

Classification of Breast Cancer from Digital Mammography Using Deep Learning

José Daniel López-Cabrera^[1,A], Luis Alberto López Rodríguez^[2], Marlén Pérez-Díaz^[2]

^[1] Centro de Investigaciones de la Informática, Facultad de Matemática, Física y Computación, Universidad Central “Marta Abreu” de Las Villas.

^[A] josedaniellc@uclv.cu

^[2] Departamento de Automática y Sistemas Computacionales, Facultad de Ingeniería Eléctrica, Universidad Central “Marta Abreu” de Las Villas.

Abstract Breast cancer is the most frequent in females. Mammography has proven to be the most effective method for the early detection of this type of cancer. Mammographic images are sometimes difficult to understand, due to the nature of the anomalies, the low contrast image and the composition of the mammary tissues, as well as various technological factors such as spatial resolution of the image or noise. Computer-aided diagnostic systems have been developed to increase the accuracy of mammographic examinations and be used by physicians as a second opinion in obtaining the final diagnosis, and thus reduce human errors. Convolutional neural networks are a current trend in computer vision tasks, due to the great performance they have achieved. The present investigation was based on this type of networks to classify into three classes, normal, benign and malignant tumour. Due to the fact that the miniMIAS database used has a low number of images, the transfer learning technique was applied to the Inception v3 pre-trained network. Two convolutional neural network architectures were implemented, obtaining in the architecture with three classes, 86.05% accuracy. On the other hand, in the architecture with two neural networks in series, an accuracy of 88.2% was reached.

Keywords: Deep Learning, Image Processing, Breast Cancer.

1 Introduction

Breast cancer is a serious global health problem due to its morbidity and mortality, with approximately 1.5 million new cases each year. Also, is the most frequent cancer in women worldwide, with 16% of the total. Mammography has proven to be the most effective method for early detection of breast cancer, allowing to diagnose the disease in its initial stage. On the other hand, techniques as ultrasound, magnetic resonance imaging and computational tomography, have proven be useful for the study and identification of these anomalies, but they are not as effective for early detection of the breast cancer [1].

Breast cancer is manifested basically by some kinds of characteristic lesions. There are masses, which are large tumors of very low contrast image with respect to normal tissue that can be malignant or benign. There are also micro-calcifications, these are tiny points of hydroxyapatite or calcium oxalate, which when present are a signal of a sick breast or who will soon get sick. These are so tiny that, if the image presents a lot of noise, the contrast is not good or the spatial resolution of the equipment is not enough, they are not noticeable. In addition, there are galactophores ducts with abnormalities, such as carcinogenic infiltration [2]. The non-detection of a breast lesion depends not only on the size of the breast or the level of contrast, but also on the sensitivity of the mammograph, as well as the ability of the radiologist who detects them and the conditions under which the radiologist analyze the image [3].

Computer aided diagnosis (CAD) systems have been developed to increase the sensitivity of mammographic exams. These systems are capable of marking suspicious regions on mammograms to help human specialists in making decisions [4]. These systems have now reached greater benefits thanks to the renewed interests of artificial intelligence in medical imaging. There is an increase in related research in the last decade on machine learning (ML) and a special type of this, deep learning (DL). This has been possible due to the development of computing capacity, storage capacities and data availability [5].

The CADs initially had four stages: pre-processing of images, segmentation of the lesion, extraction of characteristics, and classification. The traditional methods used in the classification were based on ML algorithms. The work carried out in [6] stands out, in which several classifiers with the MIAS database are analyzed, obtaining the best result with a support vector machine (SVM) with an accuracy of 84.57%, a sensitivity of 0.8261 and a specificity of 0.8889. In [7] the use of Local Binary Pattern (LBP) for the extraction of features in mammography images using the MIAS database is evaluated, the best results are achieved for a Bayesian classifier with 75% effectiveness.

A special type of DL applied to computer vision are Convolution Neural Networks (CNN) [8]. These have a special connection structure in their hidden layers from their convolution and pooling layers, making them invariant to the displacement, scaling and rotation of the input images. This means that CNNs are robust to small variations in the input images. The use of this technique implies that features extraction is an internal part of the network. In fact, CAD systems have benefited from this approach. The DL has transformed the computer-assisted image analysis approach, fundamentally distinguishing it from traditional CAD methods [9]. In [10] a study was conducted with about 29,000 mammographic images from 5 different institutions using as classifier a CNN. The result was a sensitivity of 76.1%, a specificity of 88.6% and an AUC of 0.906. In [11] a CNN was proposed for the automatic detection and classification of breast cancer using the Digital Database for Screening Mammography (DDSM) with a specificity of 96.9%, a sensitivity of 98.9%, obtaining 3 false negatives and 1 false positive. In [12] a CAD system was implemented based on one of the most successful object classification platforms, the Faster R-CNN. From this network architecture, they reached second place in the Digital Mammography DREAM Challenge competition with an AUC of 0.85. In [13], an AUC of 0.92 was achieved in the DDSM database. In [14] a CAD system based on CNN was developed, achieving an AUC of 0.98 in the DDSM database, 0.97 in the INbreast database, 0.96 in the BCDR and 0.99 in the MIAS database.

CNNs automatically learn the important features from images to perform classifications with high discriminatory power. The goal of a CNN is to find the coefficient of the kernel that minimize the error in the classification task. The processes of training require a large amount of labeled image to get a good performance, which is not frequent in the medical domain. Furthermore, training a CNN from scratch requires high computational cost, large amount of memory and time. On the other hand, to train a CNN with a small data set lead the algorithm to suffer overfitting. One of the variants to overcome these inconveniences is to use a technique known as transfer learning. This technique has been proved effectively in the medical domain, where the data set is limited [15]. A widespread way to carry out the training in CNN is using the ImageNet set [16] which has more than 1.2 million of images divided in 1000 categories. Later, a fine-tuned strategy to adjust the weights of the convolutional kernels using the new images is used.

Acquisition and processing of mammograms has greatly evolved. However, due to the low contrast in images and the small size of lesions such as microcalcifications and masses, in many cases they are still difficult to detect and even more difficult to classify as benign or malignant. CAD systems try to mitigate this problem, and specifically artificial intelligence techniques, such as CNNs, which are a current trend in image classification. The objective of this research is to develop a CNN capable of classifying the main mammary lesions detected using the transfer learning technique. This approach has not been widely addressed in the mammography imaging domain [17]. This study evaluates the performance of the Inception v3 network [18], which won the ImageNet challenge in 2016 [16].

2 Materials and Methods

2.1 Set of Images Used

The mammographic images were downloaded from Mammographic Image Analysis Society Database (MINI MIAS) [19]. This set has 322 images obtained from film radiograph, digitalized with a scanner Joyce-Loebl in the optic range of 0 and 3 and each pixel is represented with 8 bits. The images were classified by radiologists. The set has a resolution in pixels of 1024x1024. The dataset adds a file which contain a whole information about the images. This file is constituting by seven columns. In the first column appears the name of the image, in the second, the characteristics of the tissue. In the third column appear the kind of lesion, in the fourth column is the classification of the lesion in malign or benign. The fifth and sixth columns present the (x,y) axis of the lesion.

Finally, in the seventh column is the radius in pixels of the circumference that enclose the lesion. Considering this information, the regions of interest (ROI) of each image are obtained. In the case of images that did not show lesions (classified as normal), the ROIs are obtained randomly as explained below.

Working with CNN requires a lot of training data, issue that is not met for the MIAS database because it does not have many images. This caused the need to use techniques described in [20] for data augmentation, which will be explained later.

2.2 Dataset Division: Training, Validation, and Testing

The total ROI set (which from now on will be the new images) was divided into a training set and a testing set. 70% of the images with malignant lesions were selected, which is the class with fewer cases and the same number of normal images and images with benign lesions were added to form the set of three classes. The rest of the images were included in the test set. This procedure was done to keep the number of images per class balanced in the training process of the network. Something similar was done for the architecture of two classes. First, the set is divided into normal and anomalous images. Likewise, 70% of the class with fewer cases is selected (images with benign and malignant lesions) and the same amount of the other class is added (normal images). The same is done in the next stage, this time between images with malignant and benign lesions. This last set is the least amount of training images. It is important to say that in all the case training set is also divided to adjust the hyperparameters in training phase. Thus, 80% for training and another 20% for validation. The final performance of the presented system is only evaluated using the testing set.

2.3 Extraction of Regions of Interest

A program was carried out in Matlab, which extracts the ROI from each image in the database. It was divided into malignant and benign lesions according to the information obtained from the file that describes the characteristics of the images. In those images that did not show lesions (classified as normal), ROIs were randomly selected. For this, five ROI were obtained for each normal image, randomly generating the middle point and radius of the ROI. The values used for the middle point of the ROI were generated in the range of the number of pixels in the rows and columns of the images; to obtain the radius, the value was between 5 pixels and the number of pixels in the rows (which is equal to that of the columns). In this way, no region of interest exceeded the outer edges of the images. This creates a new data set, composed not of complete images, but of ROI. The colour space was modified to the images selected for the network training. The images originally appear in grayscale. In order to increase the training set, seven new images are generated for each ROI as shown in Figure 1.

It is known that DL techniques perform better when are applied to large volumes of data. A popular way to achieve larger data sets is from the data augmentation technique. This technique can increase the size of the set more than 10 times, which helps prevent overfitting. In addition, more robust models with greater generalization capacity are obtained [21]. Within the approaches used to increase the data appears the use of geometric transformations. The use of these techniques on mammogram images makes sense, since these images are affected by the size and shape of the lesions, which appear with different orientations. These operations should boost the model to learn better. The data set of images was enlarged using a series of random transformations, so that the model never sees an image twice in the same way. Random translations between -15 to 15 pixels were applied on both axes, random rotations between 0 ° and 360 °, and the mirror technique using either of the two axes as a reference. These techniques only were applied to the training set. That is, the test set was not modified. In this way, the evaluation is carried out on the original images.

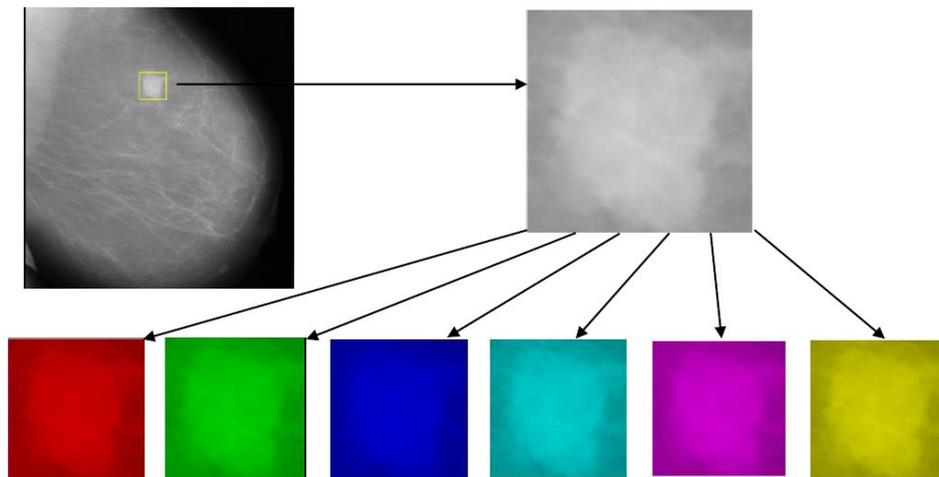


Figure 1. Converting the ROI into different colour spaces.

2.4 CNN Architectures and Parameters Used

The Inception v3 network was used. This CNN was created by the Google research group with the aim of reducing the computational cost of the previous CNN and maintaining the same level of performance. Inception v3 has several modules called Inception that, for the most part, can be described as 4 convolution filters in parallel of 1x1, 3x3 and 5x5. Thanks to the implementation of the parallel network and the descending sampling layers in each block, the execution time of the model is better than VGG and ResNet. The main modifications made to the network to reduce the computational cost were the use of 3x3 convolution filters instead of 5x5 or 7x7 as in previous modules [18]. The expansion of width in each layer was also implemented to increase the combination of characteristics in the next layer, thus obtaining a network with a balanced computational cost between its depth and width.

The first convolutional layers of a CNN learn generic characteristics and can work more like edge detectors, which should be useful for general tasks, but the following layers become progressively more specific to the details of the classes contained in the data set [4]. According to this statement, and since the mammographic images of mass lesions are very different from the ImageNet images, it is proposed to adapt the models to adjust the characteristics of the last convolutional blocks and make them more specific to the data. We adjust the weights of a pre-trained network using the new set of images modifying the last fully connected layers and using a lower learning rate in the rest of the layers.

The work of [22] shows that knowledge can be transferred to networks that were previously trained on a domain for which a large amount of data existed, and then retrained the network in a different domain through fine tuning. Training a network from scratch with a small data set is a complicated task. The current paradigm in these cases, like the one we face in this investigation, is to use a pre-trained network.

The CNN used is composed of 5 convolutional layers, each followed by a batch normalization layer, 2 grouping layers and 11 Inception modules. The Inception module is presented in the Figure 2. The first convolutional layer applies the convolution operation to the input image, which has a size of 299x299x3 pixels.

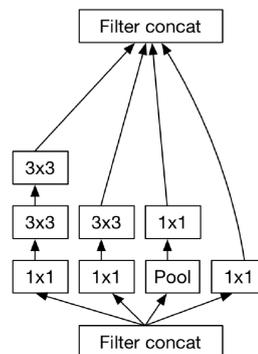


Figure 2. Inception module used by pretrained CNN.

In order to reduce overfitting, the dropout technique was used. A probability of 0.5 was used to select neurons that do not participate in training; so that a fully connected network is reduced in training at each stage. In the next stage, the removed neurons are reinserted into the network with their original weights. By avoiding training all the neurons in each of the stages, the training speed is significantly improved. Finally, the softmax function in the semi-final layer of the classifier was used to train the network under a cross-entropy regime, which gives a non-linear variant of multinomial logistic regression.

Three CNNs were created, based on the Inception v3 architecture. The original architecture of the network was maintained up to the fully connected layers, which was obtained for the ImageNet data set with 1000 outputs for 1000 class categories. Therefore, the last fully connected layers are removed; in this way the number of classes in the network output is modified. Thus, three models are built, the first one that uses three classes as output (normal, malignant and benign). The second uses only two, normal and anomalous classes. In the third model, two output classes are used as well, these are malignant and benign. The last two models are used in series, in the first stage it is decided whether or not there is an image lesion, and in the second stage the lesion is classified as malignant or benign.

In the process of fine tuning of the network the learning rate in the early stages remained small, while in the last layers the value of the learning rate was set high. In this way, the coefficients of the filters are not modified too much in the first stages which have a general knowledge in the recognition of images not being so in the last stage. In order to improve the results and avoid overfitting, data augmentation, L2 regularization, batch normalization and dropout are used. The regularization L2 guarantees the reduction of the weights. That is, it penalizes large weights by preferring smaller ones. As was explained previously, a dropout layer is used to randomly deactivate training time activations with a probability of 0.5. The subset of randomly selected activations is set to zero, which prevents some units in one layer from relying too much on a single unit in the previous layer. An NVIDIA GeForce® GTX1060 6GB graphics card with 6.1 calculation capacity was used.

2.5 Network Performance Evaluation

The performance indices to measure the effectiveness of the networks used were: accuracy, sensitivity, specificity and the area under the ROC curve (AUC). Sensitivity, specificity and AUC are only used in binary classification systems, since, in systems with more than two classes to classify, one class cannot be determined as positive and another as negative. Accuracy, calculated as the number of hits on the total of images to be classified, is used as a measure of system reliability for the architecture with three classes. Accuracy, sensitivity, specificity and AUC for binary classification systems are calculated according to [23].

3 Results and Discussion

3.1 Results for Architecture with Three Classes

Table 1 shows the composition of the test set for the 3-class architecture. The curve of training and loss for this network is presented in Figure 3. The Acc in the training phase for the training set was near to 95% and for validation set was 92.6%. However, the network archived an Acc in the test set of 86.05%. This fact occurs because in training phase the network learns the characteristics of the images and the hyperparameters are modified to reach a high Acc value.

The composition of the test set used to validate our proposed model is presented in Table 1. On the other hand, Table 2 shown the errors reaches by the network. It is observed that dense glandular tissue was difficult to correctly classify. This result is similar to those achieved by the human radiologist since this type of tissue is very difficult to identify. The CNN misclassify 6 adipose tissue, 9 glandular tissue and 15 dense glandular tissue.

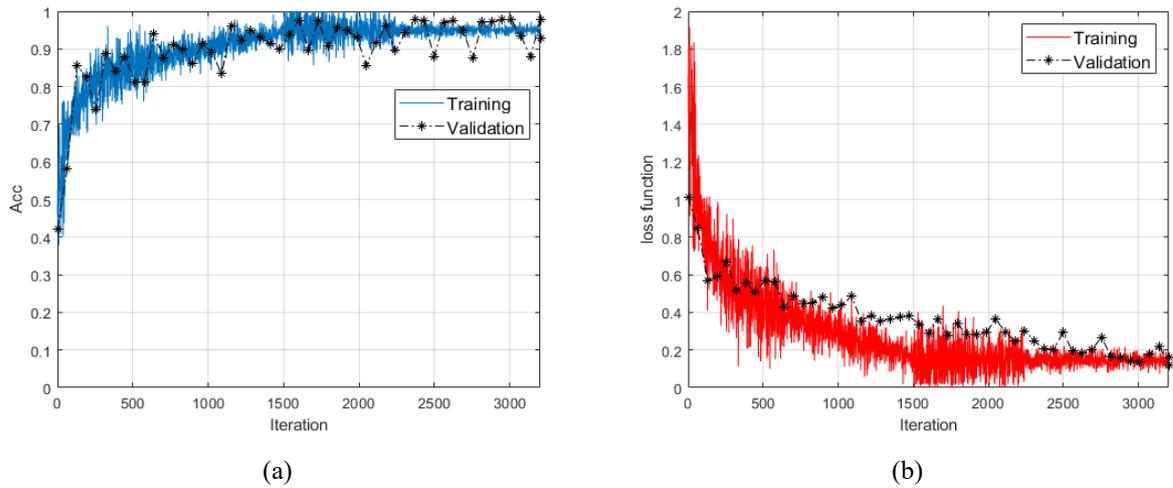


Figure 3. Training and validation curve for 3-class architecture, in (a) using Acc and in (b) using loss function.

Table 1. Composition of the test set that used to classify images with 3 classes according to the breast tissue.

Classes	Adipose	Glandular	Dense Glandular
Normal	54	53	62
Benign Lesion	6	15	10
Malignant Lesion	9	6	0

Table 2. Misclassification of the CNN with three outputs, normal, malignant and benign.

Classes	Adipose	Glandular	Dense Glandular
Normal-Benign	0	1	7
Normal-Malignant	0	2	3
Benign-Normal	1	2	2
Benign-Malignant	0	4	3
Malignant-Normal	3	0	0
Malignant-Benign	2	0	0

3.2 Results for Architecture with Two CNNs in Series

In the first part of the architecture with two classes, it is determined whether or not the image is considered a lesion. The curve of training and validation is presented in Figure 4. It is observed that the network reach almost a 100% of Acc in training phase but in the validation set the Acc is 95.8%. In this stage, and using testing set an Acc of 96.65%, a sensitivity of 88.6%, a specificity of 98% and an AUC of 93.3% in the detection of lesions in a total of 138 testing set of images were obtained. The implemented system diagnosed 2 false positives and 4 false negatives. Once the network is trained, the execution time in the classification is approximately 0.12 seconds per image.

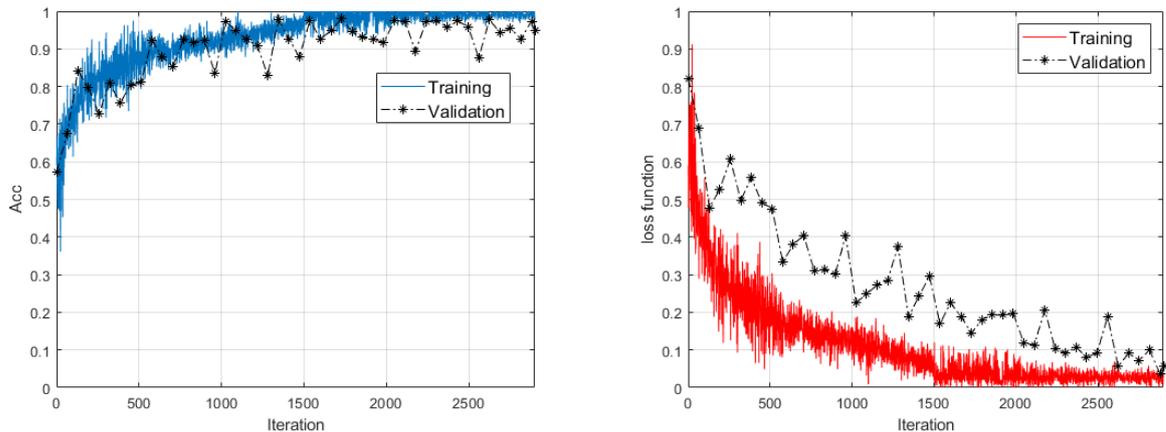


Figure 4. Acc and loss function curve for the CNN of the first stage with two class (normal vs abnormal).

The composition of the test set according to the type of breast tissue for the normal and abnormal classes used is presented in Table 3. In contrast, Table 4 shows the mistakes made by CNN according to the type of tissue. It is observed that the system made errors in 2 images of adipose tissue, 2 of glandular tissue and 2 of dense glandular tissue. A parity is observed in the amount of errors made between the different breast tissues. These results suggest that the system is independent of the type of tissue.

Table 3. Composition of the test set that divides the images into 2 classes according to the breast tissue.

Classes	Adipose	Glandular	Dense Glandular
Normal	32	33	38
Anomalous	9	12	14

Table 4. Misclassification of CNN in normal or anomalous classes for various types of breasts.

Classes	Adipose	Glandular	Dense Glandular
False Positives (FP)	1	1	0
False Negatives (FN)	1	1	2

In the second stage with two classes, it is determined whether the lesion found is malignant or benign. The curve of training and validation is depicted in Figure 5, in (a) using Acc and in (b) loss function. The value of Acc for training set is near to 99% and for validation test was 96.7%.

The network reached an accuracy of 91.3% in testing set. This was mainly due to the lower number of images in the training set, therefore, features present in the testing set were not learned by the network during the training. The system in this second part diagnosed 3 false positives and 1 false negative. The sensitivity was 93.33%, the specificity of 90.32% and the AUC of 91.82% for a total of 46 images. The computation time was below of 0.1 seconds per image classification.

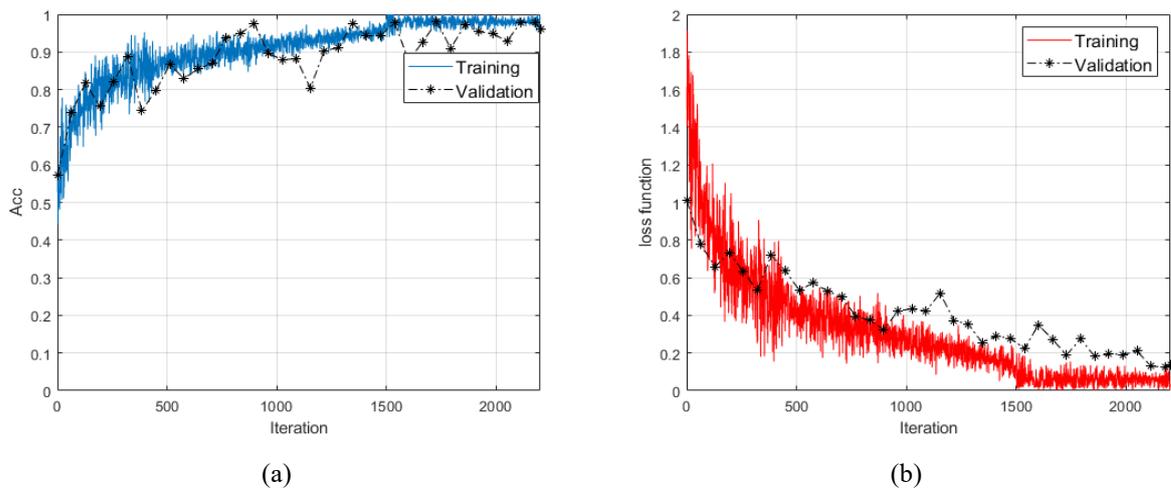


Figure 5. Curve of training phase for the classification in malignant and benign, in (a) using Acc and in (b) loss function.

Table 5 shows for the test set for the malignant and benign classes, the distribution of the images according to the type of tissue in the breast. Table 6 shows the errors made by CNN according to the type of breast tissue. It is observed that no mistake was made in the images of glandular tissue. The same number of mistakes were made in fat images, which are easier to diagnose by radiologists. In the case of images with glandular tissue, no mistakes were made. Finally, in the images of dense glandular tissues, considered the most difficult to classify by radiologists, two mistakes were made. In this sense, it can be argued that the system is helpful to radiologists as a second opinion and that it responds equally to various types of glandular tissue.

Table 5. Composition of the test set for the second stage of CNN according to the type of breast tissue.

Classes	Adipose	Glandular	Dense Glandular
Benign	6	15	10
Malign	9	6	0

Table 6. Misclassification of CNN in malignant or benign classes for various types of breasts

Classes	Adipose	Glandular	Dense Glandular
False Positives (FP)	1	0	2
False Negatives (FN)	1	0	0

The accuracy of the two parts is obtained as the probability that each one has to succeed in the classification. It is calculated as the multiplication of both probabilities. Therefore, the accuracy of the system is 88.24%, slightly higher than the classification system with three classes.

3.3 Permissible Values for Radiologists in the Identification of Lesions

According to the study conducted in [25], which analyzes the level of normal performance of radiologists to detect breast lesions from viewing mammography images, it is more common for these to diagnose false negatives than false positives, as shown in Figure 6. The results obtained in the investigation compared to the values presented in said figure are observed to have a behavior within the acceptable range for radiologists. This demonstrates the reliability of the system to implement it in practice and help, as a second opinion for radiologists.

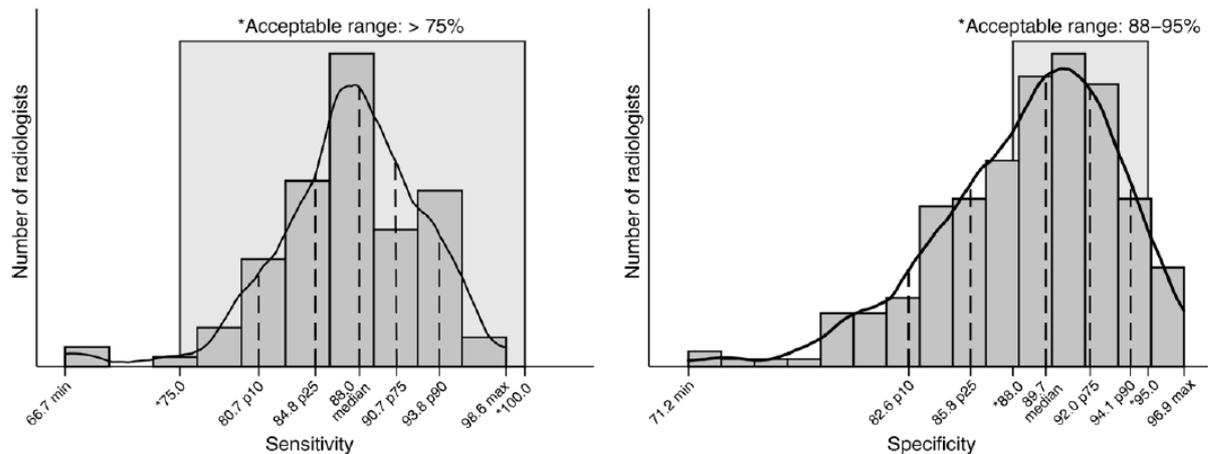


Figure 6. Performance of radiologists according to Sensibility and Specificity.

3.4 Comparison of Results with other Works and Novelty

In [26] three different machine learning algorithms with 161 images from the MIAS database were used to classify masses according to their severity, using 5 different pixel size values and 5 text analysis methods. AUC values between 40 and 78% are obtained; clearly surpassed by the networks implemented in the present investigation. Something similar happened when comparing the results achieved in the work of [27] where 3 variants were used to extract the features of the images. The highest accuracy value achieved by the classifiers was 66%. The advantage of deep learning methods, specifically the CNN used in this research to classify breast images, is observed. This is also detected in the work of [28] where the performance of the classic computer vision approach against CNNs was evaluated. In both cases, the present investigation exceeded the reported results that were for the classical approach of 82.5% and for CNN of 85.8% accuracy.

In [14] with another type of neural network and the same BD, an accuracy of 98.23% and an AUC of 99% is obtained, surpassing the work of the present investigation. However, the present work identifies any type of injury and diagnoses it, while in [14] it only classifies according to the severity of the masses. In that same work, the MIAS image database is only used for validation, the training was carried out with a total of 3 databases, using a total of 6116 images, which gives the network greater power of generalization.

The present work has combined efficiently techniques previously used separately to solve various problems, in particular the pre-trained inceptions modules using learning transfer techniques, which has facilitated that the results are independent of the type of tissue, which is not usual in practice.

It should also be emphasized as a useful aspect, that ROI image analysis has been used instead of full image, which has led to a gain in computing efficiency.

4 Conclusions

This study addressed the problem of breast cancer classification based on digital mammography images. Different pre-trained convolution neural network architectures were adjusted for the classification of these images applying the transfer learning technique. The system implemented with two binary classification architectures was the most efficient and also represents an improvement in performance compared to previous work carried out in the national and international territory. The two systems implemented classify independently of the present breast tissue, which is very useful for radiologists who have great difficulty diagnosing especially in dense glandular breasts. The results of the classification with both neural networks are within the ranges of performance accepted by experienced radiologists, which gives the system reliability at the time of its implementation in routine clinical practice. The image processing time in all cases was less than 1 second. This aspect is somewhat desired in clinical settings. Furthermore, providing radiologists with a tool for the early diagnosis of breast cancer. It is proposed as future work to train the networks implemented with other mammography image databases with a greater number of cases that provide greater power of generalization to the networks and thus acquire greater reliability.

References

- [1] F. J. Frutos Arenas, M. Seguí Azpilcueta, J. A. Nieto Parra, C. Segovia Vergel, A. Palma Gastón, and E. Olloqui Martín, “Técnicas de diagnóstico por la imagen en cáncer de mama,” *Cirugía Andaluza*, vol. 23, no. 1–2, p. 18, 2012.
- [2] V. Rodríguez López, “Análisis de imágenes de mamografía para la detección de cáncer de mama,” *Temas de Ciencia y Tecnología*, vol. 15, no. 47, pp. 39–45, 2012.
- [3] M. Perez-Diaz, “Techniques to evaluate the quality of medical images,” *AIP Conference Proceedings*, vol. 1626, no. 1, pp. 39–45, Nov. 2014, doi: 10.1063/1.4901358.
- [4] E. L. Henriksen, J. F. Carlsen, I. M. Vejborg, M. B. Nielsen, and C. A. Lauridsen, “The efficacy of using computer-aided detection (CAD) for detection of breast cancer in mammography screening: a systematic review,” *Acta Radiol*, vol. 60, no. 1, pp. 13–18, Jan. 2019, doi: 10.1177/0284185118770917.
- [5] M. L. Giger, “Machine Learning in Medical Imaging,” *Journal of the American College of Radiology*, vol. 15, no. 3, pp. 512–520, Mar. 2018, doi: 10.1016/j.jacr.2017.12.028.
- [6] S. Srivastava, N. Sharma, S. K. Singh, and R. Srivastava, “Design, analysis and classifier evaluation for a CAD tool for breast cancer detection from digital mammograms,” *International Journal of Biomedical Engineering and Technology*, vol. 13, no. 3, pp. 270–300, Jan. 2013, doi: 10.1504/IJBET.2013.058447.
- [7] M. George and R. Zwigelaar, “Breast Tissue Classification Using Local Binary Pattern Variants: A Comparative Study,” in *Medical Image Understanding and Analysis*, 2018, pp. 143–152, doi: 10.1007/978-3-319-95921-4_15.
- [8] A. Krizhevsky, I. Sutskever, and G. E. Hinton, “ImageNet Classification with Deep Convolutional Neural Networks,” in *Advances in Neural Information Processing Systems 25*, F. Pereira, C. J. C. Burges, L. Bottou, and K. Q. Weinberger, Eds. Curran Associates, Inc., 2012, pp. 1097–1105.
- [9] Y. Gao, K. J. Geras, A. A. Lewin, and L. Moy, “New Frontiers: An Update on Computer-Aided Diagnosis for Breast Imaging in the Age of Artificial Intelligence,” *American Journal of Roentgenology*, vol. 212, no. 2, pp. 300–307, Jan. 2019, doi: 10.2214/AJR.18.20392.
- [10] E.-K. Kim *et al.*, “Applying Data-driven Imaging Biomarker in Mammography for Breast Cancer Screening: Preliminary Study,” *Scientific Reports*, vol. 8, no. 1, p. 2762, Feb. 2018, doi: 10.1038/s41598-018-21215-1.
- [11] N. E. Benzebouchi, N. Azizi, and K. Ayadi, “A Computer-Aided Diagnosis System for Breast Cancer Using Deep Convolutional Neural Networks,” in *Computational Intelligence in Data Mining*, Springer, Singapore, 2019, pp. 583–593.
- [12] D. Ribli, A. Horváth, Z. Unger, P. Pollner, and I. Csabai, “Detecting and classifying lesions in mammograms with Deep Learning,” *Scientific Reports*, vol. 8, no. 1, p. 4165, Mar. 2018, doi: 10.1038/s41598-018-22437-z.
- [13] W. Lotter, G. Sorensen, and D. Cox, “A Multi-scale CNN and Curriculum Learning Strategy for Mammogram Classification,” in *Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support*, Springer, Cham, 2017, pp. 169–177.
- [14] H. Chougrad, H. Zouaki, and O. Alheyane, “Deep Convolutional Neural Networks for breast cancer screening,” *Computer Methods and Programs in Biomedicine*, vol. 157, pp. 19–30, Apr. 2018, doi: 10.1016/j.cmpb.2018.01.011.
- [15] M. Oquab, L. Bottou, I. Laptev, and J. Sivic, “Learning and Transferring Mid-Level Image Representations using Convolutional Neural Networks,” presented at the Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2014, pp. 1717–1724.
- [16] O. Russakovsky *et al.*, “ImageNet Large Scale Visual Recognition Challenge,” *Int J Comput Vis*, vol. 115, no. 3, pp. 211–252, Dec. 2015, doi: 10.1007/s11263-015-0816-y.
- [17] K. Chatfield, K. Simonyan, A. Vedaldi, and A. Zisserman, “Return of the Devil in the Details: Delving Deep into Convolutional Nets,” *arXiv.org*, May 2014.
- [18] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna, “Rethinking the Inception Architecture for Computer Vision,” presented at the Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016, pp. 2818–2826.
- [19] P. Suckling J, “The mammographic image analysis society digital mammogram database,” *Digital Mammo*, pp. 375–386, 1994.
- [20] X. Zhao, X. Wang, and H. Wang, “Classification of Benign and Malignant Breast Mass in Digital Mammograms with Convolutional Neural Networks,” in *Proceedings of the 2Nd International Symposium on Image Computing and Digital Medicine*, New York, NY, USA, 2018, pp. 47–50, doi: 10.1145/3285996.3286006.

- [21] S. C. Wong, A. Gatt, V. Stamatescu, and M. D. McDonnell, "Understanding Data Augmentation for Classification: When to Warp?," in *2016 International Conference on Digital Image Computing: Techniques and Applications (DICTA)*, Gold Coast, Australia, 2016, pp. 1–6, doi: 10.1109/DICTA.2016.7797091.
- [22] Y. Bengio, "Deep learning of representations for unsupervised and transfer learning," in *Proceedings of ICML Workshop on Unsupervised and Transfer Learning*, 2012, pp. 17–36.
- [23] F. Idrees, M. Rajarajan, M. Conti, T. M. Chen, and Y. Rahulamathavan, "PIndroid: A novel Android malware detection system using ensemble learning methods," *Computers & Security*, vol. 68, pp. 36–46, Jul. 2017, doi: 10.1016/j.cose.2017.03.011.
- [24] M. A. Al-antari *et al.*, "An Automatic Computer-Aided Diagnosis System for Breast Cancer in Digital Mammograms via Deep Belief Network," *J. Med. Biol. Eng.*, vol. 38, no. 3, pp. 443–456, Jun. 2018, doi: 10.1007/s40846-017-0321-6.
- [25] C. D. Lehman *et al.*, "National performance benchmarks for modern screening digital mammography: update from the Breast Cancer Surveillance Consortium," *Radiology*, vol. 283, no. 1, pp. 49–58, 2016.
- [26] M. Abdel-Nasser, J. Melendez, A. Moreno, O. A. Omer, and D. Puig, "Breast tumor classification in ultrasound images using texture analysis and super-resolution methods," *Engineering Applications of Artificial Intelligence*, vol. 59, pp. 84–92, 2017, doi: <http://dx.doi.org/10.1016/j.engappai.2016.12.019>.
- [27] B. Bektaş, İ. E. Emre, E. Kartal, and S. Gulsecen, "Classification of Mammography Images by Machine Learning Techniques," in *2018 3rd International Conference on Computer Science and Engineering (UBMK)*, 2018, pp. 580–585, doi: 10.1109/UBMK.2018.8566380.
- [28] J. Kurek, B. Świdorski, S. Osowski, M. Kruk, and W. Barhoumi, "Deep learning versus classical neural approach to mammogram recognition," *Polish Academy of Sciences*, vol. 66, no. 6, 2018, doi: 10.24425/bpas.2018.125930.