Image Processing and Analysis / Advanced Image Analysis

2022-2023, 2nd semester

IPA/AIA standard project

Mass Detection

Alessandro Bria

Associate Professor

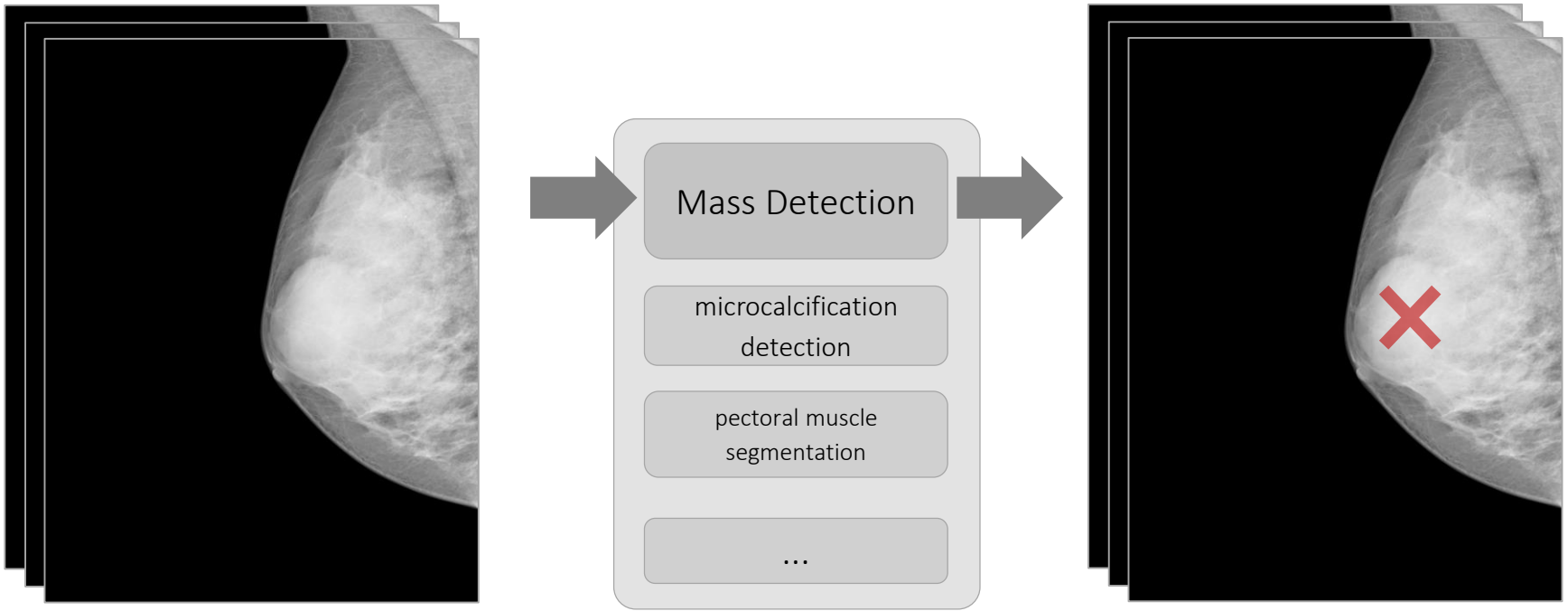
email: a.bria@unicas.it

Motivations

- **X-ray mammography** is a widely used method to screen women for early detection of **breast cancer**
- **Computer Aided Diagnosis** (CAD) helps radiologists in interpreting screening mammograms
- the two most important lesions that may be present on a mammogram are **microcalcifications** and **masses**
  - CAD often consists first in **detecting** the lesions and then **classifying** them into benign / malignant

Goal

- implement a system for automated **Mass Detection**
  - a must-have module in most **CAD** systems for mammography

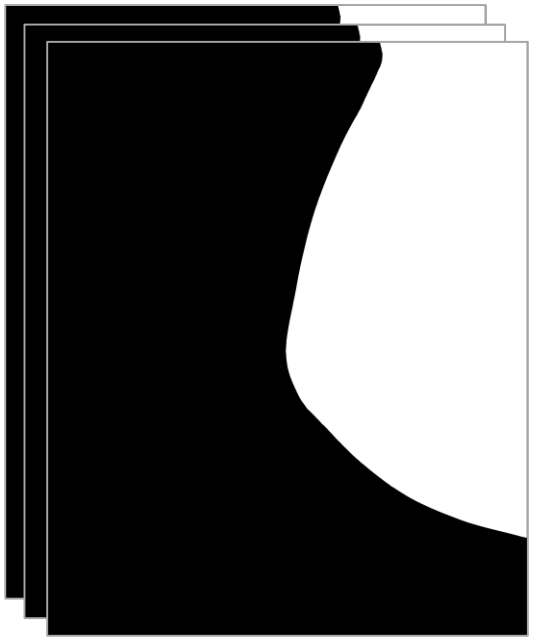


Materials

- **InBreast** dataset (410 images) containing:

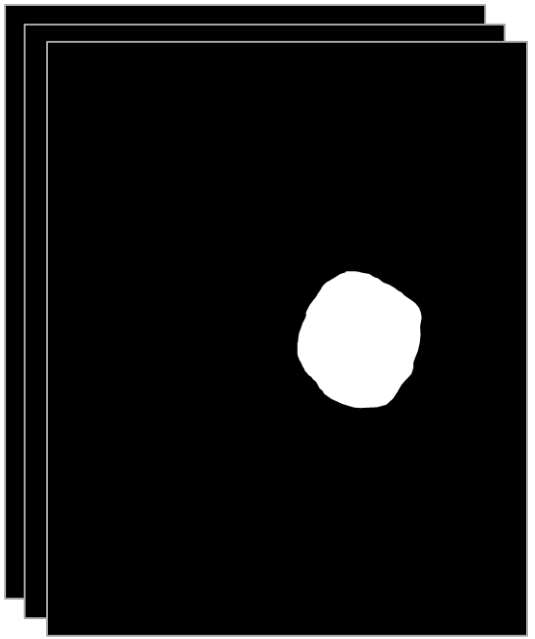
mammograms (16-bit)

`/dataset/images`



breast-air masks

`/dataset/masks`



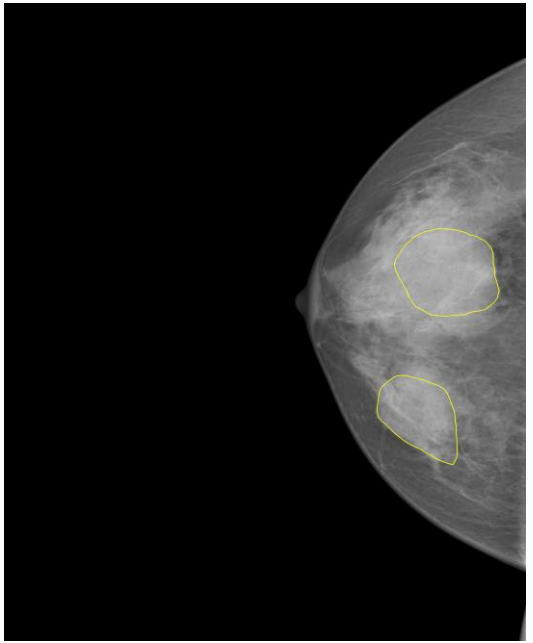
manual annotations

`/dataset/groundtruths`

Materials

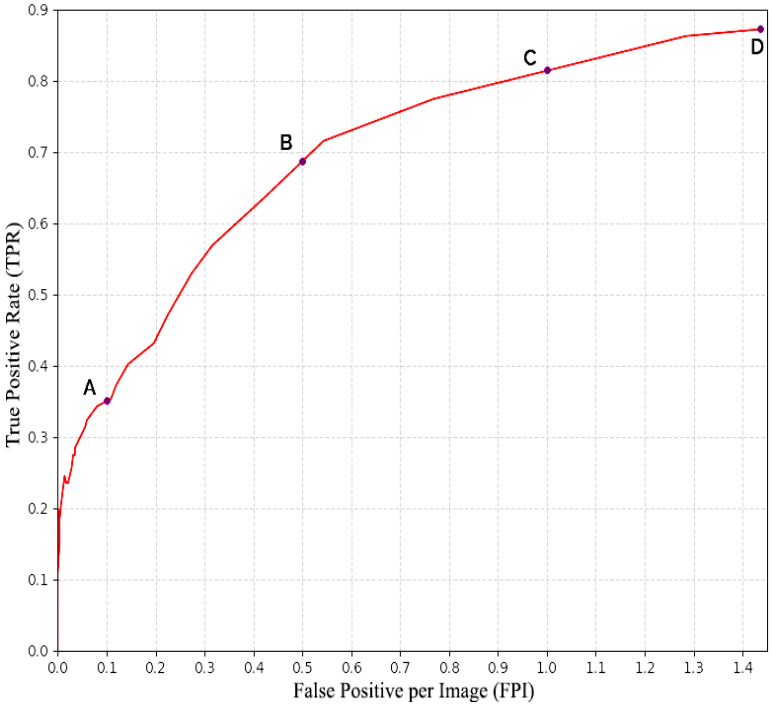
- warning: among the 410 images, only 107 contain masses (positive images)
  - there are only 107 manual annotations in the `/dataset/groundtruths` folder
  - see also the 107 overlayed annotations in the `/dataset/overlay` folder





Performance evaluation

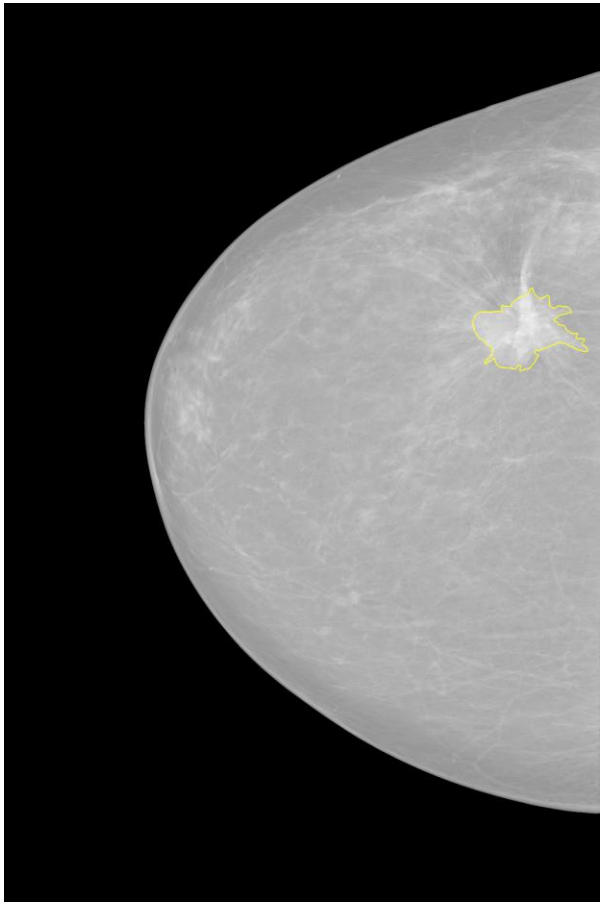
- **FROC** curve
  - True Positive Rate (**TPR**) vs. False Positive Rate per Image (**FPpI**)
    - *true positive* = a CAD finding that **matches** with a groundtruth finding
    - *false positive* = a CAD finding that *does not* **match** with any of the groundtruth findings
      - true positives should be evaluated on the 107 positive images, whereas false positives on the remaining 303 images
  - a match is found when the Dice Similarity Index between the two findings is above 0.2:
    - $match(A,B)$ if $DSI(A,B) = \frac{2|A \cap B|}{|A|+|B|} \geq 0.2$
  - the curve is obtained by varying the decision threshold of the CAD system



reasonable expectation from this course project using `ucasML` tool

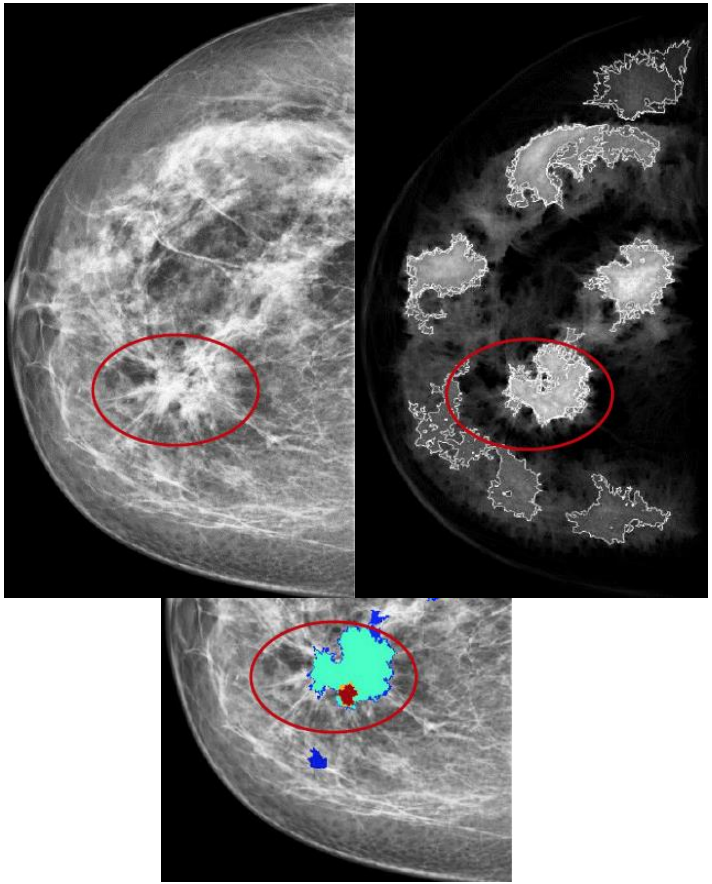
Challenges

- masses may have irregular shapes
- some low-contrasted masses
- masses can be confounded with other structures
- unbalanced data
  - it is likely you will obtain a true-positive-false-positive ratio of 1:100 or higher in your candidate lesions set
    - this will be handled by **`ucasML`**



Hints

- **preprocessing** is important
  - contrast enhancement (CLAHE, etc.)
- **candidate extraction / segmentation**
  - oversegmentation with one of the techniques learnt from *AIA*
- **feature extraction**
  - shape features, texture features (GLCM), SIFT, LBP
- **classification**
  - strongly suggested to use **`ucasML`** tool, it is robust to class imbalance
  - training set definition is crucial
    - the higher the IoU, the purer the positive samples, the easier the learning task, but the less generalizable is the classifier….this is clearly a trade-off



Constraints

- decision (classification) with machine learning (**`ucasML`**)
- 10-fold **cross validation**
  - partition the dataset image-wise in 10 subsets
  - train on 9 subsets, test on the remaining subset
  - repeat the previous step for 10 iterations with 10 different test subsets
  - within each training set, split into training and validation set as you want
  - merge predictions from all 10 test sets
  - evaluate performance (FROC)