Automated Quantitative and Qualitative Analysis of Neuroblastoma Cancer Tissue

Siamak Tafavogh

Doctor of Philosophy
University of Technology Sydney
2014

Declaration

I declare that this thesis was composed by myself and that the work contained therein is my own, except where explicitly stated otherwise in the text.

(**Siamak Tafavogh**)

To My Beloved Parents

Acknowledgements

I would like to express my sincere gratitude to my PhD principle supervisor, Associate Prof. Paul Kennedy for all the effort that he put in during my PhD. During the PhD, there were critical times which would not have been overcome without his unsparing help. His professional, academic attitude, sense of humour, conscientious working style and generous personality will be of great benefit to me in my future research work and life.

I also would like to thank my co-supervisor Dr. Daniel Catchpoole, Head of the Tumour Bank at the Children's Hospital, Westmead for his support and advice, especially in clinical domain. He provided the project with necessary clinical resources, and orchestrated the clinical team which participated in my PhD.

I would like to express great appreciation to Dr. Sedighe Vajar from Marzdaran pathology laboratory, my co-supervisor Dr. Karla Felix Navarro, Faculty of Engineering and Information Technology, Dr Nicole Graph, Head of the Department of Histopathology Children's Hospital Westmead, and Drs Amanda Charlton and Susan Arbuckle senior specialist anatomical pathologists from the Department of Histopathology, Children's Hospital, Westmead. I would like also to thank Ms. Birgit Smith for helping me to identify and correct grammar, syntax and presentation problems in my thesis.

Most of all I would like to extend my deepest gratitude to my parents Sedighe and Kambiz who made this possible with their constant support. Thank you for always being there for me during the ups and downs and for believing in me more than I do in myself. My past and my future are due to your devotion. I hope there will be a chance to compensate you for that.

Abstract

The goal of this thesis is to develop an innovative Computer Aided Diagnosis (CAD) system for the common deadly infant cancer of Neuroblastoma. Neuroblastoma accounts for more than 15% of childhood cancer deaths, and it has the lowest survival rate among the paediatric cancers in Australia. In quantitative analysis the total number of different regions of interest are counted, and qualitative analysis determines abnormalities within the tumour.

Quantitative and qualitative analysis of tumor samples under the microscope is one of the key markers used by pathologists to determine the aggressiveness of the cancer, and consequently its therapy. Because of the variety of the histological region types and histological structures in the tissue, analyzing them under the microscope is a tedious and error-prone task for pathologists. The negative effects of inaccurate quantitative and qualitative analysis have led to an urgent call from pathologists for accurate, consistent and automated approaches.

Computer Aided Diagnosis (CAD) is an automated cancer diagnostic and prognostic system which enhances the ability of pathologists in the quantitative and qualitative analysis of tumor tissues. However, there are four main issues with developing a CAD system for pathology labs: First is the fluctuating quality of the histological images. Second is a wide range of different types of histological regions and histological structures with complex morphology each adopting a

specific algorithm. Third is overlapping cells which decrease the accuracy of quantitative analysis. Fourth is a lack of utility for pathology labs when they do not follow an appropriate clinical prognosis scheme. Moreover, most of the proposed CAD systems perform either quantitative or quantitative analysis and only very few of them manipulate both types of analysis on the cancerous tumor tissue.

This thesis aims to address the issues raised by developing an innovative CAD system that assists pathologists in determining a more appropriate prognosis for the leading infant cancer of Neuroblastoma. The CAD will automatically perform quantitative and qualitative analysis on images of tumor tissue to extract specific histological regions and histological structures which are used for determining the prognosis for Neuroblastoma.

This thesis has four main contributions. Contribution 1 develops novel algorithms to enhance the quality of histological images by reducing the wide range of intensity variations. Contribution 2 proposes a series of segmentation algorithms for extracting different types of histological regions and histological structures. Contribution 3 addresses the issue of overlapping cells by developing algorithms for splitting them into single cells. Contribution 4 grades the aggressiveness level of neuroblastoma tumor by developing a prognosis decision engine.

The main outcomes of the proposed CAD system in this thesis are a series of novel algorithms for enhancing the quality of the histological images and for segmenting histological regions and histological structures of interests, introducing a prognosis decision engine for grading a neuroblastoma tumor based on a well established histopathological scheme, facilitating the process of prognosis and tumor classification by performing accurate and consistent quantitative and qualitative tissue analysis, and enhancing digital pathology by incorporating a

digital and automated system in the work flow of pathologists.

The performance of all the developed algorithms in this thesis in terms of correctly extracting histological regions, histological structures and grading the level of tumor aggressiveness, is evaluated by a pathologist from the department of histopathology in the Children's Hospital at Westmead, Sydney. Moreover, all the results are compared with state of the art methods. The results indicate that the algorithms proposed in this thesis outperform state of the art quantitative and qualitative methods of analysis.

Publications Related to this Thesis

Below is the list of the peer-reviewed journal and conference papers associated with my PhD research:

Refereed International Journals

1. **S. Tafavogh**, K. Felix Navarro, D. R. Catchpoole, and P. J. Kennedy. "Non-parametric and integrated framework for segmenting and counting neuroblastic cells within neuroblastoma tumor images." *Medical and Biological Engineering and Computing*, Springer, 51:645–655, 2013. (PMID: 23359256). This paper specifically addresses Chapter 4 and Chapter 5.

2. November 2013: submitted to journal of IEEE Transactions on Information Technology in Biomedicine, **S. Tafavogh**, D. R. Catchpoole and P. J. Kennedy, "Cellular quantitative analysis and splitting overlapping cells using shortest path between critical points on the concave regions". This paper specifically addresses Chapter 5.

3. December 2013: submitted to journal of BMC Bioinformatics, **S. Tafavogh**, K. Felix Navarro, D. R. Catchpoole and P. J. Kennedy "Segmentation and quantitative analysis of ganglion cells using histological and topological filters".

4. January 2014: submitted to journal of Microarray, A. Chetcuti, N. Mackie, **S. Tafavogh**, N. Graph, A. Charlton, T. Hinwood, D. R. Catchpoole, "Can archival tissue reveal answers to modern research questions?: Computer-aided histological assessment of neuroblastoma tumours collected over 60 years". This paper addresses Chapter 6.

Refereed International Conference Publications

1. **S. Tafavogh**, K. Felix Navarro, D. R. Catchpoole and P. J. Kennedy (2013). "Segmenting cellular regions of neuroblastoma tumor and splitting overlapping cells using shortest path between convex regions of cell contours", in *A*rtificial Intelligence in Medicine, Springer, LNCS 7885, N. Peek, R. M. Morales and M. Peleg, Eds, Springer. 171-175. This paper addresses Chapter 5.

2. **S. Tafavogh**, D.R. Catchpoole, P.J. Kennedy (2012). "Determining cellularity status of tumors based on histopathology using hybrid image segmentation." IEEE World Congress on Computational Intelligence, Australia, pp:1-8. This paper addresses Chapter 4

3. **S. Tafavogh**. Comparing clustering methods and color-spaces for histological images, in Proceedings of the IEEE International Symposium on Image and Signal Processing and Analysis, pp: 154-59, 2011. This paper addresses Chapter 4.

4. **S. Tafavogh**. "Effect of HSV and L*a*b* color spaces on segmenting histological images by expectation maximisation algorithm", Conference on Computational Molecular Biology and Bioinformatics, Moscow, vol. 22, pp: 359-361, 2011. This paper addresses Chapter 4.

5. **S. Tafavogh**. "The effects of adaptive histogram equalization and color spaces on histological segmentation quality of expectation maximization method". International Conference on Graphic and Image Processing (ICGIP), October 2011, Cairo, Egypt. This paper addresses Chapter 4.

Contents

Abstract **9**

1 Introduction **23**

1.1 Conventional Pathology and Associated Errors 25

1.2 Digital Pathology 26

1.3 How does a CAD System Analyze Tissue Samples? 26

1.4 Research Challenges 28

1.5 Research Aims and Contributions 29

1.6 Research Significance 32

1.7 Research Hypothesis 33

1.8 Thesis Structure 33

2 Literature **35**

2.1 Biological Domain 35

2.1.1 Histological Regions and Histological Structures of NT . . 36

2.1.2 Prognosis Scheme 37

2.1.3 Errors in Pathology 40

2.2 Algorithms and Automated Systems 40

2.2.1 Algorithms and Systems for Quantitative Analysis 40

2.2.2 Algorithms and Systems for Qualitative analysis 56
2.3 Research Gaps 66

3 Model of the CAD System **69**
3.1 Stages of the Model 69
3.2 The Overall Research Strategy 72
3.2.1 Awareness of Problems 72
3.2.2 Suggestions 72
3.2.3 Development 72
3.2.4 Evaluation 73
3.2.5 Conclusion 73
3.3 Data Acquisition, Deployed Software and System Validation . . . 73

4 Segmenting Histological Regions and Histological Structures in H&E Stained Histological Images **77**
4.1 Pre-Segmentation 78
4.1.1 Methodology 79
4.1.2 Proposed Approach 80
4.2 Segmenting Cellular Regions, Neuropil Regions, Ganglion Cells and Cytoplasm Components 86
4.2.1 Methodology 88
4.2.2 Proposed Approach 89
4.2.3 System Validation and Experimental Results for Segmenting Cellular Regions and Ganglion Cells 100
4.2.4 System Validation and Experimental Results for Segmenting Neuropil Regions 104
4.3 Segmenting Neuroblast Cells 107

4.3.1 Methodology 108
4.3.2 Proposed Approach 109
4.3.3 System Validation 112
4.3.4 Experimental Results 113
4.4 Related Publications 116
4.5 Contribution, Discussion and Conclusion 117

5 Post-Segmentation: Splitting Overlapping Cells 121
5.1 Methodology 122
5.2 Proposed Approach 124
5.2.1 Stage 1: Distinguishing between Types of Cells 124
5.2.2 Stage 2: Splitting Overlapped Cells 127
5.3 System Validation 138
5.4 Experimental Results 139
5.5 Related Publications 143
5.6 Contribution, Discussion and Conclusion 144

6 Neuroblastoma Classification: Quantitative Analysis 147
6.1 Methodology 148
6.2 Experimental Approach 149
6.2.1 Prognosis Decision Making Engine 149
6.3 System Validation 151
6.4 Experimental Results 152
6.5 Contribution, Discussion and Conclusion 153

7 Conclusions 155
7.1 Discussion 158

7.2 Future Study 159

Bibliography **174**

Table of Symbols

Symbols	Description
n	The total number of pixels in a image
X_i	A pixel i in a image where $i = \{1, ..., n\}$
K	Radially symmetric kernel
h	Mean-shift windows radius
$C_{k,d}$	Positive normalization constant
$\hat{f}_{h_S,h_I}(X)$	Gradient density estimator
G	Kernel window
k and g	Profiles of kernel K and G respectively
m	Mean-shift
X^s	Spatial domain of a pixel
X^i	Intensity domain of a pixel
W_w	Center of kernel w
$X^s_{w,r}$	Spatial domain of convergence points
$Y^i_{w,r}$	Intensity domain of convergence points
V_i	Luminance of pixel i
V	Luminance of white pixel
u'_i and v'_i	Chromaticity coordinates of pixel i
u' and v'	Chromaticity coordinates of white pixel

Symbols	Description
h_s	Spatial bandwidth
h_i	Intensity bandwidth
X^I	Intensity domain of pixel in 3-d vector
X^S	Spatial domain of pixel in 3-d vector
y_r	Mean-shift convergence
B	Total number of constructed mosaics in an image
$\mathbf{M}_a$	Mosaic a of size $d \times 3$ where $a = \{1, ..., B\}$ and d is the total number of pixels within the mosaic.
θ	A set of all mosaics in an image
$\bar{\mathbf{M}}_a$	The unified intensity mosaic
c	The total number of manually labeled pixels in cellular regions
t	The total number of manually labeled pixels in neuropil regions
$\mathbf{C}$	A matrix contains the manually cellular labeled pixels
$\mathbf{N}$	A matrix contains the manually neuropil labeled pixels
Z and L	The total number of pixels within the manually segmented cellular and neuropil regions respectively
$\bar{\mathbf{C}}$	Cellular filter
$\bar{\mathbf{N}}$	Neuropil filter
D_a	Euclidean distance between mosaic a and the filters
P_a	Vector which indicates cellular mosaics
AG_C	Absorption grade for cellular filter
AG_N	Absorption grade for neuropil filter
ρ	A vector of $B \times 1$ which indicates all the mosaics of neuropil regions

Symbols	Description
β	A column vector of length B which indicates all the mosaics of new nuclei regions.
ϑ	Vector which indicates the constituent mosaics of cellular regions (unrefined)
ζ	Vector which indicates the constituent mosaics of neuropil regions within ϑ
ϑ'	Vector which indicates the constituent mosaics of cellular regions (refined)
ϱ	Vector of size B which indicates cytoplasm regions
S	The total number of segmented nuclei
b	The total number of pixels in the centroid of the nuclei
$\mathbf{W}_j$	Centroid nuclei matrix for nuclei image
$\mathbf{Q}_j$	Centroid nuclei matrix for cellular image
$\mathbf{\Delta}_j$	A matrix that contains cytoplasm regions
$\mathbf{av}_j$	The average distance between centroid of nuclei j to background
$\mathbf{\bar{av}}_j$	The average distance of all b pixels of centroid j
$\mathbf{aw}_j$	The average distance of eight neighbor pixels for every pixel at centroid j to background in $\mathbf{W}_j$
$\mathbf{aq}_j$	The average distance of the eight neighbor pixels to background in $\mathbf{Q}_j$ for every pixel at centroid j
$\mathbf{\bar{aw}}$	The average distance of all b pixels of centroid j in **aw**
$\mathbf{\bar{aq}}$	The average distance of all b pixels of centroid j in **aq**
O	The total number of objects in an image
o	An object number

Symbols	**Description**
A_o	Area of object o
DE_o	The ratio of the minor diameter to the major diameter of object o
$\hat{A}_o$	Convex hull of object o
PO	Potentially Overlapping
PS	Potentially Single
SC	Single Cell
OC	Overlapping Cell
F	The total number of overlapping cells
l	Overlapping cells l where $l \in \{1, ..., F\}$
R_l	Entire region of overlapping cell l
$\hat{R}_l$	Convex hull region of overlapping cell l
H	The total number of splitting triangles for overlapping cell l
ST_x	Splitting Triangle $\forall x \in \{1, ..., H\}$
ϕ_l	A set of splitting triangles for overlapping cell l
$e_l(ST_x)$	The edge of splitting triangle x for overlapping cell l
$e_l(\hat{R})$	The edge of convex hull for overlapping cell l
$chord_l(ST_x)$	The edge of chord of splitting triangle x for overlapping cell l
$arc_l(ST_x)$	The edge of arc of splitting triangle x for overlapping cell l
EA	A matrix that contains pixels at the arc of overlapping cell l
EC	A matrix that contains pixels at the chord of overlapping cell l
DA	Distance matrix based on **EA**
DC	Distance matrix based on **EC**
DM	Distance between the pixels at the arc
	and chord of a splitting triangle
SP	The total number of initial splitting points for overlapping cell l

Symbols	Description
$\mathbf{DM}$	Distance matrix
$\mathbf{\Theta}_p$	The distance between object p from the $SP-1$ other objects in an overlapping cell l
$e(R)_l$	The edge of overlapping cell l
$\mathbf{F}_p$	A set of $SP-1$ matrices that have the smallest elements of the matrices in $\mathbf{\Theta}_p$
PS	The point sets
STL	The structuring elements
$\mathbf{T}_p$	The splitting routes for initial splitting point p
$\mathbf{J}_p^*$	A splitting route for initial splitting point p
np	The total number of pixels within the segmented neuropil regions
cp	The total number of pixels within the segmented cellular regions
NR	The amount of neuropil regions within an image
$ganglion$	The total number of identified ganglion cells in the segmented cellular regions
$neuroblast$	The total number of identified neuroblast cells in the segmented cellular regions
GR	The amount of ganglion cells within an image