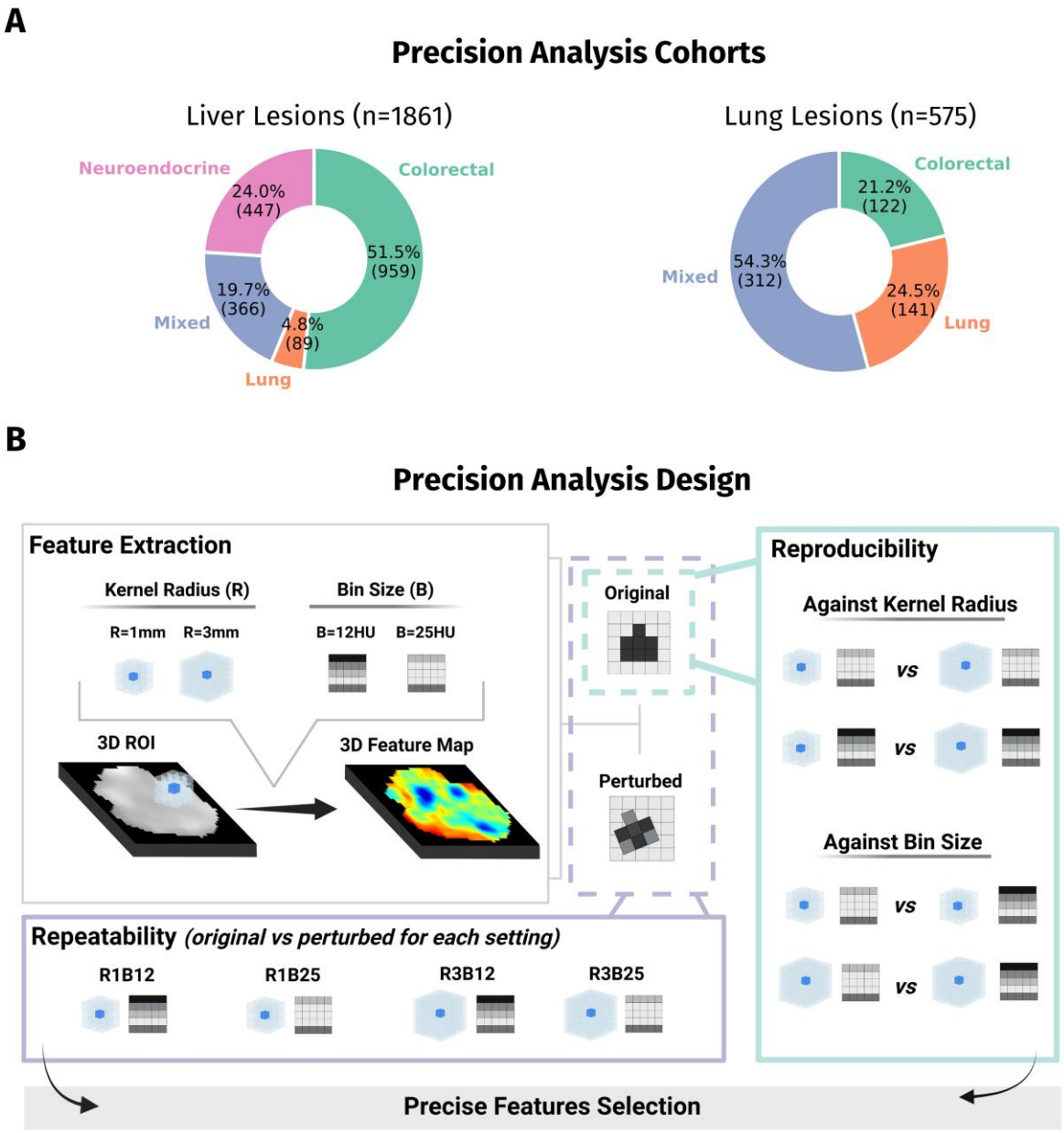

FIGURE 1. A) Distribution of lung and liver lesions across different cohorts for precision analysis. B) Precision analysis design. 3D radiomic features were computed from both original and perturbed images, four times per image, each time with a different combination of kernel radius, R (1mm/3mm), and bin size, B (12HU/25HU). To study repeatability, original-perturbed feature pairs were evaluated for every combination of extraction settings (R1B12, R1B25, R3B12 and R3B25). To study reproducibility against extraction parameters, we compared original feature pairs extracted with different extraction settings: for reproducibility against R, we first compared features extracted with fixed B=12HU and different kernel radius, and then repeated for features extracted with fixed B=25HU); analogously, for reproducibility against B, we compared features extracted from original images with fixed R=1mm and different bin size, and then repeated with those extracted with fixed R=3mm. Precise features were selected by linking reproducibility and repeatability results. R, kernel size; B, bin size.

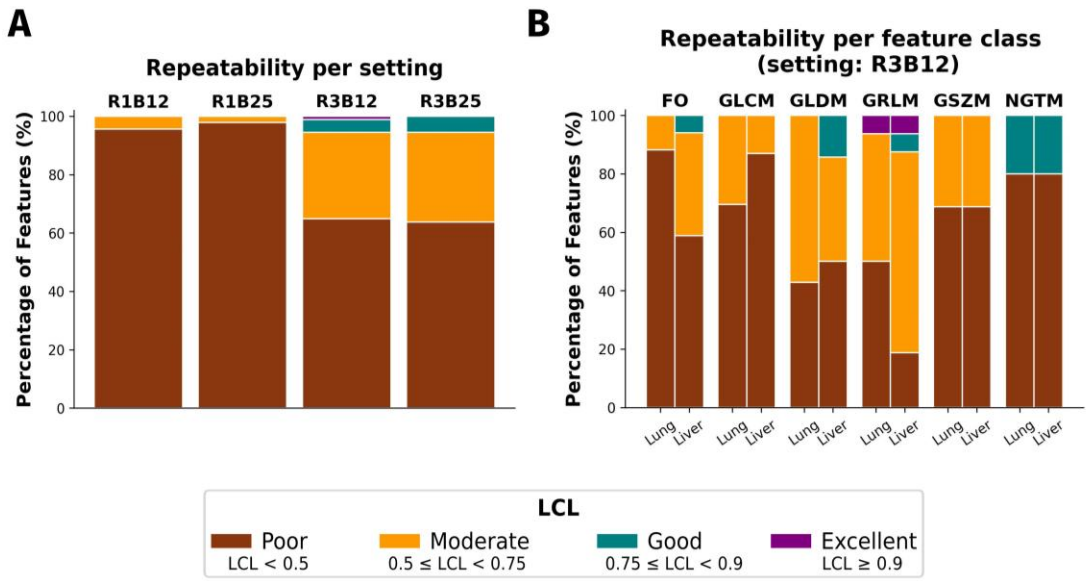

FIGURE 2. A) Repeatability distribution of radiomics features per setting. Most of radiomic features exhibit poor repeatability. Features extracted with kernel radius (R) of 3mm were more repeatable than those extractd with R=1mm. Bin size changes didn't affect repeatability. B) Repeatability distribution of radiomics features extracted with setting R3B12 per feature class for lung and liver lesions separately. First order and GLRLM features were more repeatable in liver lesions while GLCM features were more repeatable in lung lesions. LCL, 95% lower confidence limit of the Intraclass Correlation Coefficient; R3B12, features extracted with kernel radius 3mm and bin size 12HU; FO, First-Order; GLCM, Grey Level Co-occurrence Matrix features; GLDM, Grey Level Dependence Matrix; GLRLM, Grey Level Run Length Matrix; GLSZM, Grey Level Size Zone Matrix; NGTDM, Neighboring Grey Tone Difference Matrix Features.

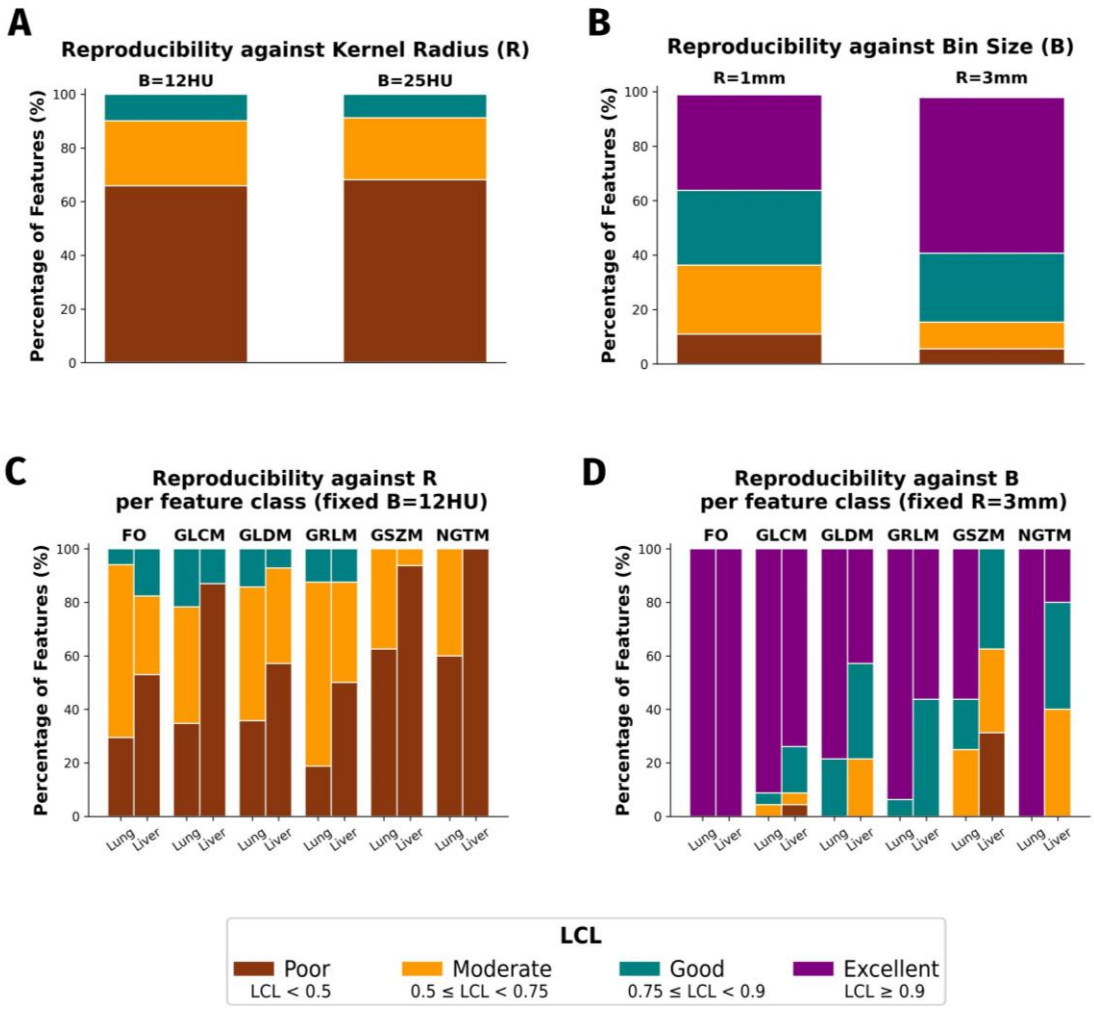

FIGURE 3. A) Reproducibility distribution against kernel radius, R, for features extracted with fixed bin size of 12HU and bin size 25HU. Most features present poor reproducibility against R. Features extracted with B=12HU were more reproducible (p<.001). B) Reproducibility distribution against bin size, B, for features extracted with fixed kernel radius 3mm and fixed kernel radius 1mm. Most features present good or excellent reproducibility against B. Features extracted with R=3mm were more reproducible (p<.001). C) Reproducibility distribution against kernel radius for features extracted with fixed bin size of 12HU per feature class for lung and liver lesions separately. Features extracted from lung lesions are more reproducible against R, especially for features belonging to GLCM and GLRLM classes. D) Reproducibility distribution against bin size for features extracted with fixed kernel radius 3mm per feature class for lung and liver lesions separately. Features extracted from lung lesions are more reproducible against B, especially for features belonging to GLCM, first-order and NGTM classes. LCL, 95% lower confidence limit of the Intraclass Correlation Coefficient; FO, First-Order features; GLCM, Grey Level Co-occurrence Matrix features; GLDM, Grey Level Dependence Matrix; GLRM, Grey Level Run Length Matrix; GSZM, Grey Level Size Zone Matrix; NGTM, Neighboring Grey Tone Difference Matrix Features.

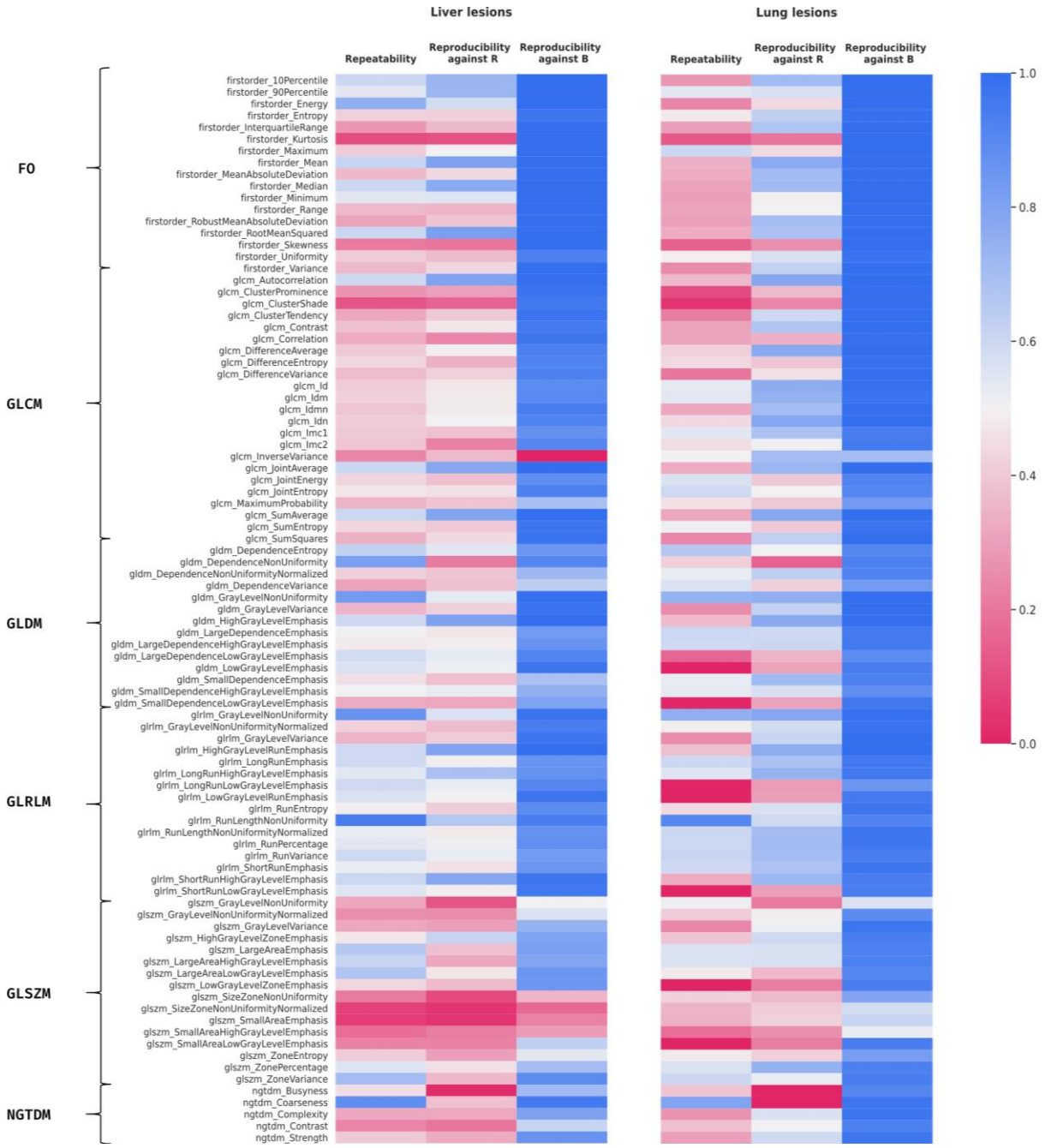

FIGURE 4. Heatmap displaying the lower 95% confidence limit (LCL) of the Intraclass Correlation Coefficient results obtained in the three experiments used to identify precise features: repeatability (setting R3B12), reproducibility against R (fixed B=12HU), and reproducibility against B (fixed R=3mm), for lung and liver lesions separately. FO, First-Order features; GLCM, Grey Level Co-occurrence Matrix features; GLDM, Grey Level Dependence Matrix; GLRLM, Grey Level Run Length Matrix; GLSZM, Grey Level Size Zone Matrix; NGTDM, Neighboring Grey Tone Difference Matrix Features.

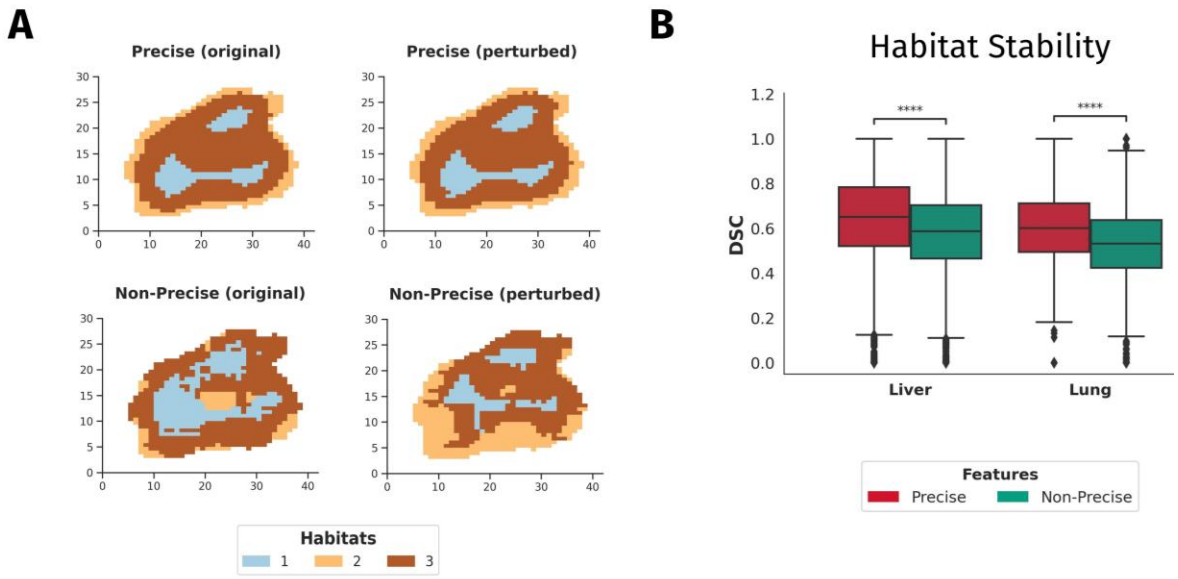

FIGURE 5. A) Example of habitats computed for one liver lesion. Top row: habitats computed with precise features extracted from original image (left) and perturbed image (right). Bottom row: habitats computed with non-precise (i.e., all extracted features) features extracted from original image (left) and perturbed image (right). Habitats computed with precise features show higher stability qualitatively (i.e. visual alignment between original-perturbed). B) Quantification of habitat stability (measured via Dice Similarity Coefficient of original-perturbed habitats) computed with precise features and non-precise features for all lung and liver lesions. Habitats computed with precise features present higher stability (Wilcoxon signed rank test, $p < 0.0001$). DSC: Dice Similarity Coefficient.

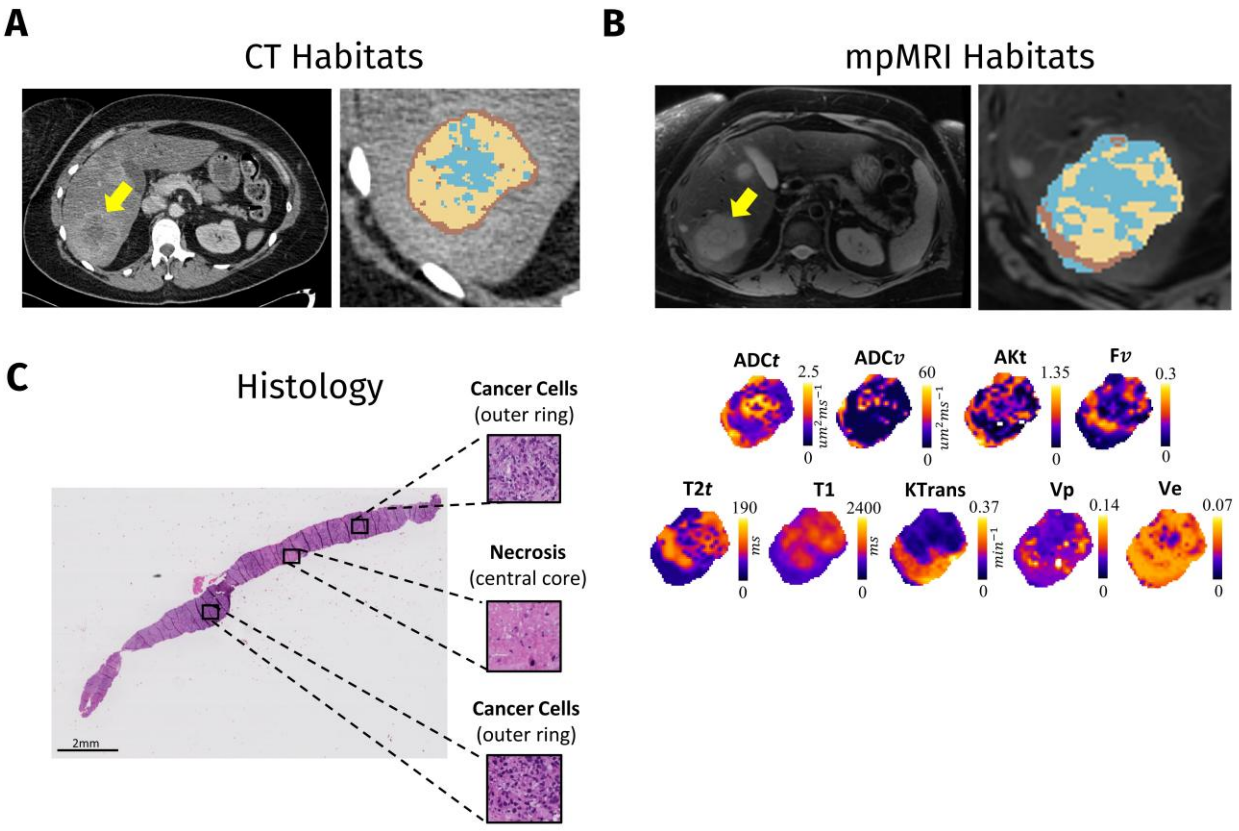

FIGURE 6. Exploration of the biological relevance of imaging habitats in an independent cohort. One exemplificatory patient (liver metastasis of melanoma). A) CT scan with visible lesion (yellow arrow) and resulting CT habitats computed with precise liver radiomic features. B) Anatomical MRI T2 scan with visible lesion (yellow arrow) and resulting mpMRI habitats computed with the 9 mpMRI maps (also shown). mpMRI and CT habitats presented consistent distributionsC) Histologic image (HE stain) with observable tissue types, annotated by a pathologist. The HE-stained histological material reveals areas of necrosis in the core of the lesion. tADCt: tissue apparent diffusion coefficient, ADCv: vascular apparent diffusion coefficient, T2t: tissue transvers relaxation time, AKt: tissue apparent kurtosis coefficient, Fv; vascular signal fraction, T1: total longitudinal relaxation time, KTrans: capillary permeability constant, Vp: plasma volume fraction, Ve: extravascular and extracellular volume fraction.

Supplemental Figures

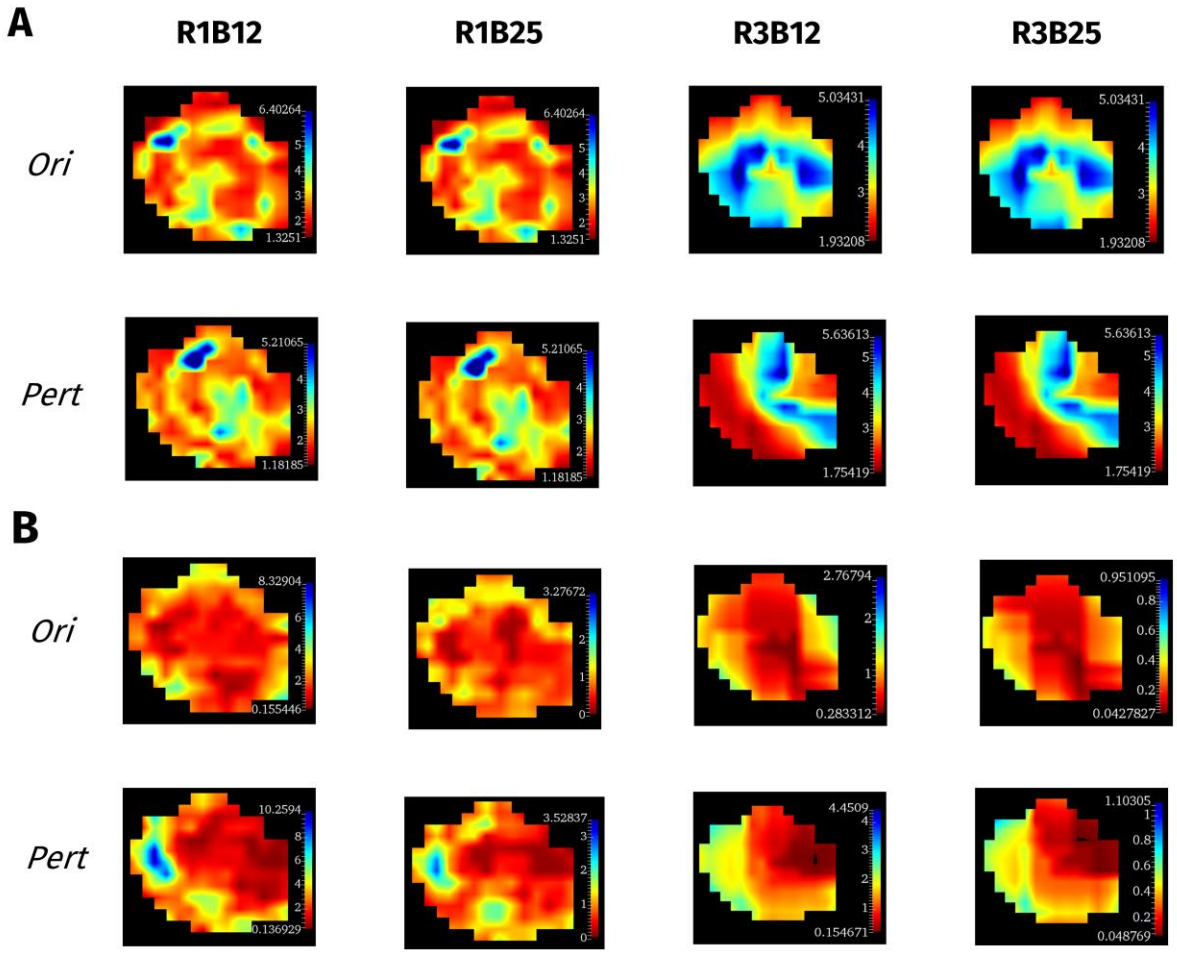

FIGURE E1. Axial section of features First-Order Kurtosis (A) and NGTDM Strength (B) from original and perturbed images of one liver lesion, extracted with 4 settings. R1B12, features extracted with kernel radius 1mm and bin size 12HU; R1B25, features extracted with kernel radius 1mm and bin size 25HU; R3B12, features extracted with kernel radius 3mm and bin size 12HU; R3B25, features extracted with kernel radius 3mm and bin size 25HU; NGTDM, Neighboring Grey Tone Difference Matrix Features; Ori, original; Pert, perturbed.

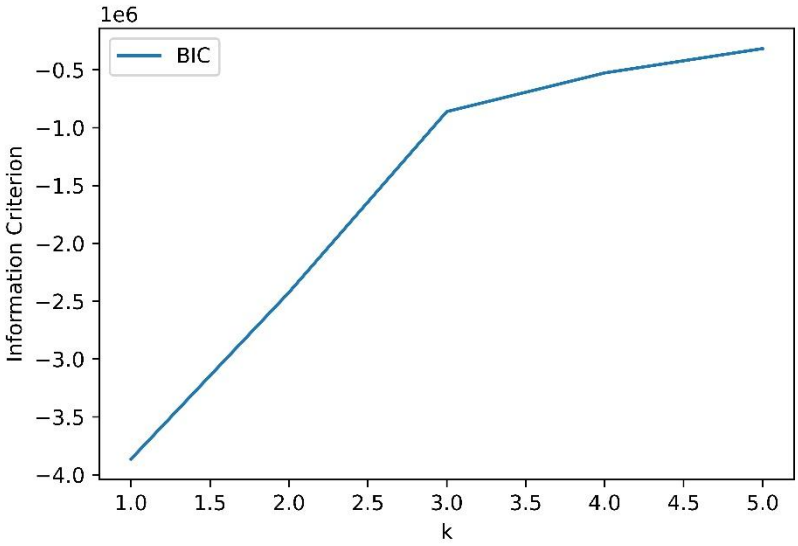

FIGURE E2. Optimal number of habitats selection. The graph displays the gradient of the Bayesian Information Criterion (BIC) against the number of clusters (k) or habitats. When fitting the data (i.e. clustering the voxel-wise features), BIC penalizes the addition of parameters that result in overfitting[30]. As it is observed, the larger the number of clusters, the larger the gradient, meaning the original BIC function keeps decreasing (and thus the likelihood increasing). However, starting from a cluster size of three the gradient of the BIC increases slower, i.e. the original function has a gentler decrease. Thus, increasing the number of clusters beyond k=3 does not result in additional information gain.

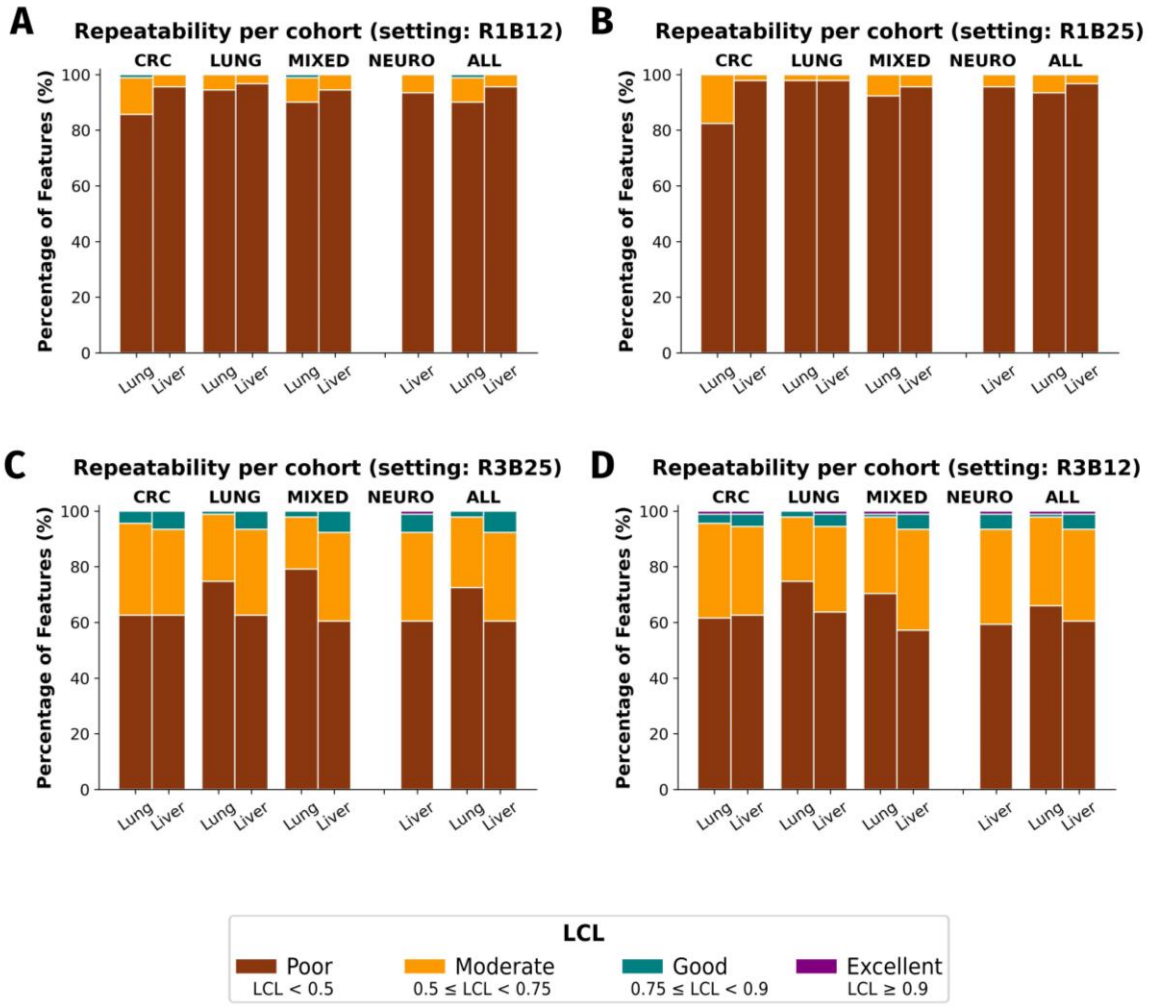

FIGURE E3. Repeatability distribution of radiomics features extracted with setting R1B12 (A), R1B25 (B), R3B12 (C) and R3B25 (D) per cohort for lung and liver lesions separately. Primary tumor has no effect on repeatability. LCL, 95% lower confidence limit of the Intraclass Correlation Coefficient; R1B12, features extracted with kernel radius 1mm and bin size 12HU; R1B25, features extracted with kernel radius 1mm and bin size 25HU; R3B12, features extracted with kernel radius 3mm and bin size 12HU; R3B25, features extracted with kernel radius 3mm and bin size 25HU; CRC: colorectal cohort; NEURO: neuroendocrine cohort; ALL: all cohorts combined.

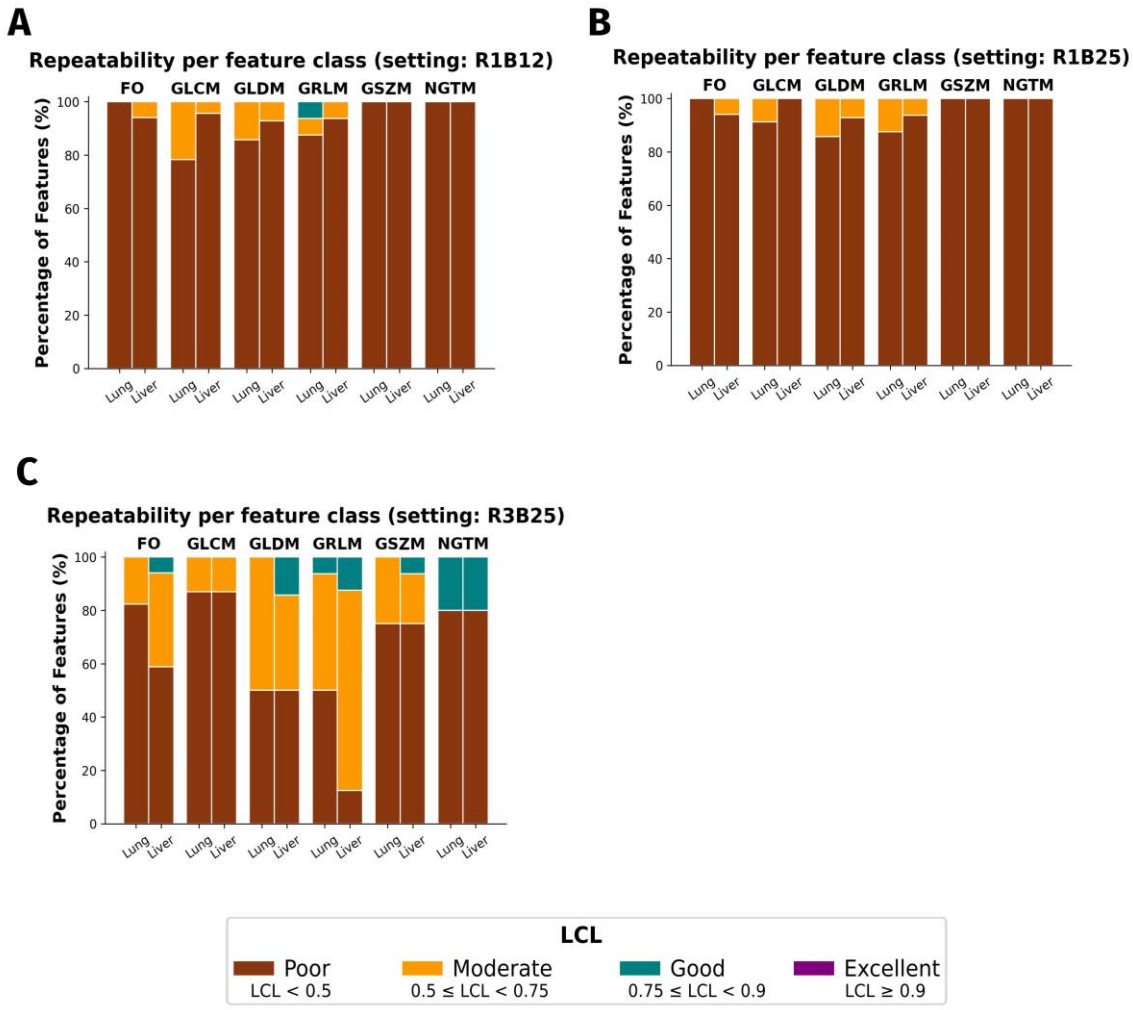

FIGURE E4. Repeatability distribution of radiomics features extracted with setting R1B12 (A), R1B25 (B) and R3B25 (C) per feature class for lung and liver lesions separately. LCL, 95% lower confidence limit of the Intraclass Correlation Coefficient; R1B12, features extracted with kernel radius 1mm and bin size 12HU; R1B25, features extracted with kernel radius 1mm and bin size 25HU; R3B12, features extracted with kernel radius 3mm and bin size 12HU; R3B25, features extracted with kernel radius 3mm and bin size 25HU; FO, First-Order; GLCM, Grey Level Co-occurrence Matrix features; GLDM, Grey Level Dependence Matrix; GLRLM, Grey Level Run Length Matrix; GLSZM, Grey Level Size Zone Matrix; NGTDM, Neighboring Grey Tone Difference Matrix Features.

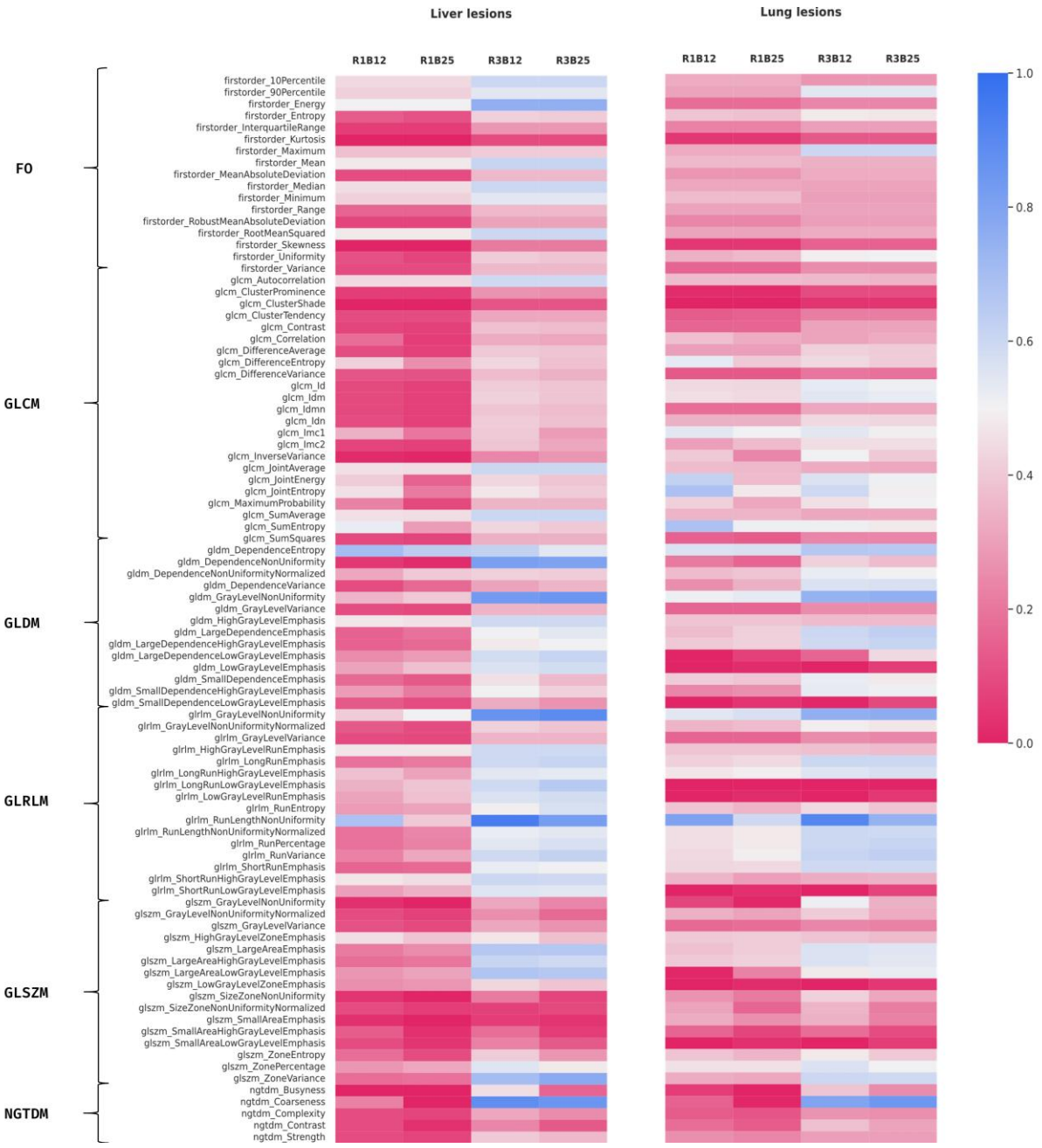

FIGURE E5. Heatmap displaying results obtained in the four repeatability experiments (one per setting) for lung and liver lesions separately. LCL, 95% lower confidence limit of the Intraclass Correlation Coefficient; R1B12, features extracted with kernel radius 1mm and bin size 12HU; R1B25, features extracted with kernel radius 1mm and bin size 25HU; R3B12, features extracted with kernel radius 3mm and bin size 12HU; R3B25, features extracted with kernel radius 3mm and bin size 25HU; FO, First-Order; GLCM, Grey Level Co-occurrence Matrix features; GLDM, Grey Level Dependence Matrix; GLRLM, Grey Level Run Length Matrix; GLSZM, Grey Level Size Zone Matrix; NGTDM, Neighboring Grey Tone Difference Matrix Features.

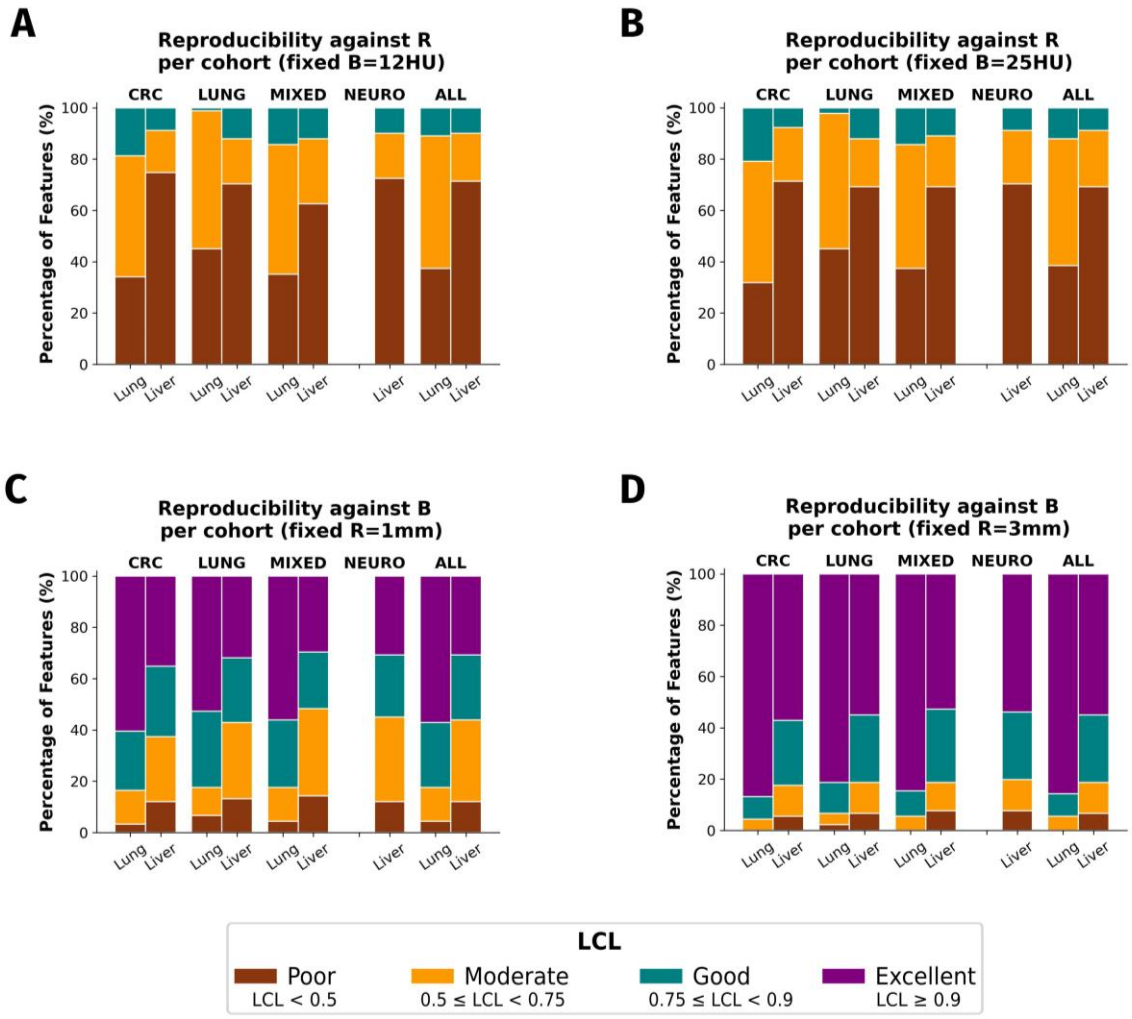

FIGURE E6. Reproducibility distribution against R of radiomics features extracted with fixed bin size of 12HU (A) and fixed bin size of 25HU (B) per cohort for lung and liver lesions separately. Similary, (C) and (D) depict the reproducibility distribution against B of radiomics features extracted with fixed radius of 1mm (C) and 3mm (D) per cohort for lung and liver lesions separately. LCL, 95% lower confidence limit of the Intraclass Correlation Coefficient; CRC: colorectal cohort; NEURO: neuroendocrine cohort; ALL: all cohorts combined.

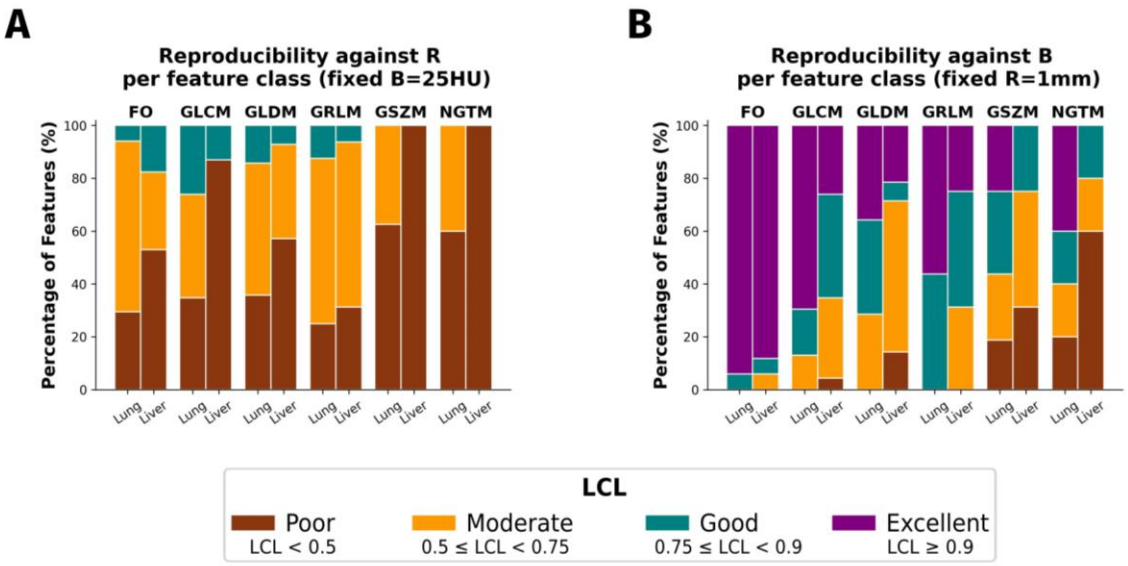

FIGURE E7. A) Reproducibility distribution against kernel radius for features extracted with fixed bin size of 25HU per feature class for lung and liver lesions separately. B) Reproducibility distribution against bin size for features extracted with fixed kernel radius 1mm per feature class for lung and liver lesions separately. LCL, 95% lower confidence limit of the Intraclass Correlation Coefficient; FO, First-Order features; GLCM, Grey Level Co-occurrence Matrix features; GLDM, Grey Level Dependence Matrix; GLRM, Grey Level Run Length Matrix; GSZM, Grey Level Size Zone Matrix; NGTM, Neighboring Grey Tone Difference Matrix Features.

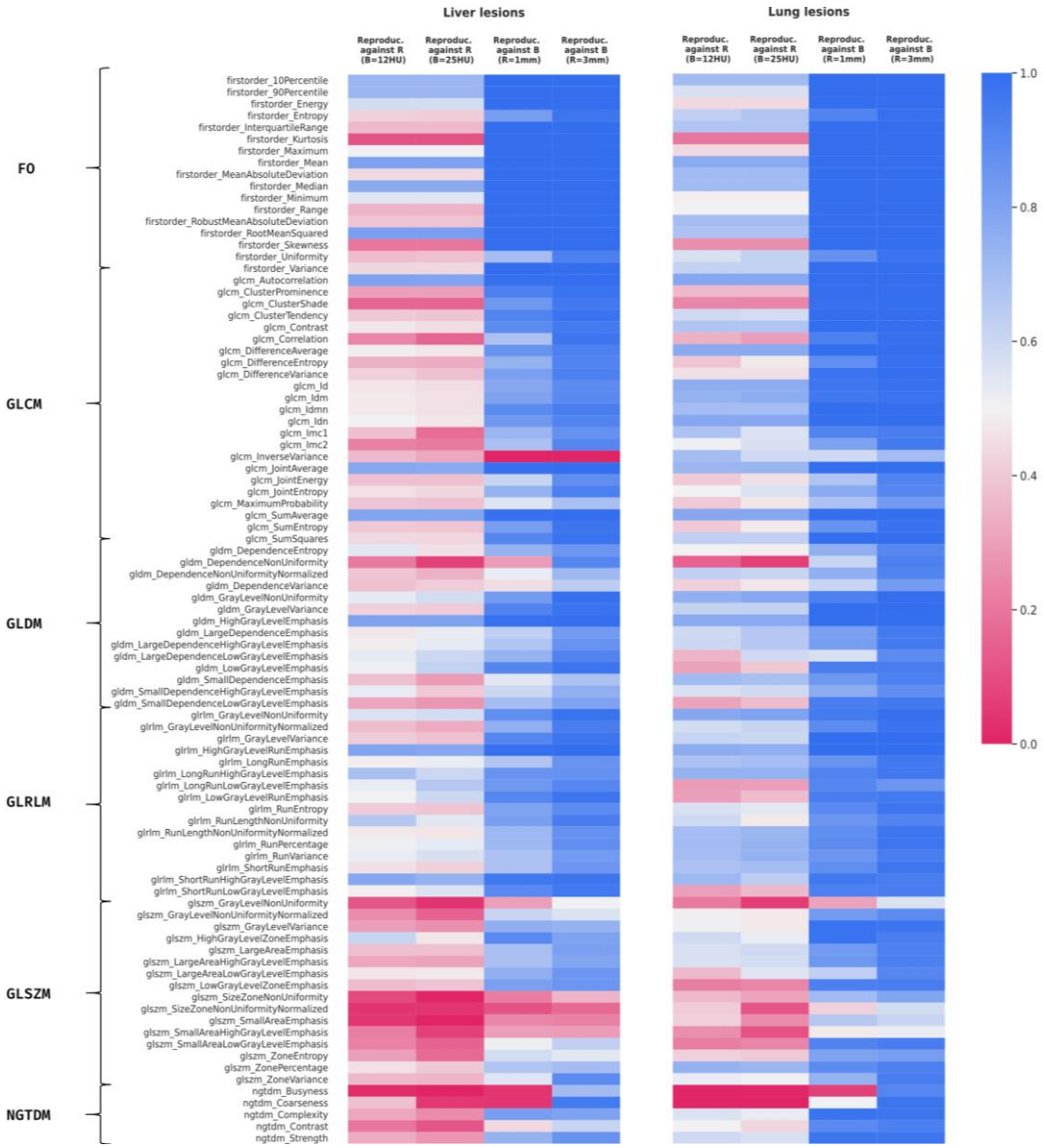

FIGURE E8. Heatmap displaying results obtained in the four reproducibility experiments for lung and liver lesions separately: reproducibility against R (fixed B=12HU), reproducibility against R (fixed B=25HU), reproducibility against B (fixed R=1mm), reproducibility against B (fixed R=3mm). LCL, 95% lower confidence limit of the Intraclass Correlation Coefficient; FO, First-Order; GLCM, Grey Level Co-occurrence Matrix features; GLDM, Grey Level Dependence Matrix; GLRLM, Grey Level Run Length Matrix; GLSZM, Grey Level Size Zone Matrix; NGTDM, Neighboring Grey Tone Difference Matrix Features.

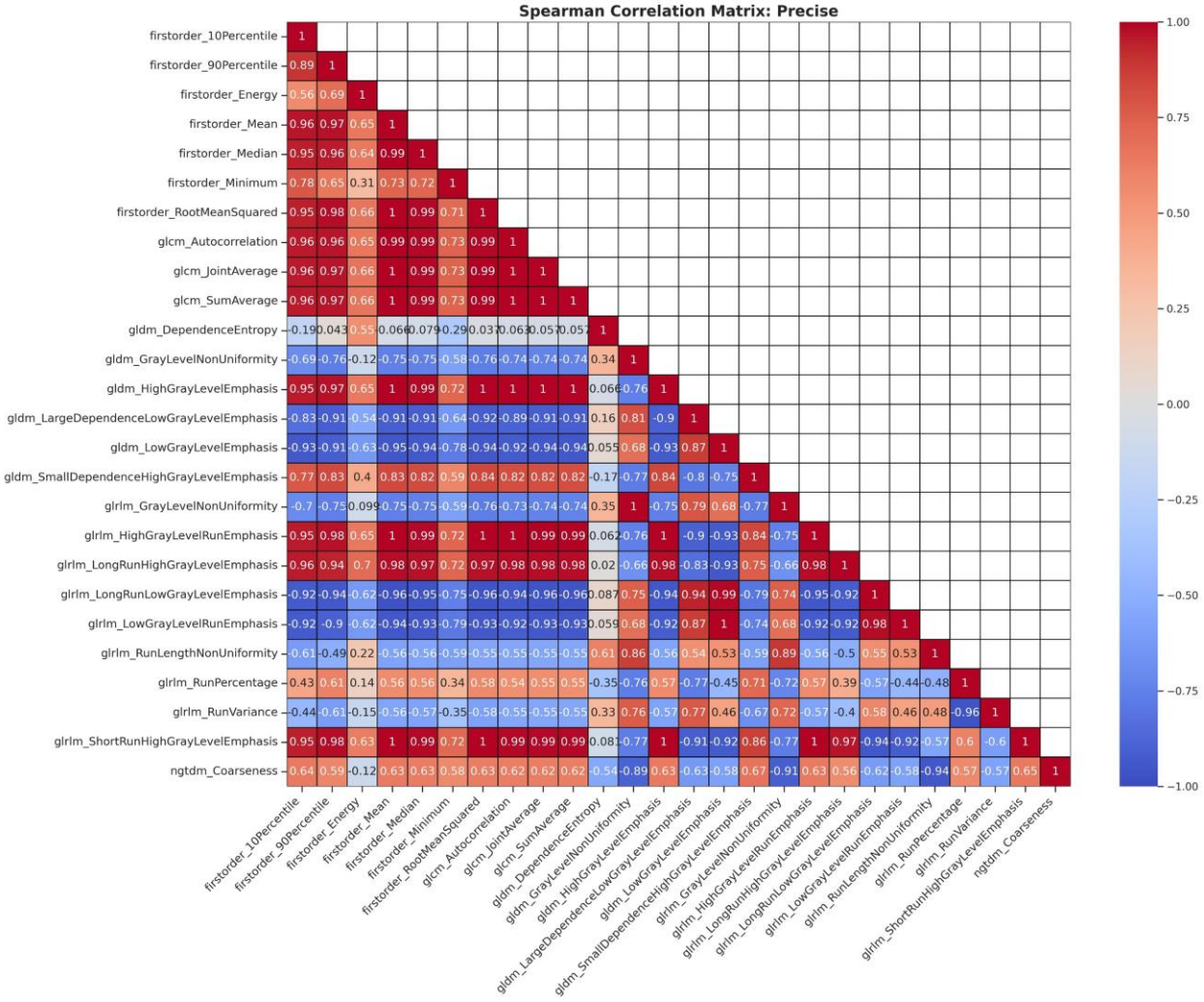

FIGURE E9. Correlation heatmap of precise RF prior to habitat computation for one liver lesion (Figure 5). Correlation was assessed using Spearman rank correlation coefficient, with p<.005.

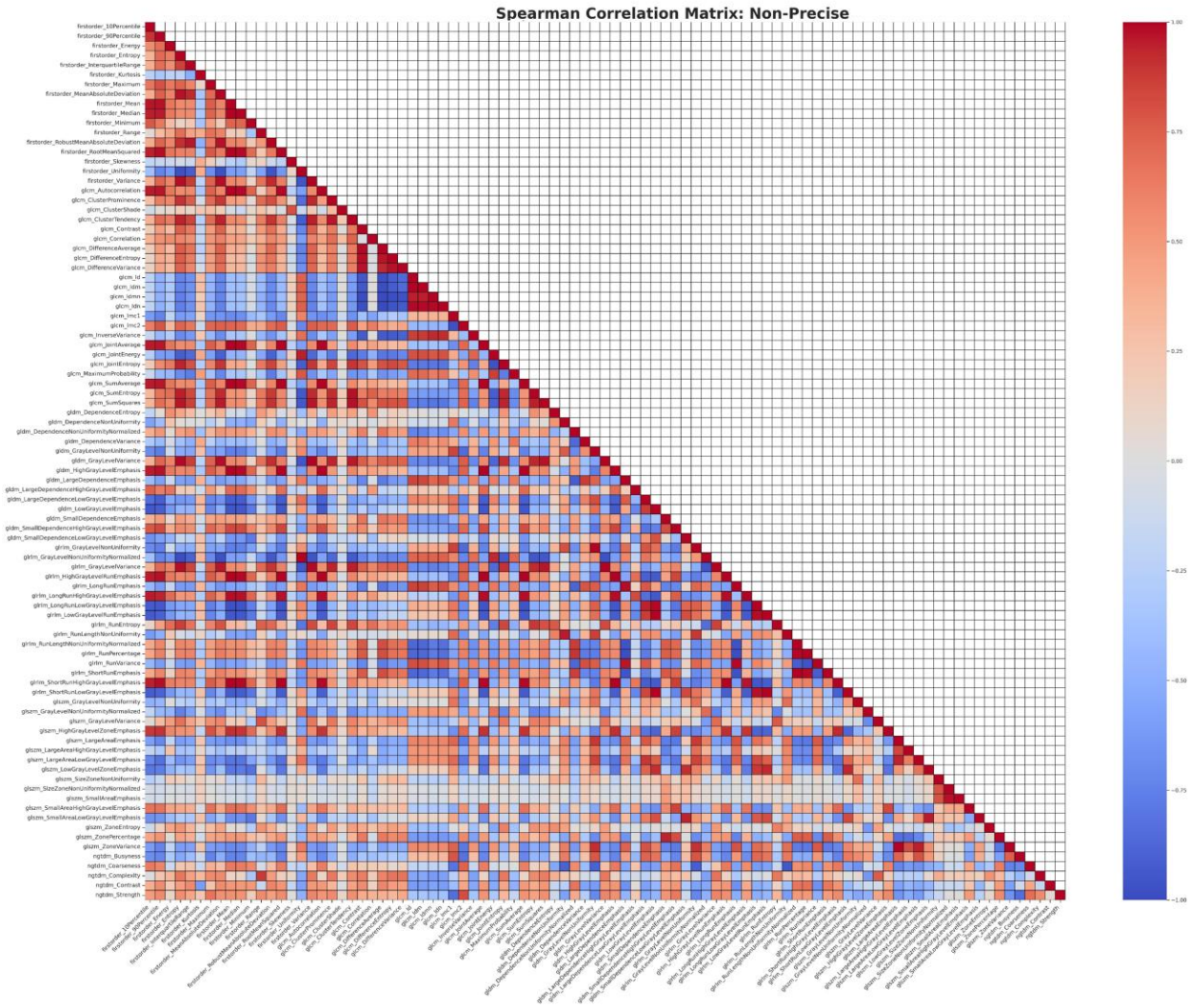

FIGURE E10. Correlation heatmap of non-precise RF prior to habitat computation for one liver lesion (Figure 5). Correlation was assessed using Spearman rank correlation coefficient, with p<.005.

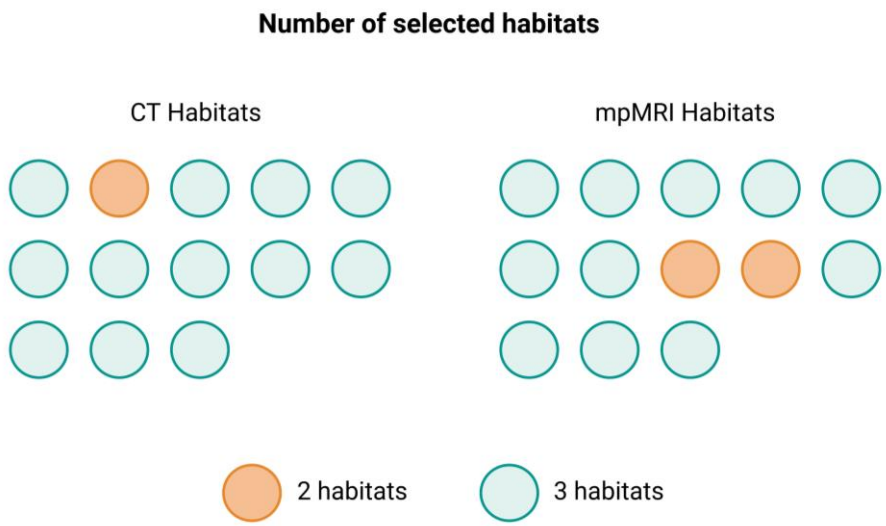

FIGURE E11. Number of selected habitats by the Bayesian Information Criterion (BIC) for both CT and mpMRI habitats in the independent cohort (13 liver lesions). Each circle represents a lesion. The model computed the same number of habitats for both modalities in 10 out of 13 patients, indicating a that the algorithm is robust in classifying regions with different phenotypes with both CT and mpMRI data.