Identifying Calcifications in Mammograms Using Texture Analysis

Mujtaba Hussain

June 27, 2008

Contents

0.1 Abstract

Cancer is one of the major diseases of the 20^{th} century. It affects majority of the population of the world and all over the world, scientists are trying to find the cure for this in all its forms. This research attempts to aid that research by attempting to automate and thereby reduce the time in cancer detection. This research attempts to identify cancerous calcifications on mammograms using texture analysis. This texture analysis is implemented using an implementation of Genetic Programming known as *RMIT-GP*. Using RMIT-GP, several programs are trained over known image sets and are then tested over unknown ones. The results showed several larger training sets are needed for improved identification of calcifications.

0.2 Introduction

0.2.1 Motivation

Cancer afflicts many men and women around the world and continues to affect more with no cure in sight. At the moment, early detection assisted with radioactive treatments is the best of hope people have of getting the cancer into remission. Often the tests involved in determining whether the patient has cancer or not take a lot of time and are very expensive. It would be of immense help to all involved in the prevention and detection of cancer if the presence of cancer among patients could be made simple and less expensive task and that is the motivation of this research.

0.2.2 Research Questions

During the course of this research, we will attempt to answer the following questions:

1. What window size will be enough to mark a cancerous pattern on a mammogram?
2. What testing size in terms of number of images is sufficient to determine whether the program has learned well or not?
3. How accurately can a cancerous pattern be identified?

0.2.3 Methodology

In this research we plan on using two sets of images. Pre-tagged mammograms and clean mammograms. We will extract a set of images from the pre-tagged mammograms of **NxN** dimensions which will include sections which are known to be cancerous and also sections which are known to be non-cancerous. We will then use RMIT-GP to evolve several programs which will have been trained to learn the difference between cancerous and non-cancerous texture using the previously mentioned set of images.

Once the program has learned from the two sets of image textures, we will choose the best program and give it a set of images to test on. Basically, this set will comprise of a bunch of non-cancerous and cancerous images and the program will try to identify which is which. Once the program has

completed evaluating the set, we check how accurately it has distinguished the cancerous set from the non-cancerous one. If the achieved accuracy is deemed insufficient by the researchers, then more training and test sets are used. The main aim of this methodology is to achieve as high an accuracy as possible.

0.3 Background

This area of research is not new as previously work has been done in the field of texture analysis.

Kurniawan D.W. [1] in 2006 submitted an honours thesis documenting texture segmentation and analysis on images of sky and grass. Given a set of images, the author was able to evaluate the image and decide up to a certain extent whether that image contained a texture of interest or not. Since the two textures of interest were only Grass and Sky, accuracy estimate of up to 88% for sky and 78% for Grass texture were achieved.The authors also used the RMIT-GP package and the images were taken from a Corel Image Database. The author used GP since it had been proved useful in previous texture analysis techniques used by his supervisor Dr.Andy Song.

Kurniawan D.W. [1] moved a window of size n x n across the image trying to identify textures in the image which had been converted from Gray scale to its appropriate image mask. If the sample inside the window identifies as a texture of interest, then count for positive retrieval is increased and the window is moved ahead one pixel. This is repeated till the bottom edge is reached. The author was able to successfully identify sky patterns to an accuracy level of 88% and grass patterns up to 79%. They also determined that even though human labelling would seem easier when compared to a GP manipulated texture segmentation, there are so many variations in textures, that it is almost impossible to be a hundred percent accurate all the time.

The author proposed a way forward with respect to identification of more complex patterns, like the ones generated by meteorological satellites in identifying patterns of rain etc. The authors propose this based on the positive result and feedback they received on their texture segmentation of two simple textures.

This approach was the main basis of this research using texture analysis using GP as its main system.

Mudigonda et al [2] tried to classify the benign and malignant cancerous masses in mammograms images using gradient and texture analysis as well bu their methodology utilised polygonal modelling of boundaries for the extraction of a ribbon of pixels across mass margins for identification.

The authors found that with texture analysis, while very good at identifying the different patterns for benign and malignant cancerous patterns, concluded that more research needed to be done across a larger scale as mass margin extraction had given positive results, just not good enough. Also, since the ribbon is not guaranteed to be continuous, edge pixels are not guaranteed to be proper, because of this, the results itself were not guaranteed.

Mudigonda et al [2] concluded in the end that the polygonal approach to cancer detection was not the best approach.

Zhang et al [3] proposed and investigates neural and statistical classifiers to classify micro-calcification(cancerous) patterns in digital mammography. The researchers used area selection, feature selection and then feature extraction to get areas where the program might have a good chance of detecting micro-calcifications. The database that they used was the one from University of South Florida. The database in particular was freely available.

Zhang et al [3] concluded that the proposed approach is able to find an appropriate feature subset and neural classifier achieves better results than two statistical models. But this approach is more complex than the GP approach and the time available in this honours research is not enough to allow such an implementation, the researchers in this case did not take up the neural approach. However, this research did lead to the database of mammograms in the University of South Florida.

0.4 Description of the System

The system being used in this research is RMIT-GP. It was developed in house by the *Evolutionary Art* and *Artificial Intelligence* branches of Computer Science department in RMIT.

The images that RMIT-GP will train and test on were obtained from the University of South Florida's Oncology Department.

0.5 Experiments

The main aim of all the experiments done was to determine whether we could increase the accuracy of detection while testing by varying the size and variety of training data. To achieve this, the training data would always be split in half. 50% of the images would always be cancerous and the other half would be non-cancerous. The training set of RMIT-GP at this point was informed about the nature of these halves and hence allowed to train on them. The question at this point confronting the researchers was what dimensions to choose so as to get sample cancerous images properly. Upon examining about 150 sample images, it was decided that a sample image of 8x8 would be enough to encapsulate a cancerous calcification. That answers our first research question.

0.5.1 Image Setup

The images were originally *GIF* format had to be converted to the proper *PGM* format using the NETPBM package available on the provided machines. Below are provided the sample formats of images used. Figure 1 is the main image:

After using the NETPB, package to convert a sample 8x8 section, it looks like:

P2
8 8
255
232 232 216 196 184 182 180 176
220 214 200 186 178 170 186 200
190 194 198 196 192 182 178 188
190 186 198 196 188 178 166 172
198 188 190 190 190 190 186 194
204 212 212 206 210 204 196 176
220 240 246 230 212 202 196 170
234 250 252 236 214 200 188 178

The first line is the identifier, the second line represents the values of the dimensions of the images, in this case 8x8 pixels and then the content of the

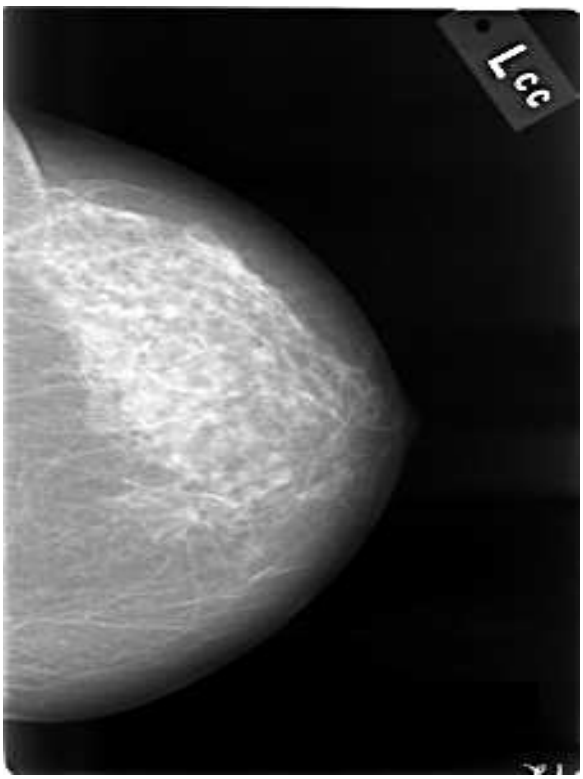

Figure 1: Sample Mammogram

image starts. These 8x8 samples were hand picked from the actual image by the researchers for cancerous sets only. Non cancerous sets were cut out en-masse using the NETPBM tools available.

0.5.2 Training

Initially we began with a sample set of 10 images and steadily progressed to 50 images for the training set. Once the training was done, we chose the best program to be used in the testing.

0.5.3 Testing

In the testing phase, we collected 50 samples in the data set out of which the percentage of cancerous and non-cancerous images was 50-50. Using the best generation obtained from the training phase, we run testing phase over this data set.

The following graphs reflect the performance of the individuals for the training set of 50 images. Figure 2 shows the accuracy of the individuals.

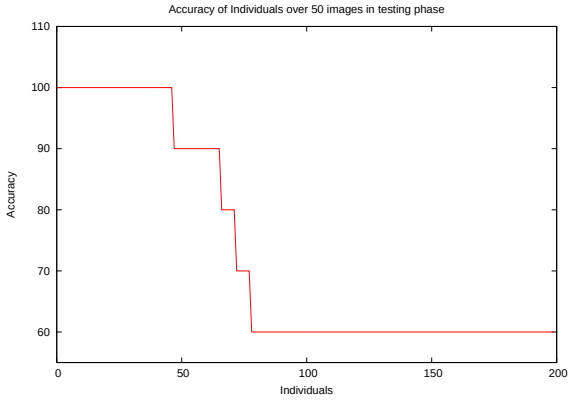

Figure 2: Accuracy of Individuals

Figure 3 shows the fitness of the individuals across the 50 training images.

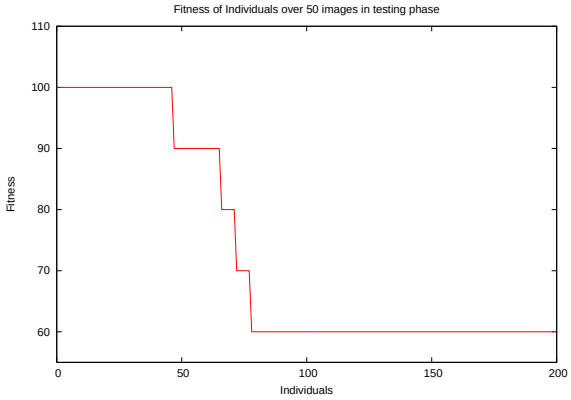

Figure 3: Fitness of the Individuals

To test the accuracy of the testing phase over a large set of images, we tested the program with 6180 non-cancerous images in order to check whether the accuracy of the program was enough to detect these as non-cancerous. As Figure 4 shows, the accuracy decreased over individuals.

The final accuracy achieved in this experiment was 88.7055 .

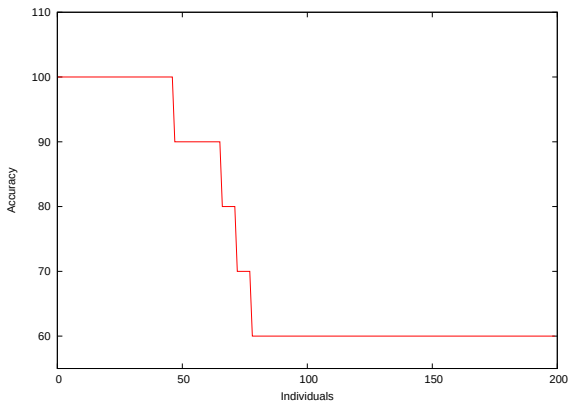

Figure 4: Accuracy over 6180 images

0.6 Discussion

From the experiments, we can see that training the GP over only 50 image sets is not enough. The accuracy achieved during the testing phase is not sufficient to properly distinguish between the textures of cancerous and non-cancerous mammograms convincingly.

0.6.1 Future Work

1. 50 images is not enough to train the classifier for images. More images and more varied images are required. Perhaps images from more sources than just one oncology department could be researched into.

2. Due to time constraints, this research could not progress further than evaluating the GP over the testing phase. Future work could involve testing over larger sections of images and then concatenating the evaluated sections together to present a 2D view of the result. This approach would highlight false positives very well.

3. In addition to Genetic Programming, we could also look at Object-Detection algorithms to identify cancerous sections of mammograms because at times, the cancerous parts are too small to be trained upon.

0.7 Conclusion

This research has shown there is great potential in trying to detect cancerous portions over mammograms using texture analysis using Genetic Programming. This approach however is too large for an honours project and should be utilised as a full fledged research project if its full potential is to be exploited.

0.7.1 Acknowledgments

I would like to thank Simon Duff, Danny Fang and especially my supervisor Dr. Vic Ciesielski, without whose patience and wisdom, I would not have been able to complete this project.

Bibliography

[1] D. W. Kurniawan. Image retrieval using texture segmentation by gp. In *Honours Thesis*. Computer Science and Information Technology, RMIT, 2006.

[2] N. R. Mudigonda, R. Rangayyan, and J. E. L. Desautels. Gradient and texture analysis for the classification ofmammographic masses. In *Medical Imaging*. Calgary University,Atlanta, IEEE, 2000.

[3] P. Zhang, B. Verma, and K. Kumar. Neural vs. statistical classifier in conjunction with genetic algorithm based feature selection. In *Evolutionary Computation*. Central Queensland University,Australia, IEEE, 2004.