

# PHOC Descriptor Applied for Mammography Classification

## Descritor PHOC Aplicado na Classificação de Mamografias

Gilmário Barbosa dos Santos<sup>1</sup>

**Abstract:** This paper describes experiments with PHOC (Pyramid Histogram of Color) features descriptor in terms of capacity for representing features presented in breast radiograph (also known as mammography). Patches were taken from regions in digital mammographies, representing benign, cancerous, normal tissues and image's background. The motivation is to evaluate the proposal in perspective of using it for execution in an inexpensive ordinary desktop computer in places located far from medical experts. The images were obtained from DDSM database and processed producing the feature-dataset used for training an Artificial Neural Network, the results were evaluated by analysis of the learning rate curve and ROC curves, besides these graphical analytical tools the confusion matrix and other quantitative metrics (TPR, FPR and Accuracy) were also extracted and analyzed. The average accuracy  $\approx 0.8$  and the other metrics extracted from results demonstrate that the proposal presents potential for further developments. At the best effort, PHOC was not found in literature for applications in mammographies such as it is proposed here.

**Keywords:** Mammography — Image Processing — Computer Aided Diagnosis — Image Classification — Spatial Pyramid of Features

**Resumo:** Este artigo descreve experimentos com descritor PHOC em termos da sua capacidade em representar características apresentadas na radiografia da mama (também conhecida como mamografia). Foram retirados recortes de regiões presentes em mamografias digitais representando tecidos benignos, cancerosos, normais e fundo da imagem. A motivação é avaliar a proposta na perspectiva de sua utilização para execução em computadores comuns de custo acessível em localidades distantes de especialistas médicos. As imagens foram obtidas do banco de dados do DDSM e processadas produzindo uma base de dados usado para treinar uma Rede Neural Artificial, os resultados foram avaliados com base na curva de aprendizado e curvas ROC, além destas ferramentas analíticas gráficas a matriz de confusão e outras métricas quantitativas (TPR, FPR e Acurácia) também foram extraídas e analisadas. A acurácia média  $\approx 0.8$  bem como as outras métricas extraídas dos resultados demonstraram que a proposta tem potencial para desenvolvimentos futuros. No melhor dos esforços, o PHOC não foi encontrado na literatura para aplicações em mamografias, como é proposto neste trabalho.

**Palavras-Chave:** Mamografia — Processamento de Imagens — Diagnóstico Auxiliado por Computador — Classificação de Imagem — Pirâmide Espacial de Características

<sup>1</sup> Computer Science Department, Santa Catarina State University, Brazil  
gilmario.santos@udesc.br

**DOI:** <http://dx.doi.org/10.22456/2175-2745.89115> • **Received:** 20/12/2018 • **Accepted:** 11/11/2019

CC BY-NC-ND 4.0 - This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License.

## 1. Introduction

Breast cancer is one of the most lethal types of carcinoma presenting a great number of occurrences around the world, indeed, statistics of the year 2018 pointed that approximately 15% of all cancer deaths among women are related to breast cancer (WHO - World Health Organization [1]).

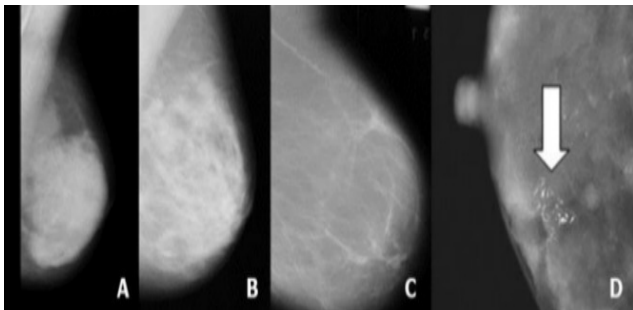
Despite the advances in research on strategies for breast cancer prevention, in practical terms one of the best approaches for preventing this disease have been the adoption of screening tests aiming for detection of breast cancer as soon as possi-

ble [2]. This early detection strategy maximizes the chances of starting a treatment in the very beginning of the disease, increasing the probability of survival.

The early breast cancer detection is the motivation for public health agencies in promoting campaigns for screening the female population. For example, the recommendation of INCA (Brazilian National Cancer Institute José Alencar Gomes da Silva) is the screening by mammography for women between 50 and 69 years old [3]. By its turn the American Cancer Society strongly recommends that the women between the ages of 45 and 54 should get mammograms every

year [4]. These campaigns are mainly based on inspection of breast radiographs, also known as mammographies, which need to be examined by an expert for diagnosis/prognosis. This examination is basically a visual inspection that can be influenced by human factors such as the visual fatigue.

Since the mammography acquisition process is based on RX radiation penetrating through the tissues, the mammary gland anatomy strongly influences the resultant mammogram. For example, the more dense is the breast, the more brighter is the resultant mammography (Fig. 1). Important finds can be camouflaged by dense breast tissues, e.g., the microcalcifications. These are thin white spots on the mammography (Fig. 1-D), they are signs scrutinized by specialists during the visual inspection [5] [6] [7] [8]. As a dense breast tissue corresponds to white regions, it practically hides the microcalcifications causing difficulties for detecting them [9] [10]. Therefore, the breast anatomical features associated to the image acquisition process based on RX penetration usually produces low-contrast and noisy mammographies [11].



**Figure 1.** Mammographies presenting density variation: cases A and B are denser than case C (fatty tissue). In the right side there is a case presenting microcalcifications (D). Composed by images from DDSM [12]

CAD (Compute Aided Diagnosis) is a general designation of software developed for improving the medical performance specially in diagnostic process. Particularly in mammography analysis, this sort of software can help the clinicians in overcoming physical limitations (visual acuity, visual fatigue, etc) that disturb the diagnosis/prognosis. In expectation of contributing to the studies on the development of CAD systems, specially in clinical classification of cases, this paper reports experiment on mammography image classification for categorizing tissues in labels BE (benign tissue), NO (normal tissue) or CA (cancerous tissue). For this sake, datasets of patches (image samples) were extracted from DDSM mammography database [12] and experiments were conducted by using a neural network model and PHOC feature descriptor.

CNN (Convolutional Neural Net) has been considered the state-of-art in general image classification. In the field of medical image applications it also presents high scores in classification tasks, such as for automated diabetic retinopathy detection [13] and for mammography image classification [14]. Nevertheless, this outstanding performance comes at

the cost of high computational complexity [15] [16]. The CNN demands intense parallel architectures such as available in FPGAs or GPUs, which compromises the application of CNN embedded in devices like ordinary/affordable desktop computers without any special configuration.

Due to different reasons, including affordability, these ordinary desktop computers are frequently found in remote areas and isolated communities, specially in thirty world countries. Places with poor medical resources. Such situations are relatively usual in Amazon forest, which presents lots of fishermen's villages and indigenous peoples living in places far and/or isolated from the nearest medical expert (e.g., an oncologist), but counting with the kind of desktop computer described just above.

Although being inexpensive desktops, these computers can present enough computational power for running a CAD-like system. Which justifies studies focused on identifying a Deep Learning (DL) model and feature descriptor suitable for this kind of computer system.

Lets suppose a person living at a remote place seeking for medical attention. Running in an desktop computer and operated by a trained person, such as a nurse or paramedic personnel, a CAD-like system (embedding a Deep Learning model using an image descriptor method) would provide the level of information necessary for patient triage.

It is important to emphasize: the proposal does not aim to provide the final assessment about the medical case, only an initial information to guide the health care personnel in charge, orienting him/her for taking measures about a medical situation happening in a remote place and/or far from any specialist.

Since CNN is too complex for running in ordinary computers, the alternative is to use a classical, relatively simpler and flexible Deep Learning method. It is also important to use the same "low profile" principle to choose a descriptor method. The Multi-Layer Perceptron named ANN (Artificial Neural Network) and the descriptor PHOC (Pyramid Histogram of Color) were the choices for the experiments described in this paper.

The proposal's scope does not include to classify the different variations of human breast cancer or varieties of breast masses since the database used for training and testing the model does not includes so detailed information (DDSM described in [12]).

The main motivation for this paper is testing a classical/simple neural net model (ANN) in classification of mammography using a simple descriptor (PHOC) in perspective of contributing for future works on CAD system development, aiming low cost platforms like ordinary computer available in places such as a municipality office or at a local bureau representing the government.

The descriptor PHOC (Pyramid Histogram of Color) was based on spatial pyramid firstly described by Lazebnik et al. [17], extended by Bosch et al. [18] [19] [20] and Hiulin Gao [21]. At the best effort, PHOC method was not found in

literature for applications in mammographies, such as it is proposed here.

PHOC is based on an very simple and classical gray-tone histogram, so that it is sensitive to the distribution of bright intensities of pixels, one of the aspects tested here is how accurately a so simple descriptor like PHOC can describe the different tissues presented in gray-tone mammographic images. In fact, the “knowledge” learned by the ANN, and consequently the proper accuracy of the classification accomplished by ANN, depends on how accurately the descriptors based on PHOC method represent the features in corresponding images patches.

The results were evaluated by analysis of the learning rate curve and ROC curves, besides these graphical analytical tools the confusion matrix and other quantitative metrics (TPR, FPR and Accuracy) were also extracted and analyzed.

This paper is organized as follows, in Section 2 a list of related works is described. The theoretical framework about the methods and models applied in this work is discussed in Section 3 and the Section 4 is dedicated to describe the material and methods applied in the experiments. The results are presented and analyses are conducted in Section 5. The conclusions and future works are described in Section 6.

## 2. Related work

The descriptor PHOC is deeply-rooted in the concept of spatial image decomposition by spatial pyramid introduced by Lazebnik et al. [17], Bosch et al. [18] [19] [20] and Hiulin Gao [21].

The processing of images in levels of details is not a new idea, in fact the Wavelet Transform (WT) already presents a similar characteristic [22], [23] and even some examples of applications can be seen in the field of mammography analysis, such as it is found in Kocur et al. [24] or Kalman et al. [25].

Nevertheless, the principles that govern WT are very different from the foundations of the spatial pyramid as proposed by Lazebnik et al. [17], Bosch et al. [18] [19] [20] and Hiulin Gao [21]. Since WT is conceived in the frequency domain its image decomposition is determined by the response of frequency-filters recursively applied. In another way, the spatial pyramid is founded on spatial domain, the decomposition is performed by taking in consideration the global and local distribution of information (features) of the image, no frequency filters are used for this sake.

Artificial Neural Networks has been successfully applied in mammography classification since years 90's [24] and presenting itself in recent works by authors such as Mehdy et al. [26] and Dheeba et al. [25]. Gardezi et al. [27] developed a recent systematic review on techniques for detection and diagnosis based on mammographies, including ANN model and CNN, emphasizing the high computational cost of the latter.

Carneiro et al. [28] describe an approach conceptually similar to Bag of Features (a step in PHOW determination

[19]) classifying mammographic images based on breast density using an Artificial Neural Network (ANN). They extracted features and applied K-means for determine the best feature for each class of mammography, then trained a ANN for classification.

Although the classical gray tone histogram seems to be too simple, it presents a good potential for mammography analysis. For example, Zhou et al. [29] worked on rules for breast density classification based on gray levels histograms and Gao et al. [21] processed images from Caltech-101 (database containing images of 101 categories not medical) concatenating PHOC with other descriptors.

## 3. Spatial Pyramid of Features

Lezibnik et al. [17], Bosch et al. [18] [19] and Hiulin Gao [21] developed their work using the same basic concept of partitioning the image into increasingly fine sub-regions and computing histograms of local features found inside each sub-region. Since this process can be interpreted as an hierarchical structure composed by levels ( $L$ ) each one presenting  $2^L$  equally sized image's regions, Lezebnik et al. [17] called it as a “spatial pyramid”. The final descriptor is a weighted concatenation of the histograms obtained for the different levels of division of the initial image.

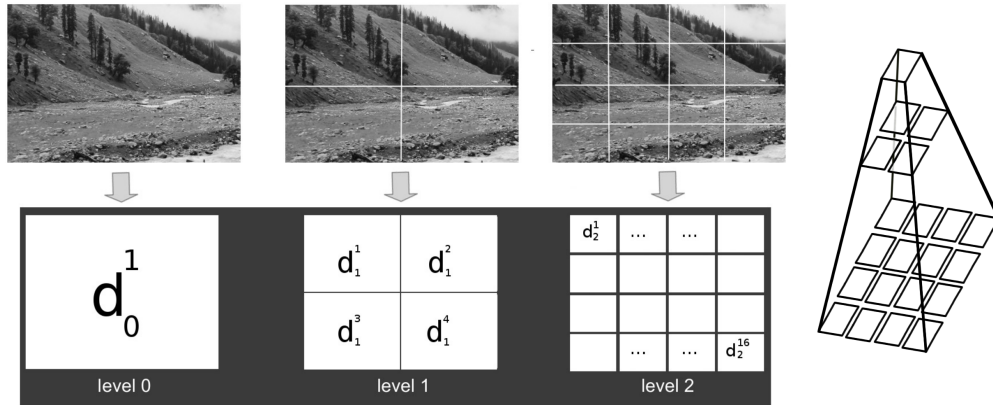
Instead of using a weighted combination of regional descriptors, as it were conducted by Lezibnik et al. [17], Bosch et al. [18] [19] and Hiulin Gao [21], here the overall descriptor is a simple unweighted concatenation of regional descriptors.

The Fig. 2 shows this concept, a descriptor that represents local image shape and its spatial layout. The spatial pyramid, in terms of the partition levels, is found on the right side of the same figure. The main idea was to capture global and local aspects of the image and represent them as part of an overall descriptor  $D_{image}$  composed by concatenation of (non-weighted) regional descriptors obtained by image partitioning. As described in Eq. 1, since  $\sqcup$  represents a concatenation of sets,  $D_L$  corresponds to concatenation of the  $2^{2L}$  regional descriptors  $d$  at the level  $L$ . This concept was applied for implementation of the five descriptors discussed, as follows.

$$\begin{aligned} D_{image} &= (D_0 \sqcup D_1 \dots \sqcup D_L) \\ D_L &= \bigsqcup_{k=1}^{2^{2L}} d_L^k, L \in \mathbb{Z}^+ \cup 0 \end{aligned} \quad (1)$$

The same basic concept (spatial pyramid of histograms) was applied for PHOC feature descriptor. In this case, the overall descriptor ( $D_{image}$ ) is composed by the classical histograms based on pixels' brightness in regions of the image.

One important difference about the PHOC strategy as used by Gao et al. [21] and the present report is that they do not processed medical images but applied PHOC on HSV images from Caltech-101 database, a renowned general purpose image database containing pictures of objects from 101



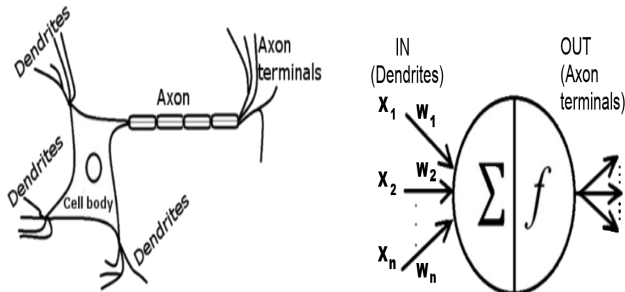
**Figure 2.**  $D = \text{concat}(d_0, d_1, d_2)$  as a concatenation of the histograms at each level of partitioning the original image. In terms of the partition levels, the correspondent spatial pyramid is on the right side of the figure. Adapted from Lazebnik et al. [17]

categories. On the other hand, here the experiments are performed on patches of mammographies, which are gray-tone low contrast images from DDSM, a very specific database in comparison to Caltech-101. At the time the current paper was written, it was not found any other PHOC application such as conceived here for mammographies.

### 3.1 Artificial Neural Network (ANNs)

The ANN is a bio-inspired Deep Learning method composed by artificial neurons, which were conceived in analogy to its biological counterparts, mainly in terms of the information flow. The Fig. 3 depicts a representation of real and artificial neuron.

The natural neuronal tissue is composed by a dense network of neurons, each one capable to receive input-stimuli coming from various different sources via dendrites and to send an output-stimulus (sending via axons terminals). The output-stimulus depends on the state assumed by the neuron's cell body under the influence of input-stimuli. As seen in the Fig. 3, the artificial neuron is an analogy to the natural one presenting an output according to an activation function ( $f$ ) dependent on a set of inputs ( $w_i, x_i$ ).



**Figure 3.** A real neuron (left) and its artificial analogy (right).

The activation function is represented by  $y$  in Eq. 2, let  $X$  is a vector of input-stimuli,  $W$  a vector of weights representing the respective strength of each element in  $X$  and  $b$  is a vector of bias values for tuning. There are different types of activation

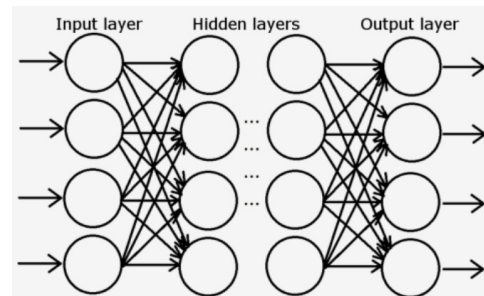
functions (e.g. linear, sigmoid, etc) presenting different curve profiles for neuronal responses to incoming stimuli.

$$y = f(X.W + b) \quad (2)$$

since:  $X = \{x_1, x_2, \dots, x_n\}$  and,

$$W = \{w_1, w_2, \dots, w_n\}$$

The ANN is composed by layers of artificial neurons, as seen in Fig. 4. The activation function represents the neuronal instance of the overall learning instantiated in the whole network.



**Figure 4.** ANN as a network composed by layers of artificial neurons.

According to Geron [30] there are four main categories for training a model: supervised learning, unsupervised learning, semi-supervised learning, and reinforcement learning. The first three strategies are founded respectively on labeled training data, unlabeled training data or partially labeled training data.

The method adopted here is based on supervised learning, which means that the dataset contains the labels for correct classification of each example and the ANN learn by associating example  $\leftrightarrow$  label throughout the dataset. The latter strategy (reinforcement learning) is applied on the field of intelligent agents, which learn by experimenting situations in a certain context/environment by penalties/rewards techniques during training [30].



The dataset is split into non-overlapping parts: training-set, validation-set and test-set. The first two sets are used throughout the training of the ANN, during which the training algorithm needs to configure the weights in ANN's internal connections in order to raise an optimal condition between fitting and generalization [31]. The training algorithm looks for fitting the model (high accuracy when the training data is entered), maintaining the capacity of generalization (high accuracy when test data is entered).

The training-set and validation-set are inputs to the training algorithm in cycles known as epochs [31]. After each training iteration (epoch), the ANN is validated, the error is evaluated and used by a back-propagation algorithm, which goes through each layer in reverse order (from the output to the first layer) to measure the error contribution from each connection (reverse pass), and finally makes adjusts in the connection weights to reduce the overall error [30]. In general, the training process proceeds until the predetermined number of epochs is reached or if the accuracy begins to decrease or simply does not change. After finished the training, the ANN is tested by using the test-set.

The breast is composed by different tissues: milk glands, milk ducts and the breast density can vary due to hormonal factors [28]. This plethora of elements influences in image formation based on RX producing a complex image. Such characteristic is extended to the patches extracted from it and to the feature descriptors obtained from the patches.

An image feature descriptor encodes image data formatted as vector of numbers representing important discriminant aspects. In certain classes of problems the feature vector can be highly dimensional and presents nonlinear relationships between the variables (e.g., problems involving object recognition). According to Buduma and Locascio [31], these kind of feature dataset can be well treated by Deep Learning models, particularly the CNN and the multi-layer ANN. The latter was chosen for the work reported in the current paper because it demands a simpler computer configuration than the former, in accordance to the requisites discussed during the introduction of this paper.

## 4. Material and Methods

### 4.1 DDSM pre-processing

The source of the data used in this work was the database DDSM - Digital Database for Screening Mammography [12] which presents thousands of gray-tone well documented images divided in three classes of cases: Benign, Cancer and Normal.

The experiments were executed in a Oracle Virtualbox environment (10 GB RAM) running SO Ubuntu 16.04 LTS. The host computer was a notebook Vaio, processor Intel, 16 GB RAM. The main programming environments were Python 2.7™ and the ANN toolbox of Matlab (R2015b)™.

The Fig. 5 shows the main steps applied for pre-processing the database, they are described as follows:

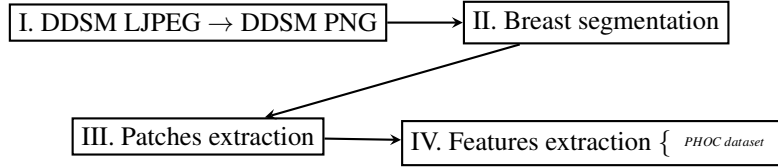
- I. DDSM database needs to be processed by adequate tool in order to extract images in a graphical format readable and useful for practical uses, the Anmol's library DDSMUtility [32] was applied for this sake. Since the database is composed by compressed LPEG images and ICS metadata files, Amno's library parses the ICS files taking metadata of the original LJPEG files, processing them to raw decompressed files and into other known formats. The PNG graphic format was the kind of image file used here;
- II. For each image (classes Benign, Normal and Cancer): the region of the breast was separated from the image for avoiding artifacts. A thresholding method (e.g. classical Otsu's binarization) was applied resulting in a binary mask which biggest region corresponded to the breast, after that a segmentation of the breast region was performed based on the area in the binary mask correspondent to the pixels of the breast;
- III. based on DDSM documentation, squared patches (256 x 256 pixels) were extracted from the images obtained from the step II. The idea was to get samples of each kind of the four possible regions in an digital mammography, resulting in 8,000 patches as follows:
  - (a) 2000 benign-patches (label BE),
  - (b) 2000 patches representing background (label BkG),
  - (c) 2000 cancer-patches (label CA) and
  - (d) 2000 normal-patches (label NO);

Note: it is important to train the ANN not only to recognize objects of interest (examples of patches representing benign, cancerous or normal tissues), but also what is not interesting for the final application, i.e., the image's background.
- IV. Features extraction using two levels spatial pyramid: PHOC dataset was obtained by features extracted from 8000 labeled patches (BE, BkG, NO, CA) using scripts in Matlab™.

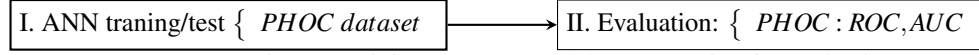
### 4.2 The Experiment

The pre-processing generated PHOC dataset composed by 8000 vector-descriptors equally distributed by label (BE, BkG, NO, CA). The next step consisted in training the ANN and evaluation of the results, as shown in the scheme in Fig. 6 and explained below:

- I. It was applied a Matlab's ANN toolbox [33] trained and tested by each descriptor dataset. The same ANN's configuration was used for all descriptors.
  - (a) Layers: empirically determined as 2 layers and 100 neurons per layer;



**Figure 5.** DDSM pre-processing and generation of features datasets.



**Figure 6.** Experiment: training the ANN and data generation for evaluation.

- (b) Neuron activation function: log-sigmoid [28]
  - (c) Training function: although Carneiro et al. [28] used gradient descent with momentum and adaptive learning rate backpropagation, here was applied scaled conjugate gradient backpropagation as training function;
  - (d) Data percentage for training/test/validation: 70% for training, 15% for validation and 15% for testing.
- II. The MSE (Mean Squared Error) is used for evaluating the ANN training performance.
- III. By measuring an ANN's classification performance one is capable to measure the quality of features descriptor used for training it.
- The metrics used to evaluate the performance were classical CM (Confusion Matrix), TPR (True Positive Rate, also known as Recall or Sensitivity), FPR (False Positive Rate) and Accuracy, as described in Eq. 3, Eq. 4 and Eq. 5 respectively, where:  $i \in \{BE, BkG, CA, NO\}$ , TP: True Positive predictions, FP: False Positive predictions and FN: False Negative predictions.

$$TPR_i = \frac{TP_i}{TP_i + FN_i} \quad (3)$$

$$FPR_i = \frac{FP_i}{FP_i + TN_i} \quad (4)$$

$$Acc_i = \frac{TP_i + TN_i}{TP_i + TN_i + FP_i + FN_i} \quad (5)$$

Neural net libraries (including the Matlab's<sup>TM</sup> toolbox used for this paper) usually apply softmax method in output layer, resulting in output vectors interpretable as class probabilities. Each input case is assigned to a class that presents the highest probability in output.

Construction of CM is based on relation between the set of predictions (neural net assignments) and targets (the known real assignments). CM synthesizes the overall performance of the classifier, its main diagonal shows the true positive predictions.

TPR, FPR and Accuracy are obtained from CM and are more quantitative, a Low TPR points out a high number of false negatives, for instance. By its turn, a low FPR indicates a high number of true negatives.

About Accuracy, it determines the level of correctness. Low accuracy indicates high number of incorrect predictions (false positives and false negatives).

It is also used the ROC (Receiver Operating Characteristic) curve and AUC (Area Under the ROC Curve). A medical diagnosis is based on a pre-determined threshold value applied to clinical variables, which determines if the case is a diseased or non-diseased one. ROC curve is a graphic relation between TPR and FPR, it shows the overall performance of supervised classification procedure at various decision thresholds [34]. In complement, AUC (Area Under the Curve) is an important metric that quantifies the area under respective ROC curve permitting quantitative comparisons.

The AUC varies in the interval  $[0.0, 1.0]$ :

- (a)  $0.0 \leq AUC \leq 0.5$ : indicates a practically useless classifier, which performance is worse than random [35];
- (b)  $0.5 < AUC \leq 1.0$ : the performance increases inasmuch as AUC gets closer to the value 1.0 (interval in which TPR is higher and FPR is lower) [35].

## 5. Results and Analysis

Figure 7 depicts the training performance of the ANN. Validation, Train and Test are described by respective MSE (Mean Squared Error) curve. As noticed in the same figure, in general, the three curves present a descent trajectory showing that respective MSE decays during training as expected.

Particularly, about the relation between test and validation curves they present a very similar profile. The test curve never increases significantly before the validation curve, which means that there is no sign of (undesirable) overfitting to the training set.

Confusion matrix is exhibited in Tab. 1 showing a main diagonal higher valued than secondary diagonal, which is an good signal in terms of classifier performance.

Considering  $i \in \{BE, BkG, CA, NO\}$ ,  $TPR_i$ ,  $FPR_i$  and  $Accuracy_i$  were extracted from CM and presented in Tab. 2:

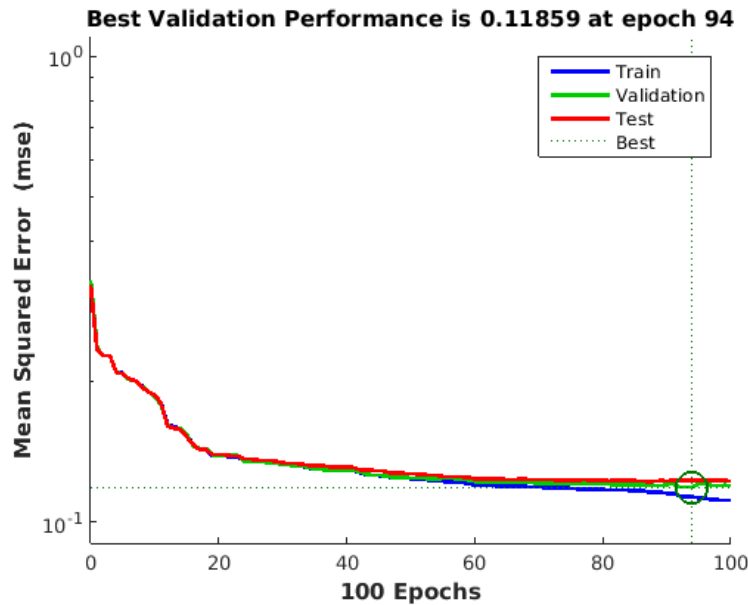


Figure 7. ANN's performance (mean squared error).

1. The  $TPR_i$  values are always higher than the correspondent  $FPR_i$ ;
2. The  $Accuracy_i$  present values always near to 1.0.

These metrics confirm the strength of the main diagonal in CM and suggests an acceptable classifier performance.

Table 1. Test confusion matrix

	Targets				
		BE	BkG	CA	NO
Predictions	BE	117	0	68	43
	BkG	5	305	5	36
	CA	99	0	118	50
	NO	89	0	87	178

Table 2. TPR, FPR and Accuracy by labels.

	TPR	FPR	Accuracy
BE	0.38	0.12	0.75
BkG	1.00	0.05	0.96
CA	0.42	0.16	0.74
NO	0.58	0.20	0.75

It was calculated the ROC curve and respective AUC for results in classifying the patches representing benign (label BE), background (label BkG), cancerous (label CA) and normal (label NO) samples. Considering  $i \in \{BE, BkG, CA, NO\}$ , the  $ROC_i$  curves and respective  $AUC_i$  are depicted in Fig. 8-left.

The classification of background patches was almost perfect ( $AUC_{BkG} \approx 1.0$ ) what is good since indicates that ANN is very capable to distinguish background from more relevant

structures/tissues for diagnosis. The ANN-PHOC's performance in classifying the other labels (BEs, CAs and NOs) is near to 0.8, which is a good indicator in terms of classifier's performance:  $AUC_{BE} = 0.792$ ,  $AUC_{CA} = 0.804$  and  $AUC_{NO} = 0.812$ .

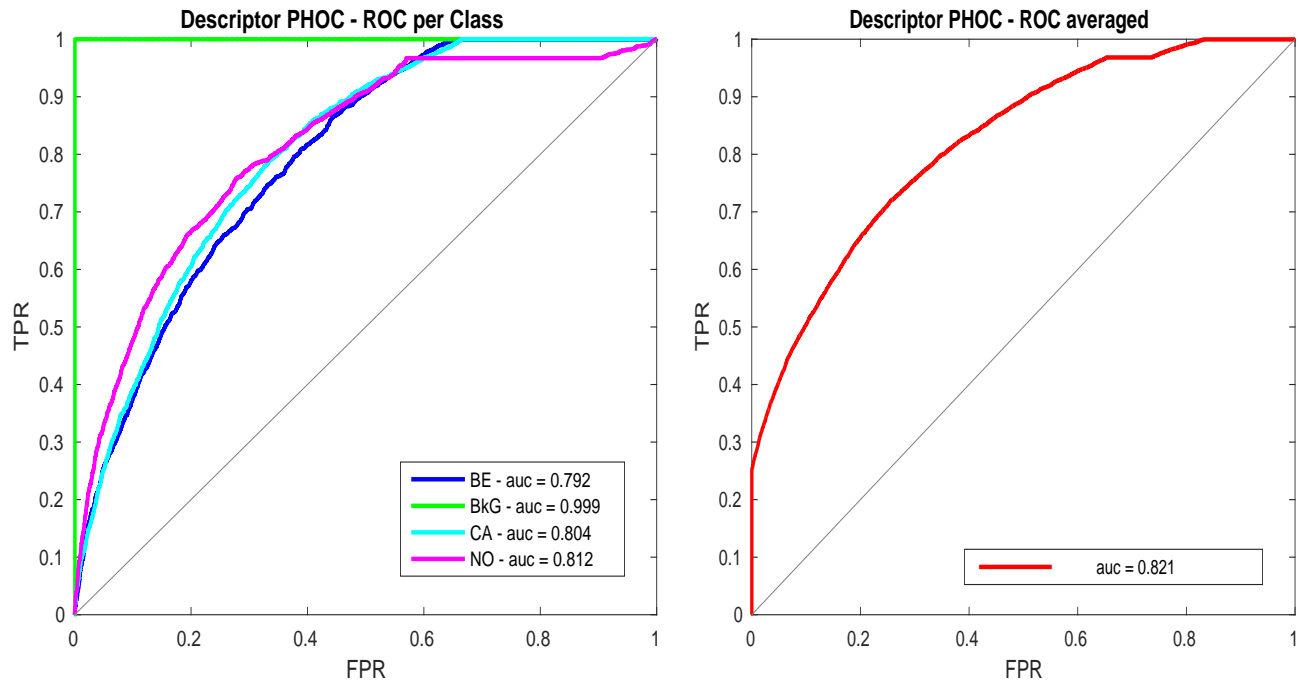
Also, the ROC-average curve and respective AUC was calculated ( $\overline{ROC}, AUC$ )<sub>PHOC</sub> depicting an overall view on the results, as it can be seen at Fig. 8-right this general performance was represented by  $AUC_{avg} \approx 0.8$ .

The analysis of results was based on three aspects: training performance, Confusion Matrix (TPR, FPR and Accuracy) and AUC-ROC curves. First two analysis indicated an acceptable performance in training and testing steps of the ANN with PHOC. The AUC-ROC shows classifier's "behaviour" for different thresholds and also indicated a relatively good performance.

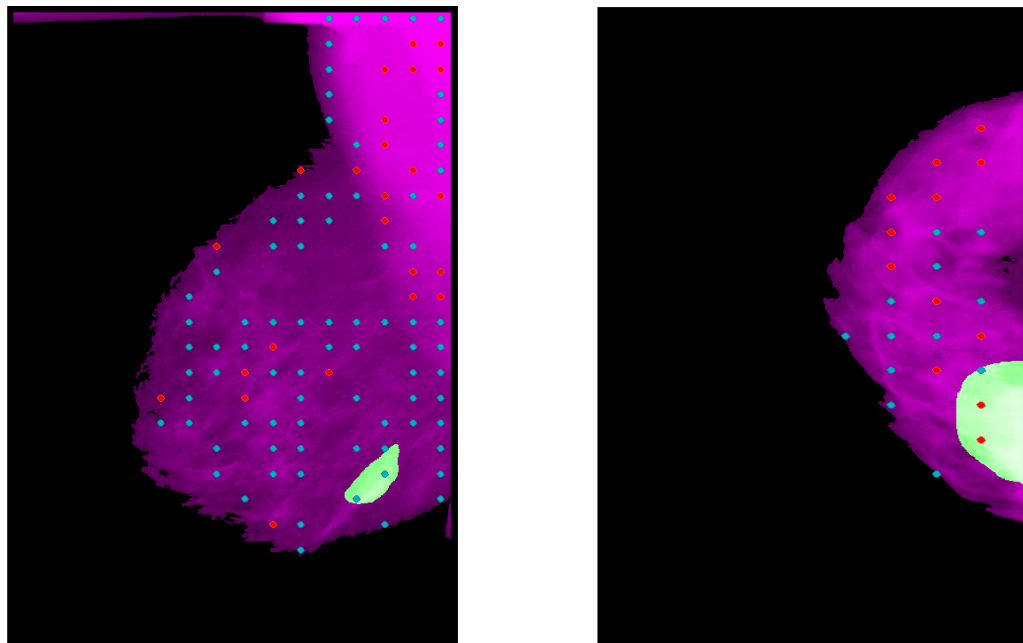
Even though these results points that the proposal (ANN and PHOC) is not ready for a CAD system, it presented relevant results in classification of important tissues represented in mammographies.

Figure 9 presents two cases resultant of application of the trained ANN. In both cases, for each patch the respective PHOC descriptor was inputted into the pre-trained ANN and labeled as benign (blue dots) ou cancerous (red dots). At the same figure, the lesions are greenish regions which location were determined by respective metadata file provided by DDSM.

Although there are lots of false positives in both of depicted cases, the image on the left side of the Fig. 9 shows that the benign lesion was correctly labeled by two blue dots. By its turn, on the right side of the same figure there is a cancerous case and also the malignant lesion was correctly detected, as shown by the red dots upon it.



**Figure 8.** ROC averaged and respective AUC.



**Figure 9.** A red dot represents a patch classified as cancerous and the blue dot is a patch classified as benign by the ANN. The greenish areas are regions marked by the experts and obtained from DDSM's metadata. The left side of the picture shows a benign case (*A-1691-LEFT-MLO* in DDSM), on the right side there is a cancerous case (*C-0098-RIGHT-CC* in DDSM).

## 6. Conclusion

The main motivation for this paper is testing a classical/simple neural net model (ANN) in classification of mammography using a simple descriptor (PHOC) in perspective of contribut-

ing for future works on CAD-like system development for breast cancer early detection, aiming low cost platforms like ordinary computer.

Image patches representing benign, cancerous, normal



and background samples were obtained from DDSM image database and processed for extraction of features by application of cited descriptor producing the wanted features dataset.

An ANN was configured and trained by using each features dataset. The interest was to evaluate PHOC descriptor in its capacity of representation the data providing a good training set to the ANN and consequently for obtaining the best ANN performance in classification of tissues. With this assumption an evaluation of the trained ANN's performances was done based on analysis of metrics ROC and AUC.

For the proposed scope, these results show that there is a potential to be explored. New settings to the ANN's hyperparameters, there is a large space for experimentation in this field) can be tested. Different patches size can be also used combined with other mammography datasets. Also, higher levels (parameter L) for spatial pyramid (PHOC) can be experimented.

## Acknowledgment

The author would like to show his gratitude to UDESC (Santa Catarina State University) and the FAPESC (Research Support Foundation of the State of Santa Catarina) for supporting this work.

## References

- [1] WHO-GENEVE-SWITZERLAND. *Breast Cancer*. [S.l.], 2019. Available at <http://www.who.int/cancer/prevention/diagnosis-screening/breast-cancer/en/>. Disponível em: <http://www.who.int/cancer/prevention/diagnosis-screening/breast-cancer/en/>.
- [2] SUN Z. ZHAO, Z. Y. F. X. H. L. Z. Z. W. S. J. J. P. Y. H. Z. Y. Risk factors and preventions of breast cancer. *International journal of biological sciences*, v. 13, n. 11, p. 1387–1397, 2017.
- [3] INCA. *INCA - Instituto Nacional de Câncer - José A. G. da Silva. Detecção Precoce*. [S.l.], 2018. Available at <http://www2.inca.gov.br/wps/wcm/connect/tiposdecancer/site/home/mama/deteccao-precoce>.
- [4] ACS. *American Cancer Society Recommendations for the Early Detection of Breast Cancer*. [S.l.], 2019. Available at <http://www.cancer.org/cancer/breast-cancer/screening-tests-and-early-detection.html>.
- [5] CHAN, H. P. et al. Improvement in radiologist's detection of clustered microcalcifications on mammograms - the potential of computer aided-diagnosis. *Invest. Radiol.*, v. 25, n. 10, p. 1102–1110, 1990.
- [6] AMARASINGHE, D. et al. Segmentation and sklearn numerical analysis of microcalcifications on mammograms using mathematical morphology. *The British Journal of Radiology*, v. 70, n. 837, 1997.
- [7] CHAN, H. P. et al. Image feature analysis and computer-aided diagnosis in digital radiography. i. automated detection of microcalcifications in mammography. *Medical Physics*, v. 14, n. 4, 1987.
- [8] SUCKLING, J. et al. The mammographic image analysis society digital mammogram database. *Excerpta Medica. International Congress Series*, p. 375–378, 1994.
- [9] YAGHJYAN, L. et al. Tissue-based associations of mammographic breast density with breast stem cell markers. *Breast Cancer Research*, v. 19, p. 100–, 2017.
- [10] TITUS-ERNSTOFF, L. et al. Breast cancer risk factors in relation to breast density (united states). *Cancer Causes Control*, v. 17, p. 1281–1290, 2006.
- [11] MENCATTINI, A. et al. Mammographic images enhancement and denoising for breast cancer detection using dyadic wavelet processing. *IEEE Transactions on Instrumentation and Measurement*, v. 57, n. 7, p. 1422–1430, July 2008.
- [12] DDSM. *DDSM: Digital Database for Screening Mammography*. 2019. Available at <http://www.eng.usf.edu/cvprg/Mammography/Database.html>.
- [13] ZENG, X. et al. Automated diabetic retinopathy detection based on binocular siamese-like convolutional neural network. *IEEE Access*, v. 7, p. 30744–30753, 2019.
- [14] Zhang, X. et al. Whole mammogram image classification with convolutional neural networks. In: *2017 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*. [S.l.: s.n.], 2017. p. 700–704.
- [15] *Efficient processing of deep neural networks: A tutorial and survey*, v. 105, n. 12. 2295–2329 p.
- [16] STOUTCHININ, A.; CONTI, F.; BENINI, L. Optimally scheduling cnn convolutions for efficient memory access. *IEEE TRANSACTIONS ON COMPUTER-AIDED DESIGN OF INTEGRATED CIRCUITS AND SYSTEMS*, v. 38, n. 6, p. 1095–1108, 2019.
- [17] LAZEBNIK, S.; SCHMID, C.; PONCE, J. Beyond bags of features: Spatial pyramid matching for recognizing natural scene categories. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. [S.l.: s.n.], 2006. p. 2169–2178.
- [18] BOSCH, A.; ZISSERMAN, A.; MUNOZ, X. Representing shape with a spatial pyramid kernel. In: *Proceedings of the 6th ACM international conference on Image and video retrieval*. [S.l.: s.n.], 2007. p. 401–408.
- [19] BOSCH, A.; ZISSERMAN, A.; MUNOZ, X. Image classification using random forests and ferns. In: *ICCV*. [S.l.: s.n.], 2007.
- [20] BOSCH, A.; ZISSERMAN, A.; MUNOZ, X. Scene classification using a hybrid generative/discriminative approach. *IEEE - PAMI*, v. 30, n. 4, p. 712–727, 2008.

- [21] GAO, H.; CHEN, W.; DOU, L. *Image classification based on support vector machine and the fusion of complementary features*. [S.l.]. Available at <http://arxiv.org/abs/1511.01706>. Disponível em: <http://arxiv.org/abs/1511.01706>.
- [22] K.P.SOMAN; RAMASHANDRAN, K.; RESMI, N. *Insight into Wavelets - from Theory to Practice*. 3rd. ed. [S.l.]: PHI, 2010.
- [23] GARGOUR, C. et al. A short introduction to wavelets and their applications. *IEEE Circuits and Systems Magazine*, v. 9, n. 2, p. 57–68, 2009.
- [24] KOCUR, C. M. et al. Using neural networks to select wavelet features for breast cancer diagnosis. *IEEE ENGINEERING IN MEDICINE AND BIOLOGY*, v. 15, n. 3, p. 95–102, 1996.
- [25] DHEEBA, J.; SINGH, N. A.; SELVI, S. T. Computer-aided detection of breast cancer on mammograms: A swarm intelligence optimized wavelet neural network approach. *Journal of Biomedical Informatics*, v. 49, p. 45–52, 2014.
- [26] MEHDY, M. M. et al. Survey: Artificial neural networks in image processing for early detection of breast cancer. *Hindawi Computational and Mathematical Methods in Medicine*, v. 2017, 2017.
- [27] GARDEZI A. ELAZAB, B. L. T. W. S. Breast cancer detection and diagnosis using mammographic data: Systematic review. *J Med Internet Res*, v. 21, n. 7, 2019.
- [28] CARNEIRO, P. C. et al. Breast density pattern characterization by histogram features and texture descriptors. *RESEARCH ON BIOMEDICAL ENGINEERING*, v. 33, n. 1, p. 69–77, 2017.
- [29] ZHOU, C. et al. Computerized image analysis: Estimation of breast density on mammograms. *Med. Phys.*, v. 28, n. 6, p. 1056–1069, 2001.
- [30] GERON, A. *Hands-On Machine Learning with Scikit-Learn and TensorFlow, Concepts, Tools, and Techniques to Build Intelligent Systems*. [S.l.]: O'Reilly Media, 2017.
- [31] BUDUMA, N.; LOCASCIO, N. *Fundamentals of deep learning: designing next-generation machine intelligence algorithms*. [S.l.]: O'Reilly Media, 2017.
- [32] SHARMA, A. *DDSM Utility*. [S.l.]: GitHub, 2015. <https://github.com/trane293/DDSMUtility>.
- [33] BEALE, M. H.; HAGAN, M. T.; DEMUTH, H. B. *Neural network toolbox: users guide*. [S.l.]. Available at [https://www.mathworks.com/help/pdf\\_doc/nnet/nnet\\_ug.pdf](https://www.mathworks.com/help/pdf_doc/nnet/nnet_ug.pdf). Disponível em: [https://www.mathworks.com/help/pdf\\_doc/nnet/nnet\\_ug.pdf](https://www.mathworks.com/help/pdf_doc/nnet/nnet_ug.pdf).
- [34] KUMAR, R.; ANTONY, G. A review of methods and applications of the roc curve in clinical trials. *Drug Information Journal*, v. 44, n. 6, p. 659–671, 2010.
- [35] PRATI, R.; BATISTA, G.; MONARD, M. A survey on graphical methods for classification predictive performance evaluation. *IEEE Transactions on Knowledge and Data Engineering*, v. 23, n. 11, p. 1601–1618, 2011.