POCUS Lesion Classification Pipeline

1. Purpose of the Pipeline

This pipeline is designed to detect breast lesions from community-acquired POCUS ultrasound images. It integrates:

- EfficientNetV2 for feature extraction,
- Multimodal fusion of clinical metadata,
- Attention mechanisms to highlight diagnostically relevant regions,
- Grad-CAM++ for clinical interpretability.

2. Workflow Steps

Step 1 – Data Exploration

Run:

bash

python explore_data.py

Generates:

- 07_metadata_distribution.png → distribution of metadata (image quality, BI-RADS, breast composition).
- 08_sample_images.png → sample of 12 dataset images.
- 09_label_distribution.png → lesion vs non-lesion distribution.
- 10_image_properties.png → image properties (size, normalization).

Results folder: hackaton_bon/results/

Step 2 – Training with Visualizations

Run:

bash

python train_cv_with_plots.py

Process:

- 5-fold stratified cross-validation training.
- Best model saved automatically.
- Graphs generated for performance monitoring.

Outputs:

- 01_loss_curves.png → training and validation loss curves.
- 02_auc_accuracy_curves.png → AUC and accuracy per fold.
- 03_confusion_matrices.png → confusion matrices for each fold.
- 04_roc_curves.png → ROC curves per fold.
- 05_probability_distribution.png → probability histograms.
- 06_cv_summary.png → summary of metrics (mean ± std).

Results folder: hackaton_bon/results/

Best model: hackaton_bon/models/best_model_cv.pth

Step 3 – Evaluation & Interpretability

Run:

bash

python evaluate_advanced.py

Generates:

- Grad-CAM++ heatmaps showing regions most influential in predictions.
- Correct detections: heatmaps coincide with true lesions.
- False positives: artifacts misinterpreted as lesions.
- Imprecise localization: overly broad heatmaps.

Grad-CAM results: hackaton_bon/gradcam_results/

3. Key Results
   - Mean AUC: 0.936 ± 0.032
   - Mean Accuracy: 0.908 ± 0.027
   - Mean Loss: 0.063 ± 0.015
   - Interpretability: Grad-CAM++ confirms focus on clinically relevant lesion regions.
4. Validation Checklist
   - Dataset explored and balanced.
   - AUC consistently > 0.90.
   - No significant false negatives.
   - Grad-CAM++ explanations validated by clinicians.
   - Documentation complete.
   - Best model saved and ready for deployment.
5. Additional Resources
   - DOCUMENTATION.md → overview of the pipeline.
   - DOCUMENTATION_DETAILLEE.md → detailed technical documentation.
   - README.md → quick start guide.
   - Results: hackaton_bon/results/
   - Models: hackaton_bon/models/
   - Grad-CAM outputs: hackaton_bon/gradcam_results/