QIRT'2022

10.21611/qirt.2022.2019
16th Quantitative InfraRed Thermography Conference

Effects of region of interest on breast cancer detection using CNN and infrared imaging

by C. B. Gonçalves*, A. C. Prado Domingos*, B. Yousefi**,***, J. R. Souza*, H. Fernandes*,****

* *Faculty of Computing, Federal University of Uberlandia, Uberlandia, Brazil*
** *Fischell Department of Bioengineering, University of Maryland, Rockville, MD, USA*
*** *Department of Electrical and Computer Engineering, Laval University, Quebec, QC, Canada*
**** *Fraunhofer IZFP Institute for Nondestructive Testing, Saarbrucken, Germany*

Abstract

Breast cancer is one of the most common type of cancer that affects woman in the World. It affects millions of women every year killing hundreds of thousands yearly. Early detection of this disease is essential for improving chances of cure and recovery of the patients, thus, one of the most important factors in patient's treatment is early detection. Mammography, the golden standard for detection this decease is not always 100% effective, which means that it is not always recommend. In this sense, infrared imaging is a promising technique that might be used as a complementary examination technique in a computer-aided diagnosis system. In this work we used genetic algorithms and convolutional neural networks to classify images from the DMR-IR public database. We analyzed both the entire image and only the breast region. Best results were F1-score of 0.92 for entire images and 0.90 for breast regions.

1. Introduction

In Brazil, breast cancer is the second type of cancer among women only behind skin cancer. In 2021, it is estimated that 66,280 new cases of the disease will occur [1]. It is also the second type of cancer among American women [2] being the second leading cancer-related death in women, behind lung cancer [2]. In 2021, the American Cancer Society estimates that over 281,550 women will be diagnosed with breast cancer, and more than 49,290 women will lose their lives. [2].

The "gold standard" for diagnosing breast cancer is mammography, which takes an X-ray of the breast [3, 4]. However, mammography has a lower sensitivity for young patients due to denser breasts [5]. According to [5], while the sensitivity in older patients is around 90%, in younger patients, the sensitivity decreases to 60%. Infrared imaging could be used as an auxiliary imaging test to mitigate this lower sensitivity for younger patients [4, 5, 6]. The human body emits infrared radiation, which can be captured by an infrared camera. As tumor areas have higher temperature than normal areas, an infrared image from the breast skin region may reveal hot spots that might be related to a tumor.

Convolutional neural networks (CNNs) has been used in computer visual and pattern recognition very successfully for several years now. In this study we used genetic algorithm (GA) to optimize hyper-parameters and architectures for the fully connected layers of two state of the art CNNs: ResNet-50 and DenseNet-201. These optimized CNNs were then trained using images from the DMR-IR[1] public database to classify patient into two classes: with cancer and without cancer. Two sets of images were analyzed. First, the entire images were used in the training, validation and testing phase, and later, only the breasts regions, i.e. region of interest (ROI), were used in the process. Results show that using only the region of interest does not affect significantly the classification outcome.

2. Related Work

The *DMR-IR* database, presented in [7], is a thermographic breast imaging database commonly used in the literature. Originally, the database consisted of data from 141 patients collected at the *Hospital Universitário Antônio Pedro* (HUAP) of *Universidade Federal Fluminense* (UFF), however, the database included the addition of images of new patients over time. Overall, each patient has images collected with both a static protocol and a dynamic protocol. Each patient has 5 images collected in different positions using the static protocol, namely: frontal, right and left sides 45° and right and left sides 90°. For the dynamic protocol, the authors proposed a collection where the patient is acclimatized until the average temperature of the region between the breasts reaches 30° or until the maximum time has been reached (5 minutes). After acclimatization of the patient, 20 images were collected in a period of 5 minutes and 2 other right and left lateral images of 90°. The images obtained have a size of 640 x 480 pixels and are publicly available on a portal made available by the researchers.

The work of [4] uses thermographic images and CNNs for breast cancer detection. For this, they used the database of thermographic breast images *DMR-IR*, considering 57 patients from the dynamic protocol (considering 19 healthy patients and 37 sick patients and each patient containing 20 images). As pre-processing, ROIs were extracted. Several state-of-the-art CNNs were evaluated, including ResNet, SeResNet, VGG16, Inception, InceptionResNetV2 and Xception, as well as CNNs proposed by the authors themselves called *baseline*. In addition, CNNs with hyperparameter optimization were proposed. The

[1] http://visual.ic.uff.br/dmi/

surrogate model proposed by the authors aims to find the best CNN architecture for the base in question using a Bayesian optimization based on the *tree parzen* estimator. Among the parameters considered in the optimization are: minimum and maximum number of blocks; number of convolutional layers and number of filters per block, optimizer type, kernel size, pooling layer size, batch normalization type, dropout rate, number of neurons connected in the last two layers and top layer type. The optimized network presented better results than all the others evaluated, obtaining 94% of accuracy and F1-score of 0.91. The authors state that, for the base in question, smaller CNNs present a better result. Another contribution of the authors is the experiments related to the influence of data increase on classification, seeking ways to deal with the lack of large amounts of available data. Based on these experiments, the authors claim that the increase in data positively impacts the classification providing more information for the models to learn. The techniques for increasing the data evaluated by the authors were: horizontal and vertical inversion, rotation between 0 and 45°, zoom (20%) and noise.

The study presented in [8] is among the pioneers to use the DMR-IR base with CNNs. The authors considered both the static protocol and the dynamic protocol images for analysis and presented comparisons between the results obtained with both protocols. 177 images from normal patients and 42 images from cancer patients were considered for the static protocol. As for the dynamic protocol, 95 healthy patients and 42 sick patients (each patient with 20 images). As there is an imbalance between the classes, the authors propose the use of cutting and duplication to increase the number of images available for the class of patients, providing a better balance of classes. The CNN used was proposed by the authors themselves. To use dynamic images, four approaches seeking to compose a single image with the information from the available images are proposed and evaluated. In addition, experiments were carried out with both color images and grayscale images. The authors report that the best results were with color images for both static and dynamic images. The results for the still images obtained were 0.98 for ACC, 0.97 for sensitivity, 1 for specificity and F1-score of 0.98. Among the dynamic approaches considered by the authors, the best results obtained were 0.95 of ACC, 0.93 of sensitivity and 0.98 of specificity and F1-score of 0.95.

The work developed by [9] combines three bases of infrared images of the breast: DMR-IR, *Instituto Jaliciense de Cancerología (IJC)* and *Instituto de Seguridad y Servicios Sociales de los Trabajadores del Estado (ISSSTE)*. In all, the authors considered 311 images, of which 267 were healthy and 44 were sick. Using the pre-trained Resnet-101 convolutional network, the authors seek to classify the images between healthy and diseased. Experiments were carried out in two groups: one with balanced data and one with an unbalanced base, where fewer thermograms of the sick class were used. In both cases, random data augmentation was used with the following techniques: rotation between 0 and 359°, reflection in the X and Y axes and translation between 0 and 50 pixels and the combination of all these techniques. Data were split into training and testing with the proportions of 70% and 30% and only training data were increased by a factor of 67 in both groups. The results indicated that the network using the balanced data performed better than the unbalanced data. For the first set (balanced) the authors obtained 92.3% of sensitivity against 84.6% of sensitivity obtained by the second set.

Authors in [3] also used static images from the base *DMR-IR* and CNNs using *transfer learning*. The CNN chosen by the authors was Alexnet. The work aims to evaluate the impact of changing specific layers of the original network. 181 images were considered, 147 healthy and 34 patients. The authors also performed experiments with both the unbalanced and balanced base. To obtain the balanced basis, data augmentation was applied to the images of sick patients in some experiments with the following techniques: vertical translation (flip) and rotation: -30° to 30° for get a balanced database. In the balanced set, the authors considered 147 healthy and 135 sick images. For the pre-processing, the authors evaluated two ways of converting the temperature values to the *range* that the network expects: they transformed the images to the range 0-255 and replicated the data to simulate 3 channels and transformed the image to RGB using Matlab's colormap jet (with this the images would be colored). Unlike the other works, *cross validation leave one patient out* was used. The best results obtained were considering the images with RGB pre-processing and data augmentation (balanced base) changing the entire densely connected part (including increasing the number of layers) which presented 94.3% of accuracy and 0.94 of F1-score .

In [10], the authors also use the *DMR-IR* base associated with some pre-trained convolutional networks available in the literature for breast cancer detection using the *fine tuning* approach. The networks considered are: ResNet101, DenseNet, MobileNetV2, ShuffleNetV2. All of them had the densely connected layer replaced. Both static and dynamic images were used. In training, only static images were used, while for testing two distinct groups were considered, one with only static images and the other with only dynamic images, so that each experiment was tested twice. In addition, data augmentation was used in the training set. The authors reported that the network that performed best was DenseNet. With the DenseNet network, the authors reported accuracy, recall, precision and F1-score equal to 1 for both the set of static and dynamic images. Furthermore, MobileNetV2 presented good results with lower computational cost with 100% accuracy in the test set of still images and 99.6% accuracy for the dynamic images.

Both [5] and [11] also use CNNs taking advantage of *transfer learning* on the *DMR-IR* base. [5] used static images considering 88 patients (total of 440 images) and CNNs AlexNet, GoogLeNet, ResNet-18, VGG-16 and VGG-19. The authors evaluated different *learning rates* and number of epochs. The best results were obtained with VGG-16 with 77.5% accuracy, 85% sensitivity and 70% specificity. [11] considered 144 patients (88 healthy and 56 patients or suspects) and the following CNNs: ResNet50, InceptionV3, VGG16, VGG19. Different proportions of training and testing were evaluated and comparisons between their work and works present in the literature that use 'traditional' ways of extracting features are evaluated. The best results reported were with ResNet50 obtaining 88.89% of accuracy.

3. Methodology

3.1 Database

The database used in this study is the DMR-IR public database [7]. The database has images from two different protocols: static and dynamic. In the static protocol, patients have 5 images (5 different poses): frontal, left, and right lateral at 45°, and left and right lateral 90°. On the other hand, the dynamic protocol collects 20 images over a period of 5 min for each motionless patient. Additionally, two other images were collected from the left and right with a 90° angle. Therefore, each patient is expected to have 27 images, 22 from the dynamic protocol, and 5 from the static. The images are 640x480 px and were captured using a FLIR thermal camera, model SC620. The authors proposed different acclimatization processes for each protocol [7].

3.2 Image Preparation

For this work we used static images. Currently there are 176 healthy patients and 39 sick patients. Since it is not balanced and we did not considered one sick patients due to visual problems in its images, a subset of the healthy patients was selected with the same number of images of the sick class. Therefore, 376 images are considered from 38 patients in each class. We did not considered 39 (number of sick patients available) because the images from one of the sick patients were not good (have some irreversible distortions). We also considered two types of images. The first set contained the original images available in the data base. The second set considered only the region of the breast which contains the possible tumor, i.e., region of interest (ROI). Those regions were selected manually following the following steps.

3.2.1 Step 1 - Normalization

Images were normalized according to the formula:

$$255 * \frac{(P_{ij} - L)}{H - L} \tag{1}$$

where P_{ij} is the value of image pixel in the line i and in the column j to be normalized, H is the highest value among the pixels to be normalized and L is the highest value among the pixels to be normalized. Figure 1 shows one image after this step.

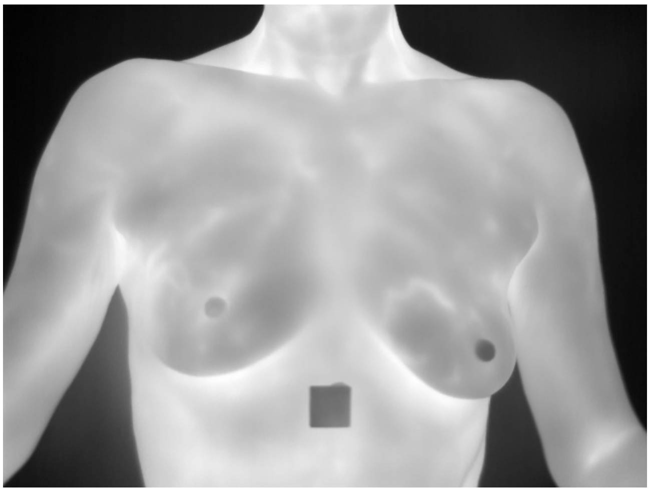

Fig. 1. *An image sample after normalization*

3.2.2 Step 2 - ROI selection

After normalization images were manually cut for the selection of the ROI. Both size and shape of these regions were chosen based on the work presented at [8]. Figure 2 presents an example of ROI.

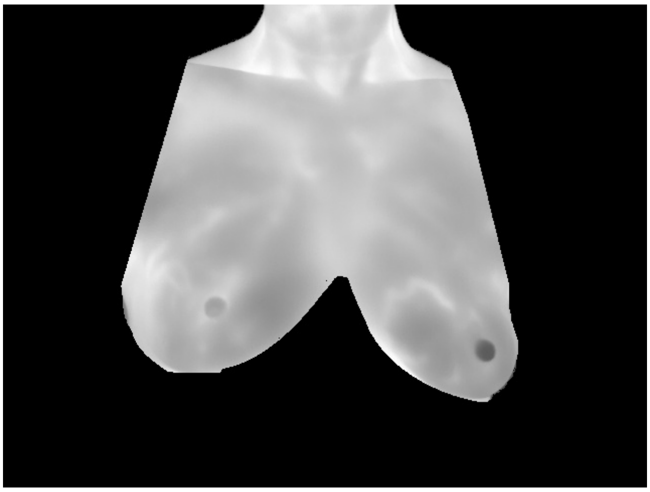

Fig. 2. *An image sample after ROI manually selection*

3.2.3 Step 3 - Resizing and Fake Coloring

Due to restrictions imposed by the used CNNs, image were resized and each pixel value was replaced by a fake color. For the size, images original size was reduce to 244 pixel by 244 pixels. For the color, since the networks expect three images for each individual being fed to the network, we used, as suggested by the work presented in [3], the 3 channels obtained by the colormap jet for Matlab applied in each image. Figure 3 shows the colored and reduced image.

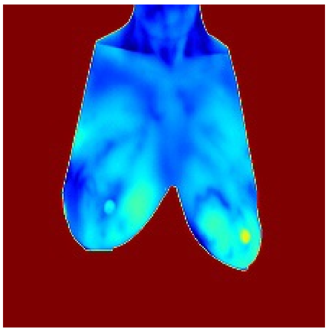

Fig. 3. *An image sample after normalization*

3.3 Classification

To classify the patients, we used stated of the art CNNs: ResNet-50 and DenseNet-201. CNNs are mainly divided in two parts: convolutional, which is composed by convolutional and pooling layers, and fully connected layer which is the part that does the classification with the data extracted by the previous part of the network [12]. Figure 4 show an example of a CNN. In order to find a good fully connected layer architecture with a good learning rate, in this study we used optimization techniques for this process. Genetic Algorithm (GA) was used to to find the following parameters: learning rate, number of fully connected layers, number of neurons of each fully connected layer, after which layers there is dropout, and dropout rate of each dropout present in the layer.

Since there is not sufficient images available for a full train, we applied a technique called transfer learning [11]. Here we keep the convolutional part of the CNNs we use as they are provided and only changing the fully connected part (provided by the GA), which is the part that does the classification. The dataset was then divided into training, validation, and testing (70% - 15% - 15%) for the classification processes. All 5 images for each patient were kept in the same group. Thus, we are certain that the network will not be trained and tested using different images from the same patient.

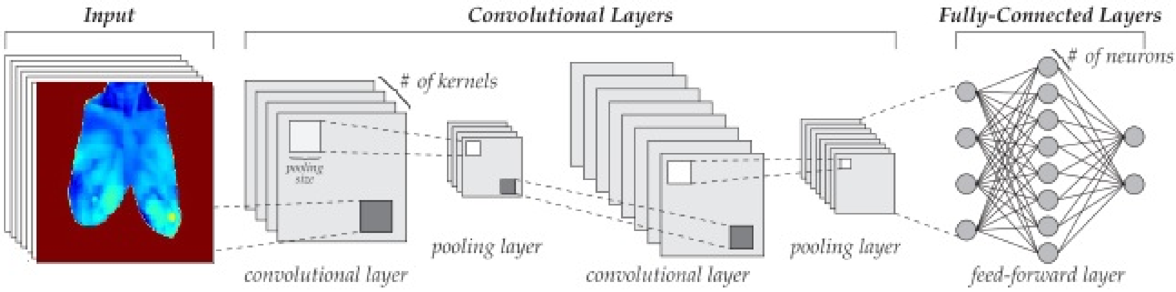

Fig. 4. *Example of CNN architecture. Adapted from [12].*

4. Results

For each set of images, i.e., entire images and ROI images, the GA ran 5 times and the best configuration was chosen. Also, several learning rates and epochs were tested. The best results were achieved with 30 epochs and 0.0001 learning rate. For the first group (entire images) the best results were: for the DenseNet-201 the highest F1-score for the test set was 0.92 while for ResNet-50 it was 0.90. For the second group (set with ROI images) the the best results were: for the DenseNet-201 the highest F1-score for the test set was 0.90 while for ResNet-50 it was 0.90.

5. Conclusions

Although we got slightly better results for the group with entire images (0.92 against 0.90), results are very close and, at this point of the research, one cannot say if there is more advantageous to use the entire image or only the ROI. Next steps of this research include the use of surrogate models to evolve GA populations more efficiently and consequently be able to to optimize more hyper-parameters simultaneously, and the use transient images of the patients to improve classification process.

Acknowledgment

This study was financed in part by the *Coordenação de Aperfeiçoamento de Pessoal de Nível Superior* – Brazil (CAPES) – Finance Code 001, by The National Council for Scientific and Technological Development (CNPq) project number 407140/2021-2, and by The Minas Gerais Research Funding Foundation (FAPEMIG) project number APQ-01576-18.

References

[1] Instituto Nacional do Câncer. Controle do câncer de mama: Conceito e magnitude. https://www.inca.gov.br/controle-do-cancer-de-mama/conceito-e-magnitude. Accessed: 2021-12-21.

[2] American Cancer Society. Key statistics for breast cancer. https://www.cancer.org/cancer/breast-cancer/about/how-common-is-breast-cancer.html. Accessed: 2021-12-21.

[3] Çağrı Cabıoğlu and Hasan Oğul. Computer-aided breast cancer diagnosis from thermal images using transfer learning. In *International Work-Conference on Bioinformatics and Biomedical Engineering*, pages 716–726. Springer, 2020.

[4] Juan Zuluaga-Gomez, Zeina Al Masry, Khaled Benaggoune, Safa Meraghni, and Noureddine Zerhouni. A cnn-based methodology for breast cancer diagnosis using thermal images. *arXiv preprint arXiv:1910.13757*, 2019.

[5] Esdras Chaves, Caroline B Gonçalves, Marcelo K Albertini, Soojeong Lee, Gwanggil Jeon, and Henrique C Fernandes. Evaluation of transfer learning of pre-trained cnns applied to breast cancer detection on infrared images. *Applied Optics*, 59(17):E23–E28, 2020.

[6] Alisson Figueiredo, Henrique Fernandes, and Gilmar Guimaraes. Experimental approach for breast cancer center estimation using infrared thermography. *Infrared Physics Technology*, 95:100–112, 2018.

[7] LF Silva, DCM Saade, GO Sequeiros, AC Silva, AC Paiva, RS Bravo, and Aura Conci. A new database for breast research with infrared image. *Journal of Medical Imaging and Health Informatics*, 4(1):92–100, 2014.

[8] Matheus de Freitas Oliveira Baffa and Lucas Grassano Lattari. Convolutional neural networks for static and dynamic breast infrared imaging classification. In *2018 31st SIBGRAPI Conference on Graphics, Patterns and Images (SIBGRAPI)*, pages 174–181. IEEE, 2018.

[9] Juan Carlos Torres-Galván, Edgar Guevara, Eleazar Samuel Kolosovas-Machuca, Antonio Oceguera-Villanueva, Jorge L Flores, and Francisco Javier González. Deep convolutional neural networks for classifying breast cancer using infrared thermography. *Quantitative InfraRed Thermography Journal*, pages 1–12, 2021.

[10] Roslidar Roslidar, Khairun Saddami, Fitri Arnia, Maimun Syukri, and Khairul Munadi. A study of fine-tuning cnn models based on thermal imaging for breast cancer classification. In *2019 IEEE International Conference on Cybernetics and Computational Intelligence (CyberneticsCom)*, pages 77–81. IEEE, 2019.

[11] Seyfullah Kiymet, Muhammet Yavuz Aslankaya, Murat Taskiran, and Bulent Bolat. Breast cancer detection from thermography based on deep neural networks. In *2019 Innovations in Intelligent Systems and Applications Conference (ASYU)*, pages 1–5. IEEE, 2019.

[12] Alejandro Baldominos, Yago Saez, and Pedro Isasi. Evolutionary convolutional neural networks: An application to handwriting recognition. *Neurocomputing*, 283:38–52, 2018.