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 neural networks in digital mammographic images.(working title)

por

Erick Michael Cobos Tandazo
1184587

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

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 (16×16), 12 units in the first hidden layer (12×12), 12 units in the second hidden layer(8×8) and two output nodes(one for YES and one for NOT). The input size (16), number of hidden layers (2) and kernel size (5×5) was obtained via cross validation, altough 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 efective 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. [1]. 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 proccess 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. [1].

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 succesive 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 an incremental improvement and it may be hard to know to which part of the system dedicate more time.

This project will center around using Convolutional Networks [**?**], a recent development in Computer Vision, to tackle some of these limitations, especifically 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 succesion.

[1]See [**?**] for an example of a CAD system developed in this institution.

3 Objectives

Yet to write

improve the results obtained with more traiditional methods Dejar el sistema aqui y el codigo de las convolutional networks so that it could be use don some other tasks or in 3d tommography comenzar en deep learning en la institucion Kickstart the work on convolutiop-nal netowkr or deep learning in the intitution. generate reslts tha culd result in an conference or journal article Perfomr a careful evaluationn of the convnets to determine what can be improved and work on it. Test the different hypothesis and give a concise answer to Sauy if this is a mehtid worht to put the resources on, . If it is yes, point to some directions wher eit could be imoporved. if it is not, poit to some of the problems that are preventing it from doing it.

4 Hypotheses

Yet to write

Can we do better than what has alreayd been reported using convnets. can we do better than what has been eported using other methods Can we simplify the task of image recognition for this task what are the best parameters ofr teh neural network (number of hidden networks, maxout vs pool, RELus vs logistic, kernel sizes)Is there a big improvement on refining and tuning the nertwork paraameters for the task in hand How good are the resutls on the enrtire mammogram image?. Is there a way to join the results on the small patche to make a prediction on the patient? Is the GPu optimization neccesary. Will the data be eough or willht network overfit to the small amount of data. Can the features obtained from a convolotional network trainedon a different database(like the imageNet database) be used ot o btain results on our images. are thos results better than using a shallow convnet trained on medical images Are convolutionla netowkrs traine don pixel images better at this task than non convolutional neural networks or other non linear classifiers (SVMs, k-means) trained on handcarfted features? Is this a good path to keep working on to try to solve these bproblem or sjould we put resources on other methods? can we achieve human-like performance

5 Background

Yet to write

We offer an overview of some essential concepts for Cancer ref(Cancer subsection), Neural Netrowkrs and the mamografic database used in this document refSeciton 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`.

[2] 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/`.

[3] 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.

[4] National Cancer Institute. Mammogram fact sheet. electronic, March 2014. Available on `http://www.cancer.gov/cancertopics/types/breast/mammograms-fact-sheet`.