ADVANCED IMAGE ANALYSIS PROJECT: A PYTHON PACKAGE (IMMAS) FOR MASS DETECTION IN DIGITAL MAMMOGRAMS

Brianna Burton, Doiriel Vanegas, Mahlet Birhanu, Marcio A. B. C. Rockenbach, Oleh Kozynets

Supervior: Professor Alessandro Bria
MAIA Master: University of Cassino and Southern Lazio
Via Di Biasio, 43
03043 Cassino (FR)

ABSTRACT

Breast cancer ranks among the most deadly diseases worldwide and is a constant threat to women's health. Imperative to survival is early and accurate diagnosis through screening. As the medical world evolves into the digital era, radiologists can be aided by a second opinion using computer aided diagnosis (CAD) systems. As such, this report presents an Intelligent Mammogram Mass Analysis and Segmentation (IMMAS) module for mass detection. It explains the optimal advanced image processing and pattern recognition techniques required for segmentation and classification of masses in breast mammogram images. Experiments are performed on the dataset INbreast, a publicly available dataset. Images are first preprocessed using intensity based transforms and then segmented using multithresholding. The resulting binary image is used to extract geometric and intensity-based features from regions, as well their local binary patterns (LBP). Two classifiers, support vector machine (SVM) and random forest (RF), are used to detect the masses and are compared using a free-response receiver operating characteristic curve (FROC) with a optimal area under the curve (AUC) from 0 to 1 of 0.59. This investigation results in the IMMAS module, which lays the basis for a competitive CAD system.

1. INTRODUCTION

1.1. Breast Cancer

Breast cancer is the most commonly diagnosed cancer among women, with approximately 182,000 women diagnosed with the condition annually in the United States [1]. Every year, around 40,000 women die of breast cancer, making it the second-leading cause of cancer death among American women after lung cancer.

Clinical symptoms of breast cancer include change in breast size or shape, skin changes, nipple abnormalities, single-duct discharge, and axillary lumps [2]. However, those symptoms usually appear late in the course of disease, reflecting advanced staging and less chance of cure. To be able to detect cancer in earlier stages, mammograms must be used.

1.2. Mammograms

Mammography is a radiographic technique for imaging of the breast. It is used for screening and diagnostic purposes. In both cases, the technique is similar. Four images are acquired, two for each breast, one being medio-lateral oblique view and another craniocaudal view. If necessary, more specific views can be obtained, according to the orientation given by the radiologist. Mammograms have an overall sensibility around 88-93% and specificity between 85-94% [3]. Therefore, they form an excellent method for early diagnosis of breast cancer.

An exam report is based on the BI-RADS system [4]. BI-RADS was designed to standardize breast imaging reporting and to reduce confusion in breast imaging interpretations. It is also an important tool that allows outcome monitoring and quality assessment. The classification by the BI-RADS varies between categories from 0 to 6. Category 0 indicates the necessity for further imaging investigation, and 6 indicates a mammogram already known to be positive for cancer. The categories between 1 and 5 indicate the risk for cancer, category 1 being very low and 5 very high.

The definition of the category relies on the imaging findings. The most common features evaluated by the radiologist include masses, calcifications, architectural distortions, asymmetries, lymphadenopathy. Those findings might be very easy or very hard to be detected, according to their characteristics and the breast composition. Therefore, it is of great clinical value to develop systems that can automatically detect suspicious areas, drawing the radiologist's attention to possible abnormalities. In the presented project, a computer aided diagnosis (CAD) system was developed to detect masses in mammograms.

2. MATERIALS

2.1. Database

This project uses a mammographic research database called INbreast [5]. The use of a publicly known database adds legitimacy to our system, since the results can easily be compared to current and future work with this INbreast.

The database was acquired between 2008 and 2010, using MammNovation Siemens FFDM. It consists on images from screening, diagnostic, and follow-up exams. In total, the database has 410 images. It also contains samples with various breast densities, one of the main characteristics that affects abnormality detection (breasts with high density are harder to evaluate). In 56 cases, a biopsy was performed, of which 11 were benign and 45 were malignant.

The dataset contains examples of normal mammograms, mammograms with masses, mammograms with calcifications, architectural distortions, asymmetries and images with multiple findings. A mass, the subject of this project, is defined by the BI-RADS as a three-dimensional structure demonstrating convex outward borders, usually evident on two orthogonal views [4].

The most significant characteristic of this database is the groundtruth annotation. While most of the databases only give a circle around the region of interest (ROI), INbreast provides the exact contour of each mass, made by a specialist in the field and validated by a second specialist. There are 115 masses among 107 images ($\sim$1.1 masses per image). The average mass area is 479 mm^2 (standard deviation: 619 mm^2, range: 15 mm^2 to 3689 mm^2). The careful and precise annotation of the contours of the masses is a great tool to simplify performance evaluation.

The dataset contains 16-bit images in `.tif` format, with matrix size ranging from 2560 x 3328 to 3328 x 4084 pixels. For every image, there are two binary masks: one with the entire breast tissue (to remove the background) and another with the pectoral muscle. For the images with masses, there is also a binary groundtruth image.

3. METHODS

The proposed system operates in four stages: (1) preprocessing based on morphological enhancement and wavelet transform, (2) mass candidate generation through segmentation, (3) feature extraction, and (4) false positive reduction using support vector machines (SVM) and random forest (RF) classifiers.

3.1. Preprocessing

Since masses are embedded in the mammary tissue, preprocessing is necessary to separate them from the surroundings. Different preprocessing methods have been considered in the literature to enhance mass like patterns such as Morphological Enhancement, Wavelet Transforms, and Contrast Limited Adaptive Histogram Equalization (CLAHE). For this work, a combination of these techniques was applied. In a first step, bright areas are added to the image (top hat transform) while subtracting the dark areas (bottom hat transform). As a result, there is an enhancement in the contrast between bright and dark areas. To improve the contrast even more, CLAHE is applied. At the end of the preprocessing step, we perform single level 2D wavelet transform followed by median filtering and reconstruction by inverse wavelet transform.

3.2. Segmentation

Once the image has been enhanced, multithresholding is applied to distinguish between breast tissue, potential masses, and the background, and the image is thresholded using the largest threshold. Through this, mass candidates are found. To improve the segmentation, morphological operations are performed. The morphological parameters are chosen to maximize the Jaccard Index calculated for a segmented mammogram and the corresponding groundtruth image. In later steps, the mass candidate will be classified as positive or negative.

3.3. Feature Extraction

In the third step, mass candidates are described using six different groups of features as stated in Table 1. A more detailed explanation of the shape, margin, texture and statistical descriptors is provided by Dong [6].

The use of Local Binary Patterns (LBP) has been reported as a good method for feature extraction in facial detection. Oliver et al. [7] proposed a method for applying LBP for false positive reduction in mass detection, by testing the performance of the LBP descriptor with different parameters for the number of neighbors P, the radius R and for different window sizes to divide each ROI (grid size). Berbar et al. [8] performed experiments on the effect of varying the grid size in the accuracy of the classifier. Based on their best results an operator was designed with $P = 8$, $R = 1$, $LBP_{8,1}^{riu2}$ and a 5×5 grid size. For each grid, 10 features are obtained, and the total number of features depends on the ROI size. Berbar suggests to use a ROI size of 256×256, obtaining 26,010 features, a value that is too high considering the low number of images available in our dataset. It was decided to limit the number of features to 1,000. To accomplish this, a ROI size of 50×50 was chosen.

3.4. Machine Learning

In the last step, two classifiers are built: a support vector machine (SVM) and a random forest (RF), to be compared later. First, the dataset is divided randomly into two subsets (train and test) in equal proportions. The division is done so that regions belonging to the same image were assigned to the same subset (training or test). Then, a 2-fold-cross-validation strategy is used.

The classifier is also run under two different conditions: one using all the features and another using all features except LBP to test the effect of the dominating feature, LBP, on the resulting classifier.

Table 1: List of image descriptors

Category	Features
Shape: Geometric	Perimeter, area, circularity, area/circularity ratio, shapefactor, NCPS
Shape: NRL	NRL mean, NRL SD, SD NRL/mean radial length (RL) ratio, NR entropy
Margin Gradient	Gradient: mean, SD, skewness
Texture: Grey Level Histogram	Intensity: mean, standard deviation, smoothness, skewness, kurtosis
Statistical: Grey Level Co-occurance Matrix	Correlation, contrast, uniformity, homogeneity, energy, dissimilarity
Whole image	Local Binary Patterns (LBP)

NRL normalized radial length, NCPS normalized central position shift, RL radial length, SD standard deviation

3.4.1. Support Vector Machine (SVM)

Support vector machines consist of a learning model for classification and regression problems. Given a set of features and their corresponding classes, this classifier builds a hyperplane to separate features related to different classes. The best result is given by the maximum margin hyperplane, named the optimal separating hyperplane (OSH). The linear discriminant function of this OSH is called support vector machine.

The scikit-learn library is used to create and test an SVM classifier. The performance of the classifier is initially evaluated by the area under the ROC curve. This score is used in several grid searches to find the best possible hyperparameters, using different kernels (linear, rbf, sigmoid, polynomial), values of C (penalty to errors), classes weight and values of gamma. The hyperparameters with the best performance are used to build the final classifier.

3.4.2. Random Forest (RF)

The random forest classifier creates ensemble of independent decision trees, where every tree is built by random selection of features [9]. Each tree is a weak classifier. In the process of classification, the final decision is obtained as a result of the majority vote. This classifier is more robust with respect to noise than other ensemble classifiers.

The random forest classifier implementation is provided by the scikit-learn library. Because the classification of mammogram ROIs is a quite unbalanced problem, class weights were automatically adjusted inversely proportional to class frequencies in the input data [10].

3.5. Evaluation method

The overall performance of our system is evaluated by a Free Response ROC Curve (FROC curve). This curve plots the correct detection of masses (true positive rate) versus the negative regions considered as masses for the classifier in each image (false positive regions per image - FPPI). The aim is to maximize the area under the FROC curve between 0 and 1 FPPI (partial area under the FROC curve).

Table 2: Results of mass detection for the two classifiers

Features	Classifier	Partial Area FROC	TPR at FPPI = 1
All	SVM	0.31	0.47
All	RF	0.13	0.21
No LBP	SVM	0.58	0.75
No LBP	RF	0.59	0.75

Partial area FROC for FPPI between 0 and 1

4. RESULTS

The 410 images of the database were preprocessed and segmented according to the methods described above, resulting in 6376 regions. Those segmented regions were compared to the groundtruth for labeling. A region was considered positive if the Dice Similarity Index between the candidate and the groundtruth was equal or higher than 0.2. This resulted in 110 regions labeled as positive (out of 115 masses in the groundtruth - 95.6% of the total) and 6266 regions labeled as negative (15.28 negative regions per image).

The results of classification using the features of these regions can be seen in Table 2, showing both the SVM and RF classifiers. The corresponding FROC curves can be found in Figs. 1-2.

5. DISCUSSION

As mentioned in the previous section, out of 115 masses in the database, the employed prepossessing and segmentation method correctly identified 110 regions. Therefore, the FROC curve never reaches its full potential as one, instead reaching a maximum of 0.956. When performing segmentation, detection of all positive regions and the increase of false positive regions was found to be a major trade-off. A slightly different approach on the prepossessing stage is anticipated to improve the segmentation in future works. Moreover, post-processing of segmented regions can be applied later.

Fig. 1: FROC Curve using all features including LBP comparing the RF and SVM classifiers.

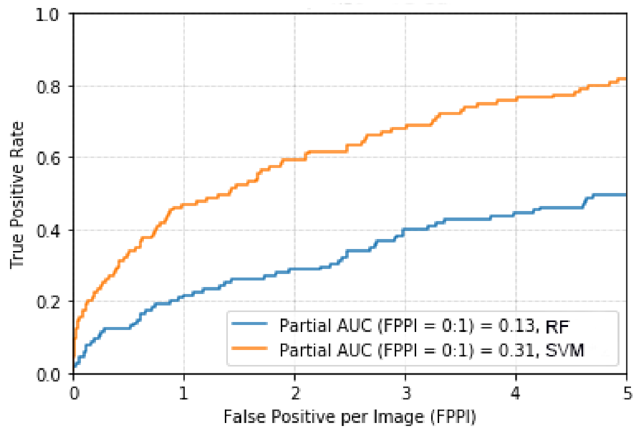
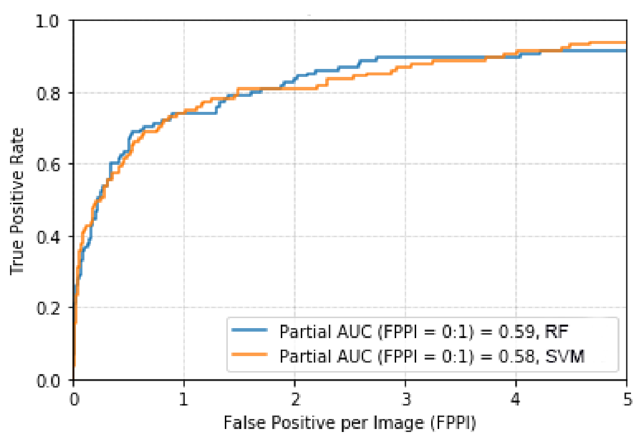


Fig. 2: FROC Curve using all features except LBP comparing the RF and SVM classifiers.



The performance of both SVM and RF classifiers was carefully scrutinized for features obtained with and without LBP. As can be seen in 2, both classifiers perform similarly (Partial AUC ≈ 0.59 for RF and ≈ 0.58 for SVM) on all features with the exception of LBP. But for the set of features including LBP, SVM proved to have a significantly higher performance in comparison to RF even though the Partial AUC was much lower than the previous case (see Figure 1) for reasons stated above.

Additionally, the limited number of the positive mass images and diagnostic reports posed a great challenge for the classifier, as it created an imbalanced class problem. Consequently, it was difficult to raise the AUC of both classifiers, while overcoming the class imbalance. Although basic classifiers are generally encouraged for such problems, a more complex approach might give improved results. Ideally, obtaining a larger image database with more positive samples is the best way to improve the classification model.

6. CONCLUSION

In this project, a system was proposed for automated mass detection in mammogram images, with the potential to provide crucial assistance to radiologists. In our method, we realized preprocessing of full images, segmentation of the ROIs, and successive classification based on predefined features. The source code of our method implementation was can be found on `GitHub` [11]. In the future, improvement of this module could be made using a larger dataset with more positive cases, as well as an additional, more precise segmentation technique, perhaps using the VFC Snake model. Moreover, in real radiology practice, physicians do not analyze images as unrelated entities, but a full diagnostic comprises the comparison between different mammogram views and previous studies. Implementing this kind of reasoning in CAD systems could represent a major improvement in their performance. The development of CAD systems that can facilitate and assist in real time mammogram analysis is challenging but has the potential to make a vast impact on radiology practice.

7. REFERENCES

[1] Ahmedin Jemal, Rebecca Siegel, Elizabeth Ward, Yongping Hao, Jiaquan Xu, and Michael J Thun, "Cancer statistics, 2009.," *CA: a cancer journal for clinicians*, vol. 59, no. 4, pp. 225–249, 2009.

[2] Pavani Chalasani, "Breast Cancer," 2018.

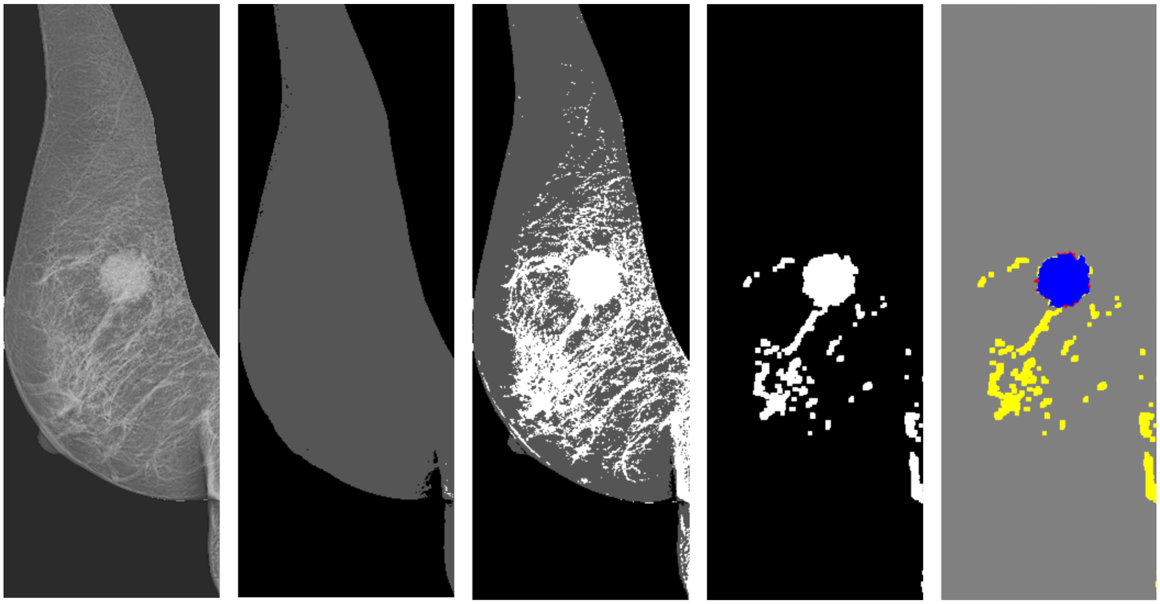
[3] Ministério da Saúde, *Mamografia - Da Prática ao Controle. Recomendações para profissionais de saúde. Instituto Nacional do Câncer (INCA)*, Brasil, 2007.

[4] EA Sickles, CJ D'Orsi, LW Bassett, and Et Al., *ACR BI-RADS® Mammography*, American College of Radiology, Reston, VA, 2013.

[5] Ines C Moreira, Igor Amaral, Ines Domingues, Antonio Cardoso, Maria Joao Cardoso, and Jaime S Cardoso, "INbreast: toward a full-field digital mammographic database.," *Academic radiology*, vol. 19, no. 2, pp. 236–248, feb 2012.

[6] Min Dong, Xiangyu Lu, Yide Ma, Yanan Guo, Yurun Ma, and Keju Wang, "An Efficient Approach for Automated Mass Segmentation and Classification in Mammograms," *Journal of Digital Imaging*, vol. 28, no. 5, pp. 613–625, 2015.

[7] Arnau Oliver, Xavier Lladó, Jordi Freixenet, and Joan Martí, "False Positive Reduction in Mammographic



(a) (b) (c) (d) (e)

Intermediate Result (IR) 1. Effect of optimized preprocessing and segmentation on the regions. (a) original preprocessed image, (b) multithresholding without preprocessing, (c) multithresholding with preprocessing, (d) full segmentation method, including multithresholding and then cleaning on a preprocessed image, and (e) overlap of the segmented image from (d) with the groundtruth image. Blue represents true positive, red false negative, yellow false positive, and gray true negative. The overlapping blue region gives the maximum Jaccard index for this image.

Mass Detection Using Local Binary Patterns," in *MICCAI 2007*, Nicholas Ayache, Sébastien Ourselin, and Anthony Maeder, Eds., Berlin, Heidelberg, 2007, pp. 286–293, Springer Berlin Heidelberg.

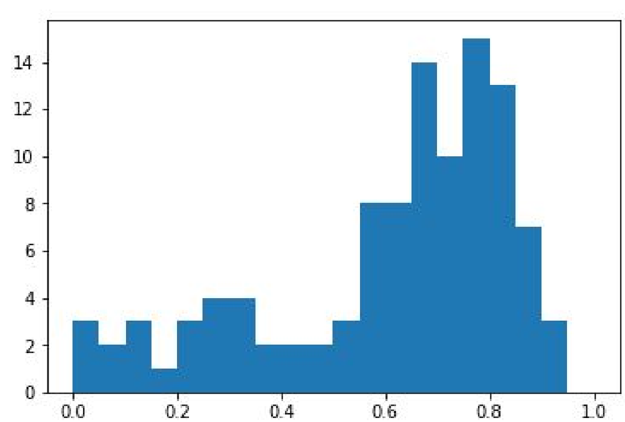
[8] M A Berbar, Y A Reyad, and M Hussain, "Breast Mass Classification using Statistical and Local Binary Pattern Features," in *2012 16th International Conference on Information Visualisation*, 2012, pp. 486–490.

[9] Leo Breiman, "Random Forests," *Machine Learning*, vol. 45, no. 1, pp. 5–32, 2001.

[10] Fabian Pedregosa, Gaël Varoquaux, Alexandre Gramfort, and Et Al., "Scikit-learn: Machine Learning in Python," *J. Mach. Learn. Res.*, vol. 12, pp. 2825–2830, nov 2011.

[11] IMMAS, "Intelligent Mammogram Mass Analysis and Segmentation," in *https://github.com/OlehKSS/IMMAS.* 2018.

IR 2. Histogram of Maximum Jaccard Index per image for segmented mass regions. Frequency vs. Jaccard Index



IR 3. Maximum area under the ROC curve obtained while optimizing hyperparameters for the SVM classifier

Kernel	No LBP	All Features
RBF	0.9436	0.8954
Sigmoid	0.9442	0.8905
Linear	0.9429	0.8666
Polynomial	0.9422	0.8969