

Breast cancer diagnosis based on mammary thermography and extreme learning machines

Maíra Araújo de Santana¹, Jessiane Mônica Silva Pereira², Fabrício Lucimar da Silva³, Nigel Mendes de Lima¹, Felipe Nunes de Sousa¹, Guilherme Max Silva de Arruda¹, Rita de Cássia Fernandes de Lima⁴, Washington Wagner Azevedo da Silva¹, Wellington Pinheiro dos Santos^{1*}

¹ Department of Biomedical Engineering, Federal University of Pernambuco, Recife, PE, Brazil.

² Department of Electrical Engineering, University Center Maurício de Nassau, Recife, PE, Brazil.

³ Polytechnic School, University of Pernambuco, Recife, PE, Brazil.

⁴ Department of Mechanical Engineering, Federal University of Pernambuco, Recife, PE, Brazil.

Abstract Introduction: Breast cancer is the most common cancer in women and one of the major causes of death from cancer among female around the world. The early detection and treatment are the major way to healing. The use of mammary thermography in Mastology is increasing as a complementary imaging technique to early detect lesions. Its use as a screening exam to identify breast disorders has been investigated. The aim of this study is to investigate the behavior of different classification methods while grouping the thermographic images into specific types of lesions. **Methods:** To evaluate our proposal, we built classifiers based on artificial neural networks, decision trees, Bayesian classifiers, and Haralick and Zernike attributes. The image database is composed by thermographic images acquired at the University Hospital of the Federal University of Pernambuco. These images are clinically classified into the classes cyst, malignant and benign. Moments of Zernike and Haralick were used as attributes. **Results:** Extreme Learning Machines (ELM) and Multilayer Perceptron networks (MLP) proved to be quite efficient classifiers for classification of breast lesions in thermographic images. Using 75% of the database for training, the maximum value obtained for accuracy was 73.38%, with a Kappa index of 0.6007. This result indicated to a sensitivity of 78% and specificity of 88%. The overall efficiency of the system was 83%. **Conclusion:** ELM showed to be a promising classifier to be used in the differentiation of breast lesions in thermographic images, due to its low computational cost and robustness.


Keywords Breast cancer early diagnosis, Thermographic images, Mammary thermography, Artificial neural networks, Extreme learning machines.

Introduction

For decades, breast cancer has been the most common type among women. In Brazil, the breast cancer mortality rates remain high, as the disease is still diagnosed in advanced stages. Even though Mammography, Ultrasonography,

Magnetic Resonance and clinical breast examination (ECM) are the most widely used and indicated methods in mastology, there are still many problems associated to them. Sometimes they are not enough to identify breast lesions in women with dense and surgically altered breasts or in women under the age of 40 years. In addition to it, some of these exams are extremely uncomfortable to the patient and there is concern about the risk associated to the use of ionizing radiation (American..., 2015; Instituto..., 2015).

In search for imaging techniques complementary to the above mentioned, thermography started being used in mastology in 1982, but at the time specialists discredited the method and therefore it was not recommended for breast diseases diagnosis. With the technological improvement of the thermographic cameras, many tools using image processing and image analysis could be developed to facilitate the detection of changes in the breast's images, so thermography became more popular and continued to be explored as a complementary screening test in mastology (Milosevic et al., 2015; Walker and Kaczor, 2012).

 This is an Open Access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

How to cite this article: Santana MA, Pereira JMS, Silva FL, Lima NM, Sousa FN, Arruda GMS, Lima RCF, Silva WWA, Santos WP. Breast cancer diagnosis based on mammary thermography and extreme learning machines. Res Biomed Eng. 2018; 34(1):45-53. DOI: 10.1590/2446-4740.05217.

*Corresponding author: Department of Biomedical Engineering, Federal University of Pernambuco, Av. da Arquitetura, Área 2, Sala 17, Cidade Universitária, CEP 50670-901, Recife, PE, Brazil. E-mail: wellington.santos@ufpe.br

Received: 22 August 2017 / Accepted: 08 February 2018

Thermography uses infrared technology to create a temperature map of the surface. When applied to medicine, the distribution of temperature gives several physiological information in a way that highly metabolic tissues appears in the images as warmer spots, so lesions such as cancers and places where angiogenesis is happening may be seen through thermograms. Regarding to the identification of lesions in the breast, the lack of depth has not been considered to be a limitation of this technique since these accelerated metabolic activities tends to increase the surface temperature of the breast (Etehadtavakol and Ng, 2013).

According to Etehadtavakol and Ng (2013), breast thermography has been shown to be efficient during early stages of tumor growth, since physiological changes usually precedes anatomical changes. Moreover, it is a completely non-contact method, with no form of radiation and compression and may be used for all women of all ages, including pregnant and breastfeeding women. This technology also works better to women with dense/fibrocystic breasts than the other screening methods vastly used nowadays.

A limitation of this method is the fact that it is easily influenced by changes in the environment, so aspects such as room temperature and humidity have to be severely controlled to guarantee exam validity.

In view of the above, several studies have been carried out on the application of thermographic images in mastology. Resmini et al. (2012), which perform several feature extractions, these features were analyzed using Support Vector Machines (SVM), k-Nearest Neighbors (KNN) and Naïve Bayes classifiers to detect the existence of lesions in thermographic images of the breast. In this work, the authors reach an approximate accuracy of 90%, and an area below the ROC curve close to 0.9. Aguiar et al. (2013) report several extracted features and the multilayer perceptron classifier was used for the detection of breast lesions in thermographic images and presented 75% of correctly classified regions. Belfort et al. (2015) perform feature extraction using the Artificial Crawlers model. The SVM classifier was used and the process presented 78% accuracy, 50% sensitivity and 84% specificity. Another work, from Acharya et al. (2012), describes the extraction of sixteen features, but uses only four, as the authors defined these as clinically significant in comparison with the others. The results obtained were 88.10% accuracy, 85.71% sensitivity and 90.48% specificity.

The aim of this work is to investigate the performance of different classification methods while grouping the thermographic images into one of the groups: cyst, benign lesion and malign lesion by using Haralick and Zernike descriptors for attributes extraction. Classifiers based on artificial neural networks, decision trees and Bayesian classifiers were used to perform the classification. To assess classification, rates of correctly classified instances and kappa indexes were compared.

Related works

Acharya et al. (2012) evaluated the feasibility of using thermal imaging as a potential tool for detecting breast cancer. Field data were collected from the Department of Diagnostic Radiology, Singapore General Hospital using non-contact thermography. Infrared thermograms were acquired using NEC-Avio Thermo TVS2000 MkIIST System 3.0-5.4 μm short wavelength (30 frames/sec), Stirling cooler, InSb detector with (256 \times 200) elements (Japan), which has a measuring accuracy of $\pm 0.4\%$ (full scale) and temperature resolution of 0.1 $^{\circ}\text{C}$ at 30 $^{\circ}\text{C}$ black body, with the instrument placed 1 m away from the chest with lens (FOV 15 $^{\circ}$ \times 10 $^{\circ}$, IFOV 2.2 mrad) attached. 90 patients were chosen at random to undergo the thermography examination. Examination was done in a temperature-controlled room with the temperature range of 20-22 $^{\circ}\text{C}$ (within ± 0.1 $^{\circ}\text{C}$). Humidity of the examination room was maintained at 60 \pm 5%. The patients were required to rest for at least 15 min to stabilize and reduce the basal metabolic rate, which will result in minimal surface temperature changes, and therefore, satisfactory thermograms. Also, the patients were asked to wear a loose gown that does not restrict airflow. Furthermore, it was ensured that the patients were within the recommended period of the 5th to 12th and 21st day after the onset of menstrual cycle since during these periods the vascularization is at basal level with least engorgement of blood vessels. In this work, we have used a total of 50 thermograms, where 25 thermograms were from cancer patients (age: 51 \pm 8 years) and 25 were from normal subjects (age: 46 \pm 10 years).

In the malignant class, 15 patients had stage III cancer and rest had stage II cancer. 50% of the lumps were found in the upperouter quadrant, 35% in the area behind the nipple, and 15% were located in the upper-inner quadrant. We have analyzed the cancerous breast in each of the 25 malignant cases and one normal breast in each of the 25 normal cases.

Acharya et al. (2012) demonstrated the utility of breast surface temperature as an indicator for malignancy. Since a thermogram presents a visual representation of 'hot spots' of the breast, and hence, the interpretation may be subjective. Therefore, Acharya et al. (2012) extracted texture features from the thermograms in order to feed into classifiers for automatic classification. This makes the interpretation more objective and automatic, and therefore, inter-observer variability of diagnostic prediction is highly reduced.

Acharya et al. (2012) have extracted 16 texture features: homogeneity, energy, entropy, moment1, moment2, moment3, moment4, entropy, angular second moment, contrast, mean, short runs emphasis, long runs emphasis, run percentage, gray level non-uniformity, and run length non-uniformity. But, only four features: moment1, moment3, run percentage, and gray level non-uniformity were selected as they were clinically significant.

By using the SVM classifier and the texture features, Acharya et al. (2012) obtained a classification accuracy of 88.10% in differentiating normal and malignant breasts. The sensitivity and specificity were also considerably high (85.71% and 90.48%, respectively).

Hankare et al. (2016) present color analysis as per the classification on the basis of segmentation. The distinguishable features which are used to detect abnormalities are based upon the variations shown as per the image shape of the hottest regions and it is confirmed by comparing with professional diagnoses. The authors claim their results demonstrate the suitability of infrared thermography as a diagnostic tool in breast cancer detection.

Hankare et al. (2016) employ an image segmentation approach using K-means clustering technique based on color features from the images. Segmentation of hot region is carried out into two steps. In first step, the pixels are clustered based on their color and spatial features, where the clustering process is carried out. They claim the advantages of their proposed method are: 1) It can segment the cancer regions from the image accurately; 2) It is useful to classify the cancer images for accurate detection; 3) Early stage detection of cancer from images. However, Hankare et al. (2016) present only qualitative results, based on color distribution. Since pseudo-color maps are not unique, this approach could not be generalized.

Araújo et al. (2014) evaluated the feasibility of using interval data in the symbolic data analysis (SDA) framework to model breast abnormalities (malignant, benign and cyst) in order to detect breast cancer. SDA allows a more realistic description of the input units by taking into consideration their internal variation. In this direction, a three-stage feature extraction approach is proposed. In the first stage, four intervals variables are obtained by the minimum and maximum temperature values from the morphological and thermal matrices. In the second one, operators based on dissimilarities for intervals are considered and then continuous features are obtained. In the last one, these continuous features are transformed by the Fisher's criterion, giving the input data to the classification process. This three-stage approach is applied to a Brazilian's thermography breast database and it is compared with a statistical feature extraction and a texture feature extraction approach widely used in thermal imaging studies. Different classifiers are considered to detect breast cancer, achieving 16% of misclassification rate, 85.7% of sensitivity and 86.5% of specificity to the malignant class.

The thermograms used by Araújo et al. (2014) were acquired with a FLIR S45 infrared (IR) camera. The analysis was performed using a data set obtained from a patient group (size $n = 50$) of the Hospital of

the Federal University of Pernambuco (UFPE), Recife, Brazil. This data set consists of patients aged greater than 35 years with a suspected mass, whose diagnoses were confirmed by clinical examination and followed by ultrasound, mammographic and biopsy exams. A standardized protocol was used for the infrared image acquisition. For this purpose, an apparatus was designed and constructed. A protocol for image acquisition was generated and it is described in Bezerra et al. (2013). This apparatus consists of two rails used for the displacement of a small carriage that supports the tripod, that is attached to the infrared camera. A support for the patient's arms made of steel, aluminum, and wood was fitted to a swivel chair. This support has a movable horizontal bar designed to move up and down. The bar is used to position the patient's hands allowing four different positions so as to comfortably accommodate patients of different heights (Bezerra et al., 2013).

Thermographic imaging should be performed in a controlled temperature room to avoid or minimize the thermal interference from external sources. To achieve better thermal conditions, the patients were subjected to an acclimatization period at least of 10 min, in order to their bodies reach the thermal equilibrium with the room. Considerations for the environment conditions as well for the patients are described in Bezerra et al. (2013). The infrared images used in this work were obtained from the frontal planes of each patient.

Belfort et al. (2015) used the same thermogram database employed by Araújo et al. (2014), but limited to 34 images, where 15 images for mammary lesion (benign or malignant) and 19 for healthy patients. Colored JPEG images were converted to grey levels and, afterwards, the regions of interest are manually extracted in left and right mammary regions. These two ROIs are then registered using b-splines (Klein et al., 2007) and used to generate a dissimilarity map. From this dissimilarity image, Belfort et al. (2015) used Artificial Crawlers Model for feature extraction (Gonçalves et al., 2014). The generated feature vectors are then classified using linear Support Vector Machines, giving an accuracy of 78%, sensitivity of 50%, and specificity of 84%.

Our proposal is based on the investigation of texture and shape descriptors to represent mammary thermograms. We used the same database studied by Araújo et al. (2014). Since we are interested in lesion classification, as Araújo et al. (2014), we also considered the following classes: malignant, benign and cyst. Acharya et al. (2012), Belfort et al. (2015) and Hankare et al. (2016) are interested in lesion detection. However, differently from Araújo et al. (2014), our feature extraction is based on the combining texture and shape features using Haralick moments and Zernike features, respectively, extracted from grey-level temperature matrices generated from pseudocolor JPEG

images. We also tested more sophisticated classifiers, like multi-layer perceptrons, random forests, and support vector machines. Our proposal returned 88.10% accuracy, 85.71% sensitivity and 90.48% specificity, without manual intervention, against the results of Araújo et al. (2014), which returned 84% of accuracy, 85.7% of sensitivity and 86.5% of specificity for the malignant class.

Methods

The images that feed the system came from thermographic images acquired at Hospital das Clínicas, Federal University of Pernambuco, where cyst, malignant and benign classes are selected (Araújo et al., 2014; Bezerra et al., 2013). For the pre-processing step of the images, the RGB-JET conversion was performed to temperature gray levels and the post-processing step was performed to balance the classes. The Zernike and Haralick moments are used to extract attributes based on geometry and texture. The next stage performs the training and subsequent classification with several classifiers based on artificial neural networks, decision trees and Bayesian classifiers. Finally, the performance of the system was evaluated through accuracy and the Kappa index. Figure 1 is a flowchart of the proposed system.

Images acquisition

The thermographic images used in this study were acquired at Hospital das Clínicas da Universidade Federal de Pernambuco (University Hospital of the Federal University of Pernambuco, HC-UFPE, Brazil) by using a FLIR infrared camera of the model S45.

In order to avoid significant changes in patients positions during the acquisition process, a mechanical device was built, this device is shown in Figure 2 and is further described in Oliveira (2012).

The car is connected to the rails in order to move the camera closer or further away from the patient; furthermore, the arms support is connected to the chair through two (2) horizontal bars, so they rotate together to change the position of the patient.

Eight (8) JPG images were obtained for each patient, each image was acquired from a different position, such as follows: T1 (frontal with hands on waist), T2 (frontal with hands raised, holding the bar located above the head (Figure 2)), MD (right breast only), ME (left breast only), LIMD (internal lateral of the right breast), LIME (internal lateral of the left breast), LEMD (external lateral of the right breast) and LEME (external lateral of the left breast). Figure 3 illustrates examples of images in each of the positions.

The image acquisition protocol was first described in Oliveira (2012) and is illustrated in Figure 4, below.

Creation of the thermographic breast image database

In this work, the images in all the positions were used (T1, T2, MD, ME, LIMD, LIME, LEMD and LEME). These images were divided into malignant, benign, cyst and normal classes, according to specialists diagnoses, which were given by using consolidated methods for each case. The malign class comprises of all cases of breast cancer proven by biopsy. The benign class refers to cases of benign tumors, also proven by biopsy. The cyst class includes cases with this diagnosis proven by fine needle aspiration (PAAF) or ultrasonography (Silva, 2015). The final database contains 1052 images.

Considering that the purpose of this approach is to verify the classification of an existing lesion, the normal class (227 images) was removed from the database for this study. Therefore, only three classes were used: malign, benign and cyst. For this study, images were taken from 100 female patients; 219 cyst images were used, 371 images with benign lesions and 235 images containing malignant lesions.

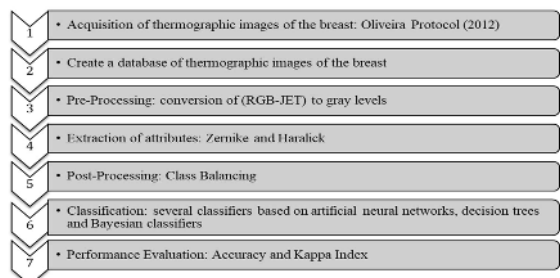


Figure 1. Flowchart of the system.



Figure 2. Mechanical device to place patient in the right position during images acquisition. (1) the trails used to move the camera support car, there are two (2) of them and they are placed on the floor; (2) plate-shaped car to support camera's tripod; (3) swivel chair where the patient is placed on; (4) arms support, which consists of a horizontal bar that moves vertically so the patient can put the hands up during the exam.

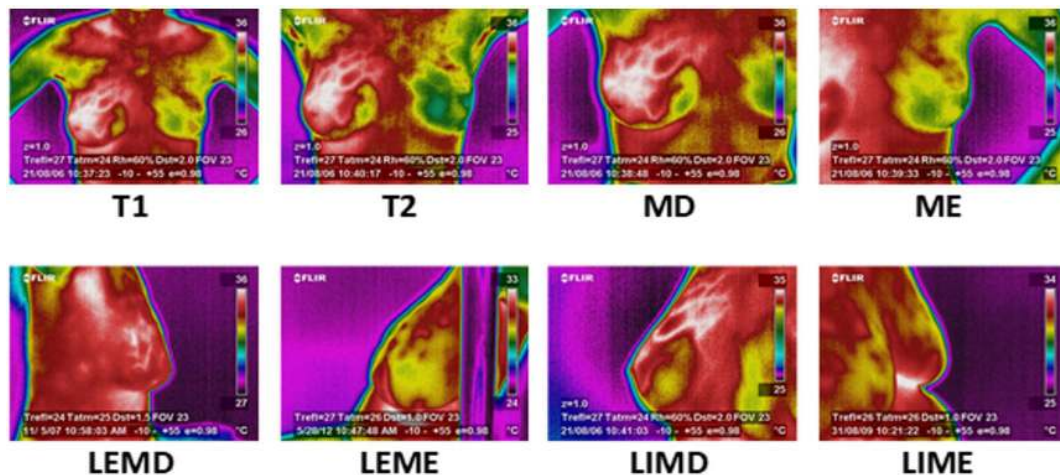


Figure 3. Example of image positions: T1 and T2 are associated to frontal acquisition with arms curved down and up, respectively; MD and ME corresponds to frontal acquisition from center to right, and from center to left, respectively; LEMD and LEME corresponds to right and left medio-lateral acquisition, in this order; LIMD and LIME are almost the same as LEMD and LEME, respectively, but closer.

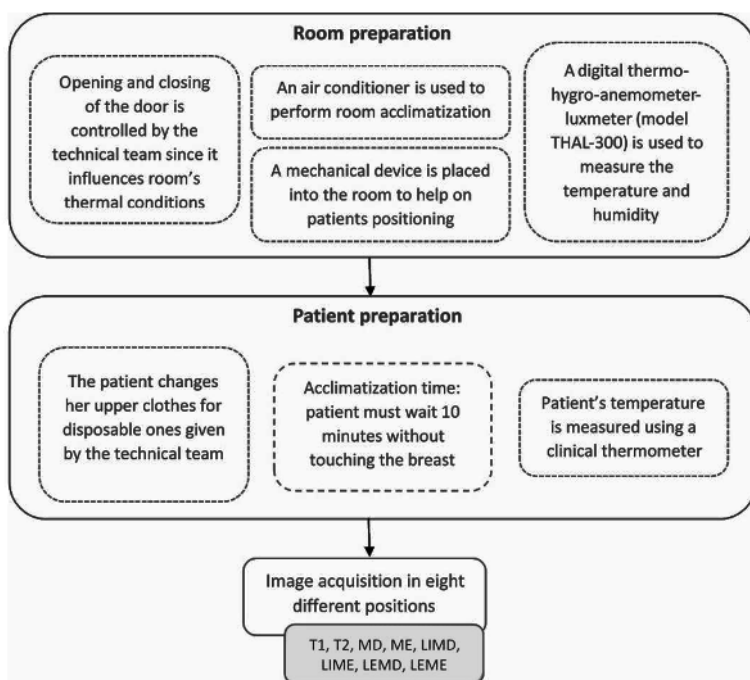


Figure 4. Scheme of acquisition protocol.

Preprocessing

The thermal image uses pseudo-coloring techniques which, in this case, were used in the acquisition of the JET color palette. Therefore, it was necessary to use RGB-JET conversion to Grayscale.

Attributes extraction

The definition of the feature extraction method is one of the most important factors for computational system performance in support of the Diagnostic (Cheng et al.,

2006). According to the characteristics, the attributes were based on geometry or texture. We used the Zernike moment attribute extractors based on the extraction of geometry and the Haralick moment based on the extraction of texture features. The first are projections of the image function in orthogonal basis functions and only the rotation is invariant (Shanthi and Bhaskaran, 2013). The second one is a value calculated from the co-occurrence matrix of the image, which quantifies some characteristics of the variation of the gray levels of these images (Cheng et al., 2006; Shanthi and Bhaskaran, 2013).

Post-processing

After the extraction of the attributes, we performed a class balancing, due to the thermal images database having varying amounts of images from the different classes. Therefore, it is necessary to use the linear balancing technique.

Classification

After the extraction of attributes and class balancing, these attributes are used as input for the classifiers that will be trained and then later perform the classification of breast lesions (malignant, benign and cyst). In this article we present a comparison between eight classifiers in order to verify their capacity to classify lesions in the breast in thermographic images. The classifiers used were Bayes Network, Naive Bayes, Support Vector Machines (SVM), Knowledge Tree J48, Multi-Layer Perceptron (MLP), Random Forest, Random Tree, and Extreme Learning Machines (ELM) (Breiman, 2001; Cheng and Greiner, 2001; Geurts et al., 2006; Haykin, 1999; Librelotto, 2014).

During the tests, we perform the training using a percentage split approach, in which part of the database is used for training while the rest is just used for test, to verify the quality of the training step. For all the classifiers above mentioned, tests were performed using percentage split and a k-folds cross validation method (Jung and Hu, 2015). To first tests the database was randomly divided in a way that 75% of the database was used for training and 25% for testing. In a second time, cross-validation method with k equals to 10 folds was used to perform the tests; in this method the dataset is randomly divided into k samples and these samples are used one by one to perform both training and testing. At the end, all database end up being used for both steps of classification.

The classification stage was performed using the free software Weka (Waikato Environment for Knowledge Analysis), version 3.8, developed at the University of Waikato, New Zealand. We used the configuration of the Bayes Net, Naive Bayes, SVM, J48, MLP, Random Forest, and Random Tree classifiers as available in the Weka 3.8 library; Table 1 shows the parameters we chose to change to each classifier.

For the ELM classifier we performed tests using the following configurations: 100, 200, 300, 400 and 500 neurons in the hidden layer with linear kernel, grade 2 polynomial, grade 3 polynomial, grade 4 polynomial, and grade 5 polynomial.

There were performed 20 tests per configuration of each classifier.

Performance evaluation

Finally, the system performance is evaluated through the average accuracy and average Kappa index for each configuration. Accuracy is the percentage of correctly classified data considering the classes also used correctly (Landis and Koch, 1977). The Kappa index is a statistical

method to assess the level of agreement or reproducibility between two sets of data; it can vary between -1 and 1. We used Cohen's Kappa index. The interpretation of the Kappa index is show in Table 2.

Extreme Learning Machine (ELM)

ELM consist of a training approach for single-tiered neural networks. This proposed learning technique is for training single-layer feedforward neural networks that accelerates learning through the random generation of input weights and the hidden layer (Huang et al., 2006).

Results

We acquired results using percentage split of 75%. Classifiers performance was assessed through the values for accuracy and Kappa indexes, which may be seen at the tables below. For all the configurations, we performed tests using Haralick extractor only (Table 3), Zernike only (Table 4) and using both extractors at the same time (Table 5).

Table 1. Configuration of the classifiers used to perform the tests.

Classifier	Parameters
BayesNet	-
NaiveBayes	-
J48	-
SVM	Linear kernel
MLP	Hidden Layers: a * Learning Rate: 0.3 Momentum: 0.2 Iterations: 500
Random Forest	Trees: 100
Random Tree	-
ELM	Number of neurons in the hidden layer: 100, 200, 300, 400 and 500; Polynomial kernel: degrees 1, 2, 3, 4 and 5

* 'a' = (attribs + classes) / 2 = 85 hidden layers.

Table 2. Interpretation of the Kappa index.

Kappa Values	Level of Agreement
< 0	Poor
0-0.20	Slight
0.21-0.40	Fair
0.41-0.60	Moderate
0.61-0.80	Substantial
0.81-1.0	Excellent

Table 3. Results of the classifiers with Haralick attribute extractors.

Classifier	Accuracy	Kappa Index
BayesNet	51.80%	0.7690
Naive Bayes	51.44%	0.2736
MLP	60.97%	0.4138
SVM	56.47%	0.3480
J48	50.54%	0.2577
Random Forest	59.17%	0.3876
Random Tree	48.38%	0.2256
ELM*	65.95%	0.4892

*Using 500 neurons and linear kernel.

Table 4. Results of the classifiers with Zernike attribute extractors.

Classifier	Accuracy	Kappa Index
Bayes Net	47.30%	0.2103
Naive Bayes	54.14%	0.3138
MLP	72.12%	0.5817
SVM	62.41%	0.4365
J48	43.88%	0.1581
Random Forest	66.01%	0.4904
Random Tree	44.42%	0.1663
ELM*	70.43%	0.5564

*Using 500 neurons and linear kernel.

Table 5. Results of the classifiers with Haralick and Zernike attribute extractors.

Classifier	Accuracy	Kappa Index
BayesNet	51.80%	0.2780
Naive Bayes	51.62%	0.2786
MLP	73.38%	0.6007
SVM	67.81%	0.5173
J48	52.70%	0.2904
Random Forest	64.57%	0.4688
Random Tree	52.16%	0.2820
ELM*	72.94%	0.5940

*Using 500 neurons and linear kernel.

Tables 6 to 8 show the results of accuracy and Kappa index obtained for the tests from 10-folds cross-validation method using only Haralick extractor, only Zernike extractor and combining both extractors, respectively.

Based on the tables above, it was verified that when using only Haralick as attributes extractor we obtained better results using ELM as the classifier, which were 65.95% of accuracy and a Kappa index of 0.4892, for the tests using percentage split and accuracy of 71.22% and 0.6676 for Kappa, for tests from cross-validation method. On the other hand, MLP classifier showed to be more efficient in the cases in which we used only Zernike as extractor and when we combined Haralick and Zernike.

The best result was obtained when we associated Haralick and Zernike attributes extractors. In this situation, 73.38% of the instances were correctly classified using MLP as the classifier, resulting in a Kappa index of 0.6007 when the percentage split approach was used, and we obtained the maximum value of 76.01% correctly classified instances and Kappa of 0.6402, also using MLP.

Qualitatively, this result showed sensitivity around 78% and specificity of 88% in the identification of malignant lesions through thermographic images. Overall, these values indicate that the system had an efficiency of 83%, which is close to the maximum value of 1 (one) implying in a satisfactory performance.

Table 6. Results of the classifiers with Haralick attribute extractors using 10-fold cross-validation.

Classifier	Accuracy	Kappa Index
BayesNet	51.12%	0.2668
Naive Bayes	49.87%	0.2480
MLP	62.26%	0.4340
SVM	56.38%	0.3457
J48	51.48%	0.2722
Random Forest	59.88%	0.3982
Random Tree	47.71%	0.2156
ELM*	71.22%	0.6676

*Using 500 neurons and linear kernel.

Table 7. Results of the classifiers with Zernike attribute extractors using 10-fold cross-validation.

Classifier	Accuracy	Kappa Index
Bayes Net	50.13%	0.2520
Naive Bayes	51.66%	0.2749
MLP	69.32%	0.5398
SVM	60.47%	0.4070
J48	48.65%	0.2298
Random Forest	64.87%	0.4730
Random Tree	47.84%	0.2177
ELM*	67.41%	0.6267

*Using 500 neurons and linear kernel.

Table 8. Results of the classifiers with Haralick and Zernike attribute extractors using 10-fold cross-validation.

Classifier	Accuracy	Kappa Index
BayesNet	51.66%	0.2749
Naive Bayes	50.99%	0.2648
MLP	76.01%	0.6402
SVM	65.54%	0.4832
J48	52.92%	0.2938
Random Forest	64.87%	0.4730
Random Tree	49.78%	0.2466
ELM*	70.06%	0.6566

*Using 500 neurons and linear kernel.

Discussion

The results presented in Tables 3 to 8 showed that the best values of accuracy and Kappa index were obtained by classifiers based on artificial neural networks. The classifiers used were selected because they were able to achieve good results according to the nature of the data. Bayes' naive classifier achieves good results when attributes are statistically independent. Thus, decision boundaries can be modeled through products of one-dimensional Gaussian distributions. Thus, evaluating the performance of Bayes' naive classifier also implies indirectly evaluating the degree of independence of

attributes. Bayesian networks are important to investigate how decision boundaries can be modeled by fairly complex rules. Connectionist learning machines, such as artificial neural networks and support vector machines, return good results when the classification problem is easily generalizable. Decision trees, in turn, model the situation in which data are difficult to generalize, requiring more ad hoc classifiers, composed of many complex rules. Random Forest classifiers are in an intermediate position, and can be used both when the data are easier to generalize (many trees) and more specific (few trees), since they are based on knowledge tree sets.

The generalization capacity of the classifiers is best measured when using cross-validation, since the random division of the data set into training and testing allows to evaluate the generalization capacity without subjecting the classifier to overfitting. Table 6, with results for use only of texture attributes (Haralick) shows that Bayesian and decision tree-based classifiers had similar accuracy scores, around 50%, while support vector machines and neural networks (MLP and ELM) had a performance of 60% and 71%, respectively. For the kappa index, the performance difference is even more evident, with clear advantage for ELM networks. This shows that, from the clinical point of view, although they are still not enough to diagnose breast lesions, the texture attributes have a great contribution to the results.

Analyzing only the Zernike attributes, the classifiers performance was proportionally similar, but with the accuracy of the MLP greater than that of the ELM. The situation is reversed for the Kappa index: higher for ELM than for MLP. When we combine texture and shape attributes, joining moments of Haralick and Zernike, the situation repeats, but with little advantage to MLP over ELM in the case of accuracy. However, the advantage of ELM in relation to MLP considering the Kappa index is quite reasonable. Considering that the ELM has the advantage of rapid training, the results point to the use of neural networks of random weights as important tools for the construction of intelligent systems to support the diagnosis of breast lesions.

This article presented a proposal of a classification method of breast lesions, using features extracted from the texture and geometry of lesions in thermal images, and making comparisons with several classifiers. The use of Zernike alone proved to be very promising in this application and the less satisfactory results occurred when only Haralick attributes were used. However, the best results were obtained by combining Haralick and Zernike moments, what indicates that both texture and geometry information are relevant to differentiate breast lesions through thermographic images. In general, ELM and MLP proved to be quite efficient classifiers for classification of breast lesions in thermographic images.

Using 75% of the database for training, the maximum value obtained for accuracy was 73.38%, with a Kappa index of 0.6007. These results increased to 76.01% of accuracy and Kappa of 0.6402 when using 10-fold cross-validation method to perform the tests. The overall efficiency of the system was 83%.

Furthermore, this study obtained significant and promising findings using ELM as the classifier, which is a much less computational costing machine, and its use may decrease the time to perform the classification without losing classification quality. Future studies may optimize the obtained results by testing other configurations for the classifiers, specially the extreme learning machine, which may become more efficient for the classification of breast lesions in thermographic images than the most commonly used classifiers.

Acknowledgements

We are grateful to Fapema, CNPq and Facepe for the partial support of this research.

References

- Acharya UR, Ng EYK, Tan JH, Sree SV. Thermography based breast cancer detection using texture features and support vector machine. *J Med Syst.* 2012; 36(3):1503-10. PMID:20957511. <http://dx.doi.org/10.1007/s10916-010-9611-z>.
- Aguiar PS Jr, Belfort CNS, Silva AC, Diniz PHB, Lima RCF, Conci A, Paiva AC. Detecção de regiões suspeitas de lesão na mama em imagens térmicas utilizando Spatiogram e redes neurais. *Cad Pesq.* 2013; 20:56-63.
- American Cancer Society. *Global cancer facts & figures.* 3rd ed. Atlanta; 2015.
- Araújo MC, Lima RC, Souza RM. Interval symbolic feature extraction for thermography breast cancer detection. *Expert Syst Appl.* 2014; 41(15):6728-37. <http://dx.doi.org/10.1016/j.eswa.2014.04.027>.
- Belfort CNS, Silva AC, Paiva AC. Detecção de lesões em imagens termográficas da mama utilizando Índice de Similaridade de Jaccard e Artificial Crawlers. In: *Anais do XV Workshop de Informática Médica;* 2015; Recife. Porto Alegre: Sociedade Brasileira de Computação; 2015.
- Bezerra LA, Oliveira MM, Rolim TL, Conci A, Santos FGS, Lyra PRM, Lima RCF. Estimation of breast tumor thermal properties using infrared images. *Signal Process.* 2013; 93(10):2851-63. <http://dx.doi.org/10.1016/j.sigpro.2012.06.002>.
- Breiman L. Random forests. *Mach Learn.* 2001; 45(1):5-32. <http://dx.doi.org/10.1023/A:1010933404324>.
- Cheng HD, Shi XJ, Min R, Hu LM, Cai XP, Du HN. Approaches for automated detection and classification of masses in mammograms. *Pattern Recognit.* 2006; 39(4):646-68. <http://dx.doi.org/10.1016/j.patcog.2005.07.006>.
- Cheng J, Greiner R. Learning Bayesian Belief Network classifiers: algorithms and system. In: *Proceedings of 14th*

- Biennial Conference of the Canadian Society for Computational Studies of Intelligence; 2001; Ottawa, Canada. Berlin: Springer-Verlag; 2001. p. 141-51. Vol. 2056.
- Etehadtavakol M, Ng EY. Breast thermography as a potential non-contact method in the early detection of cancer: a review. *J Mech Med Biol.* 2013; 13(02):1330001. <http://dx.doi.org/10.1142/S0219519413300019>.
- Geurts P, Ernst D, Wehenkel L. Extremely randomized trees. *Mach Learn.* 2006; 63(1):3-42. <http://dx.doi.org/10.1007/s10994-006-6226-1>.
- Gonçalves WN, Machado BB, Bruno OM. Texture descriptor combining fractal dimension and artificial crawlers. *Phys A: Stat Mech App.* 2014; 395:358-70. <http://dx.doi.org/10.1016/j.physa.2013.10.011>.
- Hankare P, Shah K, Nair D, Nair D. Breast cancer detection using thermography. *Int Res J Eng Technol.* 2016; 4(3):1061-4.
- Haykin S. *Redes neurais: princípios e prática.* 2nd ed. Porto Alegre: Bookman; 1999.
- Huang GB, Zhu QY, Siew CK. Extreme learning machine: theory and applications. *Neurocomp.* 2006; 70(1-3):489-501. <http://dx.doi.org/10.1016/j.neucom.2005.12.126>.
- Instituto Nacional de Câncer José Alencar Gomes da Silva. Estimativa 2005/2014 [internet]. Rio de Janeiro: INCA; 2015. [cited 2017 July 24]. Available from: <https://mortalidade.inca.gov.br/MortalidadeWeb/pages/Modelo02/consultar.xhtml#panelResultado>
- Jung Y, Hu J. A k-fold averaging cross-validation procedure. *J Nonparametr Stat.* 2015; 27(2):167-79. PMID:27630515. <http://dx.doi.org/10.1080/10485252.2015.1010532>.
- Klein S, Staring M, Pluim JP. Evaluation of optimization methods for nonrigid medical image registration using mutual information and B-splines. *IEEE Trans Image Process.* 2007; 16(12):2879-90. PMID:18092588. <http://dx.doi.org/10.1109/TIP.2007.909412>.
- Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics.* 1977; 33(1):159-74. PMID:843571. <http://dx.doi.org/10.2307/2529310>.
- Librelotto SR. Análise dos algoritmos de mineração J48 e a priori aplicados na detecção de indicadores da qualidade de vida e saúde. *RevInt.* 2014; 1(1):26-37.
- Milosevic M, Jankovic D, Peulic A. Comparative analysis of breast cancer detection in mammograms and thermograms. *Biomed Tech.* 2015; 60(1):49-56. PMID:25720034. <http://dx.doi.org/10.1515/bmt-2014-0047>.
- Oliveira MM. Desenvolvimento de protocolo e construção de um aparato mecânico para padronização da aquisição de imagens termográficas de mama [dissertation]. Recife: Federal University of Pernambuco; 2012.
- Resmini R, Conci A, Borchardt TB, de Lima RDCF, Montenegro AA, Pantaleão CA. Diagnóstico precoce de doenças mamárias usando imagens térmicas e aprendizado de máquina. *Reavi.* 2012; 1(1):55-67.
- Shanthi S, Bhaskaran VM. A novel approach for detecting and classifying breast cancer. *Int J Intell Inf Technol.* 2013; 9(1):21-39. <http://dx.doi.org/10.4018/jiit.2013010102>.
- Silva ASV. Classificação e segmentação de termogramas de mama para triagem de pacientes residentes em regiões de poucos recursos médicos [dissertation]. Recife: Universidade Federal de Pernambuco; 2015.
- Walker D, Kaczor T. Breast thermography: history, theory, and use. Is this screening tool adequate for standalone use? *Nat Med J.* 2012; 4(7).