Medical Imaging Project: BIRADS Breast Classification

Savinien Bonheur
Student
11th VIBOT Promotion
Universitat de Girona
Girona, Catalunya, Spain
Email: savinien.bonheur@gmail.com

Thomas Drevet
Student
11th VIBOT Promotion
Universitat de Girona
Girona, Catalunya, Spain
Email: dreve2thom@hotmail.fr

***Abstract*—The problem of breast classification is a key subject of public healthcare. In the following, we developed a classification technique based on 4 features easily extractables. Those features are based on the analysis of the different tissues present in the breasts which we extract using entropy. While being fast those features also show adaptability to any DBT (2D or 3D) and robustness.**
Our code can be found on Github at this address: $github.com/Totohm/MedImg\_Project\_BIRADS$

***Keywords*—BIRADS, Segmentation, Project, Medical Imaging, VIBOT.**

I. Introduction

In the 2nd semester of the VIBOT master, we studied Medical Imaging: we worked mainly on the different techniques used to acquire the images, but also on how to process them (segmentation for instance). To finish the semester, we had to do a small project: we had to implement a code that is able to compute the density of fat in a breast and then, thanks to it assign him the corresponding BIRADS labels. The BIRADS labels go from 1 to 4 and they represent the risk to develop a breast cancer: bellow 25% of fat in the breast the risk is low so it's labeled as 1, between 25% and 50 the risk is medium (label 2), between 50% and 75 the risk is high, labeled as 3 and above 75% we label as 4 because the risk is very high. All this project is coded using Matlab for it's prototyping speed (most of the function are already implemented) . In this report, we will present you the how we decided to treat the project: choose our segmentation method, present our results.

II. Pre-work

Before doing the segmentation, we have to develop other codes to do different 'annex' works in this project (creating the main file, opening the DICOM files and store their data...). The first thing we implemented here is the code to read and store the DICOM files. What we did here is creating a small function that convert all the images from the MRI from the breast stored in DICOM file as a simple 3 dimension matrix to store black and white images representing 'slices' of the breast. This conversion is not mandatory by it permits to simplify the code later and avoid to use higher dimensional matrices : The MRI images that we have, have 4 dimensions and one of them is useless because one represent the time but here all the images are captured in the same timing frame and even if the capture was taken in time, we won't see any changes along it. Then, we built the $main$ code. This code does the 'outside' actions: everything which is not directly linked to the segmentation of the fat from the image and the computation of the BIRADS. First, the $main$ will search all the DICOM that are presents in our files from our computers. To do that, the algorithm will use the dir and ls functions to navigate and for all the DICOM files we found, we will apply the function presented before to convert to a 3 dimensions matrix and then, we apply the our method to have the BIRADS label. Finally, we just create a text file (.txt) to return all the results in a very simple way (this work is done with the $fprinf$ function). Now, having an overview of how our main code is working, we can start to explain our approach to obtain the BIRADS label.

III. Work Flow

The following approach was used to tackle the problem:

- We simulates a cropped 2D DBT
- We remove the breast skin
- We extract the first two features
- We apply a power stretching on the images
- We extract the dense tissue from the breast
- Based on them, we extract the two last features
- We input those features into a Bayes classifier

IV. Processing

A. Simulating a cropped 2D DBT

To simulate a 2D cropped DBT we project on the same plane all the images. A 2D DBT is formed by placing a photon emitter and a photon sensor on each side of the breast. The obtained image is thus the sum of attenuation of all the photon alongs the breast . Hence, to obtain this images from a 3D DBT, we sum all the breast slices along the depth axis and then we normalize the obtained image.

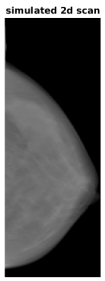

Fig. 1: cropped 2D projected scan

Remark : the breast images illustrating the rest of the paper where made with only a non cropped single slice of the breast as it illustrate more clearly the different operations.

B. Cropping the Breast's skin

To extract the breast skin, we binarize the breast (1 in the breast and zero the background) and perform a closing (morphological operation). Then,we simply multiply our original image by this mask (pixel wise) to obtain an image where the breast skin is removed.

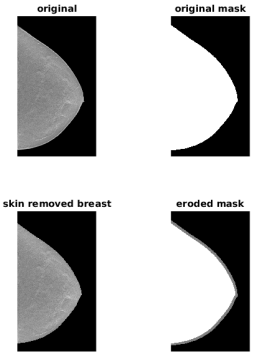

Fig. 2: Breast skin's cropping

C. Extracting the first two features

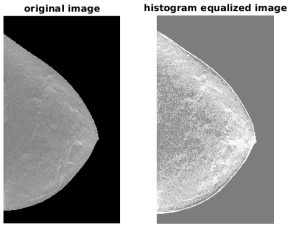

Fig. 3: Breast's images used to extract the first two features

Our first two features are :

- The mean intensity of the histogram equalized images
- The mean intensity of the original breast

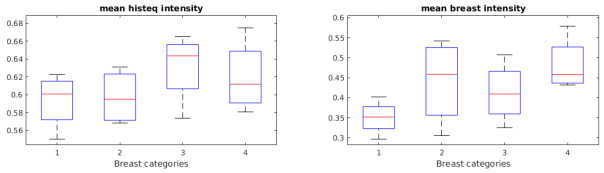

Fig. 4: first two features

D. Power stretching

The original images are low on contrast (in the breast regions): we decided to use a power stretching. By boosting the brightest area of the breast (dense tissue), more than the dark region of the breast (fat tissue); we are able to extract the different tissue more easily.

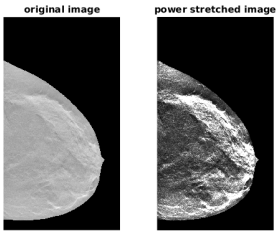

Fig. 5: Contrast augmented pictures

E. Extracting the dense tissue approaches

To extract the tissues based on theirs density, we tried the followings approach :

1) Random walker: We first tried to use a random walker for its precision and human like segmentation. However, its computation time (4 minutes to segment an images) is too long for our objectives.

2) Fuzzy C Mean: We tried to apply a fuzzy C mean segmentation using 10 clusters. Out of those ten clusters we were only keeping the 3 smallest ones. Thanks to this, we were able to extract the denser tissue quite precisely. However, the time needed for this approach was, once again, too big (approximately 18s to perform the segmentation) and the extracted features too few and too instable.

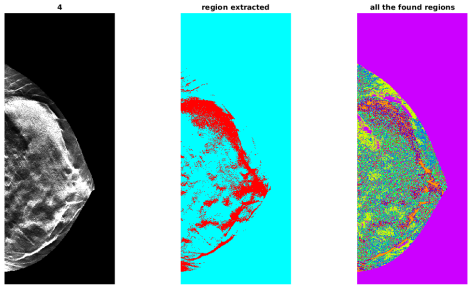

Fig. 6: FMC segmentations results

As seen before its difficult and computationally expensive to extract the breast dense regions using solely the pixels intensity. The solutions chosen, based on the good results obtained by Albert and Alpha, was to use a textures based approach.

3) Fourier Transform (FFT): As texture and frequency are closely related we tried to use a FFT to extract the regions. However, the result obtained weren't discriminating enough.

4) Entropy: Finally, our final and kept approach is to use the entropy. To segment the tissue, we calculate the entropy of each pixel in its 9 neighbors.Then, we threshold this value to obtain a mask containing the denser area of the breast.

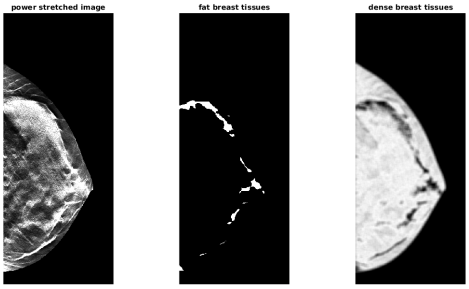

Fig. 7: entropy segmented images

F. Extracting the last two features

Our last two features are :

- The mean intensity of dense tissues/ the mean intensity of the fat tissues
- The area of dense tissues/ the area of the fat tissues

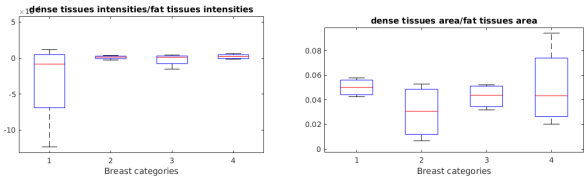

Fig. 8: Last two features

G. Bayes Classifier

After extracting all the features, we feed them to a Bayes classifier. The interest of the Bayes classifier when applied to our features (which are overlapping), it is that they do not consider the features independently but also the relationship in between the different features.

V. Settings the parameters

Our algorithm present 2 parameters which need to be optimized. One approach could be to use an genetics algorithms to find the optimal parameters. The risk of such an approach would be for the algorithm to get stuck in a local maxima. However, we know empirically than the power stretching value has to be ¡20 and we know that the threshold value is between 0 and 1. Due to limited number of value possible we decided to compute the result of our algorithm for each possible configuration (power[1:0.5:20] and thresh=[0.05:0.05:0.95]).

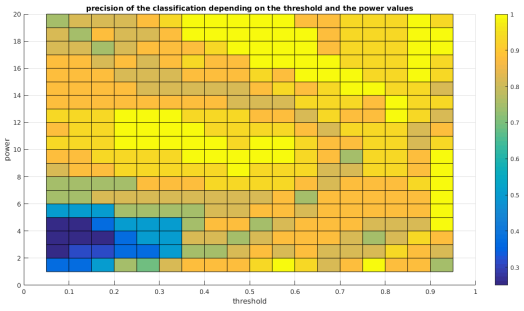

Fig. 9: precision of the algorithm based on the parameters

We then chooses a power stretching of 9 and a threshold of 0.45. In fact those value are in a stable area (the values around are still giving goods result) and appears to work in different cases with a precision of 1 (2d projected scan, one slices of the breast, mean of different slices, ...).

VI. Robustness

Our algorithm show robustness both in the parameters (the mean precision of figure 9 is 0.86) and in its input (a 2d cropped simulated DBT) as it can be obtained from any DBT machines existing. In fact, the cropping will suppress the machines fields of view variations and a 2D DBT can be obtained from both a 3D DBT scanning machines and a 2D DBT scanning machines. It is also interesting to note that, if we take wrongs thresholds on purpose (giving a precision of 0.86), the confusion matrix shows that mislabelled pictures are labelled in the neighbors of their true labels (no B4 will be misclassified as B1).

True Labels	Estimated Labels 1	2	3	4	Totals
1	4	0	0	0	4
2	1	3	0	0	4
3	0	0	3	1	4
4	0	0	0	4	4
Totals	5	3	3	5	16

Fig. 10: confusions matrix for wrongly chosen parameters

VII. Our Results

To test our algorithm we used the "leave one out" strategies. We built 16 groups of images. Each groups is composed of 15 train images and 1 test images (unique for each groups). For each groups, we train a classifier with the train images and test it with the test images. The precision is then calculated by dividing the number of correct matched picture by the

number of pictures tested. We obtained a precision of 1 with the following parameters; power=9 and threshold=0.45 in 7.8s (features extractions and classifiers operations).

VIII. Conclusion

Our approach, while privileging speed (0.46 s to extract the features), show perfect result on our data set. Despite the limited size of it (we could have over fitted our parameters, and so our, results), we have shown that our algorithm still perform well. It is also important to remark than wrongly classified pictures are classified into groups showing adjacency with their true group (B3 will not be classified as B1). The speed, the precision and the way the algorithm behave while misclassifying , could allow this method to be used to pre-classified breast upon doctors analysis. A bigger number of scan would allow to tuned the algorithm more precisely.

Acknowledgment

We would like to thank to thanks our Medical Imaging professor and supervisor of the project, Robert Marti for his help and his advices during the whole project.

IX. Sources

Grady L., " Random Walks For Image Segmentation ", IEEE Transactions On Pattern Analysis And Machine Intelligence, Vol. 28, No. 11, 2006

Despina Kontos et al., "Parenchymal Texture Analysis in Digital Breast Tomosynthesis for Breast Cancer Risk Estimation: A Preliminary Study", PMC 2010 Mar 1.