Visualisation and Topological Aspects of Higher Dimensional Data

Final Report for CS39440 Major Project

Author: Samuel Jackson (slj11@aber.ac.uk)
Supervisor: Prof. My Supervisor (rrz@aber.ac.uk)

March 31, 2015
Version: 1.0 (Draft)

This report was submitted as partial fulfilment of a MEng degree in Software Engineering (G601)

Department of Computer Science
Aberystwyth University
Aberystwyth
Ceredigion
SY23 3DB
Wales, UK

Declaration of originality

In signing below, I confirm that:

- This submission is my own work, except where clearly indicated.
- I understand that there are severe penalties for plagiarism and other unfair practice, which can lead to loss of marks or even the withholding of a degree.
- I have read the sections on unfair practice in the Students' Examinations Handbook and the relevant sections of the current Student Handbook of the Department of Computer Science.
- I understand and agree to abide by the University's regulations governing these issues.

Signature .....

Date .....

Consent to share this work

In signing below, I hereby agree to this dissertation being made available to other students and academic staff of the Aberystwyth Computer Science Department.

Signature .....

Date .....

Acknowledgements

I am grateful to...

I’d like to thank...

Abstract

Include an abstract for your project. This should be no more than 300 words.

CONTENTS

LIST OF FIGURES

LIST OF TABLES

Chapter 1

Background & Objectives

1.1 Mammography

Breast cancer is the leading cause of death among women and is the most common form of cancer found in women [9]. Early screening of breast cancer using mammography has been shown to reduce the mortality rate of women [7, 10].

Mammography is the analysis of female breast tissue through the use of X-ray radiology with the goal of producing high resolution images of the structure within the female breast. The composition of the parenchymal patterns and tissue density revealed by in a mammographic evaluation can be used in the early detection of breast cancer.

Qualitatively speaking the composition of breast tissue can be split into four distinct categories. These are Nodular densities (corresponding to Terminal Ductal Lobular Units (TDLUs), linear densities (corresponding to ducts, vessels, and fibrous strands), homogeneous, structureless densities (corresponding to fibrous supporting tissue), and radiolucent areas (corresponding to adipose tissue) [11]. Typical markers used in the detection of cancer can are the presence of clusters of micro- calcifications, masses, architectural distortions, breast density and parenchymal patterns [5, 8].

1.1.1 Risk Assessment

Mammograms provide a non-invasive means to assess the risk of a patient developing cancer given a set of mammographic images. Several different systems have been developed to aid the classification of mammographic risk based on the parenchymal patterns visible using X-ray mammography.

1.1.1.1 Wolfe

The earliest attempt to classify mammographic risk using parenchymal patterns was suggested the by Wolfe [12]. Wolfe proposed a classification system which split patients into four categories depending on the relative visible density of fat, ducts and connective tissue. The four categories are described, in order of lowest to highest risk, in ref. [12] as:

- **N1** - Breast is mostly composed of fat with no visible ducts and very little amounts of dysplasia present.
- **P1** - The parenchyma is primarily composed of fat with up to one quarter of the breast density being composed of visible ducts in the anterior position which may extend into a quadrant.
- **P2** - Breast indicates prominent duct pattern beyond one quarter of the breast that can occupy the entire parenchyma.
- **DY** - Characterised by a severe increase in breast density and often appear as homogenous, missing the duct pattern present in P2 breasts.

1.1.1.2 Boyd

Boyd et al. [2] proposed a quantitive assessment of risk based on increasing classes of mammographic density, know as the six class categories (SCC). These classes are based on the proportion of dense tissue relative to the area of the breast. The six classes are:

- $<10\%$
- 10 to $<25\%$
- 25 to $<50\%$
- 50 to $<75\%$
- $\geq 75\%$

1.1.1.3 Tabár

Tabár et al. [4] proposed as classification scheme which classifies a breast based on the percentage presence of the four building blocks of breast composition [4, 11]. The description of each of the five patterns is given as:

- **Pattern I** - Breast corresponding to pattern I exhibit scalloped contours and cooper's ligaments with evenly scattered TDLU's.
- **Pattern II** - Complete fatty replacement of both
- **Pattern III** - Prominent retroareolar duct pattern and fatty involution.
- **Pattern IV** - Extensive linear and nodular densities present throughout the parenchyma.
- **Pattern V** - Homogeneous, structureless fibrosis with a convex contour.

1.1.1.4 BI-RADS

The Breast Imaging Report and Data System (BI-RADS) [1, 3] was developed by the American College of Radiology (ACR) in an attempt to standardise the lexicon used to describe mammography reports during standard screening. BI-RADS has classifies the breast based on density into four categories [1].

1. Fatty Breast ($<$10% of dense tissue)
2. Fibroglandular ($<$0 - 48% of dense tissue)
3. Heterogeneously dense ($<$49 - 90% of dense tissue)
4. Homogeneously dense ($\geq$90% of dense tissue)

A radiologist will then classify the breast according to one of 7 categories after interpretation [1]. These are one of:

- Incomplete. Additional evaluation needed
- Normal.
- Typically benign.
- Probably benign. A shorter interval follow-up is recommended
- Suspicious Abnormality. Biopsy considered
- Highly suggestive of malignancy. Biopsy should be performed.
- Histologically proven malignancy.

1.2 Features

Features are higher level descriptive abstractions computed from lower level structure such as areas of high intensity, edges, and corners present within an image.

1.2.1 Shape Features

1.2.2 Intensity Features

1.2.3 Texture Features

1.3 Dimensionality Reduction

1.3.1 Linear

1.3.2 Non Linear

1.4 Visualisation

1.5 Analysis

1.6 Research Method

Chapter 2

Experiment Methods

2.1 Overview

2.2 Techniques

2.2.1 Features

2.2.2 Dimensionality Reduction

2.2.3 Visualisation

2.3 Datasets

2.3.1 Synthetic Data

2.3.2 Real Data

2.4 Implementation

2.4.1 Languages

2.4.2 Libraries

Chapter 3

Results and Conclusions

3.1 Comparison of Real and Synthetic Datasets

3.2 Investigation of Mapping

Chapter 4

Critical Evaluation

4.1 Conclusions

4.2 Evaluation of the Project

4.3 Future Work

Appendices

Appendix A

Third-Party Code and Libraries

Appendix B

Code samples

Annotated Bibliography

[1] C. Balleyguier, S. Ayadi, K. Van Nguyen, D. Vanel, C. Dromain, and R. Sigal, "Birads classification in mammography," *European journal of radiology*, vol. 61, no. 2, pp. 192–194, 2007.

[2] N. Boyd, J. Byng, R. Jong, E. Fishell, L. Little, A. Miller, G. Lockwood, D. Tritchler, and M. J. Yaffe, "Quantitative classification of mammographic densities and breast cancer risk: results from the canadian national breast screening study," *Journal of the National Cancer Institute*, vol. 87, no. 9, pp. 670–675, 1995.

[3] C. J. D'orsi, A. C. of Radiology, A. C. of Radiology, B.-R. Committee, *et al.*, *Illustrated Breast Imaging Reporting and Data System:(illustrated BI-RADS)*. American College of Radiology, 1998.

[4] I. T. Gram, E. Funkhouser, and L. Tabár, "The tabar classification of mammographic parenchymal patterns," *European journal of radiology*, vol. 24, no. 2, pp. 131–136, 1997.

[5] V. A. McCormack and I. dos Santos Silva, "Breast density and parenchymal patterns as markers of breast cancer risk: a meta-analysis," *Cancer Epidemiology Biomarkers & Prevention*, vol. 15, no. 6, pp. 1159–1169, 2006.

[6] A. C. of Radiology. BI-RADS Committee and A. C. of Radiology, *Breast imaging reporting and data system*. American College of Radiology, 1998.

[7] I. U. P. on Breast Cancer Screening *et al.*, "The benefits and harms of breast cancer screening: an independent review," *The Lancet*, vol. 380, no. 9855, pp. 1778–1786, 2012.

[8] M. P. Sampat, M. K. Markey, A. C. Bovik, *et al.*, "Computer-aided detection and diagnosis in mammography," *Handbook of image and video processing*, vol. 2, no. 1, pp. 1195–1217, 2005.

[9] R. Siegel, J. Ma, Z. Zou, and A. Jemal, "Cancer statistics, 2014," *CA: a cancer journal for clinicians*, vol. 64, no. 1, pp. 9–29, 2014.

[10] R. A. Smith, D. Manassaram-Baptiste, D. Brooks, V. Cokkinides, M. Doroshenk, D. Saslow, R. C. Wender, and O. W. Brawley, "Cancer screening in the united states, 2014: a review of current american cancer society guidelines and current issues in cancer screening," *CA: a cancer journal for clinicians*, vol. 64, no. 1, pp. 30–51, 2014.

[11] L. Tabár, T. Tot, and P. B. Dean, *Breast cancer: the art and science of early detection with mammography: perception, interpretation, histopathologic correlation*. Thieme, 2005.

[12] J. N. Wolfe, "Breast patterns as an index of risk for developing breast cancer," *American Journal of Roentgenology*, vol. 126, no. 6, pp. 1130–1137, 1976.