Evaluating the robustness of features generated by a foundation model from lung nodule regions in CT with different reconstruction parameters

Stephen Park[1,2], Anil Yadav[1,2], Hossein Tabatabaei[2], Denise R Aberle[1,2], Ashley E Prosper[2], William Hsu[1,2]

1 Department of Bioengineering, Samueli School of Engineering, University of California, Los Angeles, CA 90095, USA
2 Department of Radiological Sciences, David Geffen School of Medicine at UCLA, Los Angeles, CA 90095, USA

Foundation models are deep learning models trained on large datasets that can be used as the basis for various downstream applications. Recently, a foundation model specifically developed for cancer imaging has been shown to generate task-agnostic features that are highly predictive in diagnostic and prognostic tasks, highlighting its potential for discovering cancer biomarkers.[1] However, imaging-based biomarkers often encounter challenges with reproducibility due to heterogeneous image parameters, impeding their translation into clinical settings. Hence, this study aimed to assess the robustness of features extracted from low-dose computed tomography (LDCT) scans using the pretrained foundation model across varying reconstruction kernels and slice thicknesses.

59 LDCT scans containing at least one nodule were obtained from individuals who underwent lung cancer screening. For scans with multiple nodules, one nodule was selected at random. Each scan consisted of 6 variations of image conditions, representing combinations of three reconstruction kernels (smooth, medium, sharp) and two slice thicknesses (1.0 mm, 2.0 mm). The pretrained foundation model, which is a ResNet50 encoder, runs inference on a 50x50x50 mm image volume and outputs a 4,096-feature vector characterizing the volume. Using this model, 59 nodule volumes at each of the 6 different image conditions were processed. Then, the model's robustness was evaluated by computing the agreement between 4,096 features extracted from a reference condition and the other 5 conditions. The reference condition was set as medium kernel and 1.0 mm slice thickness (Medium/1.0), which approximates the recommended protocol for lung cancer screening. The metric used to quantify agreement was the concordance correlation coefficient (CCC), with a CCC of 0.9 indicating a "strong agreement." Pyradiomics, a widely used method for radiomic feature extraction, served as a baseline for comparison. Pyradiomics was used to extract 92 intensity and texture features from corresponding nodule volumes, followed by an identical analysis of robustness. The mean CCC of features and the percentage of features with CCC $\geq$ 0.9 for each condition was compared between the foundation model and Pyradiomics.

The foundation model consistently yielded high CCC values ranging from 0.937 (Smooth/2.0) to 0.984 (Sharp/1.0). In comparison, Pyradiomics resulted in considerably lower CCCs, with values between 0.674 (Sharp/1.0) and 0.843 (Medium/2.0). The percentage of features achieving "strong agreement" was constantly higher for the foundation model (range: 85.7–100%) compared to Pyradiomics (range: 38.5–59.3%).

This study demonstrates that the pretrained foundation model has the potential of extracting robust features from lung nodule regions in CT, despite variations in reconstruction kernels and slice thicknesses. Further analysis is needed to test the robustness of its performance on clinically relevant downstream tasks.

---

[1] Pai, S., Bontempi, D., Hadzic, I., Prudente, V., Sokač, M., Chaunzwa, T. L., ... & Aerts, H. J. (2024). Foundation model for cancer imaging biomarkers. *Nature machine intelligence*, 1-14.