyper-parameters profoundly influences the efficacy of our proposed approach.

6 VISUALIZATION OF THE PLANNING PROCESS

We choose one volume from each of the six datasets to illustrate the planning process of the planner across multiple rounds, as depicted in Figure S8 (Organ3D), Figure S9 (Nodule3D), Figure S10 (Fracture3D), Figure S11 (Adrenal3D), Figure S12 (Synapse3D), and Figure S13 (Vessel3D). This demonstration mirrors the sequential visual inspection undertaken by a clinician, thereby enhancing the interpretability of the proposed VR-DiagNet.

7 DISCRIMINABILITY OF RADIOMIC FEATURES

We evaluate the discriminability of raw static radiomic features and refined radiomic features using an SVM classifier and present the results in Table S7. Our findings reveal an overall performance enhancement with refined radiomic features. Specifically, the model accuracy with refined radiomic features shows improvement in 50% of the datasets, remains consistent in one dataset, and exhibits comparable accuracy in the other two datasets, indicating promise for medical research.

Table S6: Assignment of hyper-parameters for the VR-DiagNet across different datasets.

Hyper-parameters		Task 1			Task 2		
		Organ3D	Nodule3D	Synapse3D	Fracture3D	Adrenal3D	Vessel3D
f _{PL} -related	# Tree layers (L)	6	3	7	3	3	3
	Balancing factor (c)	$\sqrt{2}$	$\sqrt{2}$	$\sqrt{2}$	$\sqrt{2}$	$\sqrt{2}$	$\sqrt{2}$
	# MCTS iterations (N_{Mc})	84	84	140	84	56	28
	Momentum (m)	0.8	0.8	0.6	0.8	0.8	0.8
f _{CL} -related	# Round (N_{Ro})	10	10	10	10	10	10
	# Epoch in each round (N_{Ep})	20	20	20	20	20	20
	Learning rate (lr)	$1e^{-4}$	$1e^{-4}$	$1e^{-4}$	$1e^{-4}$	$1e^{-3}$	$1e^{-4}$
	Weight decay	$3e^{-5}$	0	$1e^{-3}$	$3e^{-4}$	$1e^{-7}$	$3e^{-4}$
	Batch size	64	64	64	64	64	64
	Scale coefficient r	8	4	16	64	4	8
	Temperature τ	1	1	0.3	1	1	1
	Loss balancing factor λ	0.1	0.1	0.1	0.1	0.1	0.1
	Scale in RandomCropResize()	(0.08, 1)	(0.08, 1)	(0.08, 1)	(0.08, 1)	(0.56, 1)	(0.08, 1)
	# Window size (n in $N(n)$)	5	5	5	5	5	5
	Dropout rate	0.1	0.1	0	0	0.05	0.1
	Granularity of input	Coarse-grained	Coarse-grained	Coarse-grained	Fine-grained	Fine-grained	Fine-grained
	Mean	0.5004	0.2686	0.5098	0.0229	0.0158	0.0193
	Std	0.2805	0.2734	0.2376	0.1104	0.1247	0.1377

Table S7: Discriminability of different radiomic features reported as ACC (†) using an SVM classifier.

Fea. Type	Org.	Nod.	Fra.	Adr.	Ves.	Syn.	Avg.
Static	0.669	0.836	0.550	0.836	0.895	0.776	0.760
Refined	0.667	0.837	0.540	0.836	0.927	0.790	0.766

8 LIMITATIONS AND FUTURE RESEARCH

This section highlights several limitations inherent in our study and potential avenues for future research. Firstly, it is imperative to acknowledge the visible escalation in computational overheads observed on CPUs as the number of slices, MCTS iteration times, and tree depth increases, which may impede scalability. Secondly, we encourage for further convergence analyses to enhance search efficiency and algorithmic stability. Thirdly, broadening the scope of

data viewpoints holds promise for bolstering robustness and achieving closer alignment with clinical practice. Lastly, as elucidated in Table S6, our VR-DiagNet demonstrates significant sensitivity to hyper-parameter assignments across diverse datasets, signaling a trajectory for future exploration aimed at devising a solution less reliant on hyper-parameter selection.

REFERENCES

[1] James C Korte, Carlos Cardenas, Nicholas Hardcastle, Tomas Kron, Jihong Wang, Houda Bahig, Baher Elgohari, Rachel Ger, Laurence Court, Clifton D Fuller, et al. 2021. Radiomics feature stability of open-source software evaluated on apparent diffusion coefficient maps in head and neck cancer. *Scientific reports* 11, 1 (2021), 17633.

[2] Joost JM Van Griethuysen, Andriy Fedorov, Chintan Parmar, Ahmed Hosny, Nicole Aucoin, Vivek Narayan, Regina GH Beets-Tan, Jean-Christophe Fillion-Robin, Steve Pieper, and Hugo JWL Aerts. 2017. Computational radiomics system to decode the radiographic phenotype. *Cancer research* 77, 21 (2017), e104–e107.

[3] Jiancheng Yang, Xiaoyang Huang, Yi He, Jingwei Xu, Canqian Yang, Guozheng Xu, and Bingbing Ni. 2021. Reinventing 2d convolutions for 3d images. *IEEE Journal of Biomedical and Health Informatics* 25, 8 (2021), 3009–3018.

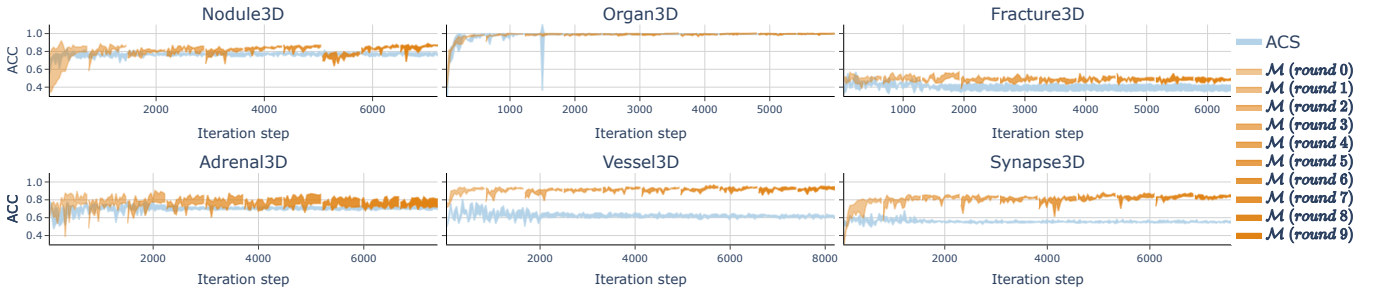


Figure S1: The convergence comparison between VR-DiagNet and the ACS method [3]. Our method demonstrates superior convergence compared to ACS [3] across six validation datasets, albeit with more significant oscillations. The symbol \mathcal{M} in the figure legend denotes our MCTS-based approach.

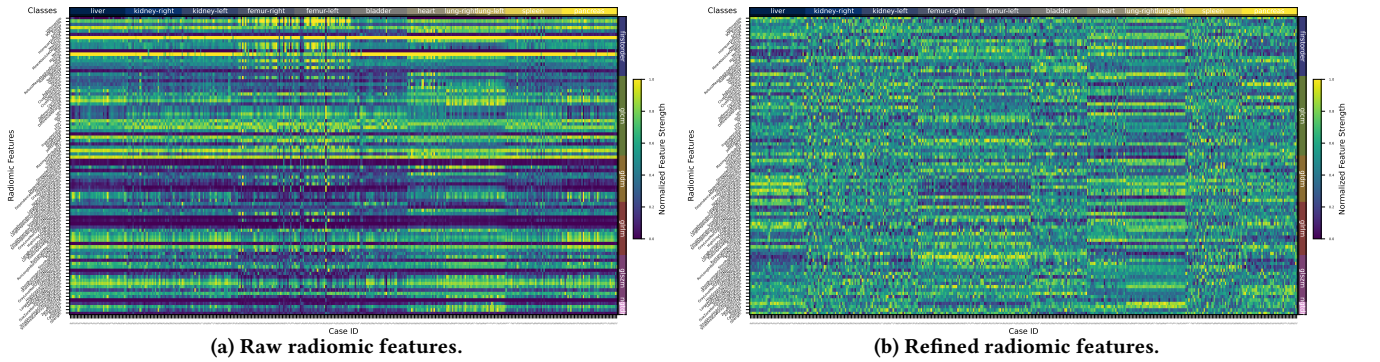


Figure S2: Comparison of normalized radiomic feature strength. Drawing method references [1, 2]. Both panels depict two identical subsets of volumes from the Organ3D test set, each containing 40 cases per class. Each row represents one of the radiomic features post-feature selection, with their respective names listed to the left of each map, while each column represents a volume, with their corresponding IDs listed below each map. The volumes are organized by class, with class names indicated at the top of each panel. Additionally, radiomic features are grouped by class, with their class names listed to the right of each map. We suggest zooming in to better observe the names and selected volume IDs for improved clarity.

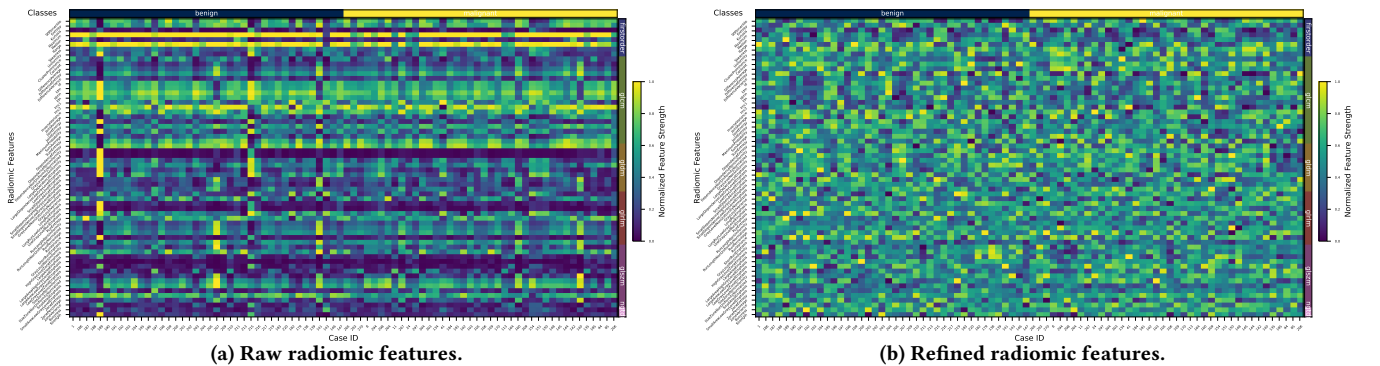


Figure S3: Comparison of normalized radiomic feature strength. Drawing method references [1, 2]. Both panels depict two identical subsets of volumes from the Nodule3D test set, each containing 40 cases per class. Each row represents one of the radiomic features post-feature selection, with their respective names listed to the left of each map, while each column represents a volume, with their corresponding IDs listed below each map. The volumes are organized by class, with class names indicated at the top of each panel. Additionally, radiomic features are grouped by class, with their class names listed to the right of each map. We suggest zooming in to better observe the names and selected volume IDs for improved clarity.

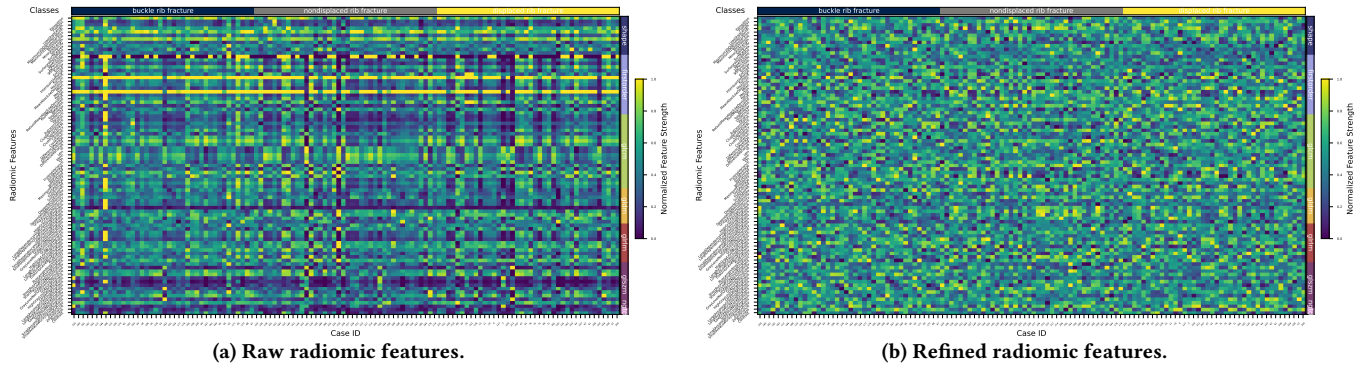


Figure S4: Comparison of normalized radiomic feature strength. Drawing method references [1, 2]. Both panels depict two identical subsets of volumes from the Fracture3D test set, each containing 40 cases per class. Each row represents one of the radiomic features post-feature selection, with their respective names listed to the left of each map, while each column represents a volume, with their corresponding IDs listed below each map. The volumes are organized by class, with class names indicated at the top of each panel. Additionally, radiomic features are grouped by class, with their class names listed to the right of each map. We suggest zooming in to better observe the names and selected volume IDs for improved clarity.

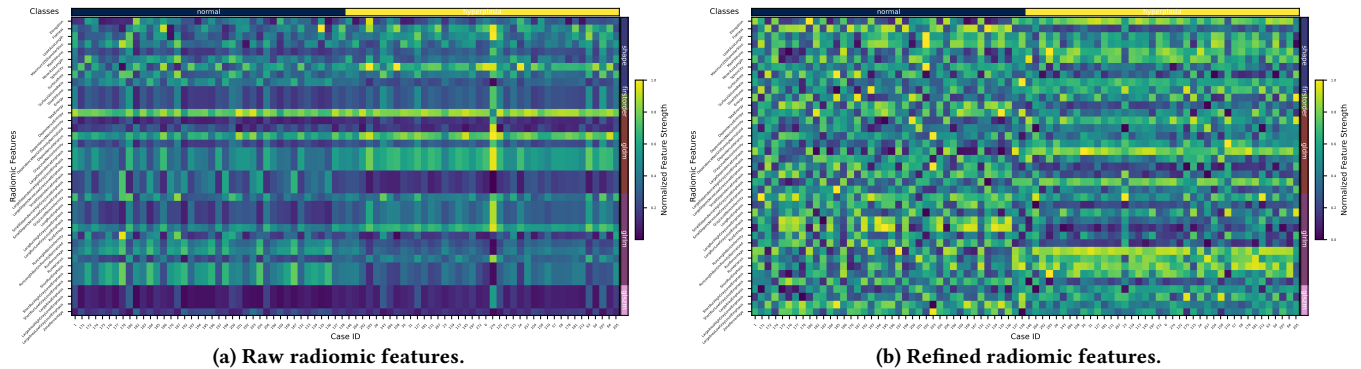


Figure S5: Comparison of normalized radiomic feature strength. Drawing method references [1, 2]. Both panels depict two identical subsets of volumes from the Adrenal3D test set, each containing 40 cases per class. Each row represents one of the radiomic features post-feature selection, with their respective names listed to the left of each map, while each column represents a volume, with their corresponding IDs listed below each map. The volumes are organized by class, with class names indicated at the top of each panel. Additionally, radiomic features are grouped by class, with their class names listed to the right of each map. We suggest zooming in to better observe the names and selected volume IDs for improved clarity.

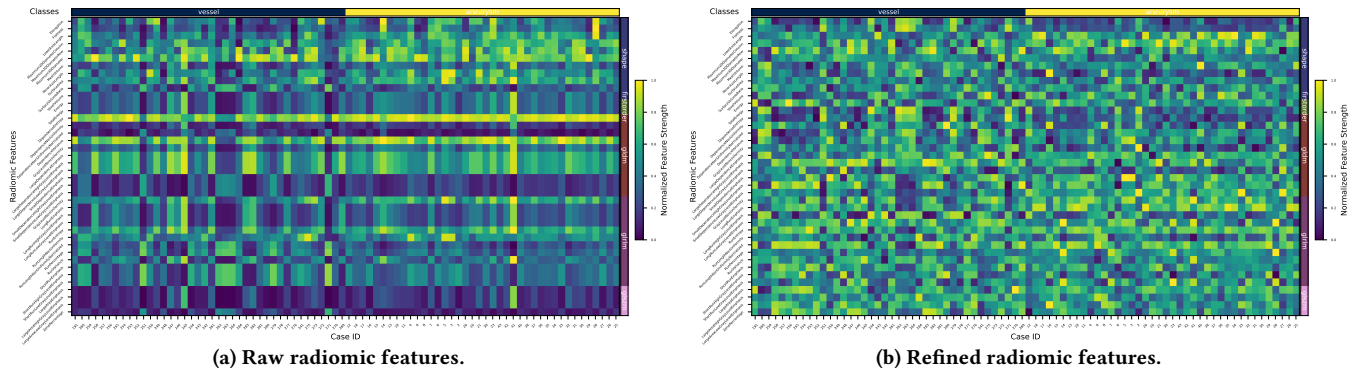


Figure S6: Comparison of normalized radiomic feature strength. Drawing method references [1, 2]. Both panels depict two identical subsets of volumes from the Vessel3D test set, each containing 40 cases per class. Each row represents one of the radiomic features post-feature selection, with their respective names listed to the left of each map, while each column represents a volume, with their corresponding IDs listed below each map. The volumes are organized by class, with class names indicated at the top of each panel. Additionally, radiomic features are grouped by class, with their class names listed to the right of each map. We suggest zooming in to better observe the names and selected volume IDs for improved clarity.

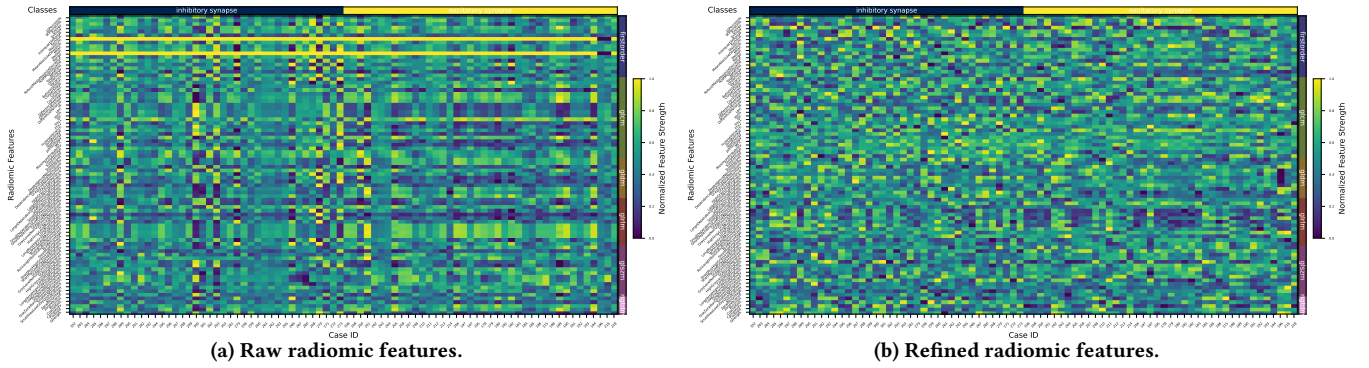


Figure S7: Comparison of normalized radiomic feature strength. Drawing method references [1, 2]. Both panels depict two identical subsets of volumes from the Synapse3D test set, each containing 40 cases per class. Each row represents one of the radiomic features post-feature selection, with their respective names listed to the left of each map, while each column represents a volume, with their corresponding IDs listed below each map. The volumes are organized by class, with class names indicated at the top of each panel. Additionally, radiomic features are grouped by class, with their class names listed to the right of each map. We suggest zooming in to better observe the names and selected volume IDs for improved clarity.

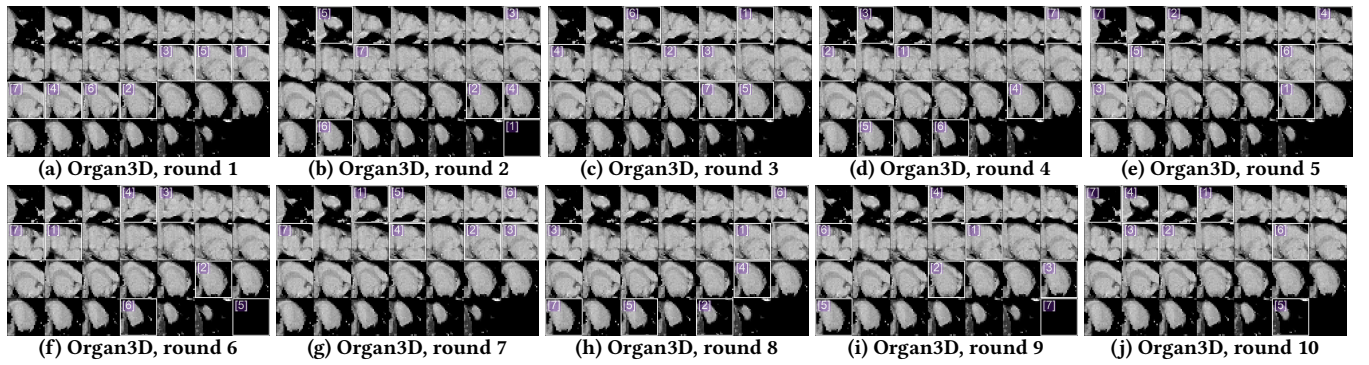


Figure S8: Illustration of the planner's iterative planning process on the 33-rd volume, belonging to the heart class, from the Organ3D dataset. The slices enclosed by white boxes represent the identified SoIs in each round, with numbers in brackets denoting their respective indices corresponding to the layer index of the tree path. This search process closely emulates the visual inspection performed by clinicians.

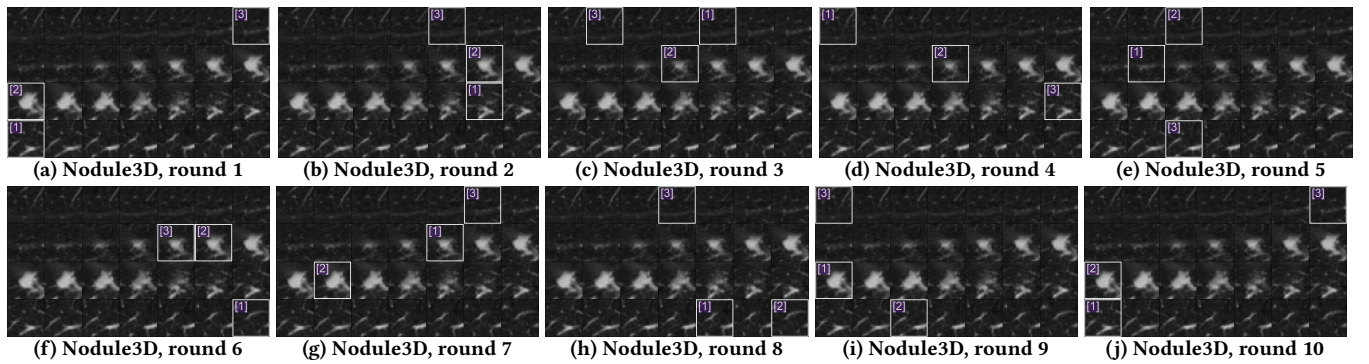


Figure S9: Illustration of the planner's iterative planning process on the 6-th volume, belonging to the malignant class, from the Nodule3D dataset. The slices enclosed by white boxes represent the identified SoIs in each round, with numbers in brackets denoting their respective indices corresponding to the layer index of the tree path. This search process closely emulates the visual inspection performed by clinicians.

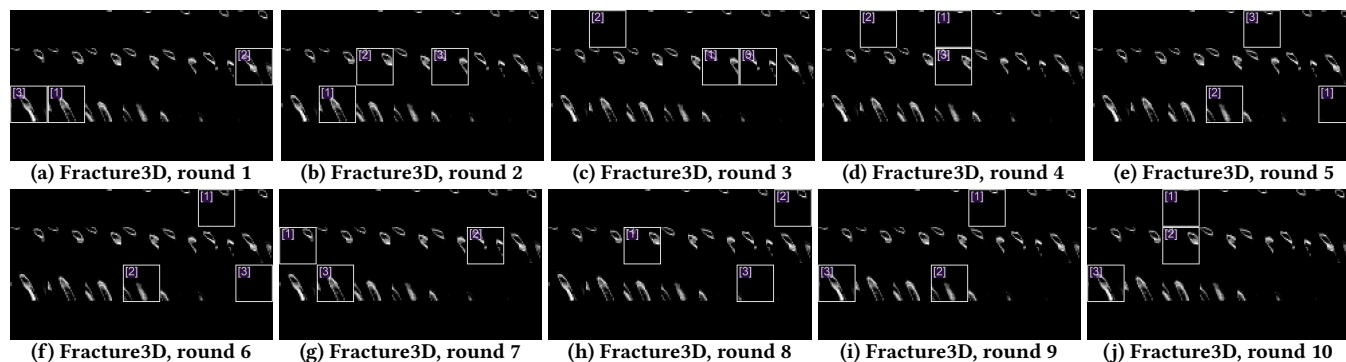


Figure S10: Illustration of the planner's iterative planning process on the 8-th volume, belonging to the nondisplaced rib fracture class, from the Fracture3D dataset. The slices enclosed by white boxes represent the identified SoIs in each round, with numbers in brackets denoting their respective indices corresponding to the layer index of the tree path. This search process closely emulates the visual inspection performed by clinicians.

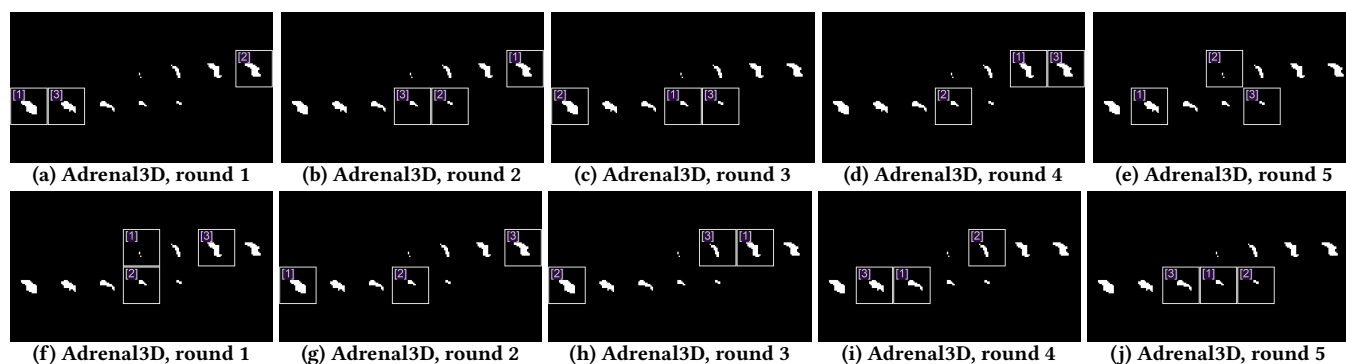


Figure S11: Illustration of the planner's iterative planning process on the 7-th volume, belonging to the hyperplasia class, from the Adrenal3D dataset. The slices enclosed by white boxes represent the identified SoIs in each round, with numbers in brackets denoting their respective indices corresponding to the layer index of the tree path. This search process closely emulates the visual inspection performed by clinicians.

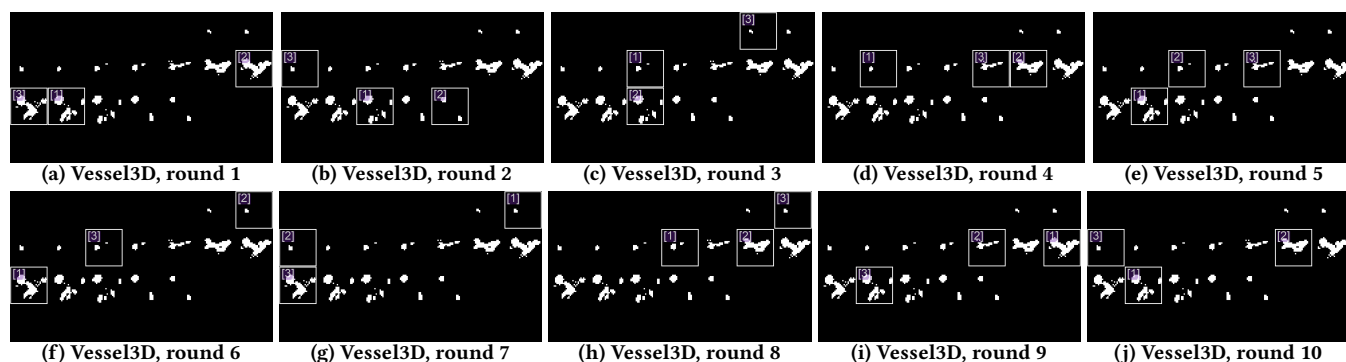


Figure S12: Illustration of the planner's iterative planning process on the 9-th volume, belonging to the aneurysm class, from the Vessel3D dataset. The slices enclosed by white boxes represent the identified SoIs in each round, with numbers in brackets denoting their respective indices corresponding to the layer index of the tree path. This search process closely emulates the visual inspection performed by clinicians.

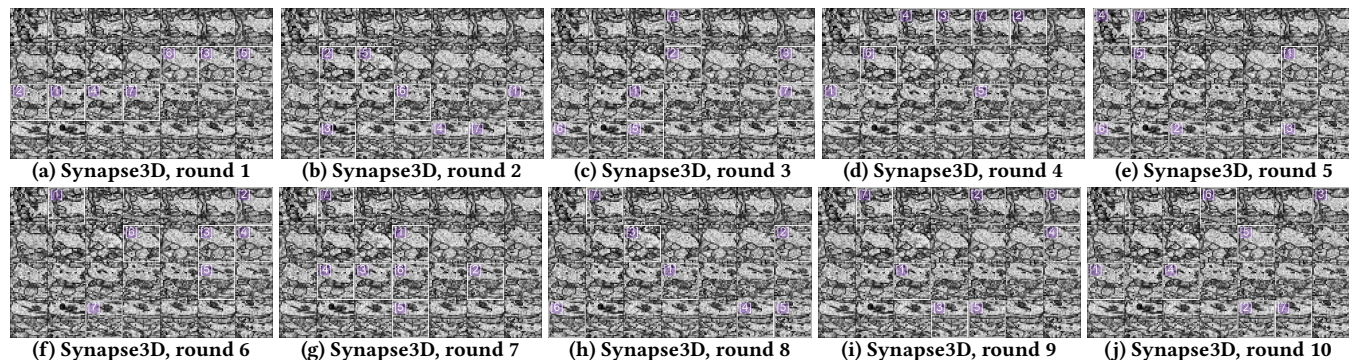


Figure S13: Illustration of the planner's iterative planning process on the 6-th volume, belonging to the excitatory synapse class, from the Synapse3D dataset. The slices enclosed by white boxes represent the identified Sol's in each round, with numbers in brackets denoting their respective indices corresponding to the layer index of the tree path. This search process closely emulates the visual inspection performed by clinicians.