Combining deep-wavelet neural networks and support-vector machines to classify breast lesions in thermography images

Maíra Araújo de Santana
Universidade de Pernambuco
Valter Augusto de Freitas Barbosa
Universidade Federal de Pernambuco
Rita de Cássia Fernandes de Lima
Universidade Federal de Pernambuco
Wellington Pinheiro dos Santos (wellington.santos@ufpe.br)
Universidade Federal de Pernambuco https://orcid.org/0000-0003-2558-6602

Research Article

Keywords: breast cancer, breast thermography, breast cancer image diagnosis, deep learning, deep-wavelets

Posted Date: March 28th, 2022

DOI: https://doi.org/10.21203/rs.3.rs-1197402/v1

Combining deep-wavelet neural networks and support-vector machines to classify breast lesions in thermography images

Maíra Araújo de Santana[1,2], Valter Augusto de Freitas Barbosa[3], Rita de Cássia Fernandes de Lima[3] and Wellington Pinheiro dos Santos[2,1*†]

[1]Núcleo de Engenharia da Computação, Escola Politécnica da Universidade de Pernambuco, Recife, Brazil.
[2]Departamento de Engenharia Biomédica, Universidade Federal de Pernambuco, Recife, Brazil.
[2]Departamento de Engenharia Mecânica, Universidade Federal de Pernambuco, Recife, Brazil.

*Corresponding author(s). E-mail(s): wellington.santos@ufpe.br;
Contributing authors: mas2@ecomp.poli.br; valter.augusto@ufpe.br; rita.lima@ufpe.br;
†These authors contributed equally to this work.

Abstract

Purpose: Breast cancer is among the leading causes of cancer death among women. The occurrence of breast cancer is similar both in developed countries and in underdeveloped and developing nations, although mortality is higher in underdeveloped countries due to late detection. Even though mammography is the most used technique for the differential diagnosis of breast cancer, breast thermography can be used as a complementary technique, more accurate than self-examination but accurate enough to guide the use of a mammogram. However, thermal imaging is still difficult for radiologists to understand. Machine learning can help improve this scenario. Deep Wavelet Neural Networks are convolutional neural networks that do not necessarily learn, as they may have predefined filter banks as their neurons. **Methods:** In this work, we propose a deep hybrid architecture to support breast thermography imaging diagnosis based on five-layer Deep-Wavelet

*Corresponding author: Wellington Pinheiro dos Santos, ORCID 0000-0003-2558-6602

Neural Networks, to extract attributes of regions of interest from mammograms, and linear kernel support vector machines for final classification. **Results:** Classical classifiers such as Bayesian classifiers, single hidden layer multilayer perceptrons, decision trees, Random Forests, and support vector machines were tested. The results showed that it is possible to detect and classify injuries with an average accuracy of 99% and an average kappa of 0.99, employing a 5-layer deep-wavelet network and a linear kernel support vector machine as the final classifier. **Conclusion:** Using a deep neural network with prefixed weights from the Wavelets Transform filterbank, it was possible to extract attributes and thus take the problem to a universe where it can be solved with relatively simple decision boundaries like those composed by linear kernel support vector machines. This shows that these new deep networks can be important in building complete solutions to improve breast thermography imaging results to support clinical diagnosis.

Keywords: breast cancer, breast thermography, breast cancer image diagnosis, deep learning, deep-wavelets

1 Introduction

1.1 Motivation and justification

Cancer, in all its forms, has become one of the greatest public health problems of the 20th century worldwide, regardless of the levels of social and economic development of different nations around the globe Groot, Baltussen, Uyl-de Groot, Anderson, and Hortobágyi (2006). Of all the forms of cancer, breast cancer is the most dangerous cancer for older and middle-aged women (Groot et al., 2006). Also the most common form of cancer among women (Groot et al., 2006). Breast cancer is among the five most common types of cancer in the world (Shrivastava, Shrivastava, Jegadeesh, et al., 2017). In Brazil alone, it corresponds to about 28% of new cancer cases per year. Although, in general, there is a good prognosis, this disease is still responsible for the highest cancer mortality rate in the female population (Gonçalves et al., 2016). According to the Ministry of Health of Brazil (MS-BR), the early detection of tumors, which consists of identifying cancer at an early stage, is essential to reduce mortality from the disease (Gonçalves et al., 2016).

Breast cancer has been proliferating both in so-called developed countries and in underdeveloped and developing countries. It follows the increase in the average life expectancy of the population, the swelling of cities, the gradual emptying of the countryside and the adoption of new, more aggressive forms of consumption (Groot et al., 2006). Even though the risk of breast cancer can be reduced through preventive strategies, such as carrying out educational campaigns that encourage visual inspection and touch of the breasts, even a good and well-accepted prevention campaign cannot eliminate most of the types of breast cancer, as they end up being diagnosed too late (Groot et al., 2006). Therefore, the existence and availability of technologies

for early detection of breast cancer in public health systems can contribute to increase the chances of cure and treatment options (Groot et al., 2006).

Currently, the main method used to identify breast lesions is mammography, which consists of a breast scan using x-rays (Maitra & Bandyopadhyay, 2017). However, despite technological advances that have resulted in improved technique and image quality, there are still situations in which mammography is insufficient to identify lesions, especially in their early stages, whether due to limitations of the method itself or inconsistencies in the specialists' diagnosis due to the great variability of clinical cases (Bandyopadhyay, 2010). For this reason, investigations using methods such as ultrasonography, magnetic resonance and clinical examinations in general have been associated with the results obtained through mammography, in order to make the diagnosis more robust (Gonçalves et al., 2016). Even with the combination of these techniques, the Ministry of Health of Brazil still states that the majority of correctly identified cases are currently of advanced stage injuries, which makes treatment difficult, when it is possible to do so, and increases the need to perform procedures invasive ones such as biopsies and mastectomies (total or partial breast removal) (Gonçalves et al., 2016).

Although mammography is still the most reliable non-invasive method in use in clinical practice, breast thermography has emerged as an interesting complement to mammographic analysis. The acquisition of breast thermography images is a non-invasive, painless process, not subject to exposure to ionizing radiation or requiring compression of the patient's breast, as occurs in the acquisition of mammographic images. Alternative methods such as breast thermography have been explored as auxiliary tools for the early diagnosis of breast cancer.

Thermography is based on the acquisition of images, recorded through an infrared camera, which show the temperature distribution in the region. The camera's general operation consists of capturing the infrared radiation emitted by the surface of interest. The technique allows the investigation of physiological effects caused by diseases from the analysis of temperature variation in the region, in the case of the existence of cancer cells, metabolic growth interferes with blood flow, resulting in an increase in the surface temperature of the injured region. Studies claim that the use of thermography can anticipate the diagnosis of breast lesions by up to 10 (ten) years, as information related to physiological changes can be extracted from it, which tend to appear before the anatomical (Etehadtavakol & Ng, 2013). Furthermore, the use of breast thermography as a screening method can greatly reduce the unnecessary exposure of patients to ionizing radiation and other tests.

In this sense, the combination of specialist knowledge with digital image analysis methods in breast thermography can contribute to the improvement of the diagnosis, prognosis and treatment of breast cancer (Bandyopadhyay, 2010; Nordin, Isa, Zamli, Ngah, & Aziz, 2008; Salmeri, Mencattini, Rabottino, Accattatis, & Lojacono, 2009). As with other approaches to supporting diagnostic imaging (Santos, Assis, Santos-Filho, & Lima-Neto, 2009; Santos, Assis, & Souza, 2009; Santos, Assis, Souza, Albuquerque, & Simas, 2008; Santos, Assis, Souza, Mendes, et al., 2009a, 2009b; Santos, Assis, Souza, & Santos-Filho, 2008, 2009; Santos, Souza, & Santos-Filho, 2007; Santos, Souza, Santos-Filho, Lima-Neto, & Assis, 2008; Santos,

Souza, Silva, Portela, & Santos-Filho, 2006a, 2006b; Santos, Souza, Silva, & Santos-Filho, 2008), the extraction of attributes is an essential aspect in obtaining good results in breast image analysis (Bandyopadhyay, 2010; Boquete, Ortega, Miguel-Jiménez, Rodríguez-Ascariz, & Blanco, 2012; Boujelben, Chaabani, Tmar, & Abid, 2009; Mascaro, Mello, Santos, & Cavalcanti, 2009; Nordin et al., 2008). The use of CBIR (*Content-Based Image Retrieval*) techniques in the representation of attributes can contribute to more accurate (Lew, Sebe, Djeraba, & Jain, 2006) analyses. Several studies have explored these aspects and achieved good results (W. Azevedo et al., 2015; W.W. Azevedo et al., 2015; F. Cordeiro, Santos, & Silva-Filhoa, 2013; F.R. Cordeiro, Bezerra, & dos Santos, 2017; F.R. Cordeiro, Santos, & Silva-Filho, 2017, 2016a, 2016b; Cruz, Cruz, & Santos, 2018; de Lima, da Silva-Filho, & dos Santos, 2016; de Vasconcelos, dos Santos, & de Lima, 2018; Lima, Azevedo, Cordeiro, Silva-Filho, & Santos, 2015; Rodrigues et al., 2019).

Overall, the accuracy of diagnosis using conventional techniques is around 70-90%, a percentage that decreases to less than 60% when dealing with women under 40 years old (Urbain, 2005). When it comes to breast thermography, although it is being used and studied in several countries around the world (Etehadtavakol & Ng, 2013), this technique is still scarcely disseminated in Brazil and, for this reason, there are few experts trained to extract relevant information to from the analysis of thermographic images. These factors, associated with the vast variability of clinical cases, make the identification and differentiation of breast lesions based on images a difficult task for human eyes, especially when it comes to small lesions or lesions that are difficult to access. Faced with these challenges, several researches have been dedicated to the study and development of intelligent classification systems to be used as assistants to specialists, in order to optimize the accuracy of the diagnosis.

Therefore, taking into account the relative success of approaches related to artificial intelligence and the need for solutions that enable the early diagnosis of breast cancer, this paper proposes a study of computational approaches for the automatic classification of lesions in thermography images of breast. The study also used a new computational tool for image attribute extraction, the *Deep-Wavelet Neural Network* (DWNN) (de Freitas Barbosa, de Santana, Andrade, de Cássia Fernandes de Lima, & dos Santos, 2020), which consists of a deep and untrained architecture, inspired by Wavelet decomposition on multiple levels.

The article is organized as follows: initially relevant papers in the field of breast cancer diagnosis will be presented using computational tools and breast thermography images. Then, we present the proposal of this study. In the next section, we present and discuss the results obtained, carrying out a quantitative and qualitative evaluation of the explored techniques. Finally, we conclude the article by highlighting the main findings and limitations, in addition to proposing some possibilities for future work.

1.2 Related Works

In the work of Ng, Fok, Peh, Ng, and Sim (2002), the authors sought to train a back-propagation neural network to identify benign or malignant lesions from a database

of 200 breast thermography images. Four different ways of representing the images were tried, through the following aspects:

- Set 1: Mean, median, mode, standard deviation and skewness.
- Set 2: Average, median and mode.
- Set 3: Age, family history, hormone replacement therapy, age at menarche, presence of a palpable nodule, previous surgery or biopsy, presence of nipple discharge, breast pain, menopause aged over 50 years, had a first child with age over 30 years.
- Set 4: Combination of attributes from sets 2 and 3.

The neural network used in this study was configured to have a learning rate of 0.5, with a momentum of 0.4 and a sigmoidal-type activation function. In this study, only one classifier configuration was used, since the aim of the authors was to analyze the different sets of attributes, comparing the effectiveness of each one of them in the representation of thermal images, according to the proposed method.

In this sense, it was observed that the accuracies of sets 1, 2 and 4 were similar to each other, with an approximate value of 61%, however, while the mean square error associated with classification using set 4 was 0.05, both the other sets (1 and 2) had an error equal to 0.12. Group 3, in turn, resulted in an intermediate error of 0.09, however, the accuracy obtained from this representation was around 53%, almost 10% lower than with the other sets.

As for system sensitivity, group 2 had the highest value, 70%, followed by groups 1 and 4, with just over 65%, and finally, group 3, with the lowest sensitivity, of almost 50%. In relation to specificity, set 3 had the best result, close to 80%, while the other sets had worse results and similar to each other, around 40%. Thus, even with less satisfactory results, group 3 was shown to result in lower rates of false positives, but this was not enough to improve the specificity of group 4, which also contained information from group 3. In general, low values of errors, but the results obtained, especially in terms of sensitivity and accuracy, were unsatisfactory, since it is an application in human beings.

Arora et al. (2008) also sought to perform a binary classification of malignant and benign lesions from breast thermography images. In their work, the authors used 94 images (320×240), acquired by themselves and whose diagnoses were previously confirmed by biopsy, of which 60 had a malignant lesion and 34 had a benign lesion.

During the acquisition, the technique of cold stress was used, in which cold air is directed to the breasts during image acquisition. In this study, 3 different image analysis techniques were used: Blinded screening mode (SBS), clinical assessment and artificial neural network (ANN). The first technique results in a risk score, which ranges from 0 (zero), minimal risk, to 7, very high risk. The two other methods give a binary result, whether it is malignant or benign. From the experiments, the authors verified that the approach using ANN stood out positively, in relation to the others, obtaining an accuracy of 81.8%, while for the SBS it was 66.7% and 71.4% for clinical analysis.

A slightly different approach was proposed by Ghayoumi Zadeh, Montazeri, Abaspur Kazerouni, and Haddadnia (2017), who chose to combine an unsupervised with a supervised learning method. The authors used a Self-Organizing Map (SOM)

to perform both the grouping process, as the images did not have a pre-established output class, and the feature extraction process, related to the morphological texture. For the classification step, it was decided to use an MLP network and cross validation with 5 folds. Two different databases were used, the first with 50 images and whose acquisition process was better standardized, and the second with 200 images without much rigor in terms of standardization.

The databases were analyzed separately and, once again, a binary classification was performed, but now of the cancer versus non-cancer type. Using the first database, the proposed method reached up to 100%, both in sensitivity and specificity. These results declined somewhat for second base, which had a sensitivity of 88%, with 99% specificity in detecting breast cancer. Despite having used cross-validation during training, the results, especially for the first database, may indicate overspecialization of the system, given the low dimensionality of the database.

In the study by Raghavendra, Rajendra Acharya, Ng, Tan, and Gudigar (2016), 50 images were used (1280×1024), equally divided into two classes: healthy and with malignant lesion. Its main objective was to evaluate the performance of several intelligent classifiers in the task of grouping images into their respective classes. To extract attributes from the images, they used the Oriented Gradients Histogram (HOG) method; then, these attributes went through two processes, the first of dimensionality reduction, using the KLPP technique (Kernel Locality Preserving Projection, and the second of selection, using selection techniques based on Student's t test, since the set of attributes extracted with the HOG was large and could contain redundancy.

Finally, the following classification methods were tested: decision tree, Linear Discriminant Analysis (LDA), Quadratic Discriminant Analysis (QDA), k-Nearest Neighbour (kNN), Fuzzy Sugeno, Naive Bayes, SVM, AdaBoost, Probabilistic Neural Network (PNN), and Breast Cancer Risk Index (BCRI). At this stage, the tests were performed using cross validation, with 10 folds. Among the evaluated methods, the decision tree achieved the best results, with 98% accuracy, 96% sensitivity and 100% specificity. Again, this system may have been overspecialized as there were only 25 instances per group and no expansion of this data was reported. The fact that the tree stood out in relation to other classifiers may also indicate a low generalizability of the results.

Fernández-Ovies et al. (2019), in turn, also evaluated the detection of breast cancer using the binary problem with healthy and cancer classes. However, in this work, the classification was performed through six different configurations of convolutive neural networks (CNNs) and using the 5-fold cross validation method. The CNNs Resnet18, Resnet34, Resnet50, Resnet152, Vgg16 and Vgg19 were used for classification. As for the image base, the authors chose to use VisionLab, which, in total, contains 5604 images (480×640), some of which were inserted synthetically, to expand the base. 2411 of these images are of healthy breasts, while 534 have cancer, the other images were not used because they did not meet the inclusion criteria established by the team. Since CNN is an extremely sensitive classifier to unbalanced classes, 500 images from each group were randomly selected to participate in the actual experiments.

With the experiments, the authors verified that, in general, the Resnet-type CNN stands out in relation to Vgg. Resnet50 and Resnet34 had the best performances, with average accuracies of 98.75% and 98.13%, respectively. However, Resnet50 proved to be less stable, as it presented a standard deviation of 1.09%, while that of Resnet34 was 0.63%. Resnet18 also presented interesting results, but not enough to surpass the other two.

2 Materials and Methods

In this study, we propose an approach for identifying and classifying breast lesions from breast thermography images. These images were obtained from extension actions carried out in partnership with the Hospital das Clínicas of the Federal University of Pernambuco, Brazil, with women from the Zona da Mata Norte and the Metropolitan Region of Recife, in Northeastern Brazil. The database contains images of both people without breast lesions and patients with cystic, benign and malignant lesions.

Initially, we convert the images from their RGB-JET pattern to grayscale, respecting the temperature information associated with the pseudocolors used in the formation of the original image. Then, we use the DWNN method to extract attributes from the images. Finally, we evaluated the performance of different algorithms in positioning images in their respective classes. This proposed approach is illustrated in Figure 1.

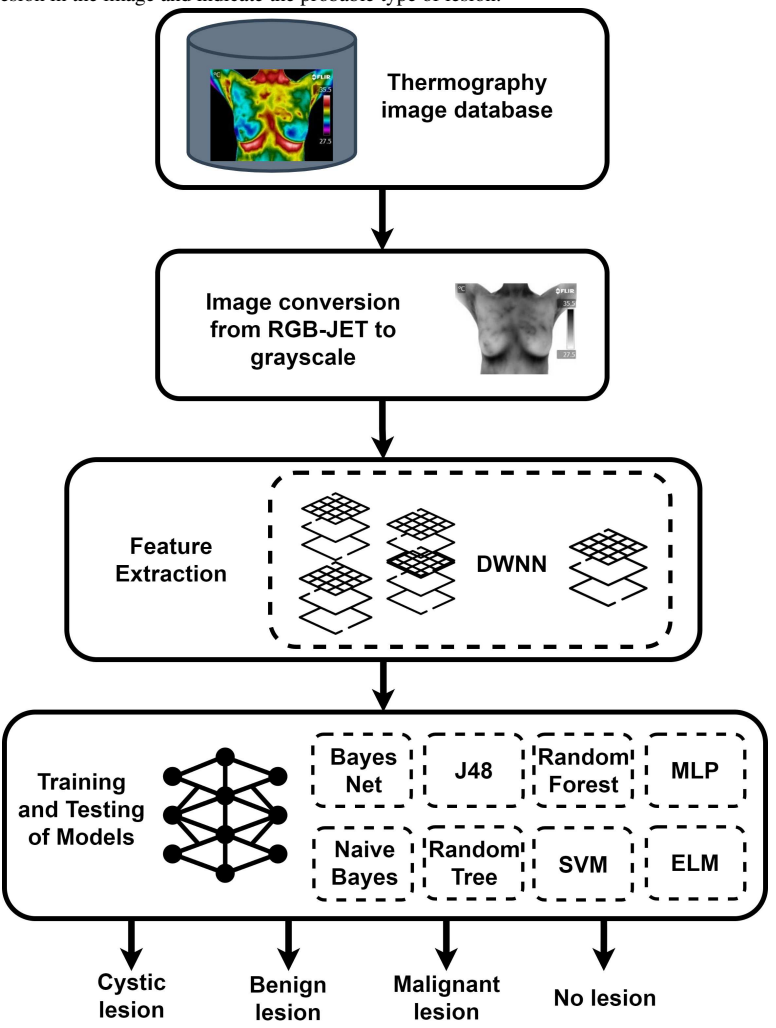
2.1 Database

The thermographic breast image database used is from the Thermal Imaging Laboratory of the Mechanical Engineering Department at UFPE, coordinated by Profa. Rita de Cássia Fernandes de Lima and obtained from extension actions carried out in partnership with the Hospital das Clínicas at UFPE with women from the Zona da Mata Norte of Pernambuco and the Metropolitan Region of Recife, under registration number 279/05 in Research Ethics Committee of the Department of Health Sciences, Federal University of Pernambuco (CEP/CCS/UFPE).

The base consists of 1052 thermographic breast images (480×640 pixels), acquired from a FLIR infrared camera, model ThermaCAMTM S45, with thermal sensitivity of 0.06°C. These images are grouped into 4 distinct classes, according to the diagnosis associated with them, which can be cyst, malignant lesion, benign lesion and no lesion, as illustrated in Figure 2. All diagnoses were established after previous investigation using specific tests for each situation, in the case of cysts, the diagnosis was confirmed by fine needle aspiration puncture (FNAB) or ultrasonography, malignant and benign lesions were confirmed through biopsies and cases from the class without lesion were verified by mammography and ultrasonography, classified as BI-RADS 1, that is, without findings (da Silva, 2015).

In the database there are also images acquired in 8 different positions, with two frontal images of both breasts (T1 and T2), a frontal image of each one of the isolated breasts (MD and ME) and four lateral images (LEMD, LEME, LIMD and LIME). The difference between the T1 and T2 images is in the position of the arms in each

Fig. 1 Proposed method for the identification and classification of breast lesions in breast thermography images. Initially the images are converted to gray scale, then we extract features through DWNN method. Finally, the feature vectors are assessed by different classification algorithms to identify the existence of a breast lesion in the image and indicate the probable type of lesion.



of them, in T1 the arms are at the waist, while in T2 they are upwards. As for the lateral images, LEMD and LEME represent records of the external sides of the right and left breasts, respectively; LIMD and LIME are images of the inner side of each breast. Examples of each of these positions are shown in Figure 3.

Images were acquired following the protocol described by Oliveira (2012). Before the acquisitions, it was necessary to carry out a series of preparations, from the patients and from the acquisition room, since the contact between parts of the

Fig. 2 Examples of images from each of the possible classes in the breast thermography database. In the upper left corner is an example of the uninjured class; right beside it, in the upper right corner, is an image containing a cystic-type lesion. The lower left and right images are, respectively, of benign and malignant lesions.

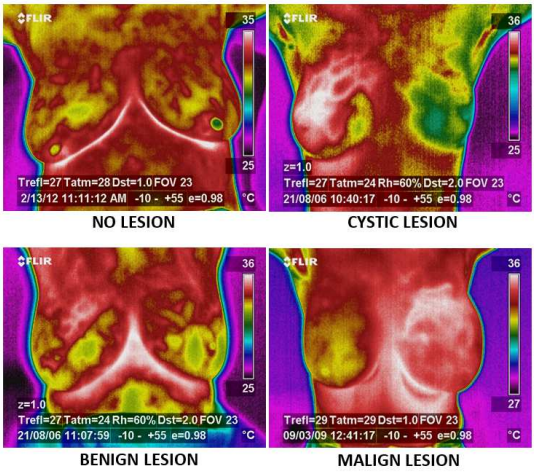
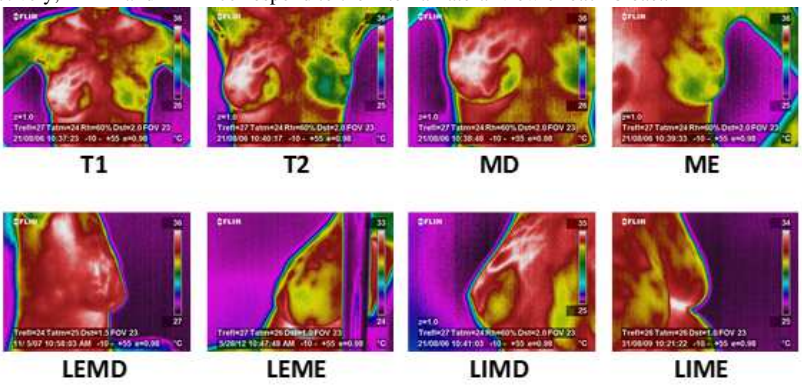


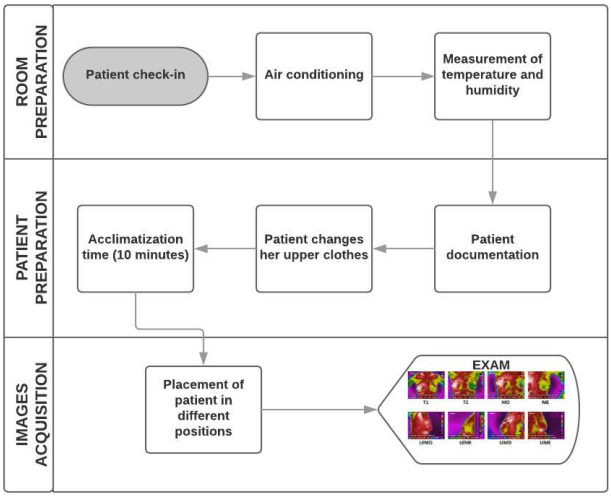
Fig. 3 Positions used in performing the breast thermography exam. In the upper left portion are T1 and T2, which are frontal images of both breasts, with the arms in different positions. MD shows the frontal image of the right breast only, while ME shows the left breast in the same position. The lower images are from lateral views, with LEMD and LEME being images of the external side of the right and left breasts, respectively; LIMD and LIME correspond to the internal lateral view of each breast.



body and the environment, or other parts of the body, changes the surface temperature through the process of heat conduction . Such procedures are described in the diagram in Figure 4.

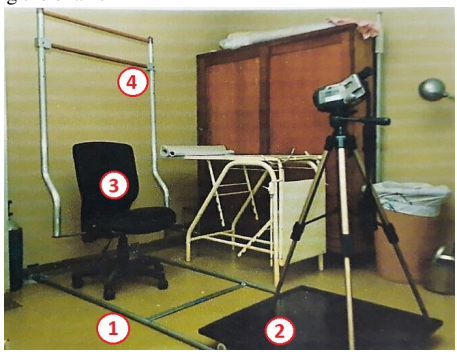
To standardize the acquisitions, in order to avoid changes in the positioning of patients during the process, a mechanical apparatus (Oliveira, 2012) was used. From this apparatus, illustrated in Figure 5, the patient remains seated in the swivel chair (at 3) during the exam, which is moved to change the image angle. The camera is placed on a tripod, which is placed on top of a support (2) that moves closer or farther away

Fig. 4 Preparation procedures for the acquisition of thermographic images



from the patient via the rails (at 1). The horizontal bars (at 4) are used for the correct positioning of the patients' arms. At least eight images were acquired per patient, one in each position; in some specific cases it was necessary to get some extra images.

Fig. 5 Mechanical apparatus used in the acquisition room. In (1) are the rails used to move the camera support (2) towards the patient, positioned on the swivel chair shown in (3); the bars in (4) are used for positioning the arms during the exam.



In this work, we used only the T1 and T2 images, as this positioning favors the visualization of both breasts and the identification of possible anomalies and issues related to asymmetries. Thus, 336 images were used, distributed in 4 classes according to what is presented in Table 1. Since the amount of images per class is different, a class balancing step was added to avoid classification biases, since, with unbalanced bases, algorithms tend to favor classes with more representatives over less populated classes. In this sense, the balance is made from the insertion of synthetic instances,

from the linear combination of real instances de Lima et al. (2016). This process resulted in a total of 968 instances, divided evenly among the classes.

Table 1 Amount of images per class

	Cystic Lesion	Benign Lesion	Malignant Lesion	No Lesion
Number of images	73	121	76	66

Thermography images use the pseudo-color technique, whereby different RGB tones correspond to different temperature values, so that it is possible to obtain the temperature matrix of a thermographic image by performing the inverse transformation of pseudo-colors into temperature . However, for image processing and standardization of pseudo-color information, avoiding errors due to the different color palettes used in image acquisition, it was necessary to perform a RGB-JET conversion to gray levels, mapping the temperature matrix to 0 to 255 levels of gray (8 bits). In this process, lighter shades of gray scale indicate higher temperatures and vice versa. In addition, the images were analyzed in their entirety, without any type of segmentation, therefore using context-based analysis. Although they contain information such as bars, numbers and labels, it was considered that, as this information is always present and in the same position, the algorithms would interpret them as redundant or not relevant to differentiate the classes.

2.2 Feature extraction

For the extraction of attributes we use the method Deep-Wavelet Neural Network (DWNN) (de Freitas Barbosa et al., 2020). DWNN is a deep, untrained network for attribute extraction, inspired by Mallat's (Mallat et al., 1989) algorithm for multilevel Wavelet decomposition. This algorithm, which emerged as a strategy for implementing the discrete wavelet transform, consists of obtaining approximations and details of an image. Approximations are the low-frequency representation of the image, conserving its general trend while smoothing out the abrupt transitions present in it. The details show the image's high-frequency components, highlighting regions of discontinuity such as edges and blemishes.

In Wavelet decomposition, low-pass and high-pass filters are applied to an image to form a set of other images that are smaller in size than the original. From low-pass filtering, approximations are obtained and details are acquired through high-pass filters. This approach allows the analysis of images in the spatial and frequency domains and, therefore, it has been widely used in pattern recognition. DWNN uses a process similar to this one, in which a neuron is formed by combining a given filter with an image size reduction process called downsampling.

The "Deep" in the name arises from the possibility of using multiple layers, making the method even deeper as new layers are added. Furthermore, as with conventional deep learning methods, such as Convolutional Neural Networks (CNN), the process consists of two basic steps, in the first, the images are submitted to filters,

and in the second, pooling is performed. However, while filters are not fixed in conventional methods, the filters used in Deep-Wavelet are fixed and refer to families of Haar wavelets.

Thus, considering a bank with n filters, an input image will be submitted to n neurons that form the first layer of the neural network. In the second layer, each of the images resulting from the first will be individually submitted to the same bank of n filters and to downsampling, in the same way as was done for the input image. The process is repeated for all network layers, according to the amount established by the user and which determines the depth of the network.

Finally, in the DWNN output layer, we have the synthesis blocks (SB), which are responsible for extracting information from the images resulting from the entire process. In these blocks, each of the reduced n^m images will be submitted to a maximum, average, minimum, median or mode function. Thus, each image is replaced by a unique value.

In this work, we used a filter bank with 4 filters and 5 layers to extract attributes, resulting, therefore, in 1024 attributes for each input image. In the synthesis block we used the averaging function to calculate the output of the network, as this one presented better results.

2.3 Classification

After extracting attributes, the set was submitted to a classification step. We evaluate the classification performance of different methods, the Bayesian network (Bayes Net), the naive Bayes classifier (Naive Bayes), the Multilayer Perceptron (MLP), the Support Vector Machine (SVM), the Extreme Learning Machines (ELM), in addition to the search tree-based classifiers: J48, Random Tree and Random Forest. The settings for each classifier are presented in Table 2.

Table 2 Classifiers parameters: SVM with linear kernel; MLP with 100 neurons in the hidden layer; ELM with sigmoid kernel; random forests with 100 trees; standard random tree and J48; and standard Bayesian network and Naive Bayes classifiers.

Classifier	Parameters
Bayesian network	Batch size: 100
Naive Bayes	Batch size: 100
MLP	Hidden layers: 1 Neurons in the hidden layer: 100 Learning rate: 0.3 Momentum: 0.2 Iterations: 500
SVM	Kernel: Linear
J48	-
Random tree	Batch size: 100 Seed = 1
Random Forest	Number of trees: 100 Batch size: 100
ELM	Neurons in the hidden layer: 100 Kernel: Sigmoid

All tests were performed using the 10-fold cross-validation method (Jung & Hu, 2015). The use of this method reduces the variability of results, providing more robustness and reliability to training and reducing the chance of overfitting occurring. In addition, each classifier configuration was tested 30 times, in order to obtain statistical information for comparing the methods.

To assess the performance of the classifiers, accuracy metrics, Kappa index, and confusion matrix were used, and from the latter, sensitivity and specificity metrics were calculated. The accuracy (Ac) corresponds to the percentage of correctly classified instances and can range from 0 to 100% (Krummenauer & Doll, 2000). This metric is calculated from the Equation 1, where $n_{corretas}$ is the number of correctly classified instances and n_{total} is the total number of instances.

$$Ac = \frac{n_{corrected}}{n_{total}} * 100\%. \tag{1}$$

The index, or coefficient, Kappa (κ) is a statistical metric to assess the agreement between the obtained and expected results (Landis & Koch, 1977; McHugh, 2012). The index can vary in the range [-1.1], where values less than or equal to 0 (zero) indicate no agreement between the results, values above 0.8 demonstrate high agreement and intermediate values represent low or moderate agreement. Kappa provides information about the degree of reproducibility of the method and is calculated as indicated in Equation 2, where $P_{calculated}$ represents the observed value and $P_{expected}$ is the expected value.

$$\kappa = \frac{P_{calculated} - P_{expected}}{1 - P_{expected}}. \tag{2}$$

The confusion matrix shows the number of instances classified into each of the possible classes. The main diagonal of the matrix presents the instances correctly classified, they are the true positives (VP) and negatives (VN). The other values represent the instances whose classes were confused, that is, they were misclassified. From this matrix it is possible to observe which classes are more confused, in addition to being possible to calculate the sensitivity and specificity associated with that classification, which are essential metrics to assess the efficiency of diagnostic systems, as they are, respectively, related to rates of false negative (FN) and false positive (FP). Given the matrix presented in Table 3, the sensitivity is calculated from the 3 Equation and the specificity through the 4 Equation.

Table 3 Example of a binary confusion matrix for classes A and B

	Class A	**Class B**
Class A	TP	FP
Class B	TN	FN

$$\text{Sensitivity} = \frac{TN}{TN + FN}, \tag{3}$$

$$\text{Specificity} = \frac{TP}{TP + FP}. \tag{4}$$

Finally, the overall efficiency of the system can be calculated from the relationship between sensitivity and specificity indicated in Equation 5.

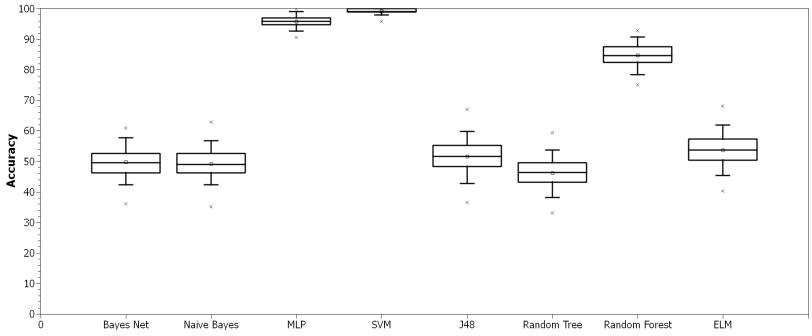
$$\text{Efficiency} = \frac{\text{Sensitivity} + \text{Specificity}}{2}. \tag{5}$$

3 Results

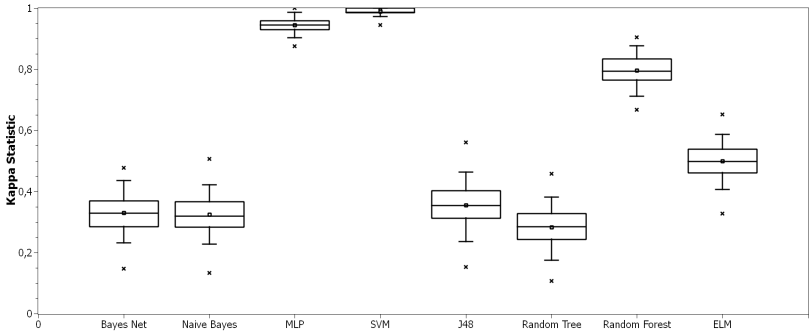
In the thermography basis used in this work, the database is divided into four classes, defined based on the diagnosis given in each case, they are: Cyst, Benign Lesion, Malignant Lesion and No Lesion.

The Deep-Wavelet Neural Network (DWNN) was used to represent the images. The DWNN used has 5 levels and median as a function of the synthesis block. The results obtained in this situation are shown in Figure 6, where graph (a) presents the accuracies for the different classification methods and in (b) the results of kappa indices are shown.

Fig. 6 Results of (a) accuracy and (b) kappa index for the thermography image base using attributes extracted by the Deep-Wavelet Neural Network



(a)



(b)

The results showed that the best performances with MLP were obtained with 100 neurons in the hidden layer and with 100 trees in Random Forest. For the SVM and ELM methods, however, the configurations that resulted in the best performances were, respectively, with linear kernel and with morphological erosion kernel.

4 Discussion

From the experiments, it is observed that the classifiers SVM and MLP stood out positively in relation to the others. In the case of SVM, both the accuracy and the kappa index were close to the maximum values, whose mean values were 99.13% and 0.99, respectively. This method also stood out for presenting the lowest standard deviation values, resulting in a low dispersion of results. There was also a good performance by Random Forest, both in terms of accuracy and in relation to the kappa index. ELM presented intermediate results, with a performance below that presented by Random Forest. The Bayesian classifiers and the J48 and Random Tree trees presented the least satisfactory results, with accuracy values less than 50% and a kappa index of, at most, 0.33.

Overall, the best results for the identification and differentiation of breast lesions on the images were with the SVM classifier. The least satisfactory results, in turn, were obtained with Random Tree. The confusion matrices for these situations (Table 4) reiterate the observations previously presented, since there is much more confusion between the classes with the worst result than in the best situation. From the matrices, it is possible to observe that, in both cases, the benign lesion class is the one that is most confused with the others, especially with the cyst and malignant lesion classes, but there is still confusion between this class and the class where there is no injury occurrence. However, as the degree of confusion is significantly reduced in the situation where the best classification performance occurred, expressive results were obtained for the metrics of sensitivity, specificity and efficiency of the system, as shown in Table 5, a from which the quality of the diagnosis can be analyzed.

Table 4 Confusion matrix for best and worst results using Breast Thermography database

		cyst	**benign**	**malignant**	**no lesion**
Best result	**cyst**	242	0	0	0
	benign	4	235	2	1
	malignant	1	0	241	0
	no lesion	0	0	0	242
Worst result	**cyst**	108	54	38	42
	benign	36	99	58	49
	malignant	36	59	109	38
	no lesion	40	45	34	123

Table 5 Sensitivity, specificity and efficiency metrics of the approach using the breast thermography basis

	Sensitivity	**Specificity**	**Efficiency**
Cyst	100.00%	99.31%	99.66%
Benign lesion	97.11%	74.92%	86.02%
Malignant lesion	99.59%	99.72%	99.66%
No lesion	100.00%	99.86%	99.93%

5 Conclusion

The present study presented a deep network-based hybrid architecture and support vector machine for solving the challenges associated with interpreting breast thermography images. Wavelet-based image descriptors and different classification methods were explored.

From the results, it is observed that the deep network approach with DWNN was able to represent the images well. When associated with support vector machines (SVM), the identification of lesions in breast thermograms reached an accuracy around 99% and more than 0.95 kappa index. Good results were also obtained with MLP, Random Forest and ELM, indicating that the problem can be generalized, but often in a non-linear way. The good performance of these methods explains the low results obtained with decision trees (J48 and Random Tree), since these methods commonly achieve good results when the basis is very specific. The relatively low performances with Bayesian methods point to a considerable dependency between the attributes.

Regarding the metrics that assess the quality of diagnosis, excellent values were obtained for sensitivity, specificity and efficiency in the diagnosis of images with cystic and non-lesion lesions, reaching up to 100% sensitivity in both cases, and values above 99% for the other two metrics. Values above 99.5% were also observed for the three measures, in relation to the diagnosis of malignant lesions. For benign lesions, the approach provided a sensitivity of 97.11%, with a specificity of 74.92% and 86.02% accuracy.

There was even greater difficulty in differentiating benign lesions from other conditions. However, even for this class, high sensitivities were obtained, that is, it can be said that the system is associated with low false negative rates, which is an essential characteristic for a system with application in the diagnostic process.

The results found highlight the power of hybrid approaches based on deep learning for solving complex and non-linear problems, recurrent features in issues associated with biomedical applications. Deep learning techniques such as DWNN, explored here, have been shown to be effective for such applications, as they increase the complexity of the decision frontier for solving the classification problem.

Acknowledgement

We thank Fundação de Amparo à Ciência e Tecnologia do Estado de Pernambuco, FACEPE, and Conselho Nacional de Desenvolvimento Científico e Tecnológico, CNPq, Brazil, for the partial financial support of this research.

Conflict of Interest

All authors declare they have no conflicts of interest.

Compliance with ethical standards

All procedures performed in studies involving human participants were in accordance with the ethical standards of the institutional and/or national research committee and with the 1964 Helsinki declaration and its later amendments or comparable ethical standards.

References

Arora, N., Martins, D., Ruggerio, D., Tousimis, E., Swistel, A., Osborne, M., Simmons, R. (2008). Effectiveness of a noninvasive digital infrared thermal imaging system in the detection of breast cancer. *The American Journal of Surgery*, *196*(1), 523–526.

Azevedo, W., Lima, S., Fernandes, I., Rocha, A., Cordeiro, F., Silva-Filho, A., Santos, W. (2015). Morphological extreme learning machines applied to detect and classify masses in mammograms. *2015 international joint conference on neural networks (ijcnn)* (pp. 1–8). Killarney.

Azevedo, W.W., Lima, S.M., Fernandes, I.M., Rocha, A.D., Cordeiro, F.R., da Silva-Filho, A.G., dos Santos, W.P. (2015). Fuzzy morphological extreme learning machines to detect and classify masses in mammograms. *Fuzzy systems (fuzz-ieee), 2015 ieee international conference on* (pp. 1–8).

Bandyopadhyay, S.K. (2010). Survey on segmentation methods for locating masses in a mammogram image. *International Journal of Computer Applications*, *9*(11), 25–28.

Boquete, L., Ortega, S., Miguel-Jiménez, J.M., Rodríguez-Ascariz, J.M., Blanco, R. (2012). Automated detection of breast cancer in thermal infrared images, based on independent component analysis. *Journal of Medical Systems*, *36*(1), 103–111.

Boujelben, A., Chaabani, A.C., Tmar, H., Abid, M. (2009). Feature extraction from contours shape for tumor analyzing in mammographic images. *Digital image computing: Techniques and applications, 2009. dicta'09.* (pp. 395–399).

Cordeiro, F., Santos, W., Silva-Filhoa, A. (2013). Segmentation of mammography by applying growcut for mass detection. *Studies in health technology and informatics*, *192*, 87.

Cordeiro, F.R., Bezerra, K.F., dos Santos, W.P. (2017). Random walker with fuzzy initialization applied to segment masses in mammography images. *2017 ieee 30th international symposium on computer-based medical systems (cbms)* (p. 156-161). Thessaloniki.

Cordeiro, F.R., Santos, W., Silva-Filho, A.G. (2017). Analysis of supervised and semi-supervised growcut applied to segmentation of masses in mammography images. *Computer Methods in Biomechanics and Biomedical Engineering: Imaging & Visualization*, *5*(4), 297–315.

Cordeiro, F.R., Santos, W.P., Silva-Filho, A.G. (2016a). An adaptive semi-supervised fuzzy growcut algorithm to segment masses of regions of interest of mammographic images. *Applied Soft Computing*, *46*, 613–628.

Cordeiro, F.R., Santos, W.P., Silva-Filho, A.G. (2016b). A semi-supervised fuzzy growcut algorithm to segment and classify regions of interest of mammographic images. *Expert Systems with Applications*, *65*, 116–126.

Cruz, T., Cruz, T., Santos, W. (2018). Detection and classification of lesions in mammographies using neural networks and morphological wavelets. *IEEE Latin America Transactions*, *16*(3), 926–932.

da Silva, A.S.V. (2015). *Classificação e segmentação de termogramas de mama para triagem de pacientes residentes em regiões de poucos recursos médicos* (Unpublished master's thesis). Universidade Federal de Pernambuco, Recife.

de Freitas Barbosa, V.A., de Santana, M.A., Andrade, M.K.S., de Cássia Fernandes de Lima, R., dos Santos, W.P. (2020). Chapter six - deep-wavelet neural networks for breast cancer early diagnosis using mammary termographies. H. Das, C. Pradhan, & N. Dey (Eds.), *Deep learning for data analytics* (p. 99-124). Academic Press. Retrieved from https://www.sciencedirect.com/science/article/pii/B9780128197646000077 https://doi.org/10.1016/B978-0-12-819764-6.00007-7

de Lima, S.M., da Silva-Filho, A.G., dos Santos, W.P. (2016). Detection and classification of masses in mammographic images in a multi-kernel approach. *Computer Methods and Programs in Biomedicine*, *134*, 11–29.

de Vasconcelos, J., dos Santos, W., de Lima, R. (2018). Analysis of methods of classification of breast thermographic images to determine their viability in

the early breast cancer detection. *IEEE Latin America Transactions*, *16*(6), 1631–1637.

Etehadtavakol, M., & Ng, E.Y. (2013). Breast thermography as a potential non-contact method in the early detection of cancer: a review. *Journal of Mechanics in Medicine and Biology*, *13*(02), 1330001.

Fernández-Ovies, F.J., Santiago Alférez-Baquero, E., de Andrés-Galiana, E.J., Cernea, A., Fernández-Muñiz, Z., Fernández-Martínez, J.L. (2019). Detection of breast cancer using infrared thermography and deep neural networks. I. Rojas, O. Valenzuela, F. Rojas, & F. Ortuño (Eds.), *Bioinformatics and biomedical engineering* (pp. 514–523). Cham: Springer International Publishing.

Ghayoumi Zadeh, H., Montazeri, A., Abaspur Kazerouni, I., Haddadnia, J. (2017). Clustering and screening for breast cancer on thermal images using a combination of som and mlp. *Computer Methods in Biomechanics and Biomedical Engineering: Imaging & Visualization*, *5*(1), 68–76.

Gonçalves, J.G., Siqueira, A.D.S.E., de Almeira Rocha, I.G., de Lima, E.F.F., da Silva Alves, L., da Silva, B.O., ... Land, M.G.P. (2016). Evolução histórica das políticas para o controle do câncer de mama no brasil. *DIVERSITATES International Journal*, *8*(1).

Groot, M.T., Baltussen, R., Uyl-de Groot, C.A., Anderson, B.O., Hortobágyi, G.N. (2006). Costs and health effects of breast cancer interventions in epidemiologically different regions of africa, north america, and asia. *The Breast Journal*, *12*(s1), S81–S90.

Jung, Y., & Hu, J. (2015). A k-fold averaging cross-validation procedure. *Journal of Nonparametric Statistics*, *27*(2), 167-179.

Krummenauer, F., & Doll, G. (2000, 06). Statistical methods for the comparison of measurements derived from orthodontic imaging. *European Journal of Orthodontics*, *22*(3), 257-269.

Landis, J.R., & Koch, G.G. (1977, 5). The measurement of observer agreement for categorical data. *Biometrics*, *33*(1), 159-174.

Lew, M.S., Sebe, N., Djeraba, C., Jain, R. (2006, February). Content-based multimedia information retrieval: State of the art and challenges. *ACM Trans. Multimedia Comput. Commun. Appl.*, *2*(1), 1–19.

Lima, S., Azevedo, W., Cordeiro, F., Silva-Filho, A., Santos, W. (2015). Feature extraction employing fuzzy-morphological decomposition for detection and classification of mass on mammograms. *Annual international conference of the ieee engineering in medicine and biology society. ieee engineering in medicine and biology society. annual conference* (Vol. 2015, pp. 801–804).

Maitra, I.K., & Bandyopadhyay, S.K. (2017). Identification of abnormal masses in digital mammogram using statistical decision making. *Hybrid Intelligence for Image Analysis and Understanding*, 339–368.

Mallat, S.G., et al. (1989). Multifrequency channel decompositions of images and wavelet models. *IEEE Trans. Acoustics, Speech, and Signal Processing*, *37*(12), 2091–2110.

Mascaro, A.A., Mello, C.A., Santos, W.P., Cavalcanti, G.D. (2009). Mammographic images segmentation using texture descriptors. *2009 annual international conference of the ieee engineering in medicine and biology society* (pp. 3653–3653).

McHugh, M.L. (2012). Interrater reliability: the kappa statistic. *Biochemia Medica*, *22*(3), 276–282.

Ng, E.-K., Fok, S., Peh, Y., Ng, F., Sim, L. (2002). Computerized detection of breast cancer with artificial intelligence and thermograms. *Journal of medical engineering & technology*, *26*(4), 152–157.

Nordin, Z.M., Isa, N.A.M., Zamli, K.Z., Ngah, U.K., Aziz, M.E. (2008). Semi-automated region of interest selection tool for mammographic image. *2008 international symposium on information technology* (Vol. 1, pp. 1–6).

Oliveira, M. (2012). *Desenvolvimento de protocolo e construção de um aparato mecânico para padronização da aquisição de imagens termográficas de mama [dissertation]* (Unpublished master's thesis). Universidade Federal de Pernambuco, Recife.

Raghavendra, U., Rajendra Acharya, U., Ng, E., Tan, J.-H., Gudigar, A. (2016). An integrated index for breast cancer identification using histogram of oriented gradient and kernel locality preserving projection features extracted from

thermograms. *Quantitative InfraRed Thermography Journal*, *13*(2), 195–209.

Rodrigues, A.L., de Santana, M.A., Azevedo, W.W., Bezerra, R.S., Barbosa, V.A., de Lima, R.C., dos Santos, W.P. (2019). Identification of mammary lesions in thermographic images: feature selection study using genetic algorithms and particle swarm optimization. *Research on Biomedical Engineering*, *35*(3), 213–222.

Salmeri, M., Mencattini, A., Rabottino, G., Accattatis, A., Lojacono, R. (2009). Assisted breast cancer diagnosis environment: A tool for dicom mammographic images analysis. *2009 ieee international workshop on medical measurements and applications* (pp. 160–165).

Santos, W.P., Assis, F.M., Santos-Filho, R.E.S.P.B., Lima-Neto, F.B. (2009). Dialectical Multispectral Classification of Diffusion-Weighted Magnetic Resonance Images as an Alternative to Apparent Diffusion Coefficients Maps to Perform Anatomical Analysis. *Computerized Medical Imaging and Graphics*, *33*(6), 442-460.

Santos, W.P., Assis, F.M., Souza, R.E. (2009). MRI Segmentation using Dialectical Optimization. *31st annual international conference of the ieee engineering in medicine and biology society* (p. 5752-5755). Minneapolis, USA.

Santos, W.P., Assis, F.M., Souza, R.E., Albuquerque, A.C.T.C., Simas, M.L.B. (2008). A Monospectral Approach for fMRI Analysis using Kohonen Self-Organized Networks and Objective Dialectical Classifiers. *International Journal of Innovative Computing and Applications*, *1*(4), 260-273.

Santos, W.P., Assis, F.M., Souza, R.E., Mendes, P.B., Monteiro, H.S.S., Alves, H.D. (2009a). A Dialectical Method to Classify Alzheimer's Magnetic Resonance Images. W.P. Santos (Ed.), *Evolutionary computation* (p. 473-486). Vukovar: InTech.

Santos, W.P., Assis, F.M., Souza, R.E., Mendes, P.B., Monteiro, H.S.S., Alves, H.D. (2009b). Dialectical Non-Supervised Image Classification. *Ieee congress on evolutionary computation (cec 2009)* (p. 2480-2487). Trondheim.

Santos, W.P., Assis, F.M., Souza, R.E., Santos-Filho, P.B. (2008). Evaluation of Alzheimer's Disease by Analysis of MR Images using Objective Dialectical Classifiers as an Alternative to ADC Maps. *30th annual international conference of the ieee engineering in medicine and biology society* (p. 5506-5509). Vancouver, Canada.

Santos, W.P., Assis, F.M., Souza, R.E., Santos-Filho, P.B. (2009). Dialectical Classification of MR Images for the Evaluation of Alzheimer's Disease. G.R. Naik (Ed.), *Recent advances in biomedical engineering* (p. 241-250). Vukovar: InTech.

Santos, W.P., Souza, R.E., Santos-Filho, P.B. (2007). Evaluation of Alzheimer's disease by analysis of MR images using multilayer perceptrons and Kohonen SOM classifiers as an alternative to the ADC maps. *29th annual international conference of the ieee engineering in medicine and biology society* (p. 2118-2121). Lyon, France.

Santos, W.P., Souza, R.E., Santos-Filho, P.B., Lima-Neto, F.B., Assis, F.M. (2008). A Dialectical Approach for Classification of DW-MR Alzheimer's Images. *Ieee world congress on computational intelligence (wcci 2008)* (p. 1728-1735). Hong Kong, China.

Santos, W.P., Souza, R.E., Silva, A.F.D., Portela, N.M., Santos-Filho, P.B. (2006a). Análise multiespectral de imagens cerebrais de ressonância magnética ponderadas em difusão usando lógica nebulosa e redes neurais para avaliação de danos causados pela doença de Alzheimer. *Xi congresso brasileiro de física médica.* Ribeirão Preto, Brasil.

Santos, W.P., Souza, R.E., Silva, A.F.D., Portela, N.M., Santos-Filho, P.B. (2006b). Avaliação da doença de Alzheimer por análise de imagens de RMN utilizando redes MLP e máquinas de comitê. *Xx congresso brasileiro de engenharia biomédica.* São Pedro, Brasil.

Santos, W.P., Souza, R.E., Silva, A.F.D., Santos-Filho, P.B. (2008). Evaluation of Alzheimer's disease by analysis of MR images using multilayer perceptrons and committee machines. *Computerized Medical Imaging and Graphics*, *32*(1), 17-21.

Shrivastava, S.R., Shrivastava, P.S., Jegadeesh, R., et al. (2017). Ensuring early detection of cancer in low-and middle-income nations: World health organization. *Archives of Medicine and Health Sciences*, *5*(1), 141.

Urbain, J.-L. (2005). Breast cancer screening, diagnostic accuracy and health care policies. *Canadian Medical Association Journal*, *172*(2), 210–211.