

**Instituto Tecnológico y de Estudios Superiores de Monterrey**

**Campus Monterrey**

**Escuela Nacional de Ingeniería y Ciencias**

**Programa de Graduados**

**Maestría en Ciencias en Intelligent Systems**

**Propuesta de Tesis**

**Early detection and diagnosis of breast cancer lessions using  
(deep) convolutional networks in digital mammographic  
images.(working title)**

por

Erick Michael Cobos Tandazo

1184587



**Tecnológico  
de Monterrey**

Monterrey, N.L., April 7 de 2015

**Instituto Tecnológico y de Estudios Superiores de Monterrey**

**Campus Monterrey**

**Escuela Nacional de Ingeniería y Ciencias**

**Programa de Graduados**

Los miembros del comité de tesis recomendamos que la presente propuesta de Erick Michael Cobos Tandazo sea aceptada para desarrollar el proyecto de tesis como requisito parcial para obtener el grado académico de **Master in Science**, especialidad en:

**Intelligent Systems**

**Comité de Tesis:**

---

Dr. Hugo Terashima Marín

Asesor Principal

---

Por definir

Sinodal

---

Por definir

Sinodal

---

Dr. Ramón Brena Pinero

Director del Programa de Maestría en  
XXXX

April 7 de 2015

**Contents**

<b>1</b>	<b>Introduction</b>	<b>1</b>
1.1	Related Work . . . . .	1
<b>2</b>	<b>Problem Definition</b>	<b>1</b>
<b>3</b>	<b>Objectives</b>	<b>3</b>
<b>4</b>	<b>Hypotheses</b>	<b>3</b>
4.1	Research Questions . . . . .	4
<b>5</b>	<b>Background</b>	<b>4</b>
<b>6</b>	<b>Methodology</b>	<b>4</b>
<b>7</b>	<b>Work Plan</b>	<b>4</b>

## Abstract

Yet to write

## 1 Introduction

Yet to write

### 1.1 Related Work

Here I offer a summary of some of the first work in using convolutional networks for breast cancer diagnosis as well as other articles that have had an influence on this thesis.

Lo et al.[3] were the first group to use convolutional networks for breast cancer detection. They used a CNN with two hidden layers to detect microcalcifications. A high sensitivity image processing technique was used to obtain a set of 2104 patches (16 by 16 pixels) of all potential disease areas from 68 digital mammograms; of these, 265 were true microcalcifications and 1821 were “false subtle microcalcifications”. Prior to training the CNN, a wavelet high-pass filtering technique was used to remove the background of these images. Each image was flipped over (left-right) and 4 rotations for each the original and flipped images were used for training (0°, 90°, 180° and 270°). The CNN was composed of one input unit ( $16 \times 16$ ), 12 units in the first hidden layer ( $12 \times 12$ ), 12 units in the second hidden layer ( $8 \times 8$ ) and two output nodes (one for YES and one for NOT). The input size (16), number of hidden layers (2) and kernel size ( $5 \times 5$ ) was obtained via cross validation, although not many other options were explored: they tried input sizes of 8, 16 or 32, one or two hidden layers and kernel sizes of 2, 3, 5 or 13. The CNN reached 0.87 average AUC when identifying individual microcalcifications and 0.97 AUC for clustered microcalcifications. Only a minimum of three calcifications was considered a detection. Sensitivity and specificity test results were not reported. This article proved that simple convolutional networks can be efficiently used for medical image pattern recognition.

## 2 Problem Definition

Breast cancer is the most commonly diagnosed cancer in woman and its death rates are among the highest of any cancer. It is estimated that about 1 in 8 U.S. women will be diagnosed with breast cancer at some point in their lifetime. [2] Early detection is key in reducing the number of deaths from breast cancer; detection in its earlier stage (in situ) increases the survival rate to virtually 100%. [2]

With current technology, a high quality mammogram is “the most effective way to detect breast cancer early”. [4] Mammograms are X-ray images of each breast used by radiologists to search for early signs of cancer such as tumors or microcalcifications. About 85% of breast cancers can be detected with a screening mammogram. [?]. This high sensitivity is the product of the careful examination of the mammograms by experienced radiologists. A computer-aided diagnosis tool (CAD) could automatically detect and diagnose these abnormalities saving the time and training needed by expert radiologists and avoiding any human error. Computer based approaches could also be used by radiologists as a help during the screening process or as a second informed opinion on a diagnostic.

CAD systems are based on image and classification techniques coming from Artificial Intelligence and Machine Learning. Traditional CAD tools for breast cancer diagnosis are composed of three steps: feature extraction, feature selection and classification. In the feature extraction phase, the system uses filters and image transformations to preprocess the mammogram and find geometric patterns which are used to produce a set of features for the image; expert knowledge is sometimes used in this phase. Feature selection or regularization is used to focus only on the important features for the classification task. Once a vector of features is obtained for each image, an standard binary classifier can be used to perform the final detection or diagnosis. These techniques have been used for many years and are standard in the industry.<sup>1</sup>.

Despite its widespread use and efficiency, systems based on traditional computer vision techniques have various limitations that should be addressed to further improve its performance:

- There is no standard way of preprocessing mammograms. Some filters are commonly used but their performances can vary.
- It uses handcrafted features. The features extracted from the image are chosen beforehand (maybe designed with the help of experts) and special filters and image techniques are used to extract them.
- It normally uses a small patch of the mammogram and makes a prediction on that patch but it does not consider the entire mammogram neither to make a prediction on the patient or to account for correlation between patches.
- To produce good results it requires knowledge in various fields such as radiology, oncology, image processing, computer vision, machine learning, etc.
- It is composed of many successive steps. At each stage, there are many techniques from which the researcher can choose and many parameters which have to be estimated. This represents a cost in time and results as it is improbable that the optimal selection of techniques and parameters is achieved.
- As it is a complex system with different subsystems involved many other issues can arise such as non desired or unknown dependencies between subsystems, difficulty to localize errors, maintainability, etc.
- The techniques currently used are complex but the improvements achieved are not substantial. Much work is needed to make only incremental improvements and it is hard to know to which part of the system dedicate more resources.

This project will center around using Convolutional Networks ??, a recent development in Computer Vision, to tackle some of these limitations, specifically automate preprocessing and feature extraction, use entire mammogram images and simplify the system pipeline by using a convolutional network as a replacement for many steps traditionally performed in succession.

---

<sup>1</sup>See [1] for an example of a CAD system developed in this institution.

### 3 Objectives

The main goal of this work is to successfully apply convolutional networks to detect and diagnose breast cancer signs, microcalcifications and masses, in mammograms.

Particularly, there are various subgoals which we expect to achieve as the project advances:

- Develop a working pipeline for processing the mammographic images from our database and training a convolutional network. Essentially, this tool could also be used for other image classification tasks.
- Kickstart the research on deep learning in the institution.
- Use a simple convolutional network to perform detection and diagnosis and study these initial results to guide further research.
- Show the viability of convolutional networks for breast cancer diagnosis.
- Use convolutional networks on an entire mammogram instead of only on small patches.
- Improve the performance of convolutional networks reported on the literature.
- Generate results that could result in a conference or journal article.
- Propose new ideas and methods for future research in the topic.

Initial exploratory research has not yet been performed and some of these particular objectives may be modified as the project progresses. Furthermore, some new research avenues could be taken if they seem promising, for instance, using convolutional networks with digital tomosynthesis images (3-dimensional X-ray images of the breast).

### 4 Hypotheses

Although a considerable amount of work on breast cancer detection and diagnosis has been done in the institution, this work intends to be the first approximation to using convolutional networks for efficiently detecting and diagnosing breast cancer. Due to its exploratory nature, we can not be certain of the results that will be obtained. Nevertheless, we have a well established idea of what results we expect to obtain.

Convolutional networks are widely used for object recognition tasks and have shown very good results [5]. They have become one of the preferred methods to perform image classification tasks (similarly to neural networks in binary classification tasks)

Convnets have shown that So it will not be crazy to expect that

As convnets have shown to be very good at object recognition tasks, even reaching above human performance, we expect them to produce good results on our tasks, probably

That convnets will work  $\hat{L}$ , achieve at least as good performance as traditional methods. It will as good a performance as in other object recognition tasks

## 4.1 Research Questions

Some of the questions which will be answered in this work are:

- Can we do better than what has already been reported using convnets. can we do better than what has been reported using other methods
- Can we simplify the task of image recognition for this task
- what are the best parameters of the neural network (number of hidden networks, maxout vs pool, RELUs vs logistic, kernel sizes) Is there a big improvement on refining and tuning the network parameters for the task in hand
- How good are the results on the entire mammogram image?. Is there a way to join the results on the small patches to make a prediction on the patient?
- Will the data be enough or will the network overfit to the small amount of data.
- Can the features obtained from a convolutional network trained on a different database (like the ImageNet database) be used to obtain results on our images. are those results better than using a shallow convnet trained on medical images
- Are convolutional networks trained on pixel images better at this task than non convolutional neural networks or other non linear classifiers (SVMs, k-means) trained on handcrafted features?
- Is this a good path to keep working on to try to solve these problems or should we put resources on other methods?
- can we achieve human-like performance
- Can we do it here, can we do it cheap and easy to implement. Is the GPU optimization necessary.

## 5 Background

Yet to write

We offer an overview of some essential concepts for Cancer ref(Cancer subsection), Neural Networks and the mammographic database used in this document refSection 3,4

## 6 Methodology

Yet to write

## 7 Work Plan

Yet to write

## References

- [1] Breast Cancer Surveillance Consortium. Performance measures for 1,838,372 screening mammography examinations from 2004 to 2008 by age-based on BCSC data through 2009. electronic, September 2013. Available on [http://breastscreening.cancer.gov/statistics/performance/screening/2009/perf\\_age.html](http://breastscreening.cancer.gov/statistics/performance/screening/2009/perf_age.html).
- [2] Jose Luis Hernandez. Selección de características y clasificación de masas por medio de redes neuronales, máquinas de vector de soporte, análisis discriminante y regresión logística en mamografías digitales. Master’s thesis, Tecnológico de Monterrey, Monterrey, Mexico, April 2014.
- [3] Nadia Howlader, Anne M. Noone, Martin F. Krapcho, J. Garshell, Denise A. Miller, Sean F. Altekruse, Carol L. Kosary, Mandi Yu, Jennifer Ruhl, Zaria Tatalovich, Angela B. Mariotto, Denise R. Lewis, Huann S. Chen, Eric J. Feuer, and Kathleen A. Cronin. SEER cancer statistics review, 1975-2011. review, National Cancer Institute, Bethesda, MD, April 2014. Available on [http://seer.cancer.gov/csr/1975\\_2011/](http://seer.cancer.gov/csr/1975_2011/).
- [4] Shih-Chung B. Lo, Heang-Ping Chan, Jyh-Shyan Lin, Huai Li, Matthew T. Freedman, and Seong K. Mun. Artificial convolution neural network for medical image pattern recognition. *Neural Networks*, 8(7–8):1201 – 1214, 1995. Automatic Target Recognition.
- [5] National Cancer Institute. Mammogram fact sheet. electronic, March 2014. Available on <http://www.cancer.gov/cancertopics/types/breast/mammograms-fact-sheet>.
- [6] Olga Russakovsky, Jia Deng, Hao Su, Jonathan Krause, Sanjeev Satheesh, Sean Ma, Zhiheng Huang, Andrej Karpathy, Aditya Khosla, Michael Bernstein, Alexander C. Berg, and Li Fei-Fei. ImageNet Large Scale Visual Recognition Challenge. *CoRR*, abs/1409.0575, 2014. Available online on <http://arxiv.org/abs/1409.0575>.