EN. 520.433 / 623 Medical Image Analysis
Final Project 1 – Mammography abnormality detection and diagnosis

03-05-2018

In this project, you are supposed to propose a method that can automatically detect and diagnose abnormalities based on the given mammography dataset.

(a) In particular, your algorithm should be able to:

1) Diagnose whether the subject is **healthy** (i.e. no mass) OR **suspicious** (i.e. has mass).
   Hint: The following methods may be useful for this task:
   - K-means
   - Histogram analysis
   - Rose diagram
   - Other
2) If suspicious, produce a mask image for the mass (i.e. intensity value = 1 for pixels belonging to mass, and 0 elsewhere), AND diagnose whether it is **malignant** OR **benign**.
   Hint: The following methods may be useful for this task:
   - Region growing
   - Rose diagram
   - Other

(b) To train your algorithm, you will be given:

1) Mammography dataset with **21** subjects. For each subject, scans of both left and right breast will be provided. **7** subjects are healthy, and 14 subjects have mass. Among the 14 subjects, **7** subjects are benign, and **7** subjects are malignant.
2) A table of truth that has:
   - Subject IDs.
   - Truth of diagnosis: 0 for **healthy** subjects, 1 for **benign** subjects, and 2 for **malignant** subjects.
3) 14 mask images for the 14 subjects with mass created by human raters. Note, since the mask images for training are not tightly segmenting the mass, you don't have to generate tight masks in your algorithm.

(c) Well before the demo day, you will be supplied evaluation code, and you should run the evaluation code on your machine to guarantee the evaluation code works well with your algorithm. After that, you will be asked to submit your code to the TAs before the demo day. The final evaluation will be done on TAs' machine based on the final code you submit.

1) To test your algorithm, the TAs will use:

Another set of mammography images of *N* subjects. n_1 subjects are healthy, and n_2 subjects have mass. Among the $n_2$subjects, a subjects are benign, and b subjects are malignant.

Hint:

- You may use the knowledge that n_1, a, and b are all non-zeros values, but you cannot expect $n_1 : a : b = 1 : 1 : 1$.

2) For each testing subject with disease, only one side (i.e. either left or right) of the breast has only one mass.
3) See README.txt for details on the output format. README.txt will be provided along with the data.

Your algorithm will be assessed based on its speed, accuracy/robustness of the mass segmentation (e.g. Dice coefficient), sensitivity and specificity of diagnoses.