International Journal of Computer Science and Engineering (IJCSE)
ISSN Print: 2278–9960;ISSN Online:2278–9979;
Vol. 8, Issue 5, Aug–Sep 2019; 13–22

International Academy of Science, Engineering and Technology
Connecting Researchers; Nurturing Innovations

THE BREAST TISSUE DENSITY CLASSIFICATION BASED ON HARALICK AND SFTA FEATURES EXTRACTION METHODS

Ashgan M. Omer[1], Frédérique Frouin[2] & AlnazierOsman[3]

[1]Research Scholar, Pierre and Marie Curie University, Paris, France & Sudan University of Science and Technology, Khartoum, Sudan

[2]Research Scholar, Université Paris Saclay, Orsay, France

[3]Research Scholar, Department of Biomedical Engineering, Sudan University of Science and Technology, Khartoum, Sudan

ABSTRACT

It has been shown that the sensitivity of any computer-aided diagnostic (CAD) system in the detection of breast cancer is impacted by breast density, thus characterizing the breast tissue type in mammograms can act as a primary step to detect cancer and reduce false positive. In this paper we introduce an approach to the classification of mammogram images according to breast tissue type based on Haralick, segmented-based fractal texture analysis SFTA, and combining Haralick and SFTA feature extraction methods. This study examines two classification tasks using support vector machines (SVM) classifiers. The first classification problem differentiate between fatty and non-fatty tissue, the second one is differentiate between glandular and dense tissue. Experiments were applied to the whole set of 322 mammogram images from the MIAS database. The experiments deal with two sides of woman breast as one case study rather than individual images. The best classification accuracies rate are 88% achieved infatty and non-fatty classification taskby using SFTA features, and 78% for glandular and dense classification using a combination of Haralick and SFTA feature extraction methods.

KEYWORDS: *Breast Density, Haralick Features, Mammogram, SFTA Features*

Article History

Received: *25 Aug 2019* **|** ***Revised:*** *10 Sep 2019* **|** ***Accepted:*** *20 Sep 2019*

INTRODUCTION

Breast cancer is a chronic disease with low incidence but high mortality and morbidity rate, ranking high among the leading cause of death among women in the world [1]. Interest in breast imaging has been fostered by the realization that approximately one in eight women will develop breast cancer over a lifetime [2]. Mammography is a primary imaging tool for screening detecting breast pathology especially breast cancer. Although technological advances have greatly improved the diagnostic sensitivity of mammography, but it is still among the most difficult medical images to be read according to the overlap of tissues, the differences in the tissue types and their low contrasts. Thus the task of the radiologist is tedious and misdiagnosis of breast cancer most of the time occur [3].

Computer Aided Detection (CAD) is pattern recognition software that identifies suspicious features on image and brings them to the attention of the radiologist, in order to reduce false negative reading [4]. Clinical studies approved that, the presence of CAD system influence and improve the diagnostic accuracy and increase assist in breast cancer detection [5].

Breast density is a radiological concept based on the proportion of radiopaque glandular tissue relative to radiolucent fatty tissue. Mammographic evaluation of dense breasts is more difficult, related to technical difficulties in mammography [6]. The dense breast more likely to develop a cancer, so breast tissue density type considers as an indicator for cancer risk, moreover the detection of breast cancer in dense breast is harder than detection cancer that surrounded by fatty tissue [7], so the researchers developed different methods using CAD system for breast tissue classification.

In this paper, we propose an approach to automatic breast tissue classification based on two feature extraction methods and SVM classifier. The mammogram images used for this study are from the Mammographic Image Analysis Society (MIAS) database [8].The database contains left and right digital mammograms for 161 patients with a total of 322 images of 50×50 micron resolution, with 8-bit pixel depth and all mammograms are in Medio lateral oblique view (MLO). The initial step of the proposed methodology is the preprocessing of the mammogram images.Preprocessing is a very important stage aid to limit the search for abnormalities on the only breast region without influence from background [9]. Preprocessing helps in reduction of mammogram size, and improves the quality of the image to make the feature extraction phase more reliable. In the second step, texture features are extracted from the region of interest (ROI). At this stage, three concurrent studies were investigated based on various sizes of ROI. The ROIswhich tested are: whole breast region, an average of five windows (50×50) pixels, and one window (128×128) pixels per image. Dimensionality reduction and classification are achieved in the subsequent step, which relies on SVM classifier. The evaluation is based on MIAS database. Results were compared with other automated breast tissue classification approaches, applied toMIAS database.

This paper is organized as follows. Section II presents a brief review on a mammogram–based CAD system for tissue classification. Section III details our proposed method. Results are presented in section IV. Finally, discussion and conclusions are provided in section V.

LITERATURE REVIEW

Breast density assessment is an important component of the screening mammography report and conveys information to referring clinicians about mammographic sensitivity and the relative risk of developing breast cancer [10]. The origins of breast density classification are due to Wolfe [11], who showed the correlation between breast density and the risk of developing breast cancer, classifying the parenchymal patterns into four categories. Subsequently Boyd et al., [12], showed a similar correlation between the relative area of dense tissue and mammo graphies risk, and developed a method to measure a percentage of breast densities from mammography using a computer-aided technique and divided mammograms into six categories. Commonly, the most used breast density classification is the BIRADS classification [13], which was developed as quality assurance tool, and covers the significant relationship between increased breast density and decreased mammography sensitivity in detecting cancer [14]. The classification of BI-RADS is divided into four categories according to their density: BI-RADS I: breast is mostly made up of fat: breast density < 25%, BI-RADS II: breast density is between 25% and 50%, BI-RADS III: breast density is between 51% and 75%, and BI-RADS IV: breast is extremely dense with breast density > 75%.

Studies concentrated on the use of the gray-level histogram are based on BIRADS classifications [15], [16]. However, many studies as [17], [18] have indicated that such histogram information was not being sufficient to classify mammograms according to BIRADS, because the four histograms of the four BIRADS classes are quite similar bothin mean gray-level and the shape of the histogram. So, classification of mammogram images is an area of active research.

Here, we highlight researches related to breast density classification and mention to part of using MIAS database.

Oliver et al. proposed several approaches focused on breast density. Examples of his researches: In [19], proposed a new approach to classification of mammography according to the breast parenchymal density, the classification based on gross segmentation and the underlying texture contained within breast tissue, the method applied on a set of 270 mammogram images of MIAS database, classifiers used are: the leave-one-out classification method and k-NN classifier. In [20], they used the whole set of MIAS database and 831 images from DDSM database, segmented the breast into fatty and dense regions based on a two class fuzzy C-means clustering approach, then extracted 10 morphological features and 216 texture features finally used number of distinct classifiers (decision tree, Bayesian, and K-nearest neighbor), the evaluation shows strong correlation between automatic and expert based Breast Reporting and Data System (BIRADS) mammographic density assessment.

Sharma et al. [21], presents a hybrid scheme for two class problem (fatty and dense) mammograms using correlation based feature selection (CFS), the classification performed using sequential minimal optimization (SMO). Texture analysis done on ROI of 322 images of MIAS database. Mustra et al. [22], based on histogram and GLCMs, 419 features were extracted from ROI of MIAS database, then used different feature selection algorithm. Different selection methods tested with different classifiers, and the best classification accuracy rate achieved by using forward - backward feature selection method and K-NN classifier.

Silva and Menotti at [23], used individual and combining various sets of texture of image histogram intensity and co-occurrence matrix features, the classification is performed using SVM with RBF kernel. They method applied on 320 MIAS mammograms images. Subashini et al., [24] based online statistic features and SVM classifier for classification, the approach evaluated by 43 mammogram images of MIAS database.

METHODOLOGY

Our experiment applied to the whole set of 322 images from MIAS database. By using Random Sample without Replacement (RSWR) method [25], the dataset was divided according to tissue types into a learning dataset of 160 images and a testing dataset of 162 images. MIAS database classifies background tissue of mammogram images into three categories, that is, fatty (F), glandular (G) and dense (D). MIAS classification was performed by experts radiologists. Figure 1 shows an example of tissue types of mammogram images as classified by MIAS. This experiment mainly restricted to cases, one case represents the combination of left and right breast images of one patient. So there are 80 cases of learning and 81 cases of testing database, the distribution of the three types of breast tissue being similar. table 1 presents the distribution of learning and testing datasets according to their tissue type.

Table 1: Tissue Types Distribution of Selected Dataset

Class	Learning (Cases)	Testing (Cases)	Total (Cases)	Total (Images)
Fatty (F)	27	26	53	106
Glandular (G)	25	27	52	104
Dense (D)	28	28	56	112
Total	**80**	**81**	**161**	**322**

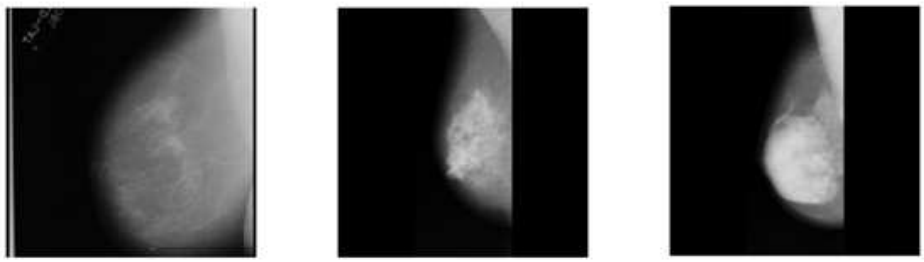

Figure 1: Sample of Different Types of Breast Tissue in the MIAS Database. a) Fatty tissue Img 135, b) Glandular Tissue Img 043, c) Dense Tissue Img 037.

Preprocessing

All MIAS mammogram images are preprocessed to identify the breast region. For preprocessing, the grayscale image converted to binary scale image. Otsu's threshold has been used for binarization process [26]. Preprocess procedure was implemented as the following steps: auto-cropped the extra parts on left and right side of mammography by sweeping in the first row of the binary image from the right corner to the left and from left corner to the right until getting the first non-zero values, then the right and left taps have cropped. Labels omitted based on morphological operations [27] and connected component labeling [28]. To gain unidirectional images, all the right-side breast images dataset are reoriented to the left, by flipping the right mammogram images. Pectoral muscle removed by segmented it using multi-level thresholding [29]. Noises removed by implement a median filter and then contrast enhanced using contrast limited adaptive histogram equalization (CLAHE) technique [30]. Figure 2 shows an example image of the preprocessing stage.

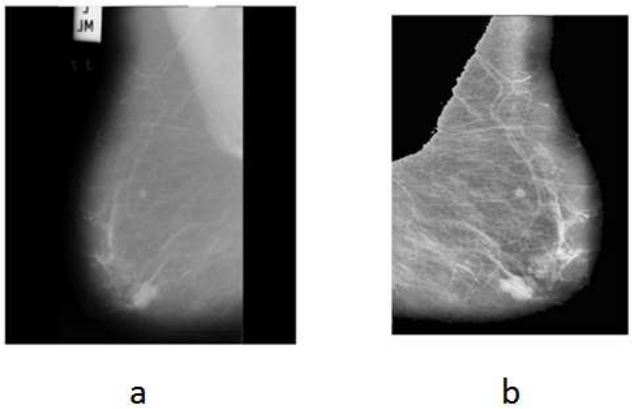

Figure 2: Preprocessing Stage (a) Original Image mdb005. (b) Output Image.

Feature Extraction

In this study two categories of textural features are extracted from different regions of interest (ROIs): Haralick features [31], and the SFTA features [32].

In this stage, three concurrent studies were investigated for extracting Haralick, SFTA, and combining Haralick-SFTA features. The key of the three approaches is a window size. In the first study (S1), features were extracted from the breast region of the preprocessed image. In the second study (S2), features were extracted from five windows (50×50) pixels per image. In the third study (S3), feature was extracted from one window (128×128) pixels per image. Windows are selected on the center of the mammogram image. For any case (left and right breast images), the average of each feature was computed.

Haralick Features

Haralick features are set of scalar textural features driven from gray-level co-occurrence matrix (GLCM), GLCM is a counting the numbers of occurrence of gray levels at a given displacement and angle. Here, A set of 13–Haralick features were extracted, the angle Θ was set to 45°, displacement (d) to 1. The following features were used: homogeneity, contrast, correlation, variance, inertia, sum-average, sum-variance, difference-variance, entropy, sum-entropy, difference-entropy, info-correlation1, and info-correlation2

SFTA Features

The SFTA extraction algorithm consists in decomposing the input grayscale image into a set of binary images from which the fractal dimensions of the resulting regions are computed in order to describe segmented texture patterns. For each resulting binary image, SFTA extraction algorithm computes three parameters (region's boundaries fractal dimension, region's mean gray level, and region's size). Thus, the SFTA feature vector dimensionally corresponds to the number of binary images. The number of resulting binary image is obtained by applying threshold segmentation. The threshold value n_t is a user defined parameter. In this experiment, the threshold was set to 4; thus, the resulting set of binary images is seven, according to$(2n_t - 1)$ as in figure 3. SFTA features are returned as $((2n_t - 1) * 3)$ vector, providing 21 SFTA features for each image denoted F1 till F21.

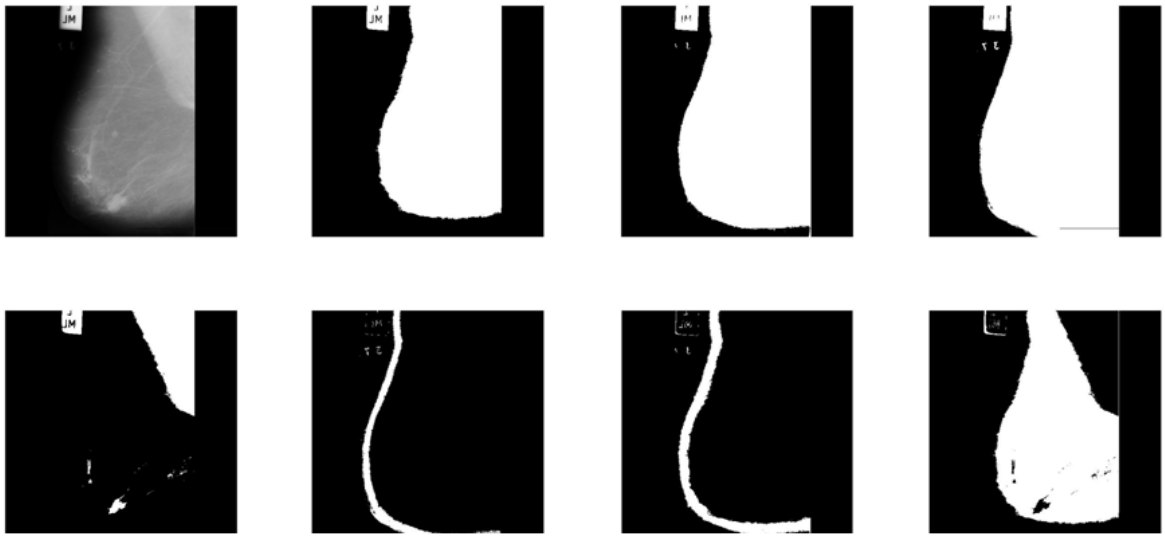

Figure 3: Example of Original Image and 7th Resultant Binary Image from SFTA Extraction Algorithm, Threshold Set to 4, (from *Left to Right* and *Top to Bottom*).

Classification

For each study (S1, S2, and S3), two classification tasks (C1, and C2) were conducted. The first task C1 used to classify fatty and non-fatty tissues. The second C2, applied to classify glandular and dense tissue. The SVM [33] classifier has been used for both C1 and C2, which built on a linear kernel function.

RESULTS AND DISCUSSIONS

"Best" features were chosen according to minimum redundancy maximum relevance (MRMR) selection method [34]. Features were computed from the average of the left and right breast images. As shown in table 2, for fatty and non-fatty classification, the best accuracy obtained is 88% by using preprocessed image approach and SFTA features, *best* SFTA features are: F3, F10, F11, F12, F15, F19, and F20. For glandular/dense classification task, the best accuracy obtained is 78% by using one window (128×128) approach and combined Haralick / SFTA features. Best combined features are: inertia, entropy, difference-entropy, info-correlation1, F2, F19, and F21.

Table 2: Classification Accuracies for Different Study Approaches

	S1(whole Breast Region)		S2(Five Windows (50×50) Pixels		S3(Only one Window (128×128) Pixels	
	C_1	C_2	C_1	C_2	C_1	C_2
Haralick	82.72%	50.91	81.48%	61.82%	83.95%	49.09
SFTA	87.65%	67.27%	86.42%	67.27%	80.25%	70.91
Combined	82.72%	49.09%	79.012	65.45%	80.25%	78.18

DISCUSSIONS

As the background tissue of left and right breast is the same for any patient, our work is restricted to both left and right breast images rather than individual images.

Our work was based on two texture features extraction methods: Haralick features, and SFTA features. Table 3, presented a brief comparison between the used methods. Combining Haralick and SFTA features, gained good result in distinguishing glandular from dense tissue. According to BI-RADS category, there is only a minimal and insignificant difference in the sensitivity of mammography between the densest breast in a lower density category and the least dense breast in the next higher density category [35].

Comparing our result with previously mentioned in literature review, Oliver [19], obtained 67% and 73% for k-NN classifier and the leave-one-out classification method for three classes classification. Moreover, Oliver in [20] obtained 77%, 72%, and 86% for four-classes categories using sequential forward selection SFS and k-NN, decision tree, and Bayesian classifiers, and 91% for two-classes categories using Bayesian classifier.

Sharma and Singh [21] obtained 96.46% accuracy rate for two classes' categories. In [22], they obtained different accuracies for different categories. Classification accuracies rate are: 91.6%, 82.5%, and 79.3% respectively for two-classes, three-class's, and four-class's categories Silva and Menotti [23] obtained 77.18% accuracy for three-classes categories. SVM classifier in [24], classify breast tissue into three classes, the classifier accuracy obtained is 95%.

Table 3: Comparison between Haralick and SFTA Features	
Haralick Features	**SFTA Features**
Widely employed, statistical approach	New employed, structural approach
Orientation and scales dependent	Threshold dependent
Contains fixed number of features	Number of features is dependent to user identifier of threshold

CONCLUSIONS

The current study provides a comparative analysis of the usage of Haralick and SFTA features for mammography breast tissue characterization. Based on fact that both breasts of the same woman have similar internal tissue, our proposed method computed the averages of extracting features from the left and right breast images. In this research we used Haralick, SFTA, and combination of Haralick / SFTA features to examine two classification tasks using three approaches which based on image window sizes. The best result obtained is for fatty/ non-fatty classification task using the whole breast region and SFTA extraction features. In future work, other features extraction methods like Gabor features, wavelets will combine to get more performance.

REFERENCES

1. *K. Ganesan, U. Rajendra, C. Kuang, L. Choo and T. Abraham, " Automated diagnosis of mammogram images of breast cancer using discrete wavelet transform and spherical wavelet transform features: A comparative study", Technology in Cancer Research and Treatment, Vol. 13, No. 6, Adenine Press 2014.*
2. *Jerrold T.Bushberg et al. "The essential physics of medical imaging", 2nd ed. by Lippincott Williams & Wilkins, ISBN 0-683-30118-7, pp. 191–193, 2002.*
3. *Fred S. Azar, "Imaging Techniques for Detecting Breast Cancer: Survey and Perspectives", Technical Report MS-BE-00-02, MS-CIS-00-11, 2000.*
4. *R. A. Castellino, "Computer aided detection(CAD): An Overview", Cancer Imaging, Vol. 5, No. 1, pp. 17–19, 2005.*
5. *JJ. Fenton, SH. Taplin, PA. Carney, L. Abraham, EA. Sickles, C. D'Orsi, EA. Berns, G. Cutter, RE. Hendrick, WE. Barlow &JG. Elmore, "Influence of computer-aided detection on performance of screening mammography". N Engl J Med, Vol. 356, pp. 1399–1409, 2007.*
6. *P. Chérel, C. Hagay, B. Benaim, C. De Maulmont, S. Engerand, A. Langer and V. Talma, "Mammographic evaluation of dense breasts : techniques and limits", Journal Radiology, Vol. 89, No. 9, pp. 1156–1168, 2008.*
7. *I. Diamant, H. Greenspan, J. Goldberger, "Breast tissue classification in mammograms using visual words", IEEE 27th Convention of Electrical and Electronics Engineers in Israel, pp. 1–4, 2012.*
8. *MIAS database http://peipa.essex.ac.uk/info/mias.html*
9. *S. Dehghani, M. Dezfooli, "A method for improve preprocessing images mammography", International Journal of Information and Education Technology, Vol. 1, No. 1, pp. 90–93, 2011.*
10. *NS. Winkler, S. Raza, M. Mackesy, RL. Birdwell, "Breast density: clinical implications and assessment methods", Radio Graphics, Vol. 35, No. 2, pp. 316–324, 2015.*

11. *JN. Wolfe, "Risk for breast cancer development determined by mammographic parenchymal pattern", Cancer, Vol. 37, No. 5, pp. 2486–2492, 1976.*
12. *NF. Boyd, JW. Byng, RA. Jong, EK. Fishell, LE. Little, AB. Miller, GA. Lockwood, DL. Tritchler, MJ. Yaffe,"Quantitative classification of mammographic densities and breast cancer risk: results from the Canadian national breast screening study", Journal of the National Cancer Institute, Vol. 87, No. 9, pp. 670–675, 1995.*
13. *I. Muhimmah,, A. Oliver, E.R.E. Denton, J. Pont, E. Pérez, R. Zwiggelaar, "Comparison between Wolfe, Boyd, BI-RADS and Tabar based mammographic risk assessment", IWDM'06 Proceedings of the 8th International Conference on Digital Mammography, pp. 407–415, 2006.*
14. *EA. Sickles, "Wolfe mammographic parenchymal patterns and breast cancer risk", American Journal of Roentgenology, vol. 188, no. 2, pp. 301–303, 2007.*
15. *N. Karssemeijer, "Automated classification of parenchymal patterns in mammograms", Phys Med Biol, Vol. 43, No. 2, pp. 365–378, 1998.*
16. *C. Zhou, HP. Chan, N. Petrick, MA. Helvie, MM. Goodsitt, B. Sahiner, and LM. Hadjiiski, "Computerized image analysis: Estimation of breast density on mammograms", Med. Phys., Vol. 28, No. 6, pp. 1056–1069, 2001.*
17. *A. Oliver, J. Freixenet, and R. Zwiggelaar, "Automatic classification of breast density", in Proc. IEEE Int. Conf. Image Process, vol. 2, pp. 1258–1261. 2005.*
18. *R. Zwiggelaar, I. Muhimmah, and E. R. E. Denton, "Mammographic density classification based on statistical gray-level histogram modelling", in Prpc. Med. Image Understanding Anal. Conf., pp. 183–186, 2005.*
19. *A. Oliver, J. Freixenet, A. Bosch, D. Raba and R. Zwiggelaar, "Automatic classification of breast tissue", Pattern Recognition and Image Analysis, IbPRIA, pp. 431–438, 2005.*
20. *A. Oliver, J. Freixenet, R. Marti, J. Pont, E. Péréz, ER. Denton, R. Zwiggelaar, "A novel breast tissue density classification methodology", IEEE Transactions on Information Technology in Biomedicine. Vol. 12, No. 1, pp 55–65, 2008.*
21. *V. Sharma & S. Singh, "CFS-SMO based classification of breast density using multiple texture models", Springer- Medical & Biological Engineering & Computing, Vol. 52, No. 6, pp. 521–529, 2014.*
22. *M. Mustra, M. Gric and K. Delac, "Breast density classification using multiple feature selection", ATKAFF, Vol. 53, No. 4, pp. 362–372, 2012.*
23. *W. R. Silva, D. Menotti, "Classification of Mammograms by the breast composition", International Conference on Image Processing, Computer Vision and Pattern Recognition (IPCV), pp. 1–6, 2012.*
24. *T. Subashini, V. Ramalingam, S. Palanivel, "Automated assessment of breast tissue density in digital mammograms", Computer Vision and Image Understanding, Vol. 114, No. 1, pp. 33–43. 2010.*
25. *J. Teuhola, O. Nevalainen, "Two efficient algorithms for random sampling without replacement", IJCM, Vol. 11, No. 2, pp. 127–140, 1982.*

26. *Ch. HimaBindu and K. Satya Prasad, "An Efficient Medical Image Segmentation Using Conventional Otsu Method", International Journal of Advanced Science and Technology, Vol. 38, pp. 67–74, 2012*
27. *K. Sreedhar and B. Panlal, "Enhancement of Image Using Morphological Transformation", International Journal of Computer Science & Information Technology (IJCSIT), Vol. 4, No. 1, pp. 33–50, 2012.*
28. *R. D. Yapa, and H. Koichi, "A Connected Component Labeling Algorithm for Grayscale Images and Application of the Algorithm on Mammograms", ACM symposium on Applied Computing, pp. 146–152, 2007.*
29. *S. Arora et al, "Multilevel thresholding for image segmentation through a fast statistical recursive algorithm", Pattern Recognition Letters, Vol. 29, No. 2, pp 119–125, 2008.*
30. *K. Zuiderveld, "Contrast Limited Adaptive Histogram Equalization", Graphics Gems, Academic Press, Cambridge, MA, 1994, pp. 474–485.*
31. *RM. Haralick, K. Shanmugam, I. Dinstein, "Textural features for image classification", IEEE Trans. System Man Cybernetics, Vol. 3, No. 6, pp. 610–621, 1973.*
32. *AF. Costa, G. H-Mamani, AJM. Traina, 2012. "An efficient algorithm for fractal analysis of textures", 25th SIBGRAP IEEE Conference, pp. 36–46, 2012.*
33. *C. Cortes, V. Vapnik, "Support Vector Machine", Machine Learning, Vol. 20, pp. 273–297. 1995.*
34. *HC. Peng, F. Long, C. Ding, "Feature selection based on mutual information", IEEE Trans on PAMI, Vol. 27, No. 8, pp. 1226–1238, 2005.*
35. *CJ. D'Orsi, EB. Mendelson, DM. Ikeda, "Breast Imaging Reporting and Data system: ACR BI-RADS", 5th ed. Reston, American College of Radiology, 2013.*