Reporte Técnico RT-C09-001

Desarrollado por. César Augusto Ramírez Gámez
Investigador asociado
Universidad Autónoma de San Luis Potosí
Facultad de Ingeniería
Área de Ciencias de la Computación

Proyecto 6005:
“Sistema automático para la interpretación inmediata de mamografías para la determinación del riesgo de cáncer”.
Autorizado por Consejo Nacional de Ciencia y Tecnología-Programas nacionales estratégicos (CONACYT-PRONACES)

Instituciones participantes: CIC-Instituto Politécnico Nacional, UASLP, Hospital Central “Dr. Ignacio Morones Prieto”.

Reporte – Algoritmos de preprocesamiento y detección para masas

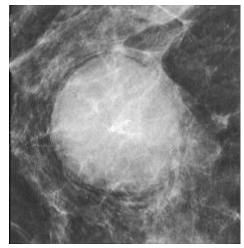
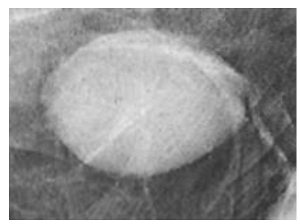
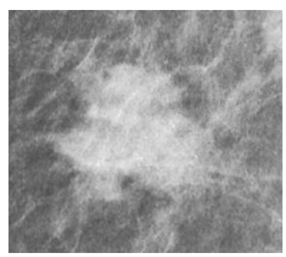
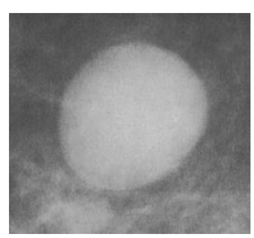
Versión 1. 19 de marzo del 2022

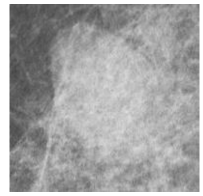
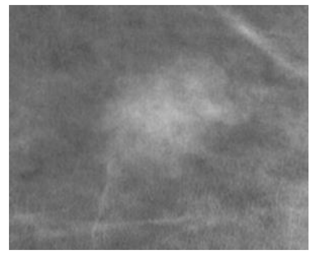
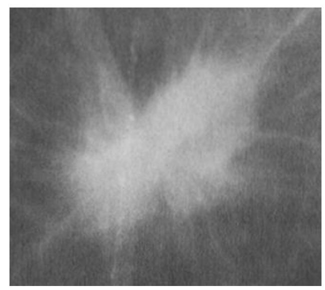
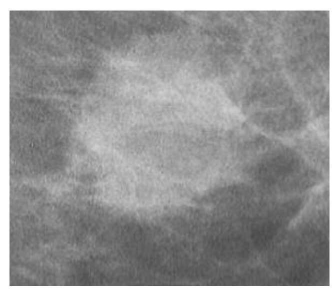
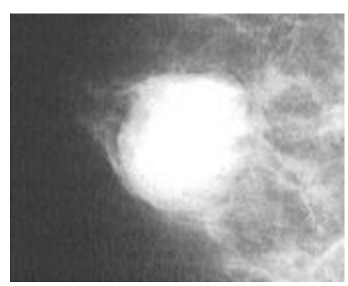
Resumen
Dentro de las posibles lesiones que es posible encontrar en la mamografía se encuentran las masas. Este documento tiene el propósito de describir las características que permiten identificar la presencia de masas en la mamografía para después mostrar los algoritmos que se han probado con la finalidad de detectar masas en la mamografías.

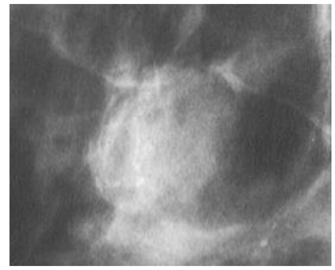
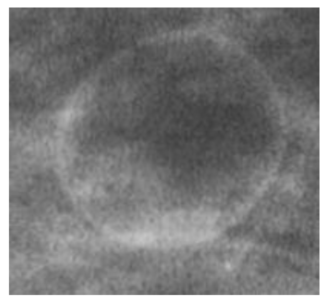
Contenido

Characterization of the masses

when it is being worked with mammography, a mass is defined as three-dimensional space-occupying lesions in the breasts and can be described by three features: shape, margin, and density. In order to verify the mass is three-dimensional, expert physicians have to locate the mass in craniocaudal and mediolateral oblique projections. Masses features are margin, shape, and density. Values for each feature are sowed in the next table base on BI-RADS 5ed [1]:

Feature	Value	Example
Forma	Round. Round masses are spherical, circular, or globular	
	Oval. Oval masses are elliptical or ovoid in shape (they may have two or three undulations).	
	Irregular. The mass is neither round nor oval. In mammography, this descriptive term usually represents a finding suspicious.	
Margen	Circumscribed. The margin is sharply outlined and an abrupt transition between the lesion and the surrounding tissue is seen. In mammography, if a part of the margin is obscured, at least 75% must be well defined so that the lesion is considered circumscribed. If any portion of the margin is indefinite, microlobulated, or spiculated, the classification should be based on this last characteristic (the most suspicious component).	

	Obscured. Any margin that is hidden behind adjacent or overlying fibroglandular tissue is considered obscured. This term is mainly used when part of the margin of the nodule is circumscribed but the rest (> 25%) is hidden	
	Microlobulated. The microlobulated margin presents short cycle undulations. In mammography, this descriptive term often represents a suspicious finding.	
	Spiculated. The spiculated margin has lines radiating from the lesion to the periphery. This descriptive term usually represents a suspicious finding.	
	Indistinct. The indistinct margin does not present a sharp delineation from the surrounding tissue, either in its entirety or in any portion. In mammography, the radiologist should not use this descriptive term -since it usually represents a suspicious finding - when you consider that the appearance is a product of the adjacent breast tissue.	
Densidad	Hyperdense.The mass has a greater degree of attenuation than expected given an equal volume of breast tissue fibroglandular.	

	Isodense. The mass has the same degree of attenuation as expected given an equal volume of breast tissue fibroglandular.	
	Hypodense. The mass has a lower degree of attenuation than expected given an equal volume of breast tissue fibroglandular. Hypodense nodules may consist of a group of microcysts.	
	Fat content. The nodules that have a density of adipose content include all those lesions that contain fat, such as oily cysts, lipomas, and galactoceles, and also those of mixed density, such as hamartomas. Fat-containing lesions are almost always benign.	

To corroborate the masses existence, following points must be taken into coun:

- The mass should be visible on both the CC and MLO views. If it is only seen in one of the two projections, it can be assumed that it was superimposed tissue and not a body with volume, such as a mass[2].
- When looking for a mass in both projections, its location must be considered, for which the distance from the nipple to the suspected mass can be used, which does not necessarily have to be the same, it can vary a little[2].
- Breasts normally maintain a high level of symmetry between them, so an element suspected of being a mass could be something that is in one of the two breasts, but not in both.
- No reference has been found that indicates the maximum and minimum size of a mass, but it can vary a lot

Mathematically, the values of that three characteristics give us 60 possible combinations, and each combination should belong to a BI-RADS assessment, but not all the combinations are physically possible. According to the qualitative analysis and physician opinion, just 31 cases can be possible. Table 2 shows the possible combinations with their BI-RADS assessment value. It can be observed that there is no combination for BI-RADS 0, 1, 4a, 5, and 6. The reasons are:

- • BI-RADS 0: This assessment should be used in case the expert physician has doubts.
- • BI-RADS 1: This evaluation should be used when there is no lesion found.

- • BI-RADS 4a: No masses features combination result in a 4a.
- • BI-RADS 5: This category needs to identify calcifications inside the mass.

Shape	Margin	Density	BI-RADS assessment
Irregular	Circumscribed	Fat-content	2
Irregular	Obscured	Fat-content	2
Irregular	Microlobulated	Hypodense	2
Oval	Circumscribed	Fat-content	2
Oval	Obscured	Fat-content	2
Oval	Circumscribed	Hyperdense	2
Oval	Circumscribed	Hypodense	2
Oval	Microlobulated	Hypodense	2
Round	Circumscribed	Fat-content	2
Round	Obscured	Hypodense	2
Round	Obscured	Fat-content	2
Round	Circumscribed	Hyperdense	2
Round	Circumscribed	Isodense	2
Round	Circumscribed	Hypodense	2
Oval	Circumscribed	Isodense	3
Oval	Obscured	Isodense	3
Oval	Obscured	Hypodense	3
Oval	Microlobulated	Hyperdense	3
Irregular	Circumscribed	Hyperdense	4b
Irregular	Spiculated	Isodense	4b
Irregular	Obscured	Hyperdense	4b
Irregular	Microlobulated	Hyperdense	4b
Oval	Obscured	Hyperdense	4b
Oval	Indistinct.	Hyperdense	4b
Oval	Indistinct.	Isodense	4b
Oval	Spiculated	Isodense	4b
Irregular	Indistinct.	Hyperdense	4c
Irregular	Spiculated	Hyperdense	4c
Oval	Spiculated	Hyperdense	4c
Round	Spiculated	Hyperdense	4c

Table 2. Possible values for masses features combination.

Methodology

This report focuses on a series of experiments carried out to translate the medical information of the masses and their characteristics into a computational situation that allows them to be detected inside the breast.

The original format for mammograms is an international standard for medical images and related information called DICOM® (Digital Imaging and Communications in Medicine). In the mammograms case, the DICOM® file contains two elements, a metadata set of the patient, type of study, institution, and devices used feature, and the second element is the gray-scale image or projection. The image inside is big, it can exceed 3500 x 2500 pixels in size and is very deep, 12 or 14 bits, while a normal image has 8 bits. Bits number means the range of values that can be represented by a pixel, in 8 bits each pixel represents a grayscale between 0 and 255 bits, in 12 bits a range between 0 and 4094, and 14 bits a range between 0 and 16383.

To detect a mass or masses, due to the large information amount, is necessary to pass for a preprocessing stage to make easier the detection process. For that reason, in this work, two kinds of experiments are presented, preprocessing and detection.

Preprocessing experiments

In preprocessing stage the objective is to remove all not relevant information to recognize masses, preserving deepness and without reducing the original image size, taking into consideration that the entrance image is in DICOM format. Figure 13 shows the steps and algorithms that are being tested to reach the mentioned objectives. The algorithms and experiments have been done over our dataset.

First, the Gray Level Slicing algorithm was used. Gray Level Slicing algorithm specifies a range of gray levels desired in an image [3]. In order to find a gray-level range where masses exist and apply a slicing to remove irrelevant information, Gray Level Slicing algorithm was modified. Experiments were done first on axillary lymph nodes, but reproducible to masses. The algorithm contains the next steps:

- From our dataset and supported by the labels, a bounding box was generated starting from the extreme upper left and extreme lower right coordinates. Fifty pixels in each direction were added to have borders and the axillary lymph node centered to create a final ROI.
- As part of the data science ap
- plication, from each ROI the pixel values belonging to the axillary lymph node were extracted, organized, and analyzed.
- From each ROI the most representative values (MRV) at the top of the histogram were extracted, organized, and analyzed.
- From the axillary lymph node values, and the fifty more representative values maximum, minimum, and average values were obtained and compared with the maximum values the maximum value in each mammogram,
- Maximum and minimum value gives a gray-level range of values to conserve, and remove the rests. shows a mammogram sliced in gray-levels values.

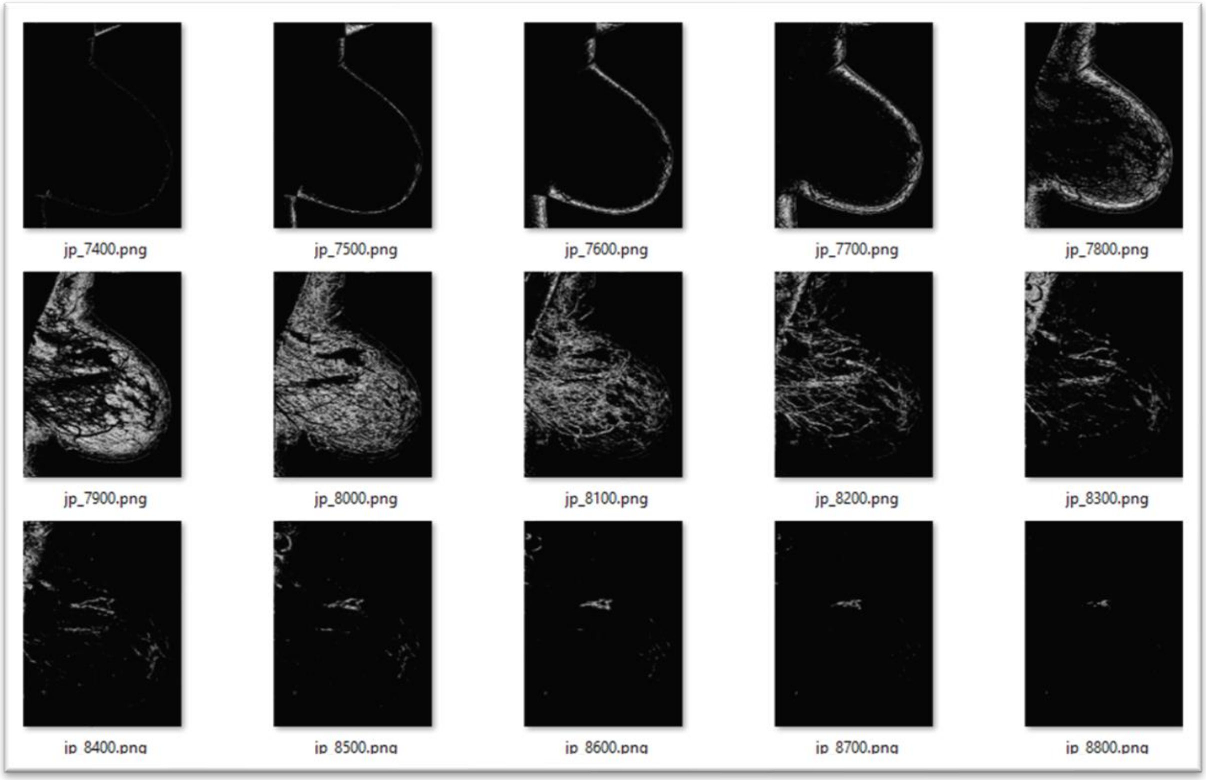


Figure 13. Gray Level Slicin algorithm. Ranges where relevant information can be located.

In this algorithm, data science was applied to find the range values where an axillary lymph node can be found. First experiments were done working with axillary lymph node values, due to they have fewer variants and are easy to be located visually on a mammogram and few have been done with masses.

The next algorithm tested was bit plane slicing. An image can be decomposed into bits planes deactivating(converting to zero) all bis in each pixel, and just letting one active(keep in one). The experiment was focused in show each plane from a set of mammography. Figure 14 shows the information that remains after hiding every bit and left active just one. In a conclusion, bits from 8 to 14 contain the relevant information.

Other algorithms in the literature were tasted. CLAHE improves the contrast, nevertheless, the information is modified and values become bigger, therefore, the image gets the16 bits of deepness.

A median filter does not prove to be useful when the mammogram preservers the size and deepness.

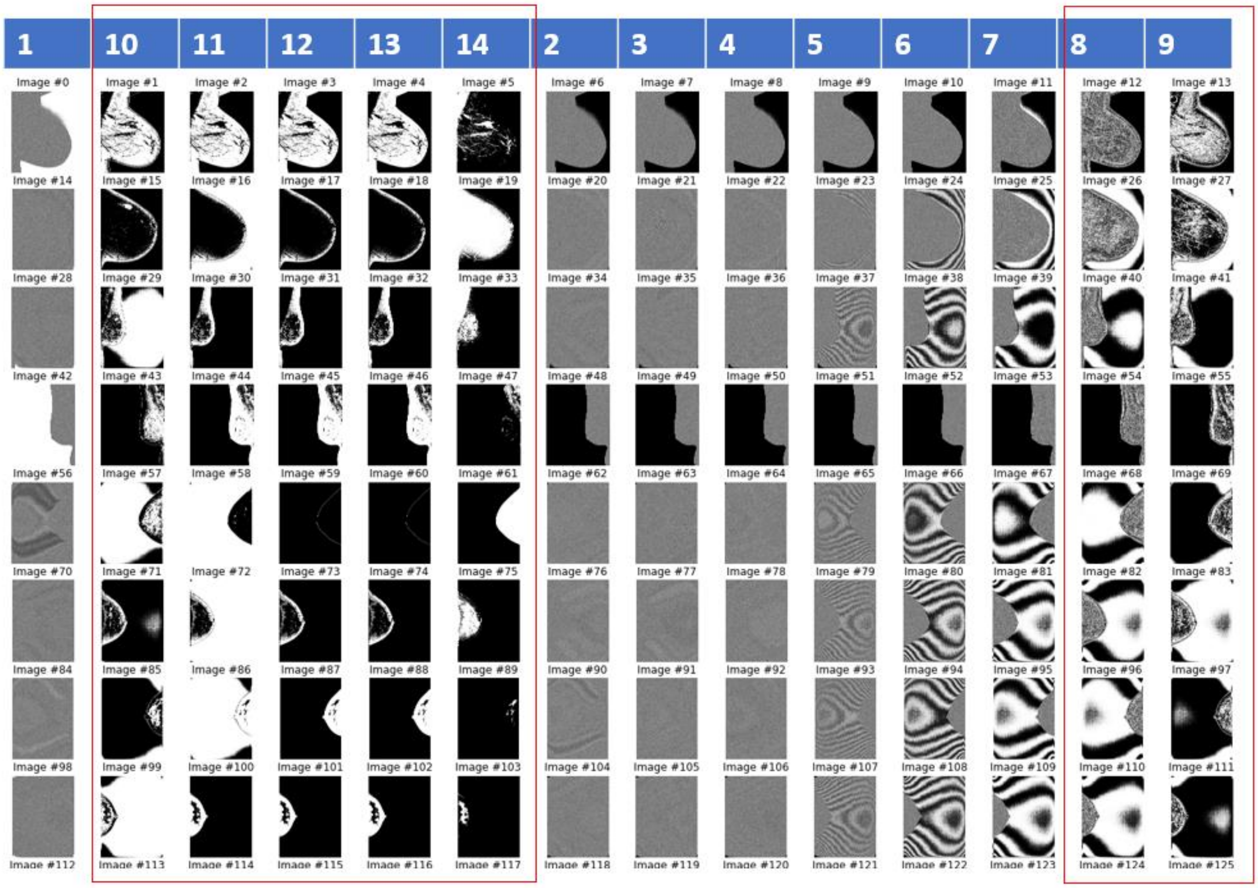


Figura 14. Bit plane slicing in mammograms

Masses recognition experiments

This stage of experiments pretends to process the preprocessed mammogram, with the original size and deepness but just the relevant information, and give a set of reference points pointing out the center of possible masses. The method chosen for the recognition is the correlation filters, supported by data science and an additional step to filter the points of reference located. Figure 13 shows the steps mentioned. In this section, each step will be described.

A correlation filter is a technique that tries to capture how similar or different a test object is from training objects [4]. The basic idea is to have a reference image and a test image. The reference image superpose over the test image at the upper left corner and the two images are multiplied (pixel-wise), and the values obtained are summed to obtain the correlation value between both images in that location. The process has to be repeated shifting the reference image for all the test images to obtain the rest correlation values. Larger correlation values will be in locations where the reference image is similar to the object in the image. Figure 15 shows an example of a correlation filter. Correlation filters have the next advantages:

- It is an integrative operation, where the result does not depend on a single pixel, thus, when the object in the test image is not completely equal, the correlation value suffers a degradation, but does not disappear.

- If the test image contains the reference object at a shifted location, the correlation output is also shifted by the same amount.
- Computing correlations via the Fourier domain is more efficient than their direct computation.

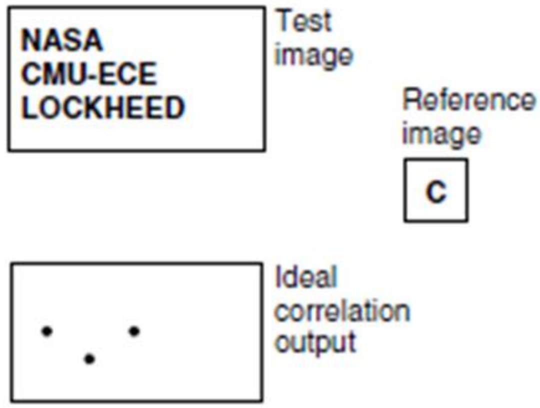


Figure 15. Correlation filter

Due to the advantages to apply correlation in the Fourier domain, experiments have been carried out with the Fourier transform applied to mammograms. To clarify this topic, an image is processed in the spatial domain when is operated over the plane containing the pixels. Manipulating an image in the frequency domain involves transforming the image using mathematical procedures to separate the image into high-frequency components(edges) and low-frequency components (smooth regions). Fourier transform is one mathematical procedure applied to images to manipulate them in the frequency domain(Figure 16).

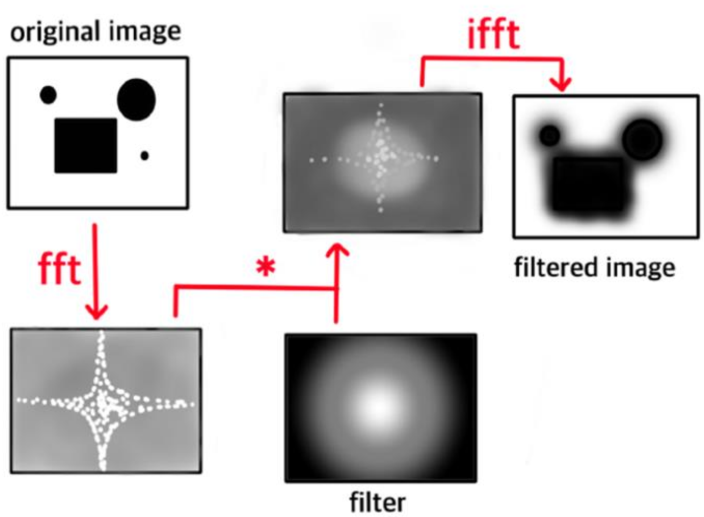


Figure 16. Fourier transform

Filtering in the frequency domain consists in modify the Fourier transform on an image and then obtaining the inverse transform to obtain the results 15. Since the Fourier transformation divides the image into low and high frequencies, a high band filter was applied to mammograms in order to show just the edges inside the breast, and try to highlight masses, results are shown in Figure 17. In c) and d) it is possible to see a black area inside the breast, that área belongs to a mass, nevertheless in a), b) and e) it is not possible to observe a significant signal to locate a mass. Experiments were also carried out with band-pass and low-pass filters, but without success in mass detection.

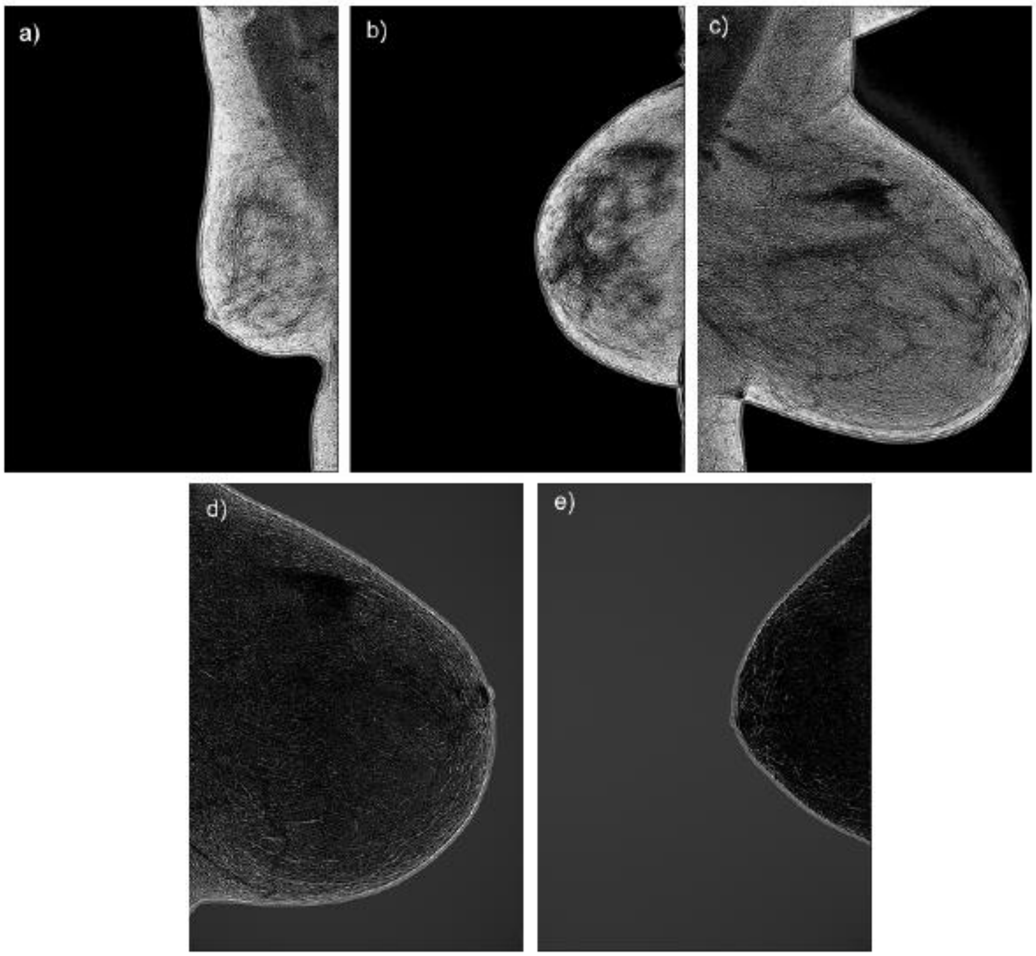


Figure 17: High band filter in Fourier transformation domain

The next experiments were focused on applying the correlation filters in the Fourier domain. These experiments were done using axillary lymph nodes first, for the same reasons explained in the previous section. The method developed to apply the correlation filter to an axillary lymph node is described below

- From each ROI the information surrounding the lesion was eliminated to create lymph node patterns.
- Patterns were grouped by the relation between the maximum value of MRV and the maximum value of the mammogram.
- Five patterns belonging to the same group were selected.

- Patterns selected were passed through an algorithm to create a correlation filter. • Correlation filter was applied to a mammogram without being preprocessed. Results are shown in Figure 18, where each small square over the mammogram means a high correlation value to indicate the possible mass existence, squares next to the mammogram are pattern examples, and down in the image is one pattern histogram.

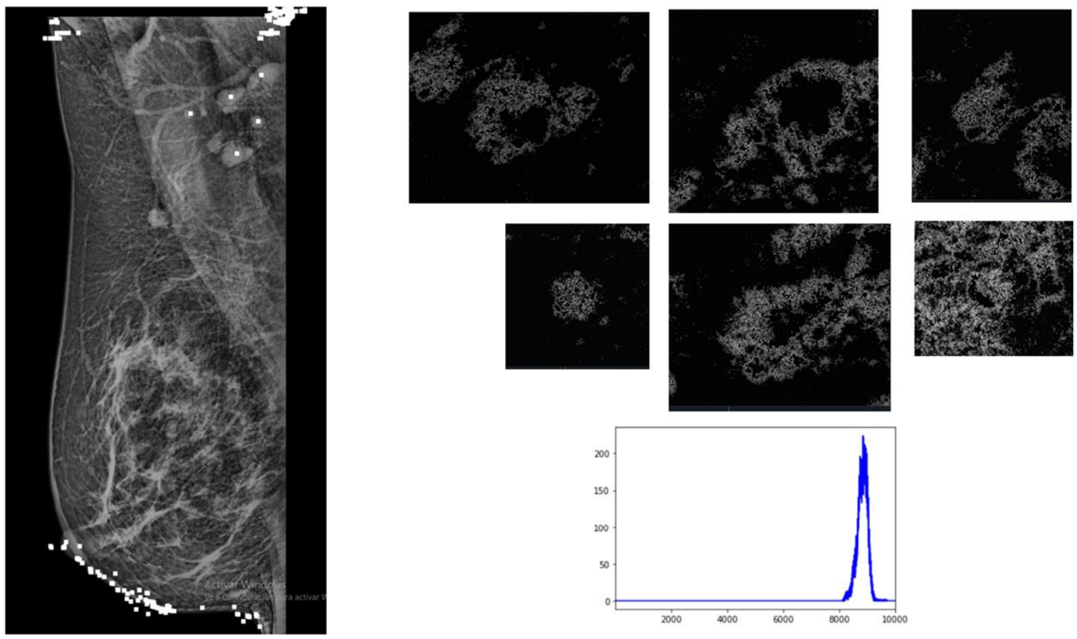


Figure 18: High band filter in Fourier transformation domain

In Figure 18 it is possible to observe multiple correlation values that are not on the lymph nodes, which can be called false positives.

To reduce the number of false positives a second experiment was carried out by preprocessing the previous image to apply the correlation filter. The result is shown in Figure 19. In this image few elements of the breast can be viewed, because the rest was removed based on the minimum value of the MRV, but the false positive number is considerably reduced.

More experiments are required in this stage. In the case of masses, 4 values are used for their density, for which, based on data science, correlation filters will be created for each density value. Filters can also be created at run time. As for the processing time, it has not yet become

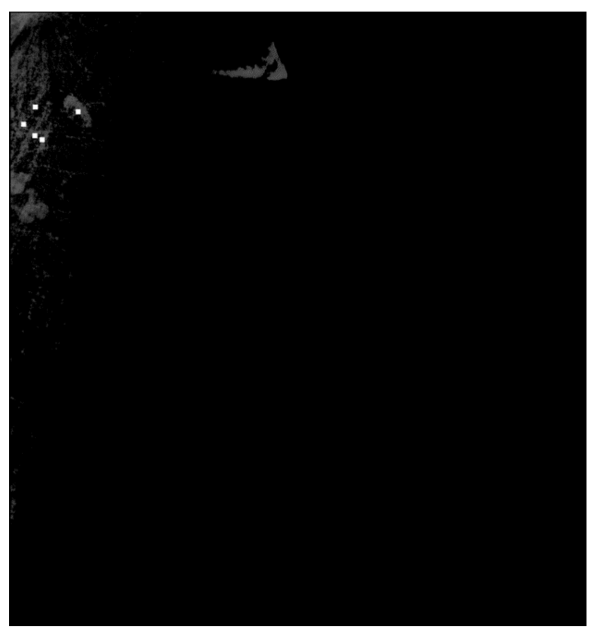


Figure 19: Correlation filter result over a preprocessed image

significant, but time measurements will begin to be taken in the experiments.

The next step for these experiments will be to apply them in masses. Currently, there is a set of masses already validated by expert physicians. The set was divided by BI-RADS categories in Table 3.

BI-RADS	
2	27
3	2
4a	0
4b	4
4c	15
Total	48

Table 3. Masses validated ditrbuted in BI-RADS category

Masses can be grouped according to their values in each feature. One objective of this research work is to apply at least a correlation filter for each density value. The first step to achieving the creation of the filters was to obtain the values contained in mass. The second step was to organize those values by a density value. And the third and final step is to obtain the máximums, minimus, and average of each density value. Table 3 shows this information.

	Most significative values of masses				Masses Values			
	Min	Max	Min Average	Max Average	Min	Max	Min Average	Max Average
Fat-containig	5474	8530	7581	7769	5229	9418	7303	8458
Hypodense	5755	8978	7996	8240	5420	10633	7685	8703
Isodense	6032	9191	7534	7664	5624	9485	7188	8417
Hyperdense	6258	9244	8204	8470	5617	9996	7755	9059

Table 3. Values for the density category

References

[1] Sickles EA et al. BI-RADS'" del ACR: Mamograf´ıa. En: Atlas BI-RADS'" del ACR, Sistema normalizado para la redacci´on de informes de estudios por im´agenes de la mama. Reston, VA, Colegio Estadounidense de Radiolog´ıa, 2013.

[2] Ikeda Debra M. and Miyake Kanae K. THE REQUISITES Breast Imaging. ELSEVIER,2017. isbn: 978-0-323-32904-0.

[3] Richard E.Woods Rafael C. Gonzalez. Digital Image Processing Second Edition. Prentice-Hall, 2001. isbn: 0-201-18075-8.

[4] Richard Juday B. V. K. Vijaya Kumar Abhijit Mahalanobis. Correlation Pattern Recognition. Cambridge University Press, 2005. isbn: 978-0-521-57103-6.