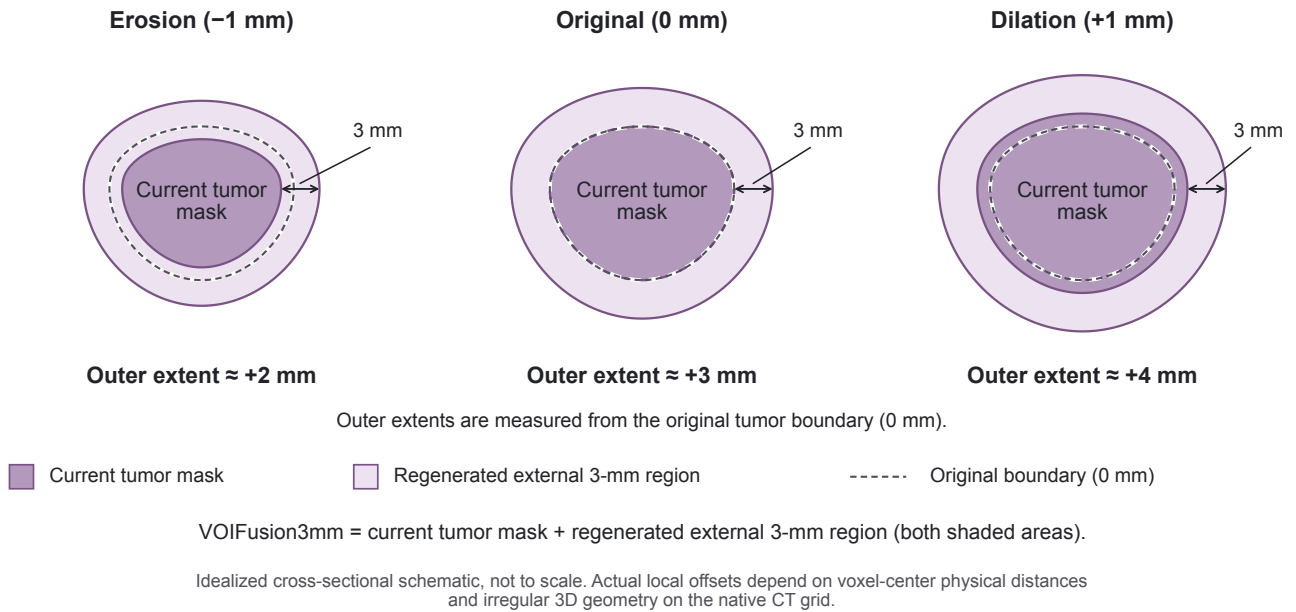
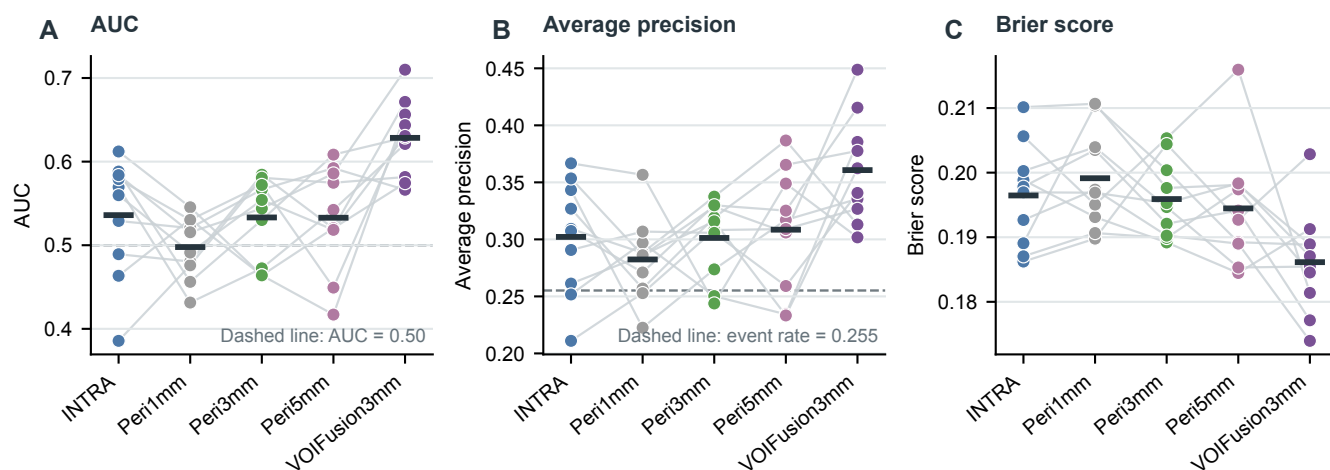


Perturb tumor mask → Recompute distance → Regenerate external 3-mm region → Form fusion

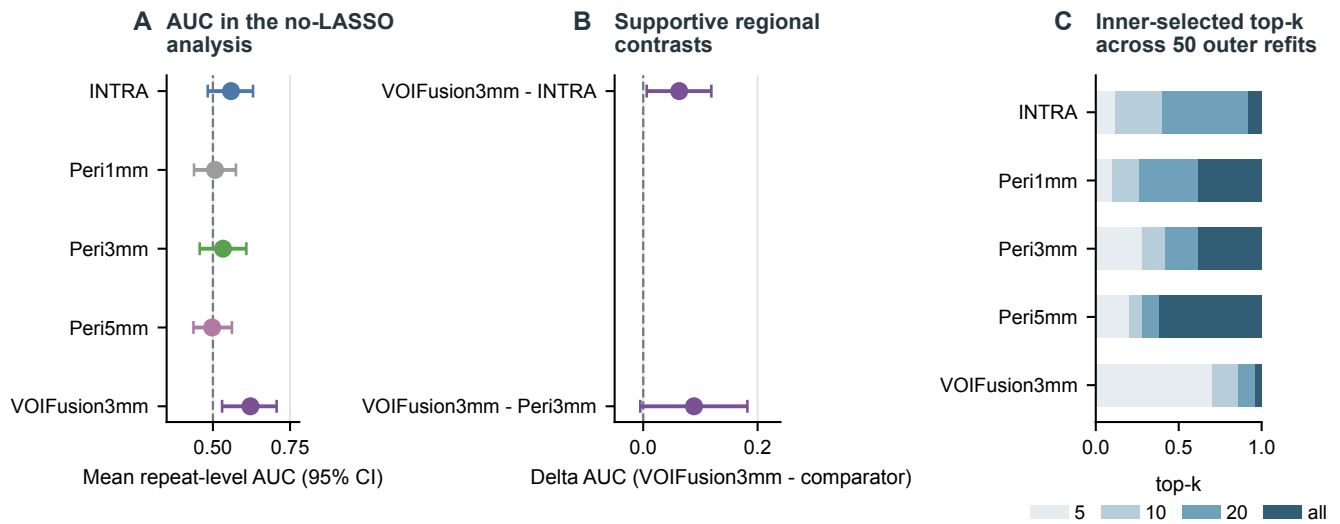


Supplementary Figure S1. Geometry of tumor-boundary perturbation and regenerated VOIFusion3mm. Schematic for the prespecified supportive fixed-model sensitivity analysis. The tumor mask was eroded or dilated by 1 mm, after which the signed-distance map was recomputed and the external 3-mm region regenerated from the perturbed boundary. Under idealized continuous geometry, the resulting outer extent of VOIFusion3mm is approximately 2 mm beyond the original tumor boundary after erosion, 3 mm for the original mask, and 4 mm after dilation. Exact local offsets depend on voxel-center physical distances and irregular three-dimensional tumor geometry on the native CT grid. The schematic is not to scale.

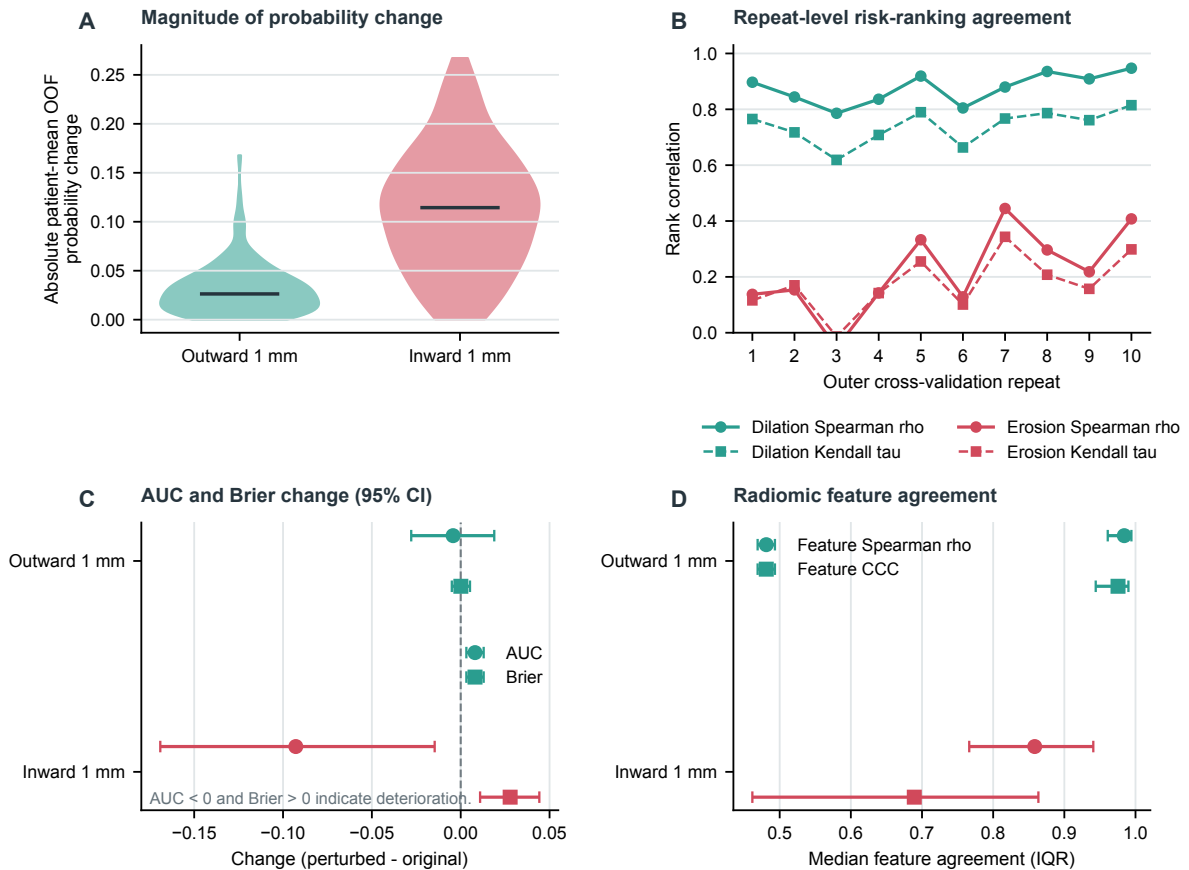


Gray lines connect the same repeat; black bars denote the mean across 10 repeats.

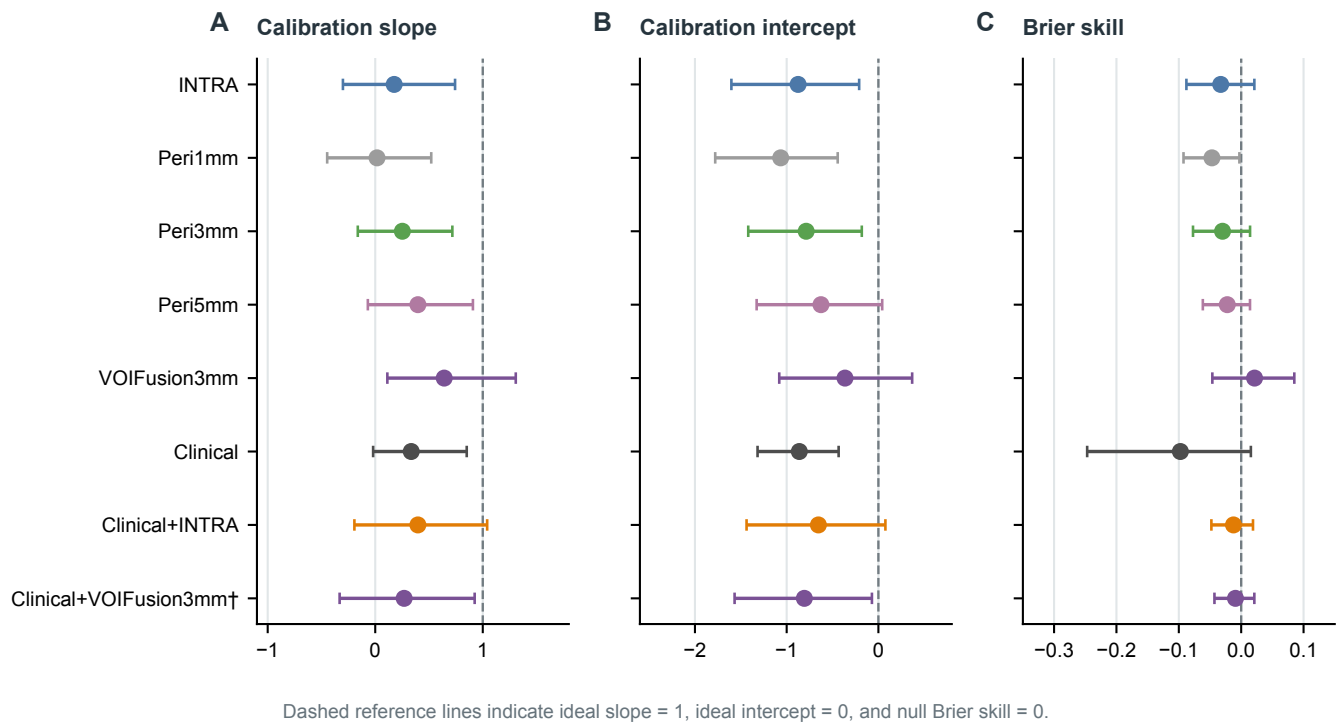
Supplementary Figure S2. Primary repeat-level discrimination and probability error. Descriptive visualization of the prespecified confirmatory primary spatial analysis; confirmatory paired comparisons are reported in Table S8. Pooled out-of-fold (A) AUC, (B) average precision, and (C) Brier score for each spatial representation in each of the 10 outer-cross-validation repeats. Gray lines connect estimates from the same repeat; colored points show repeat-specific estimates, and horizontal black bars show the mean across repeats. Reference lines denote an AUC of 0.50 and the observed event prevalence of 37/145 (0.255) for average precision. Lower Brier scores indicate lower mean squared probability error. AUC indicates area under the receiver operating characteristic curve; OOF, out-of-fold.



Supplementary Figure S3. Prespecified fallback-free no-LASSO sensitivity analysis. (A) Mean repeat-level pooled out-of-fold AUC and patient-cluster bootstrap 95% confidence interval for all five regions under the fallback-free no-LASSO specification. The dashed line denotes AUC = 0.50. (B) Supportive VOIFusion3mm-minus-INTRA and VOIFusion3mm-minus-Peri3mm AUC differences with paired patient-cluster bootstrap 95% confidence intervals. (C) Distribution of the top-k value selected within inner cross-validation for the 50 outer refits per region. All 250 ROI-specific outer refits completed without LASSO, forced minimum-feature retention, or fallback. This was a prespecified supportive sensitivity analysis, not an additional confirmatory test. AUC indicates area under the receiver operating characteristic curve; CI, confidence interval; INTRA, intratumoral region; LASSO, least absolute shrinkage and selection operator.



Supplementary Figure S4. Sensitivity of held-out predictions to physical boundary perturbation. (A) Distribution of the absolute change in patient-mean out-of-fold probability after 1-mm outward dilation or inward erosion of the reviewed tumor contour; horizontal bars denote medians. (B) Repeat-level Spearman and Kendall rank correlations between original and perturbed predictions. (C) Paired changes in AUC and Brier score, defined as perturbed minus original, with patient-cluster bootstrap 95% confidence intervals. For AUC, values below 0 indicate deterioration; for Brier score, values above 0 indicate deterioration. (D) Median and interquartile range of feature-level Spearman correlation and concordance correlation coefficient. This prespecified supportive fixed-model sensitivity analysis reused the 50 original VOIFusion3mm outer-fold models without feature reselection, retuning, probability recalibration, or refitting; it does not validate segmentation robustness. AUC indicates area under the receiver operating characteristic curve; CCC, concordance correlation coefficient; CI, confidence interval; OOF, out-of-fold.



Supplementary Figure S5. Calibration and Brier-skill summary. Mean across 10 complete repeat-level out-of-fold prediction sets and patient-cluster bootstrap 95% confidence interval for (A) calibration slope, (B) calibration intercept, and (C) Brier skill relative to an outer-training-prevalence reference. Dashed reference lines denote an ideal calibration slope of 1, ideal calibration intercept of 0, and null Brier skill of 0. These summaries are descriptive; † marks Clinical+VOIFusion3mm, whose integration was exploratory and initiated after the primary spatial comparison. OOF indicates out-of-fold.