SeleneX — Risk & Bias Log

Purpose

Track known risks/biases and planned mitigations for the SeleneX mini-prototype.

Risks & Mitigations

R1. Synthetic Biomarker Bias. CA-125 has a hard cutoff at 35 U/mL and malignant values are skewed higher.
Mitigation: Scale features (e.g., z-scoring), regularize tabular encoder, use class weights; report calibration plots; emphasize limitation in documentation.

R2. Overfitting on Small Dataset. Limited ultrasound images may cause the image encoder to memorize.
Mitigation: Early stopping on ROC-AUC, data augmentation, dropout/L2, consider frozen pretrained backbones; report learning curves.

R3. Case Leakage Risk. Multiple frames per patient can leak across splits if not grouped.
Mitigation: Case-level consistency in splitting; verify no patient overlap between train/val/test.

R4. Image Acquisition Variability. Different scanners, presets, or artifacts may degrade generalization.
Mitigation: Robust augmentations (brightness/contrast/rotation); stratify evaluation if metadata available.

R5. Imbalance and Clinical Cost. False negatives are particularly harmful.
Mitigation: Use class weights, tune threshold for higher sensitivity, report per-class metrics and ROC/PR curves.

Monitoring

- Track per-split class balance, per-class recall, confusion matrices.
- Maintain experiment logs (hashes, seeds, config) for reproducibility.

Residual Risks

Despite mitigations, synthetic tabular distributions and limited sample size may still misrepresent real-world performance.