Conference on Artificial Intelligence in Medical Imaging
OCT 20-21 | UNIVERSITY OF CALIFORNIA, IRVINE

Radiomic Sampling: A Model-Free Approach to Enhance Diversity of Validation Datasets

Alexander Knapp, Cincinnati Children's; Jonathan Dudley, PhD; Hailong Li, PhD; Jonathan Dillman, MD, MSc; Elan Somasundaram, PhD

Introduction/Background

Evaluation of medical imaging models for clinical use critically depends on data diversity in the validation set, yet conventional sampling methods often fall short. We introduce Radiomic sampling, a model-free image selection method that optimizes dataset diversity by analyzing high-dimensional radiomic features without requiring a pretrained embedding model.

Methods/Intervention

We compute 107 radiomic features for each image, which are then used to represent the image as a feature vector. Principal Component Analysis coupled with a bootstrapping strategy (50 iterations, 50-case subsamples) is applied to these feature vectors to extract the most informative components. The sampling method begins with a single randomly selected image. A greedy, Euclidean distance-based selection process then iteratively chooses images with maximal feature variation from previously sampled images, yielding a highly diverse sample set and improved data efficiency, particularly for model validation in new cohorts. Radiomic sampling was evaluated on liver–spleen segmentation tasks for CT and MRI by comparing bootstrap subsets from random and radiomic sampling using median average Dice Similarity Coefficient (DSC). Lower DSC indicates increased heterogeneity and segmentation difficulty due to anatomical variations and artifacts. Each experiment involved 50 exams per subset across 50 bootstrap iterations.

Results/Outcome

Radiomic sampling consistently yielded lower median DSC values than random sampling, indicating selection of more challenging cases. On a pediatric MRI dataset evaluated with the TotalSegmentator-MRI model, liver DSC dropped from 0.911 to 0.876 and spleen from 0.866 to 0.779. Similarly, on a pediatric CT model tested with the TotalSegmentator-CT dataset, liver DSC decreased from 0.956 to 0.933 and spleen from 0.929 to 0.888. All differences were statistically significant (Mann–Whitney U test).

Conclusion

Radiomic sampling provides a rigorous framework for stress-testing models by targeting their most challenging cases, ensuring robust segmentation evaluation. Its systematic selection of diverse samples can also improve data selection and data annotation efficiency for clinical validation datasets.

Statement of Impact

Radiomic sampling offers a novel approach to enhance the reliability of AI models by systematically identifying diverse and challenging cases. This method enables developers and clinicians to pinpoint model limitations, ultimately improving generalizability and fostering greater clinical confidence in model deployment.

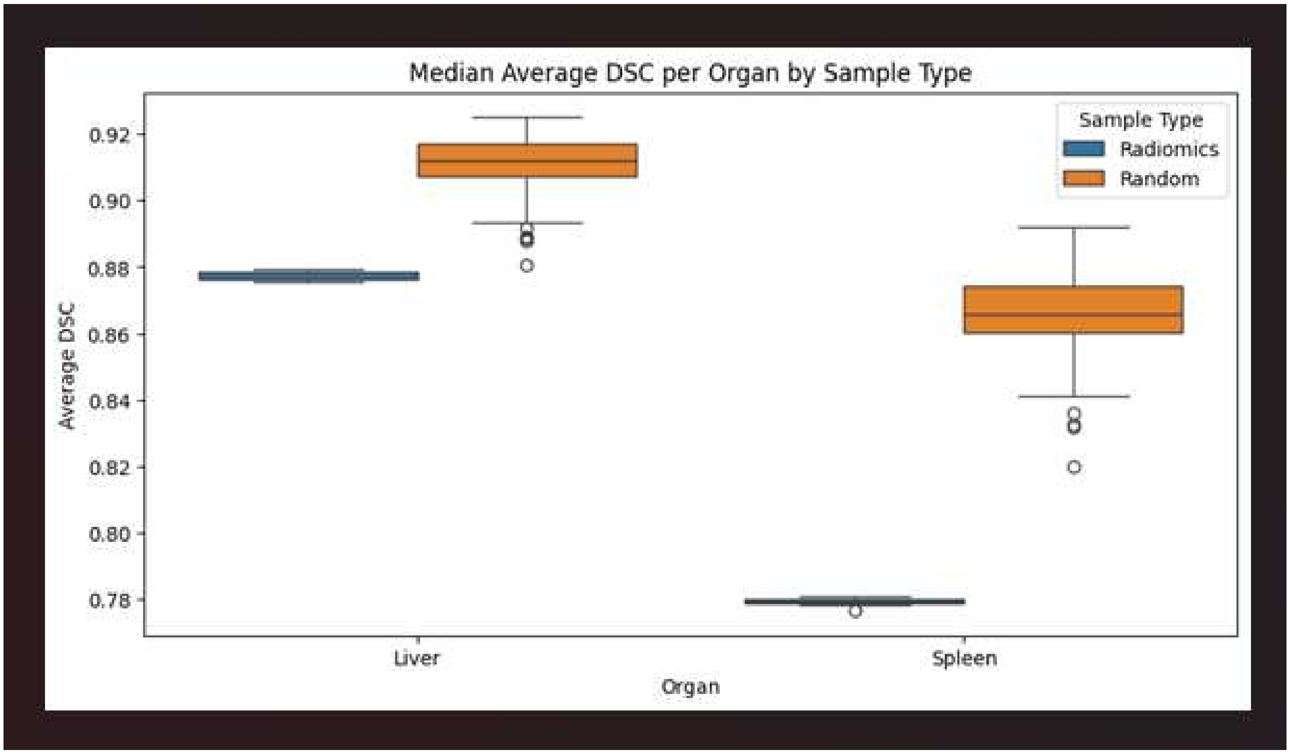

Figure 1: Box plot showing the average DSC distribution for 50 bootstrapped samples with 50 images each sampled using radiomic and random sampling for the TotalSegmentator MRI model on a Pediatric MRI dataset. The small IQR of Radiomic Sampling shows its ability to consistently pick harder cases across the bootstrap.

Keywords

Image Similarity Search; Dataset Diversity; Model Validation; Radiomics; Segmentation; Pediatrics