

**UNIVERSIDAD JUÁREZ AUTÓNOMA DE TABASCO  
DIVISIÓN ACADÉMICA DE INFORMÁTICA Y  
SISTEMAS**

---

---

**CREACIÓN DE UN BANCO DE DATOS DE  
MAMOGRAMAS PREPROCESADOS**

Trabajo recepcional bajo la modalidad de tesis  
que para obtener el grado de

**Licenciado en Sistemas Computacionales**

Presenta

**Omar Trinidad Gutiérrez Méndez**

Directores

**Dra. Juana Canul Reich  
M. en C. Guillermo de los Santos Torres**

Cunduacán, Tabasco(izquierda) —¿ Fecha(derecha)

# Contents

<b>1</b>	<b>Generalidades</b>	<b>3</b>
1.1	Antecedentes . . . . .	3
1.2	Problema . . . . .	4
1.2.1	Planteamiento del problema . . . . .	4
1.2.2	Delimitation of the investigation . . . . .	4
1.2.3	Research questions . . . . .	4
1.3	Objectives . . . . .	4
1.3.1	General objective . . . . .	4
1.3.2	Specific objectives . . . . .	5
1.4	Justification . . . . .	5
1.5	Methodology . . . . .	5
<b>2</b>	<b>Theoretical framework</b>	<b>6</b>
2.1	Frame of reference . . . . .	6
2.1.1	Bith depths . . . . .	6
2.2	Conceptual framework . . . . .	6
2.2.1	Mamogramas . . . . .	7

# Chapter 1

## Generalidades

### 1.1 Antecedentes

Breast cancer is a major cause of death in Tabasco and Mexico. It is a second cause of death for women aged 30-54. While this is a public health problem, a way to decrease the number of deaths is by improving the detection of cancer. The most common way to detect it is through the study of mammograms. This is not an easy task, there is a margin of error in the opinion of the radiologists.

In turn, it is possible to increase the success of diagnoses using CAD systems: Computer Aided-Detection (CADs) or Computer Aided-diagnosis (CADx), that kind of systems assist doctors in the interpretation of mammograms. CAD systems are built with artificial intelligence, digital image processing, and pattern recognition. An important step in building such systems is the creation of databases of medical images. Most of the mammographic databases are not publicly available, however, there are several databases of this type. There are two databases of mammograms of public domain: the USF (University of South Florida) database is a Digital Database for Screening Mammography (DDSM) [1] and the MIAS Mini-Mammographic Database (mini-MIAS) [2].

Both of them are widely used by the mammographic image analysis research community. Our main goal is to create a database similar to the above. Our project was supported by “Dr. Juan Graham Casusus” hospital who gave us a batch of raw mammograms in DICOM file format [3]. especificar, el problema DEL hospital –diagnóstico– segunda opinión, decir que no se ha implementado algo así, pero que existe

.... info on what to look at for image treatment

La introducción de [15] da una explicación buena de por qué son buenos los sistemas CAD.

## 1.2 Problema

### 1.2.1 Planteamiento del problema

It was necessary to discover which filters are the best choice in order to obtain optimized images ready for afterward phases such as segmentation, processing or analysis. Namely, it was necessary to reduce the noise and other artifacts in the image. Furthermore, we need to create an electronic file for each image. These data are obtained from each DICOM file and with the help of specialized doctors. The most important data is the medical mammogram diagnosis. Mammograms are classified according to the American College of Radiology (ACR) in BI-RADS. Mammography Atlas, edition 4th [4].

The assessment categories are: a. Mammographic assessment incomplete: 1. Category 0: Need additional imaging evaluation and/or prior mammograms for comparison. b. Mammographic assessment complete: 1. Category 1: Negative. 2. Category 2: Benign finding(s). 3. Category 3: Probably benign finding. 4. Category 4: Suspicious abnormality. 5. Category 5: Highly suggestive of malignancy. 6. Category 6: Proven malignancy.

### 1.2.2 Delimitation of the investigation

#### Scope

decir que hay un estado inicial y uno final la extensión del trabajo incluye — filtros y tipos de filtros aplicados, imágenes listas para etapas posteriores. hablar de base de datos más como un dataset que otra cosa — estructura: imagen cruda, procesada y ficha

a database of preprocessed mammograms, namely, mammograms without noise, reduction of artifacts and leveling of image quality.

#### Limitations

3. No se utilizarán clasificadores para evaluar el resultado, la evaluación será hecha principalmente por médicos, es decir, será una evaluación —.

### 1.2.3 Research questions

1. What is the best combination of filters?

## 1.3 Objectives

### 1.3.1 General objective

To create a data set with preprocessed mammograms with filters in the spatial or spectral transform domain.

### **1.3.2 Specific objectives**

To create an electronic file for each image. To store the mammograms in a database oriented to storage and retrieval of images. To create an online database of domain public. To find what are the more suitable filters in the treatment of the image.

## **1.4 Justification**

Our project is useful for the scientific community dedicated to processing and image analysis, mainly those who are dedicated to the study of medical images. As mention before our project could be a first step in creating a CAD system. In the long term, the creation of this database would benefit the Mexican radiologists. It will improve the state of the art medical image processing in the country.

## **1.5 Methodology**

We choosed a factorial design of experiments as a methodology.

x experimental and processing parameters were chosen to be analysed in this research:

1. Variable A: Operator in a filter X
2. Variable B: Operator in a filter X
3. Variable C: Operator in a filter X

Process of image preprocessing applied to our research:

1. Image digitalisation
2. Image enhancing by filtering
3. Evaluation with both histogramas and medical opinion.
4. Presentation of the results

## Chapter 2

# Theoretical framework

### 2.1 Frame of reference

The pre-processing step serves to improve the processing with other filters like “segmentation”,

best techniques for enhancement of X-ray image may not be best for enhancement for microscopic images. [6]

The enhancement methods can broadly be divided into the following two categories: Spatial Domain Methods (SDM) and Frequency Domain Methods (FDM). Spatial domain methods which are operate directly on pixels. Frequency domain which operates on the Fourier transform of an image. [6]

Many studies show that the histogram equalization is an effective method for improving the quality of medical images. [6, 8].

#### 2.1.1 Bith depths

12-bit image (4096 levels of gray) displayed with 16 allocated bits appears dark. This happens because the maximal possible amplitude of 12-bit image for each pixel is 4095 and the maximal amplitude that can be displayed is 65535 (16 allocated bits). Therefore, 4096 levels need to be linearly scaled to 65536 levels to achieve good image displaying [8].

### 2.2 Conceptual framework

Digital image is a binary representation ... Pixel Histograms are frequency distributions, and histograms of images describe the frequency of the intensity values that occur in an image [9]. Image histogram offers a graphical representation of the tonal distribution of values in a digital image [10]. Color model Grayscale Image processing Medical image processing DICOM \*\*\* Different kind of medical images Las consideraciones técnicas (físicas) de

cómo se obtienen los mamogramas están fuera del alcance de este trabajo. In screening mammography, as practiced in USA, (also in Mexico), two x-ray images of each breast, in the medio lateral oblique and craniocaudal views, are acquired. MRI CT X-ray: Mammography is radiographic examination that is designed for detecting breast pathology, particularly breast cancer. [11] \*\*\* Mammographic features characteristic of breast cancer are masses, particularly ones with irregular or “spiculated” margins; clusters of microcalcifications (tiny deposits of calcium); and architectural distortions of breast structures. \*\*\* Noise: Nuclear images are generally the most noisy. Noise is also significant in MRI, CT, and ultrasound imaging. In comparison to these, radiography produces images with the least noise. The noise can cover and reduce the visibility of certain features within the image. The loss of visibility is especially significant for low-contrast objects. \*\*\*

Noise in digital mammographies (Practical digital mammography by Beverly Hashimoto) De acuerdo a Hachimoto en las imágenes mamográficas tenemos cuatro tipos de ruido: \*quantum \*fixed electronic \*señales secundarias \*quanta secundario indirecto

### 2.2.1 Mamogramas

Un mamograma típicamente tiene de 12 a 16 bits de profundidad, con una resolución de 4000x5000 pixeles [13], [14].

#### 2.3. Technological framework

\* Recordar escribir que en las primeras pruebas sólo se mejoró la visualización del tejido graso y no del tejido mamario.

\* Recordar escribir sobre (general purpose computers)

\* Unsharp - masking to reduce noise.

Libros que ayudarán con los conceptos teóricos:

Formación física de las imágenes Algoritmos T