MODEL CARD:

Purpose:

This prototype combines ultrasound images and biomarkers (**CA-125, BRCA, Age**) to classify ovarian tumors as **benign or malignant.**

Dataset:

MMOTU ovarian ultrasound dataset with 8 classes collapsed into binary labels (benign vs malignant).
Biomarkers: CA-125, BRCA status, Age.

Models Used:

• Image-only CNN (ResNet backbone)
• Tabular-only MLP
• Fusion of CNN + MLP

Key Metrics:

Image-only:
Accuracy 97%
ROC AUC 0.80

Tabular-only:
Accuracy 97%
ROC AUC 1.00

Fused:
Accuracy 97%
ROC AUC 0.96

Explainability:

We used **two complementary techniques** to make the model's decisions more interpretable:

- **Grad-CAM:**
  Applied to the **ultrasound image** for the model. Grad-CAM produces a heatmap showing the regions of the ovary image that contributed most to the prediction (benign vs malignant).
  This helps clinicians visually validate whether the model is focusing on the tumor area.

- **SHAP:**
  Applied to the **biomarker data** (Age, CA-125, BRCA mutation status).
  SHAP values quantify how much each biomarker pushed the prediction toward benign or malignant for each patient.
  For example, a **high CA-125 value** might strongly increase the malignant score, while **younger age** might shift predictions toward benign.

Together, these provide **multi-modal transparency**:
- clinicians can see where the model looks and why certain biomarker values matter (SHAP).

Limitations:

Small dataset size:

The MMOTU dataset is limited in number of patient cases. Small datasets can cause models to learn specific data instead of generalizable patterns.

Risk of overfitting:

With deep CNNs and few samples, the model may memorize training images rather than learning true tumor features. **This may explain the unusually high accuracy but low precision/recall.**

Synthetic splits:

The train/validation/test split is created algorithmically. If images from the same patient accidentally appear in multiple splits, there is a risk of **data leakage**, artificially inflating performance.

Limited biomarker set:

Only 3 features (Age, CA-125, BRCA) are included. **In actual scenario diagnosis considers many more clinical and genetic factors**, so this pipeline is **not comprehensive**.

Not clinically validated:

This prototype is purely experimental and not approved for real-world diagnostic use.