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

Cance is this and this, bereast cancer is this and this. breats cancer has high incicdence and high mortality,

Breast cancer has the second highest incidence of any cancer in the United States, an estimated 14% of all new cancer cases in 2014 were breast cancer and the third highest mortality of any cancer accounting for 6.8% of cancer related deaths [**?**]. Among women it is by large the most commonly diagnosed cancer (30.3% of all cancers) and has the highest death rate (15.1%) besides lung cancer.[**?**]. mammograms are this and radiologist do that all the other things from problem definition (problem definition section does not exist on thesis just on plan.)

Michael: aqui comienza lo nuevo In this work we will center on using mammogrmas (2d Xray images of the breast) to detect microcalcifications and masses and diagnos its likelihood of breast cancer, i.e, make a prediction on the probability that a mmaogram signals breast cancer. this is of importance because of ... this is withr the view of form part odf abigger CAd project developed in the insitution that could help radiologists in the detection of breast cancer either by proposing zones where the radiologists could take a second look, or by giving a second informed opinion on the same patient. (why convnets?) convnets are a natural extension to artificial neural networks that deal with images and that Summary of the related work part here(this guys dd it first and this guys and i'm gteting my ideas from here and there and my contribution is this...) We will do this and this and that (methods and plan of work) Goals: We want to improve over the published results, ... and know if we can a ctuallly use this

1.1 Related work

Here i offer a summary of some of the work done on topics similar to that treated on this article and that have influenced this paper

[3] is the first article 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. Work done here at Tec

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. [1] Early detection is key in reducing the effects of breast cancer: a detection in its earlier stage (in situ) allows for better treatment and increases the survival rate to virtually 100%. [1]

With current technology, a high quality mammogram is "the most efective way to detect breast cancer early" [2]. Mammograms are x-ray images of each breast used by radiologists to search for early signs of cancer such as tumors or microcalcifications.

90% of breats cancers can be detected using ammomgram and radiologist can detect up to .Nonetheless, these tasks is not easy and requires training an dexperience to be realized effectuvely a computer aided diagnostic tool that can automatically detect these signs from digitizes images of the mammogrmas couls save the time to expert and increase the aomoun t of cancers detect or pint to experst whthe zones that could be treated woith more rcareExepret radiologists can detect up to ... but normally you do double duty... Radiologists normally get this amunt of time right. and a second decision would increase the amount of positives(?) detected and save time for the experts.

In this work we will focus on using mammograms The project developed here has already dealed with this problem using .... and .... cite. but we tried to use convnets so that we can improve the results obtained here and in the literature with some advanced image pattern algorithms. A further review was given on t eintroduction.

What is the problem/limitations? Not yet efficient, requireds handcrafted fuigures and a lot of parameter fitting an estimation to get the better results. It depends on a lot of steps and thus could be prone to errors and requires work in various differents fields (image processing, radiology, macjhine learning) to prouce good results. Furthermore, some of tit depends on expert information (like the shapes of microcalcifications and masses and important image features) to produce results while a better systmem could let the computer figure out what are those neccesary features

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 convolutiopnal 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?

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] 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. electronic, National Cancer Institute, Bethesda, MD, April 2014. Available on `seer.cancer.gov/csr/1975_2011/`.

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

[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.