

Attribution Fingerprinting Framework (Bidirectional Validation)

Key numbers
mask_free rev AUROC: 0.9905
predicted_mask deg: 28-49%
Clinical: Fwd FAIL / Rev PASS

1) Inputs: ID dataset, OOD dataset, trained U-Net segmenter

2A) Endpoint branch: predicted_mask
(mask-dependent features)

2B) Endpoint branch: mask_free
(mask-independent features)

3) Shift scoring: mean absolute z-score vs ID reference

4) Validation gates: Gate-3 (discrimination), Gate-4 (robustness), Gate-5 (clinical linkage + determinism)

**5A) Decision: mask_free
PREFERRED for deployment**

**5B) Decision: predicted_mask
INVESTIGATIONAL (needs hardening)**