

# **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 20, 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

<b>1</b>	<b>Background &amp; Objectives</b>	<b>1</b>
1.1	Mammography . . . . .	1
1.1.1	Risk Assessment . . . . .	1
1.2	Features . . . . .	2
1.2.1	Shape Features . . . . .	2
1.2.2	Texture Features . . . . .	2
1.3	Dimensionality Reduction . . . . .	2
1.3.1	Linear . . . . .	2
1.3.2	Non Linear . . . . .	2
1.4	Visualisation . . . . .	2
1.5	Analysis . . . . .	2
1.6	Research Method . . . . .	2
<b>2</b>	<b>Experiment Methods</b>	<b>3</b>
2.1	Overview . . . . .	3
2.2	Techniques . . . . .	3
2.2.1	Features . . . . .	3
2.2.2	Dimensionality Reduction . . . . .	3
2.2.3	Visualisation . . . . .	3
2.3	Datasets . . . . .	3
2.3.1	Synthetic Data . . . . .	3
2.3.2	Real Data . . . . .	3
2.4	Implementation . . . . .	3
2.4.1	Languages . . . . .	3
2.4.2	Libraries . . . . .	3
<b>3</b>	<b>Results and Conclusions</b>	<b>4</b>
3.1	Comparison of Real and Synthetic Datasets . . . . .	4
3.2	Investigation of Mapping . . . . .	4
<b>4</b>	<b>Critical Evaluation</b>	<b>5</b>
4.1	Conclusions . . . . .	5
4.2	Evaluation of the Project . . . . .	5
4.3	Future Work . . . . .	5
	<b>Appendices</b>	<b>6</b>
<b>A</b>	<b>Third-Party Code and Libraries</b>	<b>7</b>
<b>B</b>	<b>Code samples</b>	<b>8</b>
	<b>Annotated Bibliography</b>	<b>9</b>

## **LIST OF FIGURES**

## **LIST OF TABLES**

# Chapter 1

## Background & Objectives

### 1.1 Mammography

Breast cancer is the leading cause of death among women world wide and is the most common form of cancer found in women [2]. 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 breast. The composition of the internal structure can then be used to permit earlier 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).

#### 1.1.1 Risk Assessment

Mammograms provide a non-invasive means to assess the risk of a patient developing cancer. The general goal of a mammographic risk assessment is to evaluate the risk of patient having cancer. The composition of breast tissue can be categorised using the Breast Imaging Reporting And Data System (BI-RADS) [1]. BI-RADS classifies mammograms based on the density of tissue (and therefore risk) in the mammogram.



## **1.2 Features**

### **1.2.1 Shape Features**

### **1.2.2 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] A. C. of Radiology. BI-RADS Committee and A. C. of Radiology, *Breast imaging reporting and data system*. American College of Radiology, 1998.
- [2] 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.