



THE UNIVERSITY  
*of* ADELAIDE

**Automated Detection, Segmentation and  
Classification of Masses from Mammograms  
using Deep Learning**

by

Neeraj Dhungel

A thesis submitted in fulfillment for the  
degree of Doctor of Philosophy

in the

Faculty of Engineering, Computer and Mathematical Sciences  
School of Computer Science

October 2016

# Contents

<b>Contents</b>	<b>i</b>
<b>List of Figures</b>	<b>iii</b>
<b>Abstract</b>	<b>iv</b>
<b>Declaration</b>	<b>vi</b>
<b>Publications</b>	<b>vii</b>
<b>Acknowledgements</b>	<b>viii</b>
<b>1 Introduction</b>	<b>1</b>
1.1 Motivation . . . . .	4
1.2 Contributions of This Thesis . . . . .	6
1.3 Thesis Outline . . . . .	7
<b>2 Literature Review</b>	<b>8</b>
2.1 Mammogram Pre-processing . . . . .	8
2.2 Breast Mass Detection . . . . .	10
2.2.1 Problem Definition . . . . .	10
2.2.2 Background . . . . .	10
2.3 Breast Mass Segmentation . . . . .	13
2.3.1 Problem Definition . . . . .	13
2.3.2 Background . . . . .	14
2.4 Breast Mass Classification . . . . .	15
2.4.1 Problem Definition . . . . .	15
2.4.2 Background . . . . .	16
2.5 Mammogram Analysis Systems . . . . .	16
2.6 Conclusions . . . . .	18
<b>3 Methodology</b>	<b>19</b>
3.1 Overview of the method . . . . .	19
3.2 Dataset Definiton . . . . .	19

---

3.3	Image Enhancement . . . . .	20
3.3.1	Otsu's Thresholding . . . . .	20
3.4	Mass Detection . . . . .	21
3.4.1	Mass Candidate Generation . . . . .	22
3.4.2	False Positives Reduction with R-CNN . . . . .	24
3.4.3	Candidate Selection using RF . . . . .	25
3.4.4	Mass Detection Refinement using Bayesian Optimisation . . . . .	26
3.5	Mass Segmentation . . . . .	29
3.5.1	Conditional Random Field (CRF) . . . . .	29
3.5.2	Structured Support Vector Machine (SSVM) . . . . .	31
3.5.3	Potential Functions . . . . .	32
3.5.4	Mass Segmentation Refinement using Active Contour Model . . . . .	35
3.6	Mass Classification . . . . .	36
3.7	Conclusions . . . . .	38
<b>4</b>	<b>Deep Structured Learning for mass Segmentation from Mammograms</b>	<b>39</b>
<b>5</b>	<b>Tree Re-weighted belief propagation using deep learning potentials for mass segmentation from mammograms.</b>	<b>45</b>
<b>6</b>	<b>Deep learning and structured prediction for the segmentation of mass in mammograms</b>	<b>51</b>
<b>7</b>	<b>Automated Mass Detection in Mammograms Using Cascaded Deep Learning and Random Forests</b>	<b>61</b>
<b>8</b>	<b>The Automated Learning of Deep Features for Breast Mass Classification from Mammograms</b>	<b>71</b>
<b>9</b>	<b>A Deep Learning approach to fully automated analysis of Masses in Mammograms</b>	<b>81</b>
<b>10</b>	<b>Conclusion and Future Works</b>	<b>124</b>
10.1	Summary of Contributions . . . . .	124
10.2	Future Work . . . . .	126

# List of Figures

1.1	Some examples of mammogram images. . . . .	2
1.2	Types of lesions in mammograms from the INbreast dataset, where we denote masses using red contours and calcifications using green contours. . . . .	2
1.3	Mass detection, segmentation and classification. . . . .	3
2.1	Main steps of a CAD system that analyses breast masses from mammograms. . . . .	9
3.1	Breast profile segmentation using Otsu’s thresholding. . . . .	20
3.2	Our methodology for mass detection with refinement. . . . .	21
3.3	Mass candidate generation using m-DBN. . . . .	22
3.4	Structured Prediction Model for segmentation of masses. . . . .	30
3.5	Examples of unary potentials for few training images from DBN, CNN and GMM. . . . .	31
3.6	DBN model that takes the mass candidate as input and outputs a unary potential (probability map) for our segmentation algorithm (Chapters 4, 5, 6). . . . .	33
3.7	The CNN model that takes mass candidate as input and outputs a unary potential (probability map) for our segmentation algorithm (Chapter 6). . . . .	33
3.8	Example of fully automated mass detection, segmentation and classification of mass from mammograms using the proposed methodology. . . . .	36
3.9	Two stage CNN model for mass classification in mammograms (Chapter 8). . . . .	36

# *Abstract*

Breast cancer is considered to be one of the major contemporary problems affecting the lives of thousands of women worldwide. One of the most effective tools in the fight against this disease is early detection based on the manual analysis of X-ray mammograms. This manual process of interpretation of mammograms involves the detection of breast lesions (e.g., masses), the segmentation of lesions boundaries and the classification of lesions based on their shape, appearance and texture features. This manual analysis of breast lesions from mammograms presents large interpretation variability amongst radiologists. This variability can be reduced with the aid of computer aided diagnosis (CAD) systems that can act as a second reader in the analysis of breast lesions. However, for a CAD system to be useful in a clinical setting, it must effectively classify lesions as benign or malignant.

Detection, segmentation and classification of breast lesions are the main three steps involved in fully automated CAD systems that can work in the analysis of mammograms. Building a CAD system is difficult because mammograms are marred by low signal to noise ratio for the visualisation of breast lesions. In addition, breast lesions present a large variation in terms of shape, size and appearance. A large number of methods have been applied for building automated CAD systems for both types of lesions, namely mass and micro-calcification, but in this work we focus only on the analysis of masses. The major drawback of current approaches is that they generate a large number of false positives and miss a fair amount of true positive regions during the mass detection stage. Furthermore, mass segmentation is generally based on active contour models and graph-based approaches that rarely capture the large shape and appearance variations of breast masses. Finally, mass classification is generally implemented using sub-optimal hand-crafted features and machine learning classifiers such as support vector machines (SVM), linear discriminant analysis (LDA), artificial neural net (ANN), etc. One major limitation of the majority of existing CAD systems is that most of them require manual intervention to obtain mass candidates for segmentation and classification.

This thesis presents a new approach based on recently developed deep learning models to develop a fully automated CAD system for automated detection, segmentation and classification of masses from mammograms. Our proposed solution to the mass detection problem consists of three stages: 1) mass candidate generation using multi-scale deep learning and Gaussian mixture models, 2) false positive reduction with a cascade of deep learning and random forests classifiers, 3) candidate refinement with a local search algorithm based on Bayesian optimisation. Our proposed mass segmentation methods are based on two kinds of structured output learning methods, namely: 1) structured support vector machine for parameter estimation and

---

graph cut for inferring the segmentation labels, and 2) truncated fitting for parameter learning and tree re-weighted belief propagation for inference. The resulting segmentation is then refined using an active contour model. Our proposed mass classification deep learning method is modelled with a two-step training procedure, where the first step is based on a pre-training stage that estimates a large set of hand-crafted features, which is followed by a fine-tuning step that learns a classifier (that classifies masses into benign and malignant). Finally, we integrate our mass detection, mass segmentation and mass classification methods into a fully automated CAD system for the analysis of masses in mammograms. We validate our methodology on two publicly available datasets (INbreast and DDSM-BCRP) using different performance measures such as average Dice index for segmentation, free receiver operating curve (FROC) and average precision curve for detection, receiver operating curve (ROC), area under curve (AUC) and accuracy for classification. The experiments show that our methodology for detection, segmentation and classification of breast masses achieves competitive results with respect to the current state-of-the-art techniques in terms of all performance measures mentioned above.

## *Declaration*

I certify that this work contains no material which has been accepted for the award of any other degree or diploma in my name, in any university or other tertiary institution and, to the best of my knowledge and belief, contains no material previously published or written by another person, except where due reference has been made in the text. In addition, I certify that no part of this work will, in the future, be used in a submission in my name, for any other degree or diploma in any university or other tertiary institution without the prior approval of the University of Adelaide and where applicable, any partner institution responsible for the joint-award of this degree.

I give consent to this copy of my thesis when deposited in the University Library, being made available for loan and photocopying, subject to the provisions of the Copyright Act 1968.

The author acknowledges that copyright of published works contained within this thesis resides with the copyright holder(s) of those works.

I also give permission for the digital version of my thesis to be made available on the web, via the University digital research repository, the Library Search and also through web search engines, unless permission has been granted by the University to restrict access for a period of time.

Signed:

---

Date:

---

# Publications

My thesis is based on the content of the following peer-reviewed conference and journal papers:

- Neeraj Dhungel, Gustavo Carneiro, Andrew P. Bradley. Deep structured learning for mass segmentation from mammograms. In *IEEE International Conference on Image Processing (ICIP)*, 2015.  
(DOI: [10.1109/ICIP.2015.7351343](https://doi.org/10.1109/ICIP.2015.7351343))  
(This paper was selected among the top 10% of the accepted papers of the conference)
- Neeraj Dhungel, Gustavo Carneiro, Andrew P. Bradley. Tree RE-weighted belief propagation using deep learning potentials for mass segmentation from mammograms. *IEEE International Symposium on Biomedical Imaging (ISBI)*, 2015.  
(DOI: [10.1109/ISBI.2015.7163983](https://doi.org/10.1109/ISBI.2015.7163983))
- Neeraj Dhungel, Gustavo Carneiro, Andrew P. Bradley. Deep learning and structured prediction for the segmentation of mass in mammograms. *18th International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, 2015.  
(DOI: [10.1007/978-3-319-24553-9\\_74](https://doi.org/10.1007/978-3-319-24553-9_74))
- Neeraj Dhungel, Gustavo Carneiro, Andrew P. Bradley. Automated Mass Detection in Mammograms Using Cascaded Deep Learning and Random Forests. *International Conference on Digital Image Computing: Techniques and Applications (DICTA)* 2015.  
(DOI: [10.1109/DICTA.2015.7371234](https://doi.org/10.1109/DICTA.2015.7371234)).  
This paper was first presented at 3rd workshop on Breast Image Analysis (BIA) as part of 18th International Conference on Medical Image Computing and Computer Assisted Intervention (MICCAI-BIA), Munich, Germany, which gave us permission to publish this paper elsewhere.
- Neeraj Dhungel, Gustavo Carneiro, Andrew P. Bradley. The Automated Learning of Deep Features for Breast Mass Classification from Mammograms. *19th International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, 2016.  
(Accepted for Publication)  
MICCAI 2016 has awarded me with a student travel grant for this paper.
- Neeraj Dhungel, Gustavo Carneiro, Andrew P. Bradley. A Deep Learning approach to fully automated analysis of Masses in Mammograms. *Submitted to Medical Image Analysis (MedIA)* .  
(Under Review)



# *Acknowledgements*

I would like to extend my sincere gratitude to all my family, friends and supervisors who stood behind me and encouraged during my PhD research.

More importantly, I would like to thank my principal supervisor Assoc. Prof. Gustavo Carneiro for his excellent guidance and mentorship that allowed me to complete my PhD in the given time-frame. He also gave me the opportunity to work on machine learning during my PhD despite knowing my limited background in that topic.

I would like to appreciate the help and supervision provided by Prof. Andrew P. Bradley (The University of Queensland), my co-supervisor Dr. Tat-Jun Chin and Dr. Zhi Lu (University of South Australia). I would also like to thank my friends Zhibin Lao, Gabriel Maicas and Dr. Tuan Anh Ngo (Vietnam National University of Agriculture), who have shared with me the knowledge as well as the experience in other areas of life.

I would also like to thank the graduate office of The University of Adelaide for providing me with the Adelaide Scholarship International (ASI) and the Australian Research Council's Discovery Projects funding scheme (project DP140102794) for providing me financial support to attend various international and national conferences. Finally, I also thank all the staff of the Australian Centre for Visual Technologies (ACVT), School of Computer Science for academic and ethical work environment.

Finally, I would also like to acknowledge the funding provided by Undergraduate Grant Commission (UGC), Nepal, when I was working as a faculty member at the Kathmandu University, Dhulikhel, Nepal for our project "Brain Tumour Segmentation", which motivated me to pursue the PhD degree in the field of medical image analysis at The University of Adelaide. I would like to thank my dearest friend Sujan Adhikari, who was also the team leader of this project (now pursuing PhD degree at Nanyang Technological University, Singapore) for his help and support.

*August, 2016*

*Neeraj Dhungel*

*Dedicated to my family especially for my wife and friends for their  
unconditional love and support.*