Breast Cancer Detection

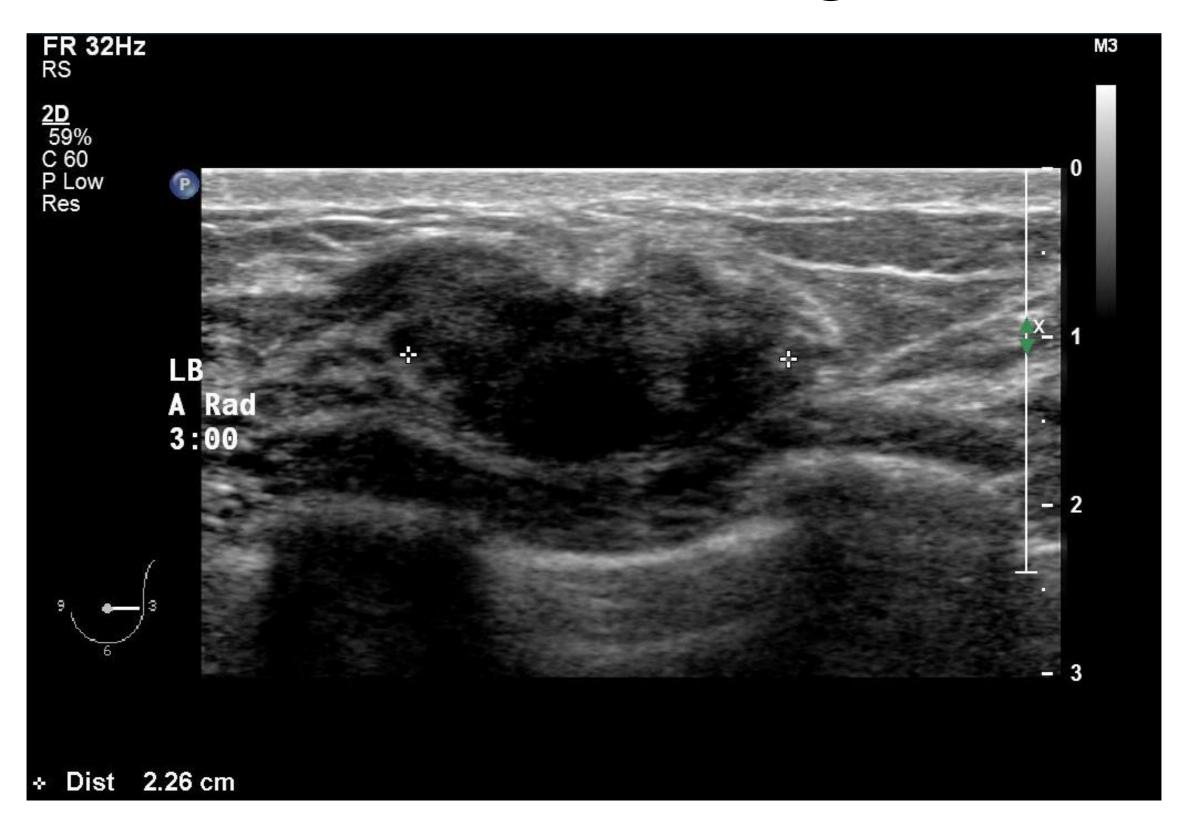Group member: Yin Yifan,   Ding Xiangyun

Problem Description

• Molecular subtyping of breast cancer has become common practice to understand prognosis of disease, and to design a treatment plan. The subtype indicates the severity of the cancer and influences the treatment plan.

• This task is to develop an automated method to classify the molecular subtype of breast cancer based on ultrasound images and clinical diagnostic data.

Dataset

• 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 images

Expected output

- Algorithm for automatic analysis of ultrasound images and clinical data.

- A framework to automatically classify the molecular subtype of breast cancer.

- An online website for medical staff.

Project plan

- Learn how to read ultrasound image and basic breast cancer knowledge.

- Read related papers about ultrasound image procession and breast cancer detection.

- Detection algorithm design and implementation. Evaluate the model.

- Build and deploy the website.

- Presentation.

Data pre-process

- Resize the images to have the same distance scale.

- Remove artifacts from the ultrasound images.

- Data augmentation to increase the size of training set, for example, picture rotation, flipping, et al.

- Split the dataset to training set and validation set.

Cancer detection

- Try different models, including traditional machine learning methods and deep learning methods, for prediction.

- Experiments on different neural networks.

- Try to mark the cancer area on each picture.

Schedule

Week	Main task
9-10	Learn related medical knowledge and coding skills. Read papers and research related methods. Determine the approximate code framework.
11-12	Implement the algorithm. Make a simple demo.
13-14	Optimize our algorithm. Try different models and methods.
15-16	Visualize the results on the website. Prepare presentation.

Division of labor

- Ding Xiangyun: method research; website deployment.

- Yin Yifan: algorithm implementation; test and experiments.

Thank you.