Features Selection Study for Breast Cancer Diagnosis Using Thermographic Images, Genetic Algorithms, and Particle Swarm Optimization

Amanda Lays Rodrigues da Silva, Federal University of Pernambuco, Recife, Brazil

Maíra Araújo de Santana, University of Pernambuco, Recife, Brazil

Clarisse Lins de Lima, University of Pernambuco, Recife, Brazil

José Filipe Silva de Andrade, Federal University of Pernambuco, Recife, Brazil

Thifany Ketuli Silva de Souza, Federal University of Pernambuco, Recife, Brazil

Maria Beatriz Jacinto de Almeida, Federal University of Pernambuco, Recife, Brazil

Washington Wagner Azevedo da Silva, Federal University of Pernambuco, Recife, Brazil

Rita de Cássia Fernandes de Lima, Federal University of Pernambuco, Recife, Brazil

Wellington Pinheiro dos Santos, Federal University of Pernambuco, Recife, Brazil
  https://orcid.org/0000-0003-2558-6602

ABSTRACT

Early detection of breast cancer is critical to improve treatment efficiency and chance of cure. Mammography is the main method for breast cancer screening; however, it has some limitations. Infrared thermography is a technique that is being studied for its benefits. The existing tumor classification systems are detailed, complex, and have low usability. Therefore, combining specialized professionals with methods of digital image analysis using thermography can help improve the diagnosis. Considering this, some computational areas are working on studies and creating methods to assess these data. The features selection plays a key role in this process, as it is a way to help solving data multidimensionality problems. This study aims to reduce the amount of features from thermographic images with mammary lesions. The authors used genetic algorithm and particle swarm optimization for features selection and compared the performance of each method to the performance using the entire set of features.

KEYWORDS

INTRODUCTION

According to the Brazilian data from the National Cancer Institute (INCA), breast cancer was responsible for about 28% of new cases of cancer in 2018. Breast cancer is, then, the second type of cancer that most affects women in Brazil. In this country, 59,700 new cases of the disease were estimated for 2019 (Instituto Nacional de Câncer, 2017). Another worrying fact is the increased occurrence in young women (Borchartt et al., 2013). This growth may be partly because of the increased incidence due to a greater exposure of women to risk factors resulting from the urbanization process and changes in lifestyle (Porter, 2008). The situation is aggravated by the aging population that occurred intensely in Brazil (Victora et al., 2011). Due to this large number of influences, early detection of the disease is critical since the sooner the disease is detected, the better are the treatments. Therefore, the early detection provides better chances of cure of the patient, which would lead to a decrease in the fatality rate due to this type of cancer (Borchartt et al., 2013). According to Lessa and Marengoni (2016), the chances of cure dramatically reduce if the disease is not diagnosed in the early stages (Lessa & Marengoni, 2016).

Nowadays, mammography is the gold standard for breast cancer diagnosis. However, this exam has some limitations, such as the deficiency in detecting the disease in the case of dense breasts, which are breasts mainly made of glandular tissue. This type of breast is the most predominant in young patients. In addition, mammography has high rates of false positives and exposures patient to ionizing radiation. The exposure to ionizing radiation may even increase the chances of developing the disease (Borchartt et al., 2013). Thus, there is a need for other methodologies to support the diagnosis of this type of cancer (Leles, 2015).

Infrared thermography is a fast, inexpensive and noninvasive screening technique that aims to record the radiation emitted by the patient's skin surface. The temperature variations of cancerous tissue in relation to healthy neighboring tissue are due to the angiogenesis process. Through this process, the lesion stimulates the creation of new blood vessels for its nutrition. With more vessels feeding the lesion, the temperature of the region is higher than the temperature of healthy regions (Dourado Neto, 2014). Keyserlingk et al. (1998) reported in their paper that sensitivity for detecting ductal carcinoma-type cancer is significantly improved when combining mammography to thermal imaging (Keyserlingk et al., 1998). The sensitivity when combining these two techniques was 95% (Andrade, Paiva & Correa, 2017). Thus, due to the facts explained above, thermography is an attractive technique to help to provide an early detection of breast cancer.

Tumor classification systems used today are detailed, complex and usually hard to be used by pathologists, so, the specialists do not feel comfortable using them (Ferreira, Oliveira & Martinez, 2011). However, it is proven that the combination of specialized professionals with digital image analysis methods applied to breast thermography can contribute to the improvement of breast cancer diagnosis, prognosis and treatment (Bandyopdhyay, 2010). As a consequence, intelligent systems are being developed as a decision support tool in many health areas. These systems may help pathologists to perform a more objective, accurate and uniform classification of the lesions. Thus, minimizing the limitations imposed by the existing classification systems, and thus accelerating the work of this professional (Ferreira, Oliveira & Martinez, 2011).

This technological advance has contributed to the generation and storage of an amount of data that is constantly increasing at a faster rate than we are able to process. From this, several areas have been dedicated to the research and the proposal of methods and processes to treat these data (Andrade, Santana & Santos, 2018; Azevedo et al., 2015; Souza et al., 2019). The features selection plays a critical role in this process, since it may overcome an important issue in machine learning: the computational cost. Features selection is often performed as a preprocessing step. Its purpose is to select the most important features, since the non-relevant or redundant features may reduce the accuracy, as well as increase the computational cost (Lee, 2005; Souza, 2017). This leads to the following issue: How many and which features are enough or necessary to describe a particular problem? For many problems, we

can use evolutionary algorithms to find the best possible solution. This work aimed to evaluate the performance of the most relevant features subsets, selected by Genetic Algorithm (GA) and Particle Swarm Optimization (PSO), in the problem of breast lesion classification in thermographic images.

BACKGROUND

The breasts are exocrine glands that participate and help in the human reproduction process. It is located in the outermost portion of the anterior chest and their functions are the production, storage and release of milk (Moore, Dalley & Agur, 2001). The breast is made by the grouping of alveoli, which are the smallest units of the gland. This set of structures constitutes the morphofunctional unit of the breast, the mammary lobes. The grouping of the lobes builds the mammary parenchyma. The parenchyma is surrounded by stroma, which is characterized by the presence of fatty and connective tissues, vessels and nerves (Menke, 2000). Externally, the parenchyma has the areola (that has a small prominences due to the opening of the Montgomery sebaceous gland ducts) and the mammary papilla (center of the areolas). There is also a tubular structures responsible for draining milk ducts (Moore, Dalley & Agur, 2001).

The hormones release that occur in the female body causes notable and important changes for puberty and development. Such as the release of estrogen and progesterone, which act on the growth of the gland (including its development in lactating structures) and breastfeeding, respectively. (Guyton & Hall, 2011). The prolactin increases during pregnancy and postpartum. During the postpartum, the prolactin acts directly on breast milk production (Guyton & Hall, 2011; Menke, 2000).

The breast tissue responds to environmental, dietary and hormonal stimuli, causing cellular changes and, thus, a state susceptible to benign and malignant diseases (Saladin & Porth, 2010). The breast cancer is a malignant neoplasm that usually relies on the formation of a mass. This mass consists in a cluster of cells that have lost their ability to structure themselves (Weinberg, 2008).

There are some risk factors linked to the increased incidence of breast cancer. Among them are:

- The gender influence, since women are more affected by this pathology;
- Women aged from 50 to 60 years;
- Statistics indicates that menarche before the age 11 and late menopause influence this type of cell mutation. As well as postmenopausal hormone replacement therapy, exposure to excess estrogen and progesterone (Brasileiro Filho, 2011; Kumar et al., 2010);
- Genetic inheritance suggests a greater predisposition to the development of breast cancer. This occurs mainly when there are mutations in the BRCA1 and BRCA2 genes. These genes are directly associated to cell metabolism, cell cycle regulation and cell control (Coelho et al., 2018).

Breast carcinoma can be classified into two types: *in situ* and invasive. The *in situ* is delimited to the ducts whereas the invasive infiltrates the stroma. These two categories may be subdivided as to the location into: ductal and lobular (both *in situ* and invasive). There are also some special cancers that are not grouped into the aforementioned types, such as tubular, metaplastic, inflammatory, mucinous, and medullary (Brasileiro Filho, 2011; Kumar et al., 2010).

It is necessary to know about the prognostic marker, since it allows the understanding of the biological characteristics of the tumor and its growth. The prognostic marker also provides useful information to select the best therapeutic intervention for the patient (Brasileiro Filho, 2011). These factors include: its classification, tumor size, histology, and lymph node involvement. Knowing which hormonal receptors the breast mammary carcinoma has is considered an important predictor of hormone treatment. There are cancers that are ER-expressive and PR-expressive, directly related to receptors of estrogen and progesterone, respectively. In such cases, the hormone treatment often provides good results and a positive short-term prognosis by reducing or blocking these hormones. It is also observed that those expressing the HER2/neu protein (which promotes the growth of breast

cells), have treatments that act directly on this protein, increasing the chances of success (Brasileiro Filho, 2011; Kumar et al., 2010).

As mentioned before, breast cancer is the most common type of cancer in women. Early diagnosis provides a positive prognosis and a high survival rate (Brasileiro Filho, 2011). Therefore, there is a huge relevance in investing on tools and methods that help to provide an efficient diagnosis. Nowadays, there are several exams used in the process of breast cancer diagnosis. One of the main exam is the self-examination of the breasts, where the patients palpate their own breasts. Some imaging exams are also used. In order to improve cost-effectiveness, accessibility and usability, people are studying and investing on breast thermography as a complementary tool for this process.

Thermography is a non-invasive, non-contact, non-ionizing, fast and totally painless exam. It does not emit ionizing radiation and has no side effects. This technique consists in the recording the irradiated temperature of the skin, based on the principle of heat transfer regulation of the human body to the environment through thermoregulation. This thermoregulation can change the cross section of blood vessels, the irrigation in the tissues and, finally, the temperature (Ring & Ammer, 2012; Sessler, 1994). Then, the presence of a pathology will trigger a strong blood supply in the affected region in order to transfer antibodies to fight the disease. Thermography would then be able to detect these irregularities due to temperature changes present in this region (Ring & Ammer, 2012).

Its procedure basically consists in three main steps. The first one is the preparation of the patient and the environment, balancing the thermal conditions, so as not to have influence of external variables. The second step is the images acquisition using an infrared camera, which captures the infrared radiation emitted by the body. Lastly, there is the identification of the cancer stage (Araújo, 2009; Cockburn, 2018).

After the exam, the images can be analyzed by a software, usually guided by artificial intelligence applications. The results have been satisfactory in different computational methods used in the representation and recognition of thermal data (Pereira et al., 2017; Santana et al., 2017; Santana et al., 2018).

Artificial Intelligence

Over time, the definition of artificial intelligence followed four lines of thought: Systems that think like humans (Haugeland, 1985), systems that act like humans (Kurzweil, 1990), systems that think logically (Charniak & Mcdermott, 1985) and logically acting systems (Poole, Mackworth & Goebel, 1998). Sortly explaining, we can categorize artificial intelligence into empirical and theoretical. In the first one, an act is performed by experimental confirmation. The theoretical uses mathematics and engineering as a tool. The machine learning technique is a way to build an artificial intelligence system, which allows the autonomy of machines through the use of algorithms (Jordan & Mitchell, 2015).

The use of artificial intelligence to aid diagnosis in the thermography field is based on the analysis of thermograms captured by an infrared camera. Since a breast pathology causes an increase in temperature at the site of the lesion, the affected region will be visible in the images as warmer spots. In some approaches, images will be segmented into two, left breast and right breast. The criterion for suspicion of an anomaly is defined by the areas of temperature differences between the breasts. In other approaches, as the one used in this chapter, the image is entirely assessed, without segmentation, thus, performing a context-based analysis.

To enable the interpretation of images by the machine, the features or attributes are extracted. These features may assume values that are often incomprehensible to humans but intelligible to the machine. The attribute set is organized as a vector for each image. Thus, for an entire image database, the attribute set becomes a matrix, where each row represents an instance/image and each column is a feature. The moments of Haralick and Zernike are some of the most widely used features extractors. The Haralick moments provide information about texture by analyzing occurrence between gray levels. And Zernike moments evaluate shape-related patterns (Felipe, Olioti & Traina, 2005; Oliveira et al., 2012; Santana et al., 2018).

The features are used as input to the training. The attribute matrix generally assumes large dimensions due to the amount of extracted features. This results in a high computational cost. For this reason, it is advisable to reduce this matrix. The dimensionality reduction of attribute can be done through features selection processes. Features selection consists of choosing a subset of relevant features from the acquired ones (Blum & Langley, 1997). In this process, all unnecessary or redundant features are removed from the database. Among the most common types of features selection techniques are Genetic Algorithms and Particle Swarm Optimization, both used in the study presented in this chapter.

Genetic Algorithms

Genetic Algorithm (GA) is a particular class of evolutionary algorithms that use techniques inspired by the Darwinian principle of natural selection and genetic reproduction. According to Darwin's theory, the selection process allows greater longevity for the fittest individuals, ie, the highest probability of reproduction. These individuals are more likely to perpetuate their genetic codes for generations to come. The genetic codes is the identity of each individual and are represented by chromosomes (Kalfels, 2017). The GAs are constructed by reproducing these principles, seeking the best solution to a given problem by evolving sets (populations) of solutions (individuals) encoded as artificial chromosomes. The chromosomes are implemented as a data structure to represent a possible solution in the search space of the problem. These chromosomes are, then, subjected to an evolutionary process consisting of the evaluation of aptitude, selection and genetic operators (crossover and mutation). Throughout many execution of this process, the resulting population tends to contain increasingly fit individuals, ie better solutions (Kalfels, 2017). Figure 1 presents a flowchart of the execution of a GA, with the stopping criterion being the value of the greatest aptitude.

Figure 1. Diagram of a GA execution (Source: Adapted from Kalfels, 2017)
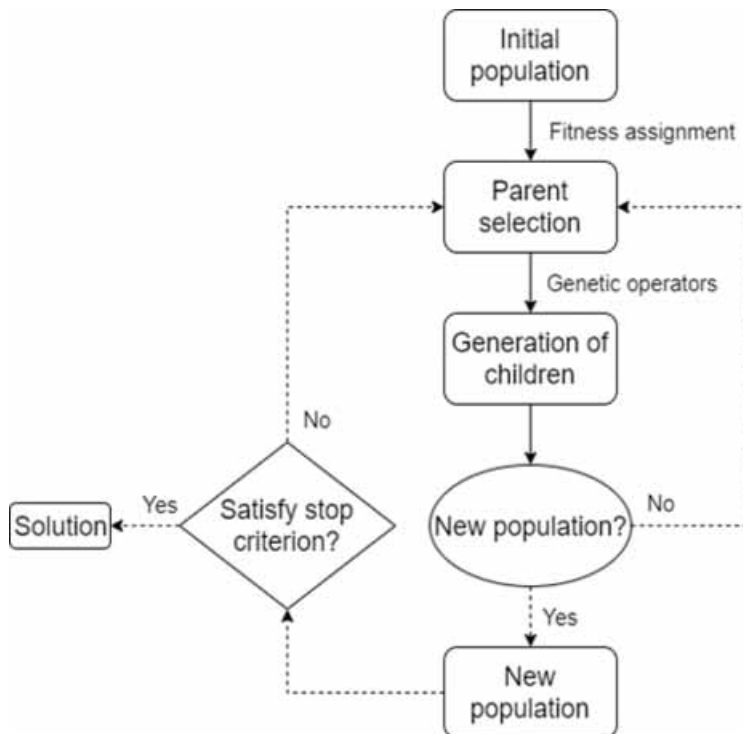
Particle Swarm Optimization

This algorithm was introduced in 1995 by Kennedy and Eberhart (Kennedy & Eberhart, 1995). It is a technique that is based on the collective movement of a group of particles: the particle swarm. Each member of this swarm is moved through the search space of the problem by two forces. One attracts them, with a random magnitude, to the best location found by oneself (pbest). The other force attracts them to the best location found among some or all of the swarm members (gbest). The position and the velocity of each particle are updated at each iteration until the whole cluster converges to obtain the best result (Nascimento et al., 2012). Figure 2 summarizes the execution of a PSO algorithm.

   The velocity vector of each particle must be updated by:

$$v_{k+1}^i = w \cdot v_k^i + c_1 \cdot r_1 \cdot \left(p^i - x_k^i \right) + c_2 \cdot r_2 \cdot \left(p_k^s - x_k^i \right) \tag{1}$$

where v_k^i is the current particle velocity, p^i is the best position found by particle i, p_k^s is the best position among all particles in iteration k. The parameter w represents the particle inertia and controls its ability to explore the solution space. Thus, a high value determines a global search while a low value defines a local search. Usually, these values range from 0.4 to 1.4. Parameters c_1 and c_2 are called trust parameters and define how much a particle trusts itself (c_1) or the group (c_2). Commonly both assume a value of two. Finally, r_1 and r_2 are random numbers between zero and one (Siciliano, 2007).

Figure 2. Diagram of a PSO execution (Source: Adapted from Menezes, 2007)
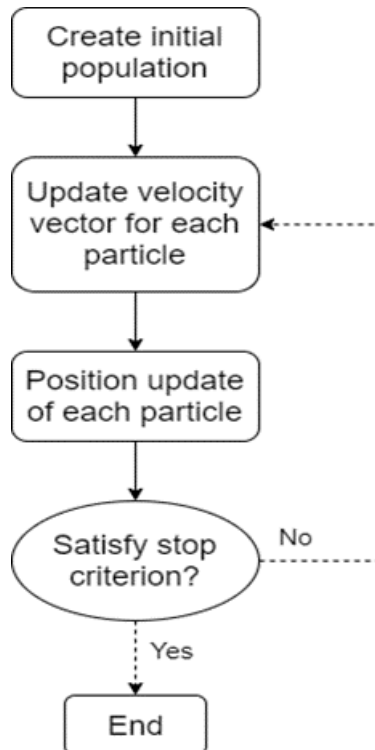
Literature Review

In the study of Gautherie et al. (1987), the images were collected from patients with suspected breast cancer referred to Imam Khomeini Hospital. The algorithms were AdaBoost, SVM, k-NN, NB and PNN. The GA classifier combined with AdaBoost achieved the best accuracy for 0° and 90° imaging on both breasts. And the best performance was achieved on 45° imaging of the left breast. In addition, the mRMR classifier combined with AdaBoost had the best accuracy in 45° images of the right breasts.

In the work of Santana et al. (2018), the thermographic images obtained at the University Hospital of the Federal University of Pernambuco were used. Haralick and Zernike descriptors were used to extract information about the texture and shape of the images, respectively. Initially, the images were classified as: cyst, malignant and benign. The behavior of several classifiers based on artificial neural networks, decision trees and bayesian classifiers was analyzed. The Extreme Learning Machine (ELM) and Multilayer Perceptron (MLP) algorithms were the most efficient, obtaining 78% sensitivity and 88% specificity.

Fernández-Ovies et al. (2019) also evaluated the performance of intelligent algorithms for breast cancer detection in thermographic images. In their study, the classification was performed through six different configurations of Convolutional Neural Networks (CNN) and using the 5-fold cross validation method. The authors used a database of 1000 images, being 500 of healthy breast and 500 images of breast cancer. The authors found that, in general, the Resnet type of CNN outperformed the Vgg type. With Resnet CNN, they obtained average accuracies of up to 98.75%.

In the study of Sánchez-Ruiz, Olmos-Pineda and Olvera-López (2020) they used texture features to describe bilateral breast infrared images and evaluated the performance of an Artificial Neural Network (ANN) architecture in identifying breast cancer. A Genetic Algorithm with population size of 100, crossing rate of 70% and 100 generations was used to perform feature selection. The use of features vector optimized by GA improved classification performance in at least 10%, achieving an average accuracy of 90.17%.

Mishra and Rath (2020) also used texture features to describe breast thermograms, however, they invested in a different feature selection approach. They compared the performance of Principal Component Analysis (PCA) and Autoencoder as feature selectors. By using PCA they tried to find the most relevant features. In a second moment, they used Autoencoder in an attempt to remove the redundant features. These methods reduced from 27 features to 7 (using PCA) and 9 (using Autoencoder). Moreover, the set of features selected by PCA resulted in a better representation. Breast classification performance into normal and abnormal using features selected by PCA was of accuracy of 95.45% with specificity of 88.07%. Using Autoencoder, the accuracy decreased to 81.93% with specificity of 65.45%.

METHOD

This study uses thermographic images captured in JPG format by a FLIR infrared camera model S45 with a 480 x 640 pixel resolution. They were locally obtained from a cooperation with the breast thermography research group of the Department of Mechanical Engineering from Federal University of Pernambuco (UFPE). Images acquisition was performed according to the protocol described in Oliveira (2012). Since thermographic images are highly sensitive to environmental changes, room temperature and humidity were measured and controlled during whole process, with no windows or door opened and no airflow directed towards the patient. All patients had their temperature measured and went through a 10 minutes acclimatization period without touching their breasts. Moreover, a mechanical device was used to control positioning and patient-camera distance (Oliveira, 2012).

Breast images were acquired for each subject in eight different positions: two frontal images of both breasts (T1 and T2) and three images of each breast, separately, right and left, at different angles, frontal (MD and ME), external lateral (LEMD and LEME) and internal lateral (LIMD and LIME) (Santana et al., 2017; Santana et al., 2018). These positions are shown in Figure 3. Dataset was

Figure 3. Positions of image acquisition per thermographic exam. T1 and T2 are bilateral frontal images, which were the positions used in this study. MD and ME are unilateral frontal images. LEMD, LEME, LIMD and LIME shows each breast in different lateral views (Source: The authors)
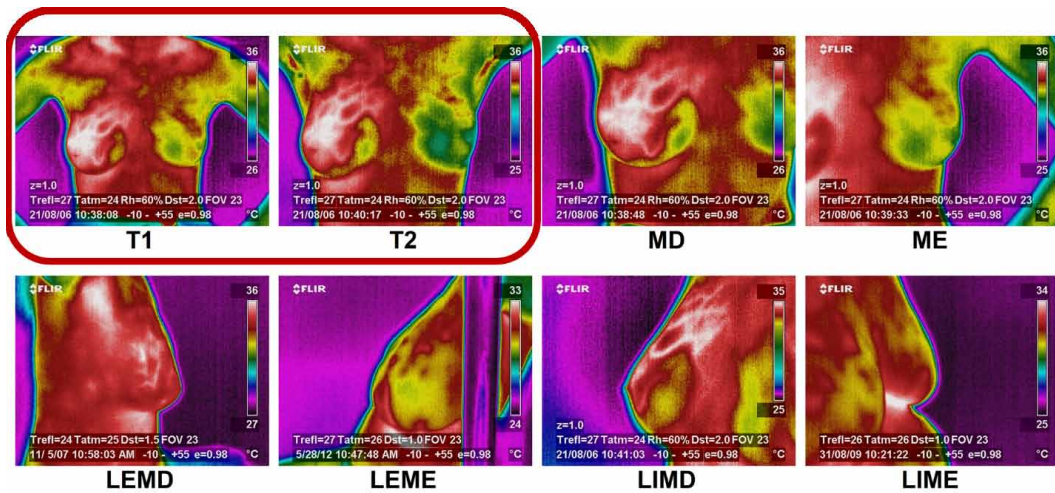
also organized according to diagnosis, which were previously established based on well established methods for each type of lesion, such as clinical examinations, biopsies, mammograms and breast ultrasounds (Dourado Neto, 2014).

In this study, the authors used images of cases of cystic lesions, benign lesions, malignant lesions and also images without identified lesions. Moreover, the experiments were performed using only the frontal images T1 and T2 because these conditions are considered to favor the identification of the region of interest (Santana et al., 2017; Sánchez-Ruiz, Olmos-Pineda & Olvera-López, 2020). The other six positions (MD, ME, LEMD, LEME, LIMD and LIME) were not used in this approach despite being acquired during the acquisition protocol. Thus, 336 images were used in the experiments, being 66 of no lesion, 73 of cysts, 121 of benign lesions and 76 of malignant lesions. The different number of images in each class may result in a biased result during training. The lesions could be more commonly classified into the class that has more instances. Thus, to avoid this problem, a class balancing was performed (Santana et al., 2017; Santana et al., 2018).

First of all, the images were submitted to a preprocessing step of conversion from RGB-JET to grayscale. So, the lighter shades of the grayscale indicated higher temperatures (Santana et al., 2017; Santana et al., 2018). In this approach, the authors used a combination of the moments of Haralick and Zernike as features extractors. Thus, the images are described by texture and shape features. Both characteristics tend to be essential for the differentiation of the types of lesions assessed here (cysts, benign lesions and malignant lesions).

The classification performance was first assessed in the dataset represented by all the features extracted by Haralick and Zernike algorithms, resulting in a total of 168 features. Then, we performed tests with different configurations of GA and PSO algorithms, to find the best subsets of selected features. All tests were empirically performed, evaluating several values for the parameters of both algorithms in order to assess changes in their behavior. Tables 1 and 2 show the parameters chosen for GA and PSO, respectively. For GA (Table 1), we tested the number of generations and the population size from 10 to 100, crossover probability from 0.1 to 0.9, mutation probability from 0.05 to 0.1 and selection operator of tournament and roulette methods. Considering PSO algorithm (Table 2), different numbers of iterations (20, 50, 100, 150 and 200) were tested. We evaluated the algorithm

Table 1. Parameters for the GA

Parameter	Values
Generation	10 to 100
Population size	10 to 100
Crossover probability	0.1 to 0.9
Mutation probability	0.05 to 0.1
Selection operator	Tournament and Roulette

Table 2. Parameters for the PSO

Parameter	Values
Individual weight	0.34
Inertia weight	0.33
Social weight	0.33
Iterations	20/50/100/150/200
Population size	10 to 100

using population sizes from 10 to 100. The weights (individual, inertia and social) were fixed for all configuration.

As previously mentioned, the parameters were empirically defined, but the minimum and the maximum values to be used were based on the literature. The mutation operator provides the algorithm with exploratory behavior, as it induces it to search for new points in the search space. So, if a genetic algorithm were developed only based on selection and crossing, the system would converge prematurely. The crossover operator generates new individuals to a very limited extent after a few generations. The mutation is, thus, essential for maintaining diversity and renewing genetic material. The mutation alters the structure of the chromosome by creating individuals with different properties than those found in most of the population. It prevents the model from remaining in an optimum location. Since it is an extremely aggressive change and, to some extent, unpredictable in results, the application rate of this operator is usually very low (Dias, 2005). Regarding the PSO algorithm, the weight values were not changed.

Finally, the generated features subsets were tested with several classifiers to assess their performance. These results were compared to the ones found when using the entire set of features. Such analysis was made through the accuracy and kappa statistic, since these metrics are essential for the assessment of classification quality.

In this study, the authors used the following classifiers: Bayes Net, Naive Bayes, Multilayer Perceptron (MLP), Support Vector Machine (SVM), J48, Random Tree, Random Forest and Extreme Learning Machine (ELM). The authors tested different models of SVM and ELM algorithms. Table 3 shows the configuration of each classifier. All experiments were conducted using the k-fold cross-validation method with 10 folds. The authors also performed 20 repetitions for each model to acquire statistical information. All experiments were performed using the Waikato Environment for Knowledge Analysis (WEKA), version 3.8, which is an open source data mining tool that aggregates a set of classification algorithms, regression algorithms, preprocessing and clustering techniques, all implemented in Java (Witten & Frank, 2005).

Table 3. Classifiers configuration

Classifier	Configuration
Bayes Net	Batch size: 100
Naïve Bayes	Batch size: 100
MLP	Number of hidden layers: 2 Number of hidden neurons: 100
SVM	Kernel function: polynomial (exponent 1-5), RBF
J48	Batch size: 100
Random Tree	Batch size: 100
Random Forest	Number of trees: 100
ELM	Number of hidden neurons: 400 Kernel function: sigmoid, polynomial (exponent 1-5)

RESULTS AND DISCUSSION

From the obtained results, the authors selected the three best ones from the group with all features and from the selections performed by GA and PSO. Table 4 shows the classification performances, regarding to average values of accuracy, kappa statistic and training time for the dataset with all features. The results for the subsets of selected features chosen by GA and PSO are shown on Tables 5 and 6, respectively. These results were achieved using the parameters described in Table 7.

In both feature selection methods, the outputs are modeled as binary vectors where each vector coordinate corresponds to an attribute. When this coordinate is equals to 1, the feature is considered relevant by the model, otherwise, it is excluded. Considering this, Figure 4 shows each feature occurrence using the best three GA (Figure 4a) and PSO (Figure 4b) configurations. From these graphs, we may confirm that more features were considered relevant by GA algorithm than PSO. However, PSO results showed to be more consistent among different configurations, so it seems more robust to slight changes in population size and number of iterations.

Table 4. The first three best classification results using all features

Classifier	Accuracy	Kappa	Time (ms)	Number of features
ELM – poly kernel (P2)	94.00 ± 2.81	0.9323 ± 0.0314	0.02 ± 0.01	168
ELM – poly kernel (P3)	93.99 ± 2.84	0.9321 ± 0.0324	0.05 ± 0.01	168
ELM – poly kernel (P4)	93.45 ± 2.88	0.9260 ± 0.0322	0.06 ± 0.02	168

Table 5. The first three best classification results using the features selected by GA

| Classifier	Accuracy	Kappa	Time (ms)	Number of features
ELM – poly kernel (P3)	87.96 ± 3.77	0.8640 ± 0.0420	0.05 ± 0.00	69
ELM – poly kernel (P3)	87.89 ± 3.35	0.8632 ± 0.0380	0.05 ± 0.00	74
ELM – poly kernel (P3)	87.74 ± 3.77	0.8613 ± 0.0437	0.05 ± 0.00	73

Table 6. The first three best classification results using the features selected by PSO

Classifier	Accuracy	Kappa	Time (ms)	Number of features
ELM – poly kernel (P3)	86.58 ± 4.23	0.8481 ± 0.0484	0.05 ± 0.01	49
ELM – poly kernel (P3)	85.16 ± 3.98	0.8325 ± 0.0447	0.05 ± 0.01	38
ELM – poly kernel (P3)	83.44 ± 4.25	0.8124 ± 0.0477	0.05 ± 0.00	39

Table 7. The first three best classification results using the features selected by PSO

Algorithm	Parameter	Value
GA	Generation	50
	Population size	30 to 60
	Crossover probability	0.3 to 0.5
	Mutation probability	0.05
PSO	Iterations	100 to 150
	Population size	10 to 20
	Inertia weight	0.33
	Individual weight	0.34
	Social weight	0.33

Figure 4. Features heatmap for (a) GA and (b) PSO methods
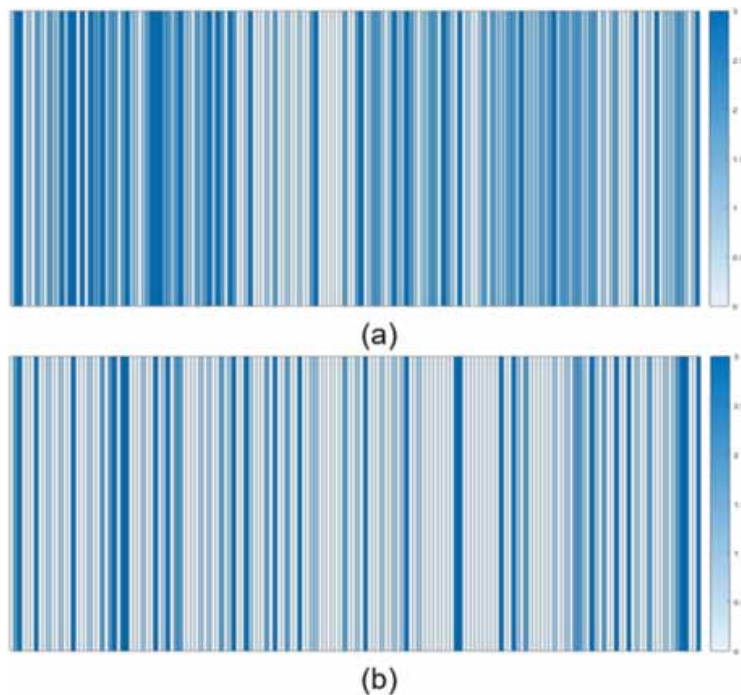
In the best scenario, the feature selection process reduced the amount of features from 168 to 69 (with GA) and 49 (with PSO). Table 8 shows lists of features for all three configurations: entire set of features, reduced set using GA and reduced set using PSO. Since this approach uses texture and shape features from Haralick and Zernike extractors, the authors randomly named the features but highlighted in orange the features related to texture and in blue the shape-related ones. It is important to notice that the meaning of an individual feature is irrelevant for human interpretation since each feature is abstract, but together they provide texture and shape information from the image.

Finally, Figure 5 illustrate the average results of accuracy and kappa statistic for the best model of each set of features. These results are the ones highlighted in Tables 4-6. For the feature selection algorithms, GA achieved the best performance when using 50 generations, population size of 60, crossover probability of 0.5 and mutation probability of 0.05. The best result for PSO was found with 100 iterations, a population size of 10, and inertia, individual and social weights of 0.33, 0.34 and 0.33, respectively.

The number of features decreased in almost 59% using GA and in around 70% with PSO. GA provided a slightly better classification performance, when compared to PSO, but only 1.4% better. Considering the amount of selected features, PSO was able to provide the smallest set with a low loss in classification performance. This scenario confirms what is observed in the literature (Tavares, Nedjah & Mourelle, 2015). Specifically concerning to the identification of breast masses, Silva Neto (2016) found that PSO had a positive performance in selecting features from mammographic images.

In the genetic algorithm, the tests with a mutation rate of 0.05 were more relevant than those with a mutation rate of 0.1. This result is in agreement with the literature, since Vellasco et al. (2014) discusses in their study that a very high mutation rate should be avoided. A high mutation rate can make the search essentially random, leading the convergence towards an optimal solution (Vellasco et al., 2014). Therefore, rates below 0.1 are more indicated. The value of the generations in this article also meets the literature. It shows that the very low generation value causes a drop in the performance. On the other hand, a high value requires longer processing time, but provides better coverage of the problem domain, avoiding convergence for local solutions. Therefore, a balance should be sought with regard to the size chosen for the generation. With regard to the population size, the studies show that very small population offers inferior search space coverage, causing a drop in performance. A sufficiently large population provides better coverage of the problem domain and prevents premature convergence to local solutions. However, with a larger population, a larger computational resources become necessary (Vellasco et al., 2014). The number of swarm particle iterations when altered, after the result is already found, was not relevant.

It is important to highlight that the relevance between the features is different. Some subsets that had a higher number of features were less accurate than subsets with fewer attributes. Regarding to the classification method, the best performing classifier was ELM with polynomial kernel of exponent 3. This configuration presented accuracies close to 88% with Kappa between 0.85 and 0.86. SVM classifier also showed a good performance, with accuracy around 87% and kappa statistic close to 0.83, being also successful in the studies from Carvalho Filho et al. (2014) and Silva Neto (2016).

CONCLUSION

This chapter proposed an approach based on GA and PSO to reduce the amount of features used to identify and classify breast lesions in thermographic images. The results showed that both approaches provided a significant reduction in the number of features, being greater for PSO. Although the number of attributes decreased, the accuracy did not drop considerably in comparison to the accuracy of the classification with all features, especially when using GA. A similar performance was observed between the feature selection methods regarding to training time. In addition, the authors observed a good performance of ELM classifier in the classification of these images.

Table 8. List of features for all set of features. Texture features are shown in orange, while shape features are in blue

Complete set of features (168 features)				Selected by GA (69 features)		Selected by PSO (49 features)	
X1	X43	X85	X127	X1	X95	X2	X151
X2	X44	X86	X128	X2	X96	X3	X156
X3	X45	X87	X129	X3	X97	X7	X160
X4	X46	X88	X130	X5	X98	X10	X162
X5	X47	X89	X131	X10	X103	X13	X164
X6	X48	X90	X132	X12	X104	X16	X165
X7	X49	X91	X133	X13	X106	X20	X168
X8	X50	X92	X134	X15	X110	X23	-
X9	X51	X93	X135	X16	X115	X25	-
X10	X52	X94	X136	X18	X116	X26	-
X11	X53	X95	X137	X20	X118	X28	-
X12	X54	X96	X138	X22	X123	X29	-
X13	X55	X97	X139	X23	X125	X36	-
X14	X56	X98	X140	X25	X129	X39	-
X15	X57	X99	X141	X28	X130	X42	-
X16	X58	X100	X142	X29	X132	X47	-
X17	X59	X101	X143	X33	X133	X49	-
X18	X60	X102	X144	X35	X135	X52	-
X19	X61	X103	X145	X36	X140	X55	-
X20	X62	X104	X146	X37	X142	X58	-
X21	X63	X105	X147	X38	X145	X65	-
X22	X64	X106	X148	X39	X147	X68	-
X23	X65	X107	X149	X41	X148	X71	-
X24	X66	X108	X150	X42	X153	X74	-
X25	X67	X109	X151	X43	X158	X85	-
X26	X68	X110	X152	X47	X161	X87	-
X27	X69	X111	X153	X48	X168	X91	-
X28	X70	X112	X154	X49	-	X94	-
X29	X71	X113	X155	X51	-	X96	-
X30	X72	X114	X156	X52	-	X97	-
X31	X73	X115	X157	X54	-	X100	-
X32	X74	X116	X158	X55	-	X109	-
X33	X75	X117	X159	X58	-	X110	-
X34	X76	X118	X160	X59	-	X120	-
X35	X77	X119	X161	X65	-	X123	-
X36	X78	X120	X162	X75	-	X124	-
X37	X79	X121	X163	X82	-	X138	-
X38	X80	X122	X164	X85	-	X139	-
X39	X81	X123	X165	X86	-	X140	-
X40	X82	X124	X166	X88	-	X142	-
X41	X83	X125	X167	X91	-	X143	-
X42	X84	X126	X168	X94	-	X148	-

Figure 5. Average (a) accuracy and (b) kappa statistics for the best result of each set of features
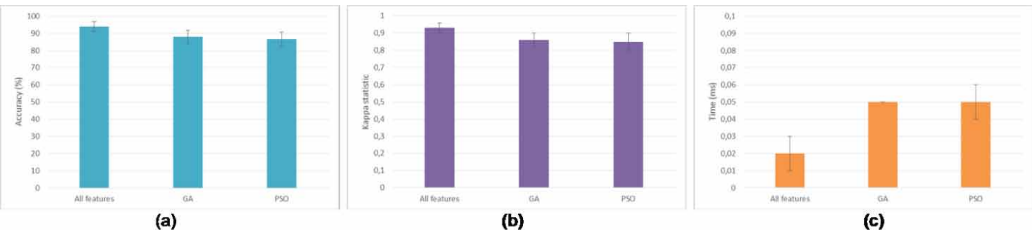
Overall, the findings of this study were compatible with those proposed in the literature. They were also interesting as an initial approach using conventional methods with low computational cost. However, more experiments are necessary before any system development or application in humans. In the future, other approaches for data representation and feature selection may be used, such as CNNs and Autoencoders. Improvements in the amount of data for each studied class (cystic lesion, benign lesion, malignant lesion, and healthy) are also being considered in order to provide better results.

ACKNOWLEDGMENT

REFERENCES

Andrade, F., Paiva, A., & Correa, A. (2017). Análise de Imagens de Termografia Dinâmica para Classificação de Alterações na Mama Usando Séries Temporais. In *Proceedings of Conference on Graphics, Patterns and Images*. Niteroi, Brazil: Elsevier.

Andrade, M. K. S., Santana, M. A., & Santos, W. P. (2018). *Avaliação do Desempenho de Classificadores Inteligentes na Detecção da Doença de Alzheimer em Imagens de Ressonância Magnética Utilizando Extratores de Forma e Textura*. Paper presented at the II Simpósio de Inovação em Engenharia Biomédica (SABIO 2018), Recife, Brazil.

Araújo, M. C. (2009). *Utilização de Câmera por Infravermelho para Avaliação de Diferentes Patologias em Clima Tropical e Uso Conjunto de Sistemas de Banco de Dados para Detecção do Câncer de Mama* (Master's thesis, Federal University of Pernambuco, Recife, Brazil). Available from https://repositorio.ufpe.br/handle/123456789/5062

Azevedo, W. W., Lima, S. M. L., Fernandes, I. M. M., Rocha, A. D. D., Cordeiro, F. R., Silva-Filho, A. G., & Santos, W. P. (2015). *Morphological extreme learning machines applied to detect and classify masses in mammograms*. In *2015 International Joint Conference of Neural Networks (IJCNN)*, Killarney, Ireland. doi:10.1109/IJCNN.2015.7280774

Bandyopdhyay, S. K. (2010). Survey on segmentation methods for locating masses in a mammogram image. *International Journal of Computers and Applications*, *9*(11), 25–28. doi:10.5120/1429-1926

Blum, A., & Langley, P. (1997). Selection of relevant features and examples in machine learning. *Artificial Intelligence*, *97*(1-2), 245–271. doi:10.1016/S0004-3702(97)00063-5

Borchartt, T. B., Conci, A., Lima, R. C. F., Resmini, R., & Sanchez, A. (2013). Breast thermography from an image processing viewpoint: A survey. *Signal Processing*, *93*(10), 2785–2803. doi:10.1016/j.sigpro.2012.08.012

Brasileiro Filho, G. (2011). *Bogliolo Patologia* (8th ed.). Guanabara Koogan.

Carvalho Filho, A. O., de Sampaio, W. B., Silva, A. C., de Paiva, A. C., Nunes, R. A., & Gattass, M. (2014). Automatic detection of solitary lung nodules using quality threshold clustering, genetic algorithm and diversity index. *Artificial Intelligence in Medicine*, *60*(3), 165–177. doi:10.1016/j.artmed.2013.11.002 PMID:24332156

Charniak, E., & Mcdermott, D. (1985). *A Bayesian Model of Plan Recognition*. Addison-Wesley.

Cockburn, W. (2018). *The Truth About Breast Thermography*. Retrieved from https://www.healingwell.com/articles/post/the-truth-about-breast-thermography

Coelho, A. S., Santos, M. A. S., Caetano, R. I., Piovesan, C. F., Fiuza, L. A., Machado, R. L. D., & Furini, A. A. C. (2018). Predisposição hereditária ao câncer de mama e sua relação com os genes BRCA1 e BRCA2: Revisão da literatura. *Revista Brasileira de Análises Clínicas*, *50*(1), 17–21.

Dias, D. M. (2005). *Refinery Scheduling Optimization Using Genetic Algorithms and Cooperative Coevolution* (Master's thesis, Pontifícia Universidade Católica do Rio de Janeiro, Rio de Janeiro, Brazil). Available from https://www.maxwell.vrac.puc-rio.br/colecao.php?strSecao=resultado&nrSeq=5969@2

Dourado Neto, H. M. (2014). *Segmentação e análise automática de termogramas: um método auxiliar na detecção do câncer de mama* (Master's thesis, Federal University of Pernambuco, Recife, Brazil). Available from Repositório UFPE.

Felipe, J. C., Olioti, J. B., & Traina, A. J. M. (2005). *Discriminação de Aspectos Malignos em Massas Tumorais de Mamografias Usando Características de Forma das Imagens*. Paper presented at the V Workshop de Informática Médica, Porto Alegre, RS, Brazil.

Fernández-Ovies, F. J., Alférez-Baquero, E. S., Andrés-Galiana, E. J., Cernea, A., Fernández-Muñiz, Z., & Fernández-Martinez, J. L. (2019). Detection of Breast Cancer Using Infrared Thermography and Deep Neural Networks. *Nature*, *11466*, 514–523.

Ferreira, J. O., Oliveira, H. C. B., & Martinez, M. R. (2011). Aplicação de uma metodologia computacional inteligente no diagnóstico de lesões cancerígenas. *Revista Brasileira de Inovação Tecnológica em Saúde*, *2*(1), 4–9.

Gautherie, M., Haehnel, P., Walter, J. P., & Keith, L. G. (1987). Thermovascular changes associated with in situ and minimal breast cancers. Results of an ongoing prospective study after four years. *The Journal of Reproductive Medicine*, *32*(11), 833–842. PMID:3430492

Guyton, A. C., & Hall, J. E. (2011). *Tratado de fisiologia médica* (12th ed.). Elsevier Brasil.

Haugeland, J. (1985). *Artificial Intelligence: The Very Idea*. The MIT Press.

Instituto Nacional de Câncer. (2017). *Estimativa 2018: incidência de câncer no Brasil*. INCA.

Jordan, M. I., & Mitchell, T. M. (2015). Machine learning: Trends, perspectives, and prospects. *Science*, *349*(6245), 255–260. doi:10.1126/science.aaa8415 PMID:26185243

Kalfels, E. H. H. (2017). *Algoritmos genéticos como uma abordagem para a alocação de grade de horários da UFSC* (Bachelor's thesis, Universidade Federal de Santa Catarina, Florianópolis, Brazil). Available from https://repositorio.ufsc.br/xmlui/handle/123456789/177693

Kennedy, J., & Eberhart, R. C. (1995). Particle swarm optimization. In *Proceedings of IEEE International Conference On Neural Networks*. Perth, Austrália: IEEE. doi:10.1109/ICNN.1995.488968

Keyserlingk, J., Ahlgren, P., Yu, E., & Belliveau, N. (1998). Infrared Imaging of the Breast: Initial Reappraisal Using High-Resolution Digital Technology in 100 Successive Cases of Stage I and II Breast Cancer. *The Breast Journal*, *4*(4), 245–251. doi:10.1046/j.1524-4741.1998.440245.x PMID:21223443

Kumar, V., Abbas, A. K., Fausto, N., & Aster, J. C. (2010). *Robbins & Cotran Patologia Bases Patológicas das Doenças* (8th ed.). Elsevier.

Kurzweil, R. (1990). *The Age of Spiritual Machines*. The MIT Press.

Lee, H. D. (2005). *Seleção de atributos importantes para a extração de conhecimentos de base de dados* (Doctoral thesis, University of São Paulo, São Carlos, Brazil). Retrieved from www.teses.usp.br10.11606/T.55.2005.tde-22022006-172219

Leles, A. C. Q. (2015). *Desenvolvimento de procedimento e análise de imagens térmicas para a identificação do câncer de mama* (Master's thesis, Universidade Federal de Uberlândia, Uberlândia, Brazil). Retrieved from https://repositorio.ufu.br/handle/123456789/15022

Lessa, V., & Marengoni, M. (2016). Applying artificial neural network for the classification of breast cancer using infrared thermographic images. In *Proceedings of the International Conference on Computer Vision and Graphics*. Warsaw, Poland: Springer. doi:10.1007/978-3-319-46418-3_38

Menezes, N. G. (2007). *Síntese Topológica Evolucionária e Otimização Numérica de Parâmetros em Estrutura CMOS* (Master's thesis, Universidade Federal do Rio de Janeiro, Rio de Janeiro, Brazil). Retrieved from http://www.pee.ufrj.br/index.php/en/producoes-academica/dissertacoes-de-mestrado/2007-1/2007033002-2007033002/file

Menke, C. H. (2000). *Rotinas em Mastologia* (2nd ed.). ArtMed Editora.

Mishra, V., & Rath, S. K. (2020). Detection of breast câncer tumours based on feature reduction and classification of thermograms. *Quantitative Infrared Thermography Journal*, *1*, 1–14. doi:10.1080/17686733.2020.1768497

Moore, K. L., Dalley, A. F., & Agur, A. M. R. (2001). *Anatomia orientada para a clínica* (6th ed.). Guanabara Koogan.

Nascimento, F. A. F., Dias, A. N., Figueiredo Filho, A., Arce, J. E., & Miranda, G. M. (2012). Uso da Meta-Heurística otimização por enxame de partículas no planejamento Florestal. *Scientia Forestalis*, *40*(96), 557–565.

Oliveira, L. F., Narloch, A. L. M., Kist, D. M., Soares Filho, M. P., Meneghello, G. E., Cavalheiro, G. G. H., & Tillmann, M. A. A. (2012). *Extração de Características de Forma utilizando matriz de co-ocorrência e Atributos de Haralick*. Paper presented at the Workshop de Visão Computacional (WVC), Pelotas, RS, Brazil.

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

Pereira, J. M. S., Santana, M. A., Lima, N. M., Sousa, F. N., de Lima, R. C. F., & dos Santos, W. P. (2017) *Método para Classificação do Tipo da Lesão na Mama Presentes nas Imagens Termográficas utilizando Classificador ELM*. Paper presented at the I Simpósio de Inovação em Engenharia Biomédica (SABIO 2017), Recife, PE, Brazil.

Poole, D., Mackworth, A. K., & Goebel, R. (1998). *Computational Intelligence: A Logical Approach*. Oxford University.

Porter, P. (2008). "Westernizing" women's risks? Breast cancer in lower-income countries. *The New England Journal of Medicine*, *358*(3), 213–216. doi:10.1056/NEJMp0708307 PMID:18199859

Ring, E. F., & Ammer, K. (2012). Infrared thermal imaging in medicine. *Physiological Measurement*, *33*(3), 33–46. doi:10.1088/0967-3334/33/3/R33 PMID:22370242

Saladin, K. S., & Porth, C. (2010). *Anatomy & physiology: the unity of form and function* (Vol. 5). McGraw-Hill.

Sánchez-Ruiz, D., Olmos-Pineda, I., & Olvera-López, A. (2020). Automatic region of interest segmentation for breast thermogram image classification. *Pattern Recognition Letters*, *135*(1), 72–81. doi:10.1016/j.patrec.2020.03.025

Santana, M. A., Pereira, J. M. S., Lima, N. M., Sousa, F. N., de Lima, R. C. F., & dos Santos, W. P. (2017). *Classificação de lesões em imagens frontais de termografia de mama a partir de sistema inteligente de suporte ao diagnóstico.* Paper presented at the I Simpósio de Inovação em Engenharia Biomédica (SABIO 2017), Recife, PE, Brazil.

Santana, M. A., Pereira, J. M. S., Silva, F. L., Lima, N. M., Sousa, F. N., Arruda, G. M. S., Lima, R. C. F., Silva, W. W. A., & Santos, W. P. (2018). Breast cancer diagnosis based on mammary thermography and extreme learning machines. *Research on Biomedical Engineering*, *34*(1), 45–53. doi:10.1590/2446-4740.05217

Sessler, D. I. (1994). Consequences and treatment of perioperative hypothermia. *Anesthesiology Clinics of North America*, *12*, 425–456.

Siciliano, A. V. (2007). Algoritmos Genéticos e Particle Swarm Optimization e suas aplicações problemas de Guerra Eletrônica. In *Proceedings of IX Simpósio de Guerra Eletrônica*. Rio de Janeiro, Brazil: ITA.

Silva Neto, O. P. (2016). *Detecção automática de massas em imagens mamográficas usando particle swarm optimization (PSO) e índice de diversidade funcional* (Master's thesis, Universidade Federal do Maranhão, Maranhão, Brazil). Retrieved from https://tedebc.ufma.br/jspui/handle/tede/tede/298

Souza, A. H. (2017). *Seleção de atributos relevantes: aplicando técnicas na base de dados do Herbário Virtual da Flora e dos Fungos* (Master's thesis, Universidade Federal do Amazonas, Itacoatiara, Brazil). Retrieved from https://tede.ufam.edu.br/handle/tede/6232

Souza, T. K., Andrade, J. F. S., Almeida, M. B. J., Santana, M. A., & Santos, W. P. (2019). *Métodos Computacionais Aplicados ao Diagnóstico de Câncer de Mama por Termografia: uma revisão de literatura*. Paper presented at the III Simpósio de Inovação em Engenharia Biomédica (SABIO 2019), Recife, Brazil.

Tavares, Y., Nedjah, N., & Mourelle, L. M. (2015). *Utilização de otimização por enxame de partículas e algoritmos genéticos em rastreamento de padrões*. Paper presented at the 12º Congresso Brasileiro de Inteligência Computacional (CBIC 2015), Curitiba, PR, Brazil. doi:10.21528/CBIC2015-049

Vellasco, P. C. G. S., de Lima, L. R. O., de Andrade, S. A. L., Vellasco, M. M. B. R., & da Silva, L. A. P. S. (2014). *Modelagem de Estrutura de Aço e Mistas*. Elsevier., doi:10.1016/B978-85-352-7197-3.00010-8

Victora, C. G., Barreto, M. L., Leal, M. C., Monteiro, C. A., Schmidt, M. I., Paim, J., Bastos, F. I., Almeida, C., Bahia, L., Travassos, C., Reichenheim, M., & Barros, F. C. (2011). Health conditions and health-policy innovations in Brazil: The way forward. *Lancet*, *377*(9782), 2042–2053. doi:10.1016/S0140-6736(11)60055-X PMID:21561659

Weinberg, R. A. (2008). *A biologia do câncer*. Artmed.

Witten, I. H., & Frank, E. (2005). *Data Mining: Practical machine learning tools and techniques* (2nd ed.). Morgan Kaufmann.

Amanda Lays Rodrigues da Silva graduated in Nursery (2014) and received a Masters' degree in Biomedical Engineering (2019) from the Federal University of Pernambuco, Brazil. Her main research interests are applied artificial intelligence, pattern recognition for image diagnosis, and breast cancer early diagnosis.

Maíra Araújo de Santana is a doctoral student in Computer Engineering at the Polytechnic School of the University of Pernambuco, Brazil. She holds a bachelor's and master's degree in Biomedical Engineering from Federal University of Pernambuco, Brazil. Develops research in Biomedical Computing, with an emphasis on machine learning and pattern recognition for early diagnosis of breast cancer and in Applied Neurosciences. Currently, she works in the area of Affective Computing and in the development of intelligent systems based on image and signal processing and analysis using artificial neural networks.

Clarisse Lins de Lima graduated with a Master's in Biomedical Engineering from the Federal University of Pernambuco (UFPE), she is currently a PhD student in Computer Engineering from the University of Pernambuco (UPE). She is interested in the areas of digital epidemiology, epidemic prediction and artificial intelligence applied to health.

José Filipe Silva de Andrade is an undergraduate student in Biomedical Engineering and member of the Research Group on Biomedical Computing.

Thifany Ketuli Silva de Souza is an undergraduate student in Biomedical Engineering and member of the Research Group on Biomedical Computing.

Maria Beatriz de Almeida is an undergraduate student in Biomedical Engineering and member of the Research Group on Biomedical Computing.

Washington Azevedo holds a PhD in Computer Science from the Federal University of Pernambuco - UFPE (2017). He holds a Master's degree in Computer Science from the Federal University of Pernambuco - UFPE (2011). He holds a degree in Systems Analysis from the Salgado de Oliveira University - UNIVERSO (2004). He has a postdoctoral degree in the Department of Biomedical Engineering of the Federal University of Pernambuco - UFPE (10/2017 until 10/2019) having as supervisor Professor. Dr. Wellington Pinheiro dos Santos. He was a Test Engineer of the CIn / Motorola project (from May 10, 2006 to October 31, 2007). He has experience in Computer Science, acting on the following subjects: Software Testing Engineering, Artificial Intelligence, Artificial Neural Networks, Hybrid Intelligent Systems, Handwriting Character Recognition, Pattern Recognition and Biomedical Engineering.

Rita de Cássia Fernandes de Lima has a Bachelor's degree in Physics from Universidade Federal de Pernambuco (1973), a Master's at Nuclear Engineering from Universidade Federal de Pernambuco (1984), and Doctorate at Nuclear Engineering from Universidade de São Paulo (1997). She has experience in Biomedical Engineering, focusing on Modeling of Biological Phenomena, acting on the following subjects: heat transfer, numerical simulation, biomechanics, heat transfer in medical procedures, termography, finite volume method, early breast cancer detection.

Wellington Pinheiro dos Santos received a bachelor's degree in Electrical Electronics Engineering (2001) and MS in Electrical Engineering (2003) from the Federal University of Pernambuco, and Ph.D. in Electrical Engineering from the Federal University of Campina Grande (2009). He is currently a Professor (exclusive devotion) of the Department of Biomedical Engineering at the Center of Technology and Geosciences - Engineering School of Pernambuco, Federal University of Pernambuco, acting in Undergraduate Biomedical Engineering and Graduate Program in Biomedical Engineering, which was one of the founders (2011). He founded the Center for Social Technologies and Bioengineering, at the Federal University of Pernambuco, NETBio-UFPE (2012). He is member of the Graduate Program in Computer Engineering from the Polytechnic School of Pernambuco, University of Pernambuco, since 2009. He also has experience in the area of Computer Science, with an emphasis on Graphic Processing (Graphics), acting on the following themes: digital image processing, pattern recognition, computer vision, evolutionary computation, numerical methods of optimization, computational intelligence, computer graphics, virtual reality, game design and applications of Computing and Engineering in Medicine and Biology. He is a member of the Brazilian Society of Biomedical Engineering (SBEB), the Brazilian Society of Computational Intelligence, and the International Federation of Medical and Biological Engineering (IFMBE).