Breast Cancer Detection

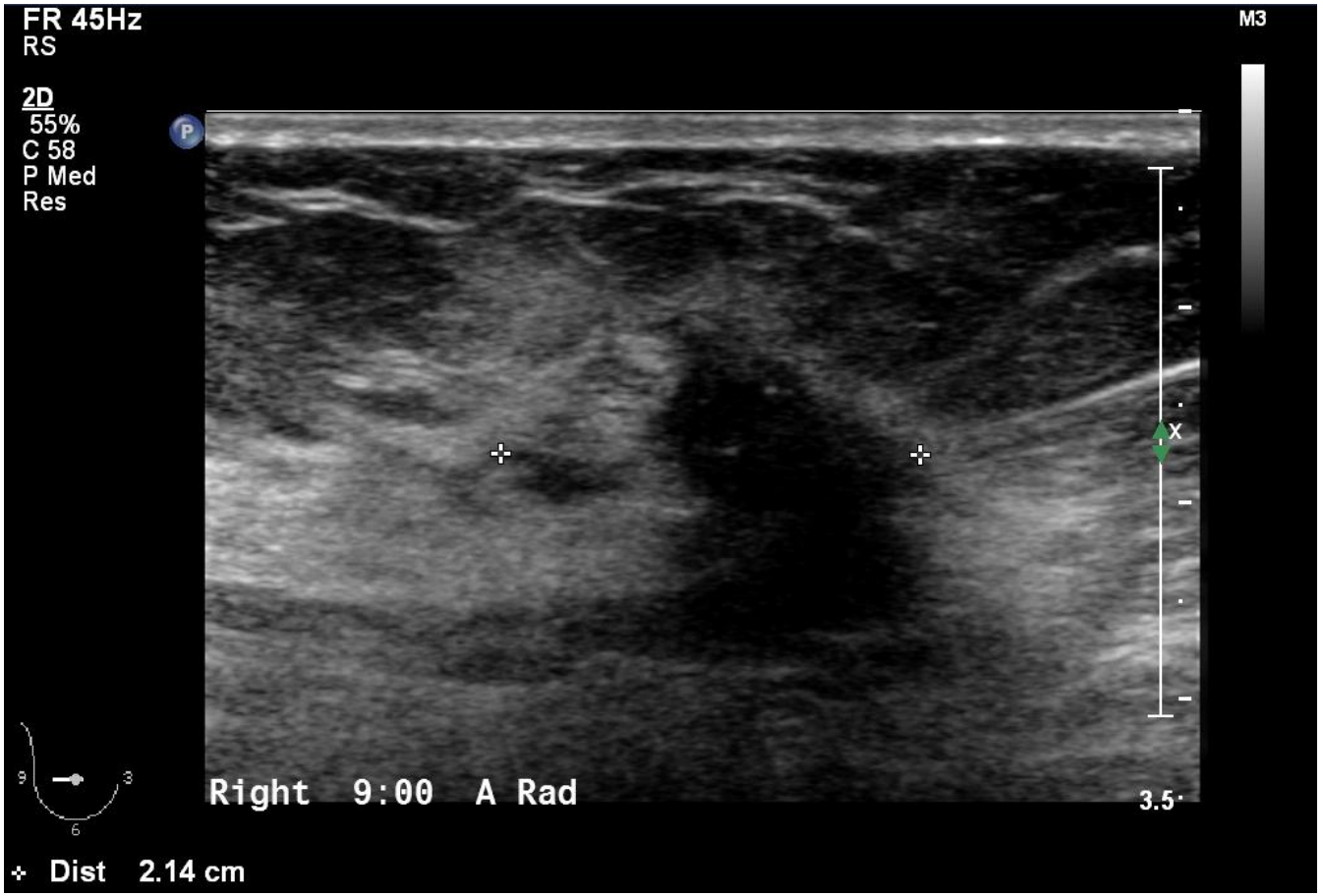
- Participate in a competition about breast cancer detection
  - This task is to develop a model to classify the molecular subtype of breast cancer based on ultrasound images and clinical diagnostic data.
  - Link: https://biendata.com/competition/detection2/
  - Submit a report and your code.
  - The report should include, but is not limit to, methods, experimental results, related work and references.
  - Your nickname and your rank in the competition should be also included in the report

Dataset

- You are given hundreds of medical records of breast cancer patients. Each medical record is associated with several ultrasound images and some clinical diagnostic data.
- The clinical diagnostic data contains the following fields:

Field	Type	Meaning
id	str	Patient ID
age	int	Age of the patient
HER2	int	Scale (0-3) of how strongly HER2 (marker for genetic predisposition for breast cancer) is detected
P53	bool	Whether P53 (marker for genetic predisposition for cancer) is positive
molecular_subtype	int	Molecular subtype of breast cancer, there are four types of molecular subtypes in the dataset (1: Luminal A, 2: Luminal B, 3: HER2-Enriched, 4:Triple Negative)

Ultrasound Image Example



Evaluation

- The dataset has been split into training/testing parts. The training data (with label) and testing data (without label) will be given.
- You can submit the predictions of testing data to evaluate your model.

- Evaluation Metric: Accuracy and Macro-F1

$$Accuracy = \frac{1}{N}\sum_{i=1}^{N}\mathbb{1}(\hat{y}_i = y_i)$$

$$Macro\text{-}F1 = \frac{1}{C}\sum_{c=1}^{C}F1_c = \frac{1}{C}\sum_{c=1}^{C}\frac{2 \times Precision_c \times Recall_c}{Precision_c + Recall_c}$$