

UNIVERSITY of CALIFORNIA  
Santa Barbara

**Object- and Spatial-Level Quantitative Analysis of Multispectral  
Histopathology Images for Detection and Characterization of Cancer**

A Dissertation submitted in partial satisfaction of the  
requirements for the degree

Doctor of Philosophy  
in  
Electrical and Computer Engineering

by

Laura E. Boucheron

Committee in charge:

Professor B. S. Manjunath, Chair  
Professor Shivkumar Chandrasekaran  
Professor Miguel Eckstein  
Doctor Neal R. Harvey  
Professor Kenneth Rose

March 2008

The dissertation of Laura E. Boucheron is approved.

---

Professor Shivkumar Chandrasekaran

---

Professor Miguel Eckstein

---

Doctor Neal R. Harvey

---

Professor Kenneth Rose

---

Professor B. S. Manjunath, Committee Chair

December 2007

Object- and Spatial-Level Quantitative Analysis of Multispectral Histopathology

Images for Detection and Characterization of Cancer

Copyright © 2008

by

Laura E. Boucheron

*To Steven,*

*who encouraged me to get my Ph.D.,*

*and put up with me while I got it.*

## Acknowledgements

Thanks to my advisor Dr. Manjunath for providing me with the opportunity to pursue my graduate studies at UCSB, for his guidance in my research, and for teaching me to keep my eye on the larger picture. Thanks also to my committee members, Drs. Chandrasekaran, Eckstein, Harvey, and Rose for their valuable inputs to and viewpoints on my research. Thanks to the IGERT IDM program for financial support during the first two years of my Ph.D. studies.

Thanks to my housemates, Chiou-Wei Tsai, Ed Hill, and Mike Quinn who made my home-life far more interesting than it ever would have been living alone. Also, thanks to our honorary housemate Colleen Schwartz with whom there is never a dull moment. You are all great friends, and I'm blessed to know you.

Thanks to my fellow VRL people, it has been an honor to get to know you all. In particular, I would like to acknowledge Nhat Vu, the first person I met at UCSB, for making me laugh and for his unique comments and suggestions about my research. Jiyun Byun has been a constant source of encouragement to me, tackling the same academic frustrations six months ahead of me. She has also been an excellent sounding board for the technical details of my research. Mike Quinn has been a wonderful friend, and (while I can't really thank him for it) the brother I never knew I didn't want. Thanks to Zhiqiang Bi for performing the SVM classifications for me and Luca Bertelli for applying his GPAC method to my data. Thanks to Dmitry Fedorov, Marco Zuliani, Elisa Drelie Gelasca, and Luca Bertelli for the always interesting conversations, as well as useful suggestions for my research. Or should it be rather: Merci pour les conversations toujours intéressantes, et suggestions utiles pour ma recherche. Special thanks to Mike, Chiou-Wei, Jiyun, and Elisa for allowing me to couch surf at their places during my many visits.

Thanks to Barbara Drescher, for introducing me to the vision side of psychology, and for her collaborations in my first year.

Many thanks to my team members at Los Alamos for their willingness to support me in my research, and providing me with suggestions, ideas, and feedback, as well as helping me get up to speed with GENIE. Particular thanks to Steven Brumby, Nancy David, Damian Eads, Kim Edlund, Karen Glocer, Neal Harvey, Rohan Loveland, Matt McCabe, Norma Pawley, Reid Porter, Lakshman Prasad, Alexei Skurikhin, and James Theiler. Special thanks to Neal (Harve) Harvey, my mentor and committee member, for his advice and guidance in my research endeavors, and to Karen Glocer for her technical input on feature selection methods.

Thanks to my many friends who supported me over the years. Special thanks to the canine search and rescue bunch, Kathryn Berchtold, Wendee Brunish, and Jennifer Young for providing me with an escape to the woods and some puppy time; thanks also to Abby, Halley, Tobi, and Maggie for providing a much needed distraction. Thanks to Will Dearholt to whom I owe uncountable dinners "sometime when it's really expensive,"

for the encouragement, Jeep and hiking excursions, and reminders that life is often easier if you don't get wound up about things. Thanks to the many others, especially Jeremy Ayers, who have at some point witnessed the wrath of my final sleepless days and still speak to me.

Of course, great thanks to my family for their unconditional love and support through all these years, even if I didn't listen to them and come back home. My parents, Susan and Edward, led by example when they decided to go to (and finish) graduate school. I hope I have served as an example for my sisters, Alina and Leandra, who are in the middle of their college careers.

Last, but certainly not least, thanks to my boyfriend Steven, who has stuck with me throughout the whole process, including my often bi-weekly assertions that I was going to drop out and my less than exemplary exhibitions of grace under pressure in the final months.

# **Curriculum Vitæ**

Laura E. Boucheron

## **Education**

<b>March 2008</b>	<b>Doctor of Philosophy</b> Department of Electrical and Computer Engineering University of California Santa Barbara Santa Barbara, CA
<b>May 2003</b>	<b>Master of Science</b> Department of Electrical and Computer Engineering New Mexico State University Las Cruces, NM
<b>May 2001</b>	<b>Bachelor of Science</b> Department of Electrical and Computer Engineering New Mexico State University Las Cruces, NM

## **Experience**

<b>2005-present</b>	<b>Graduate Research Assistant</b> Space and Remote Sensing Sciences Los Alamos National Laboratory Los Alamos, NM
<b>2003-2005</b>	<b>IGERT Fellow in Interactive Digital Multimedia</b> Department of Electrical and Computer Engineering University of California Santa Barbara Santa Barbara, CA
<b>2003</b>	<b>Summer Intern</b> Theoretical Division Los Alamos National Laboratory Los Alamos, NM
<b>2002-2003</b>	<b>Graduate Research Assistant</b> Department of Electrical and Computer Engineering New Mexico State University Las Cruces, NM
<b>2001</b>	<b>Summer Intern</b> Agilent Technologies Rohnert Park, CA

**1998-2000**

**Summer Intern**

Sandia National Laboratories  
Albuquerque, NM

**Publications**

L. E. Boucheron, N. R. Harvey, and B. S. Manjunath, “On Using Cytoplasm and Stroma Features in Automated Cancer Diagnosis,” To be presented at: *The 2008 Workshop on Bio-Image Informatics: Biological Imaging, Computer Vision, and Data Mining*, Santa Barbara, CA, Jan 2008.

L. E. Boucheron, N. R. Harvey, and B. S. Manjunath, “A Quantitative Object-Level Metric for Segmentation Performance and Its Application to Cell Nuclei,” In Proceedings: *International Symposium on Visual Computing*, LNCS 4841, pp. 208-219, Nov 2007.

L. E. Boucheron, Z. Bi, N. R. Harvey, B. S. Manjunath, and D. L. Rimm, “Utility of Multispectral Imaging for Nuclear Classification of Routine Clinical Histopathology Imagery,” *BMC Cell Biology*, 8(Suppl 1):S8, (10 July 2007), <http://www.biomedcentral.com/1471-2121/8/S8>.

L. E. Boucheron, N. R. Harvey, and B. S. Manjunath, “Utility of Multispectral Imaging for Analysis of Routine Clinical Histopathology Imagery,” Presented at: *The 2006 Workshop on Multiscale Biological Imaging, Data Mining, & Informatics*, Santa Barbara, CA, Sep 2006.

L. E. Boucheron, “Compression of digital elevation maps for fast and efficient search and retrieval,” Masters thesis, New Mexico State University, May 2003.

L. E. Boucheron and C. D. Creusere, “Lossless wavelet-based compression of digital elevation maps for fast and efficient search and retrieval,” *IEEE Transactions on Geoscience and Remote Sensing*, vol. 43, no. 5, May 2005, pp. 1210-1214.

L. E. Boucheron and C. D. Creusere, “Compression of digital elevation maps for fast and efficient search and retrieval,” In Proceedings: *2003 International Conference on Image Processing*, vol. 1, pp. 629-632, Sep 2003.

## Abstract

### Object- and Spatial-Level Quantitative Analysis of Multispectral Histopathology Images for Detection and Characterization of Cancer

Laura E. Boucheron

The main goal of this dissertation is the development and discussion of techniques for higher-level image analysis, i.e., object-level analysis, of breast cancer imagery. Established cytologic (cell) criteria can be contradictory, and even histologic (tissue) criteria (considered the gold standard for diagnosis) are subject to varied interpretation. There is thus a need to quantitatively define characteristics of breast cancer to better coordinate clinical care of women presenting breast masses. We propose here an approach for such quantitative analysis, Quantitative Object- and spatial Arrangement-Level Analysis (QOALA), using expert (pathologist) input to guide the classification process.

The main contributions in this work are four-fold. First, we quantitatively analyze the utility of multispectral imagery for classification and segmentation tasks in histopathology imagery. Second, we develop object-level segmentations for several histologic classes, as well as a quantitative object-level segmentation metric. Third, we extract a comprehensive set of both object- and spatial-level features which are used in a feature selection framework for classification of objects and imagery. Fourth, we extend the concepts of object-level features to higher-level image objects, analyze the utility of these high-level objects for image classification, and introduce the concept of a probabilistic graph-based model of imagery.

Overall, QOALA yields very good object- and image-level classification performances. More specifically, the object-level features as implemented in QOALA are versatile and general enough to elicit important information from even imperfectly segmented objects. Additionally, the use of non-nuclear features, namely features of cytoplasm and stroma have good classification performance, often exceeding that of nuclei. Higher-level features display a potential to increase both object- and image-level classification performance.



# Contents

<b>List of Tables</b>	<b>xvii</b>
<b>List of Figures</b>	<b>xxi</b>
<b>1 Introduction</b>	<b>1</b>
1.1 Motivation . . . . .	1
1.2 A Brief Introduction to Some Pathology . . . . .	2
1.2.1 Histo- and Cyto-pathology . . . . .	2
1.2.2 Imagery and Multispectral Image Acquisition . . . . .	4
1.2.3 Specimen Preparation . . . . .	4
1.2.4 Cytology Characteristics of Breast Cancer . . . . .	5
1.3 Analysis of Histo/Cytopathology Images . . . . .	6
1.3.1 Challenges . . . . .	6
1.3.2 Commonly Used Performance Metrics . . . . .	7
1.3.3 State-of-the-Art . . . . .	8
1.4 Overview of the QOALA Framework . . . . .	14
1.4.1 Related Work . . . . .	16
1.4.2 QOALA Framework . . . . .	17
1.5 Outline of Dissertation and Research Contributions . . . . .	18
<b>2 Multispectral Analysis of Pixel-Level Nuclear Classification</b>	<b>23</b>
2.1 Motivation . . . . .	23
2.2 Related Work . . . . .	24
2.3 Computation of Derived RGB Images . . . . .	24
2.4 Pixel-Level Classification . . . . .	28
2.4.1 Classifiers . . . . .	28
2.4.2 Performance Metric . . . . .	32
2.5 Pixel-Level Classification Results . . . . .	33
2.5.1 Classification Using All Available Image Bands . . . . .	33

2.5.2	Classification Using Single Image Bands . . . . .	38
2.5.3	Single Image Bands versus All Image Bands . . . . .	42
2.5.4	Analysis of FLDA Coefficients and Bands Chosen in AFE Solutions	45
2.5.5	Entropic Analysis of Image Bands . . . . .	47
2.5.6	Principal Components Analysis of Multispectral Images . . . . .	49
2.6	Summary . . . . .	50
2.7	Future Directions . . . . .	52
<b>3</b>	<b>Pixel-Level Classification of Cytoplasm and Stroma</b>	<b>53</b>
3.1	Motivation . . . . .	53
3.2	Related Work . . . . .	54
3.3	Pixel-Level Classification . . . . .	57
3.3.1	k-means Clustering and Region Merging . . . . .	57
3.3.2	Classification Results . . . . .	58
3.4	Discrimination between Stroma and Cytoplasm . . . . .	59
3.4.1	Gabor Feature-Based Discrimination . . . . .	59
3.4.2	Nuclear Proximity, Gabor, and Spectral Feature-Based Discrimination . . . . .	62
3.4.3	Genie Pro Feature-Based Discrimination . . . . .	65
3.5	Summary . . . . .	74
3.6	Future Directions . . . . .	75
<b>4</b>	<b>Nuclei Segmentation: Methods and Evaluation Metrics</b>	<b>77</b>
4.1	Motivation . . . . .	77
4.2	Related Work . . . . .	79
4.2.1	Object-Level Analysis . . . . .	79
4.2.2	Segmentation Evaluation . . . . .	79
4.2.3	Nuclear Segmentation . . . . .	80
4.2.4	Segmentation of Cytoplasm and Stroma . . . . .	81
4.3	Segmentation Metric . . . . .	82
4.3.1	Definition . . . . .	83
4.3.2	Metric Variation versus Segmentation Quality . . . . .	88
4.3.3	Application to Non-Elliptically Shaped Objects . . . . .	91
4.3.4	Comparison to Other Segmentation Metrics . . . . .	91
4.3.5	Ground Truth Image Markup . . . . .	94
4.4	Watershed-Based Segmentation of Nuclei . . . . .	96
4.4.1	Watershed on the Complemented Distance Transform (WSCDT) .	96

4.4.2	Watershed on H-minima Transforms (WSHmin) . . . . .	97
4.4.3	Watershed Using Granulometry-Based Markers (WSGran) . . . . .	102
4.4.4	Blobdetector-Based Segmentation (WSBlob) . . . . .	112
4.5	Concavity-Based Segmentation of Nuclei . . . . .	114
4.5.1	Overview of the Concavity-Based Segmentation Algorithm (Kumar)	114
4.5.2	Ground Truth Markup . . . . .	118
4.5.3	Choice of Thresholds . . . . .	119
4.5.4	Results . . . . .	122
4.5.5	Concavity-Based Segmentation as Markers for Watershed Segmentation (WSCM) . . . . .	124
4.5.6	A Note on Computational Efficiency . . . . .	128
4.6	Ideal Maximum Segmentation Performance of Nuclei . . . . .	129
4.7	Segmentation of Cytoplasm and Stroma . . . . .	133
4.7.1	Voronoi-Based Segmentation of Cytoplasm . . . . .	133
4.7.2	Connected Components-Based Segmentation of Stroma . . . . .	133
4.7.3	Voronoi-Based Segmentation of Cytoplasm and Stroma . . . . .	133
4.8	Summary . . . . .	133
4.8.1	Segmentation Evaluation . . . . .	135
4.8.2	Nuclear Segmentation . . . . .	135
4.9	Future Directions . . . . .	136
4.9.1	Segmentation Evaluation . . . . .	136
4.9.2	Nuclear Segmentation . . . . .	137
<b>5</b>	<b>Feature Extraction and Selection</b>	<b>139</b>
5.1	Motivation . . . . .	139
5.2	Related Work . . . . .	140
5.3	Object-Level Feature Definitions . . . . .	141
5.3.1	Size and Shape Features . . . . .	141
5.3.2	Radiometric and Densitometric Features . . . . .	147
5.3.3	Texture Features . . . . .	148
5.3.4	Chromatin Density Features . . . . .	153
5.3.5	Fractal Dimension Features . . . . .	155
5.3.6	Summary . . . . .	156
5.4	Spatial-Relation Feature Definitions . . . . .	156
5.4.1	Graph and Tree Structures . . . . .	158

5.4.2	Spatial-Relation Features . . . . .	161
5.4.3	Summary . . . . .	165
5.5	Region-Based Spatial-Relation Features . . . . .	167
5.6	Feature Selection and Classification Performance . . . . .	174
5.6.1	An Overview of Grafting . . . . .	174
5.6.2	Object-Level Feature Selection and Performance . . . . .	175
5.6.3	Spatial-Relation Feature Selection and Performance . . . . .	179
5.6.4	Combining Object-Level and Spatial-Relation Features . . . . .	181
5.6.5	Image-Level Performances . . . . .	181
5.7	Feature Subsets . . . . .	187
5.7.1	Object-Level Feature Subsets . . . . .	187
5.7.2	Important Object-Level Features . . . . .	188
5.7.3	Spatial-Relation and Combined Object and Spatial Feature Subsets	188
5.8	Summary . . . . .	190
5.9	Future Directions . . . . .	192
<b>6</b>	<b>Higher-Level Objects</b>	<b>193</b>
6.1	Motivation . . . . .	193
6.2	Related Work . . . . .	193
6.3	Higher Level Segmentation . . . . .	194
6.3.1	An Introduction to the Region Connection Calculus (RCC) . . . . .	195
6.3.2	Implementation of RCC . . . . .	197
6.3.3	Results . . . . .	199
6.4	Use of Higher-Level Objects for Classification . . . . .	199
6.4.1	The Plausibility of Using Feature Graphs . . . . .	202
6.4.2	Higher-Level Objects . . . . .	204
6.4.3	Results . . . . .	207
6.4.4	Higher-Level Object Feature Subsets . . . . .	209
6.5	Summary . . . . .	210
6.6	Future Directions . . . . .	210
6.6.1	Higher-Level Objects . . . . .	210
6.6.2	Introduction to Probabilistic Graph Models (PGMs) . . . . .	210
<b>7</b>	<b>Summary and Conclusions</b>	<b>213</b>
7.1	Summary of Research Contributions . . . . .	213

7.1.1	Chapter 2: Multispectral Analysis of Pixel-Level Nuclear Classification . . . . .	213
7.1.2	Chapter 3: Pixel-Level Classification of Cytoplasm and Stroma . . . . .	214
7.1.3	Chapter 4: Nuclei Segmentation: Methods and Evaluation Metrics . . . . .	215
7.1.4	Chapter 5: Feature Extraction and Selection . . . . .	216
7.1.5	Chapter 6: Higher-Level Objects . . . . .	218
7.2	Conclusion . . . . .	219
	<b>Bibliography</b>	<b>221</b>
	<b>Appendices</b>	<b>238</b>
A	Glossary of Pathology Terminology	239
B	Glossary of Machine Learning and Computer Vision Terminology	243
C	Glossary of Acronyms and Symbols	247
C.1	Acronyms . . . . .	247
C.2	Symbols . . . . .	250
D	Color- and Texture-Based Classification of Nuclei	251
E	An Introduction to GENIE and Genie Pro	253
F	Extra Plots: Classification of Individual RGB Image Bands	255
G	Ranking of Extra Pixels and Pixels Missed by Nucleus	259
H	Optical Density of Transmission Microscopy Images	269
I	Extra Plots: Random Graph Features	271
J	Grafting Feature Subsets: Nuclei	281
K	Grafting Feature Subsets: Object- and Spatial-Level	293
L	Grafting Feature Subsets: Higher-Level Objects	321



# List of Tables

1.1	General state-of-the-art histo/cytopathology image analysis. . . . .	12
1.2	State-of-the-art breast cancer histo/cytopathology image analysis. . . . .	15
2.1	DR, FAR, and P using all image bands. . . . .	34
2.2	Performance using all image bands for different training sets. . . . .	36
2.3	Wilcoxon p-values for multispectral versus RGB. . . . .	37
2.4	Maximum performance using all bands and single image bands. . . . .	42
2.5	Wilcoxon p-values for multispectral versus RGB bands. . . . .	43
2.6	Wilcoxon p-values for all bands versus single bands. . . . .	43
2.7	Entropy of image regions. . . . .	49
3.1	DR, FAR, and P for k-means classification. . . . .	60
3.2	OS DR, FAR, and P for Section 2.4.1 classifiers. . . . .	60
3.3	Gabor cytoplasm/stroma discrimination. . . . .	61
3.4	SFFS cytoplasm/stroma discrimination. . . . .	65
3.5	Genie Pro cytoplasm-stroma results. . . . .	66
3.6	Genie Pro single attribute plane results, multispectral. . . . .	73
3.7	Genie Pro single attribute plane results, ccd. . . . .	74
4.1	Wilcoxon p-values for multispectral versus RGB for WSCDT. . . . .	97
4.2	Wilcoxon p-values for multispectral versus RGB for WSHmin. . . . .	100
4.3	Wilcoxon p-values for multispectral versus RGB for WSGran. . . . .	109
4.4	Wilcoxon p-values for multispectral versus RGB for WSBlob. . . . .	113
4.5	In-sample thresholds for concavity-based segmentation. . . . .	121
4.6	Wilcoxon p-values for multispectral versus RGB for Kumar. . . . .	123
4.7	Wilcoxon p-values for multispectral versus RGB for WSCM. . . . .	128
4.8	Average computation time per image. . . . .	128
4.9	Ideal nuclear segmentation performance. . . . .	129

4.10	Ideal nuclear segmentation term performance.	132
5.1	Summary of object-level features.	157
5.2	Summary of spatial-relation features.	166
5.3	Object-level feature subset performance, $\lambda_1 = 0.05$ .	177
5.4	Object-level feature subset performance, $\lambda_1 = 0.01$ .	177
5.5	Multispectral versus ccd object-level features, $\lambda_1 = 0.05$ .	178
5.6	Multispectral versus ccd object-level features, $\lambda_1 = 0.01$ .	179
5.7	Spatial-relation feature subset performance.	179
5.8	Region-based spatial-relation feature subset performance.	180
5.9	Object-level and spatial-relation feature subset performance, $\lambda_1 = 0.05$ .	183
5.10	Object-level and spatial-relation feature subset performance, $\lambda_1 = 0.01$ .	183
5.11	AUC for OS image-level classification.	183
5.12	AUC for OS spatial-relation classification.	185
5.13	Statistics of CC object-level feature subsets.	187
5.14	Statistics of WSGran spatial-relation feature subsets.	189
6.1	Summary of high-level object features.	204
6.2	High-level object feature subset performance, $\lambda_1 = 0.05$ .	206
6.3	High-level object feature subset performance, $\lambda_1 = 0.01$ .	206
6.4	AUC for OS image-level classification for high-level objects.	207
6.5	Statistics of GT and WSHmin Cell <sub>NC</sub> feature subsets.	209
J.1	Statistics of object-level nuclei subsets.	282
J.2	Feature subset for GT nuclei, $\lambda_1 = 0.05$ .	282
J.3	Feature subset for CC nuclei, $\lambda_1 = 0.05$ .	282
J.4	Feature subset for WSHmin nuclei, $\lambda_1 = 0.05$ .	283
J.5	Feature subset for WSGran nuclei, $\lambda_1 = 0.05$ .	283
J.6	Feature subset for GT nuclei, $\lambda_1 = 0.01$ .	284
J.7	Feature subset for CC nuclei, $\lambda_1 = 0.01$ .	287
J.8	Feature subset for WSHmin nuclei, $\lambda_1 = 0.01$ .	288
J.9	Feature subset for WSGran nuclei, $\lambda_1 = 0.01$ .	290
K.1	Feature subset for CC nuclei, $\lambda_1 = 0.01$ .	294
K.2	Feature subset for CC cytoplasm/stroma, $\lambda_1 = 0.01$ .	295
K.3	Feature subset for CC cytoplasm, $\lambda_1 = 0.01$ .	298

K.4	Feature subset for CC stroma, $\lambda_1 = 0.01$ .	300
K.5	Feature subset for CC nuclei and cytoplasm/stroma, $\lambda_1 = 0.01$ .	302
K.6	Feature subset for CC nuclei and cytoplasm, $\lambda_1 = 0.01$ .	306
K.7	Feature subset for WSGran spatial nuclei, $\lambda_1 = 0.01$ .	309
K.8	Feature subset for WSGran nuclei and spatial nuclei, $\lambda_1 = 0.01$ .	310
K.9	Feature subset for WSGran nuclei, cytoplasm/stroma, and spatial nuclei, $\lambda_1 = 0.01$ .	312
K.10	Feature subset for WSGran nuclei, cytoplasm, and spatial nuclei, $\lambda_1 = 0.01$ .	316
L.1	Feature subset for GT Cell <sub>NC</sub> , $\lambda_1 = 0.01$ .	322
L.2	Feature subset for WSHmin Cell <sub>NC</sub> , $\lambda_1 = 0.01$ .	325



# List of Figures

1.1	Example cytopathology and histopathology images. . . . .	3
1.2	Example two-step iterative process. . . . .	19
1.3	Flowchart of dissertation research. . . . .	22
2.1	Example multispectral image stack. . . . .	26
2.2	RGB representations of the multispectral stack in Figure 2.1. . . . .	27
2.3	RGB responses for the MatlabCentral function <i>spectrumRGB</i> . . . . .	28
2.4	Average P, DR, and FAR using all image bands. . . . .	35
2.5	P, DR, and FAR on single multispectral bands. . . . .	39
2.6	P, DR, and FAR on single RGB bands. . . . .	41
2.7	Scatter plots of nuclei versus non-nuclei. . . . .	44
2.8	FLDA coefficients. . . . .	46
2.9	Spectral bands chosen in AFE solutions. . . . .	46
2.10	Entropy of images/image regions. . . . .	48
2.11	Logarithmic plot of PCA eigenvalues. . . . .	50
2.12	Projection of example image onto first three eigenvectors. . . . .	51
3.1	Multispectral and ccd average histologic spectra. . . . .	55
3.2	Example of stroma and cytoplasm. . . . .	56
3.3	Gabor output. . . . .	61
3.4	Genie Pro algorithm, multispectral imagery. . . . .	67
3.4	(Cont.) Genie Pro algorithm, multispectral imagery. (continued on next page) . . . . .	68
3.5	Genie Pro algorithm, ccd imagery. . . . .	70
3.5	(Cont.) Genie Pro algorithm, ccd imagery. (continued on next page) . . .	71
4.1	Example two-step iterative process (repeated figure). . . . .	78
4.2	Performance of the QS metric. . . . .	86

4.3	Metric variation versus segmentation quality. . . . .	90
4.4	Application of the QS and size metrics to an example silhouette. . . . .	92
4.5	Comparison of the QS metric and Hausdorff distance. . . . .	93
4.6	Illustration of the object-level image markup process. . . . .	95
4.7	WSCDT segmentation performance using all image bands. . . . .	96
4.8	WSCDT segmentation performance on single image bands. . . . .	98
4.9	WSHmin segmentation performance using all image bands. . . . .	100
4.10	Value of WSHmin performance metric terms for different imagery. . . . .	101
4.11	Example WSHmin segmentations. . . . .	103
4.12	Pattern spectra of dataset images. . . . .	104
4.13	Pattern spectra of example images. . . . .	105
4.14	Example residue images. . . . .	106
4.15	WSGran segmentation performance. . . . .	108
4.16	Value of WSGran performance metric terms on different imagery. . . . .	110
4.17	Example WSGran segmentations. . . . .	111
4.18	Laplacian of Gaussian. . . . .	113
4.19	WSBlob segmentation performance using all image bands. . . . .	114
4.20	Example WSBlob segmentations. . . . .	115
4.21	Illustrative examples of CC and CL alignment metrics. . . . .	116
4.22	Flowchart of the candidate split line selection process. . . . .	117
4.23	Illustration of the Kumar ground truth markup process. . . . .	120
4.24	Kumar segmentation performance. . . . .	122
4.25	Example Kumar segmentation. . . . .	123
4.26	WSCM segmentation performance. . . . .	125
4.27	Value of WSCM performance metric terms for different imagery. . . . .	126
4.28	Example WSCM segmentations. . . . .	127
4.29	Example ideal segmentations. . . . .	130
4.30	Example cytoplasm and stroma segmentations. . . . .	134
5.1	Illustration of the O'Callaghan direction constraint. . . . .	159
5.2	Example graph and tree structures. . . . .	160
5.3	Region-based spatial-relation features. . . . .	167
5.4	Random graphs, type 1. . . . .	169
5.5	Random graphs, type 2. . . . .	170
5.6	Random graph features. . . . .	171

5.7	Different levels of visual information. . . . .	182
5.8	ROC curves for OS image-level performance. . . . .	184
5.9	ROC curves for OS spatial-relation performance. . . . .	186
5.10	ROC curve for tiles. . . . .	191
6.1	The eight base relations of RCC8. . . . .	195
6.2	Example higher-level spatial reasoning using RCC8. . . . .	196
6.3	Proper part versus externally connected. . . . .	198
6.4	Histograms of RCC8 relations between histologic entities. . . . .	200
6.5	Histograms of RCC8 relations between histologic entities, benign and malignant. . . . .	201
6.6	Illustrative “cytology” images. . . . .	202
6.7	Example hierarchy for the images in Figure 6.6. . . . .	203
6.8	Illustrative higher-level objects. . . . .	205
6.9	ROC curves for OS image-level performance for high-level objects. . . . .	208
6.10	Example probabilistic graph model for the hierarchy in Figure 6.7. . . . .	211
D.1	Example outputs of the modified GPAC method. . . . .	252
F.1	Classification of individual rgbequal bands. . . . .	256
F.2	Classification of individual truecolor bands. . . . .	257
F.3	Classification of individual ccd bands. . . . .	258
G.1	Pixels missed (PM) ranked by nucleus. . . . .	260
G.2	Extra pixels (EP) ranked by nucleus. . . . .	265
H.1	Example optical density image. . . . .	270
I.1	Random graph features. . . . .	272

# Chapter 1

## Introduction

The main goal of this dissertation is the development and discussion of techniques for higher-level image analysis, i.e., object-level analysis. While this is a problem of interest in many applications, we seek the development of methods to assist in the quantification of breast cancer in histo/cytopathology imagery. In this chapter, we present motivation for the research and a brief introduction to some pathology concepts needed for an understanding of this research. We also summarize some of the problems inherent in the analysis of medical imagery and review relevant research describing the state-of-the-art in histo- and cyto-pathology image analysis, both in general and specifically for breast cancer. We present an overview of the general framework for this research, and provide a summary and the main research contributions for the rest of the dissertation.

### 1.1 Motivation

While the specific factors for developing breast cancer are diverse and not completely understood, it is estimated (based on 2000-2002 statistics) that 13.2% of women born today in the United States will be diagnosed with breast cancer [1]. This statistic, generally reported as “1 in 8,” is the cumulative risk if a woman were to live 110 years; the actual risk depends on age, a bimodal distribution with the first peak at premenopause (40-50 years) and the second at approximately 65 years [2]. Using data from 1969-2002, it was estimated that 211,240 new cases of invasive breast cancer and 58,490 new cases of non-invasive breast cancer would be diagnosed in 2005, while 40,410 women would succumb [3].

Breast cancer is now known to consist of a variety of diseases; complicating the diagnosis and treatment procedures are a large number of conditions that mimic breast cancer and may present an increased risk for developing breast cancer [3, 4]. Of particular interest are benign diseases and the possibility that these may be precursors to malignant conditions (i.e., “ premalignancy ”) [2, 5–10]. Established cytologic (cell) criteria can

be contradictory, and even histologic (tissue) criteria (considered the gold standard for diagnosis) are subject to varied interpretation [2].

It has become common clinical practice to remove all breast masses due to the possibility of breast cancer, even though 80% of these excised lesions turn out to be benign [2]. There is thus a need to quantitatively define characteristics of breast cancer to better coordinate clinical care of women presenting breast masses. In light of the incidence and mortality statistics of breast cancer, and the demonstrated need for quantitative cyto-logic and histologic analysis of breast biopsy specimens, we propose here an approach for such quantitative analysis, using expert (pathologist) input to guide the classification process.

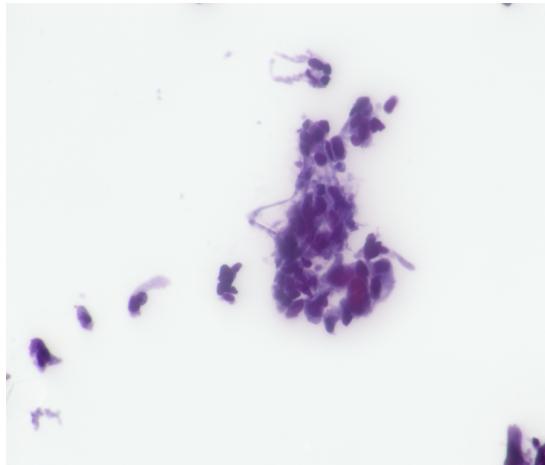
## 1.2 A Brief Introduction to Some Pathology

### 1.2.1 Histo- and Cyto-pathology

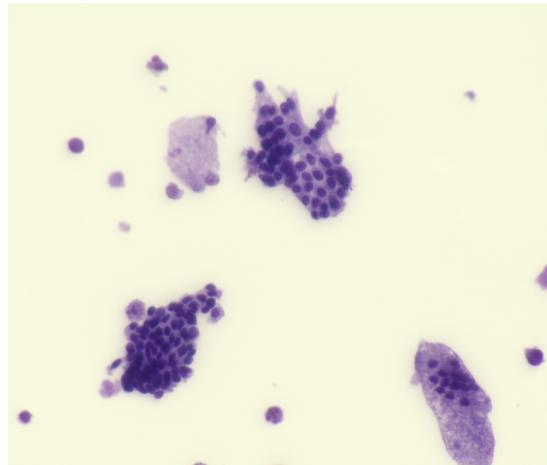
In order to understand our medical imagery, we must first lay out some of the terminology involved. The CancerWEB project, published by the Centre for Cancer Education at the University of Newcastle upon Tyne, UK, has an excellent online medical dictionary (found at <http://cancerweb.ncl.ac.uk/omd>).

- pathology: “The branch of medicine concerned with disease, especially its structure and its functional effects on the body.” [11]
- cytology: “The study of cells. Implies the use of light or electron microscopic methods for the study of morphology.” [11]
- histology: “The study of cells and tissue on the microscopic level.” [11]

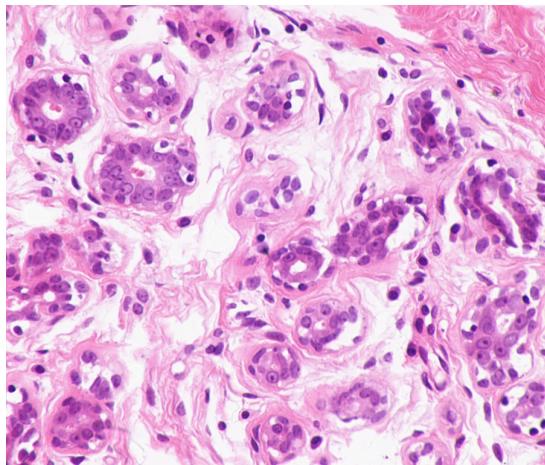
Thus, cytopathology is the study of diseased cells and histopathology is the study of diseased tissues, both at the microscopic level. The distinction between the two techniques is a bit blurred since any study of tissue will inherently include the study of constituent cells. The distinction is, then, usually determined by the technique used to harvest the specimens. Exfoliative (obtained by scraping) and aspirative (obtained via syringe) cytology are two common sources of cytopathology images; a very common example of exfoliative cytology is the routine cervical Pap smear. These images, due to the nature of the specimen retrieval, consist mainly of individual cells and cell clusters (see Figure 1.1 (a)-(b)). While these cell clusters may retain some of the underlying tissue architecture (i.e., spatial arrangement of the cells within clusters, indicative of tissue origin and disease state), there is not a global sense of tissue structure in cytopathology images. Histopathology images (Figure 1.1 (c)-(d)), on the other hand, are usually obtained from a biopsy specimen as a thin slice, and thereby do retain global tissue structure.



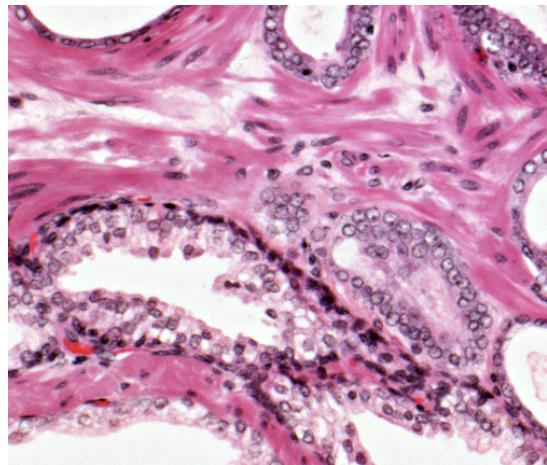
(a) Cytopathology, colon.



(b) Cytopathology, breast.



(c) Histopathology, breast.



(d) Histopathology, prostate.

**Figure 1.1.** Example cytopathology images of colon (a) and breast (b); these images consist mainly of individual cells and cell clusters which may retain some of the underlying tissue architecture. There is not, however, a global sense of tissue structure as there is in histopathology images; see for example the histopathology images of breast (c) and prostate (d).

### 1.2.2 Imagery and Multispectral Image Acquisition

We have 58 Hematoxylin and Eosin (H&E) stained histology images of breast tissue from our collaborators at the Department of Pathology at Yale University, as well as images of other cancers. All pathology images in this dissertation are courtesy of Dr. David Rimm's lab in the Department of Pathology at Yale. To achieve the best acceptance of (or least resistance to) new techniques in analyzing histo/cytopathology images, it is important to fit them, with minimal disruption, into existing processing streams in the clinical laboratory [12]. With this in mind, we choose to focus our analysis on routine H&E stained imagery from a standard light microscope. The microscope setup at Yale does, however, use a liquid crystal tunable filter (LCTF) for multispectral imaging capabilities: the VariSpec<sup>TM</sup>(CRi, Woburn, MA). The use of standard H&E stained imagery facilitates easy adoption of analysis techniques into routine pathology practice, as well as allowing for the possibility to analyze a wealth of archival samples.

Our collaborators in the Department of Pathology at Yale University use an Olympus BH-2 light microscope (Olympus America, Melville, NY) with the CRi VariSpec<sup>TM</sup> model VIS2-CM, coupled to a Retiga 1300 monochrome CCD camera (Quantitative Imaging, Burnaby, British Columbia, Canada). The VIS2-CM has a wavelength range of 400 nm to 720 nm, and can be tuned for bandwidths between 0.25 nm and 20 nm with response times on the order of 50 ms to 150 ms [13]; the filter at Yale is tuned for 10 nm bandwidths, and is most commonly used with wavelengths from 420 nm to 700 nm. Image acquisition is controlled with CRi's PanKroma software which calculates exposure times for each acquisition wavelength and corrects for uneven illumination with flat-fielding. The CCD camera acquires each spectral band as an 8 bit, 768×896 image, with spatial resolution of <1 μm per pixel.

The general image acquisition procedure is as follows. First, the slide is loaded onto the microscope stage and a low power (100x total magnification) manual scan of the slide is performed. For regions of interest, a higher power (400x) magnification is used for further analysis and image acquisition. A white cube must be acquired for the flat-fielding operation, which requires the acquisition of a full wavelength range for an empty portion of the slide; this is also used to acquire exposure times for each wavelength, such that each band will occupy the full dynamic range. Finally, the region of interest is centered and focused manually and the multispectral image stack is acquired and flat-fielded. This whole process takes on the order of several minutes.

### 1.2.3 Specimen Preparation

Histology specimen preparation follows the general process of fixation, embedding, mounting, and staining: fixation stops metabolic processes in cells and preserves cell structure; embedding allows the specimen to be sliced into thin sections (usually 5-15 μm); mounting fixes the thin section to a slide; and staining colors the otherwise

colorless cellular material for viewing under a microscope, and provides the ability to highlight certain molecular characteristics [14]. The most common pathological preparation for tissue is the formalin-fixed, paraffin-embedded, Hematoxylin and Eosin (H&E) stained section.

Hematoxylin is a blue-staining basic dye that stains genetic material; this is mainly seen in cell nuclei, although some components of cytoplasmic and extracellular material is also stained. See Figure 1.1 (c), where the nuclei are the blue-purple, roughly elliptical blobs. Eosin is a pink-staining acidic dye that stains membranes and fibers; this is most obviously seen in cytoplasm and connective tissue (in Figure 1.1 (c) these are the pink areas surrounding the nuclei and the wispy pink filaments, respectively).

#### 1.2.4 Cytology Characteristics of Breast Cancer

As with many biological specimens, breast cancers present a wide variability in appearance even among the same subtype. Some general cytology characteristics of breast cancer include (adapted from [2]):

- Slide background
  - Cellularity (abundance of cells)
  - Nuclear and/or cytoplasmic debris, necrosis
  - Poor cohesion of cells, single cells
  - Absence of naked, bipolar nuclei (elongated, dark nuclei, indicative of proper epithelial differentiation)
- Cells
  - Abundant and dispersed
  - Irregular and disordered clumps/clusters
  - Pleomorphic (containing more than one stage in the life cycle) and enlarged
- Nuclei
  - Pleomorphic
  - Hyperchromatic (overall increase in staining intensity) with indicative chromatin (nuclear material, including DNA) texture and prominent and enlarged nucleoli (small, round sub-organelles)
  - Eccentric, crowded, enlarged
  - Irregular membranes
  - Increased mitoses (dividing cells)

None of these characteristics is solely indicative of the presence of cancer; furthermore, many benign breast diseases have similar characteristics. It is the expertise and experience of the clinical pathologist that determines the diagnosis. Most cytology specimens are used to determine a suspicion of cancer and subsequent histologic specimens are used for the final diagnosis [2]. In particular, cytology specimens do not allow distinction between invasive and *in situ* cancers since there is no information about tumor margins [2].

In addition to the inherent variability in specimen characteristics, there are a wide variety of types and subtypes assigned to breast cancers (more are distinguished on a regular basis in an attempt to facilitate efficient treatment). These subtypes display subtle morphologic differences, which underscores the importance of pathologist input to the quantitative analysis process. While computer vision techniques may be able to elucidate even more subtle information than a human (e.g., chromatin texture), it is essential to provide a system in which the expert pathologist can properly train the computer vision algorithms.

## 1.3 Analysis of Histo/Cytopathology Images

### 1.3.1 Challenges

#### Biological and Staining Variability

Perhaps the largest difficulty in computer analysis of biological images is the inherent variability found in biological specimens. Even within a single cell type in the body, an amazing amount of normal variation can be found. This variation becomes more pronounced with diseases such as cancer in which the cells no longer retain proper morphology associated with their lineage. Adding to this difficulty is the variability in preparation and staining procedures. Many researchers call for standardization of these procedures and many pathology labs have protocols in place, but staining variation is still present in specimens. This may be due to differences in the staining procedure, but is also dependent on intrinsic properties of the specimen being stained. As an example, one characteristic of breast cancer cells is hyperchromasia, an overall increase in staining intensity.

#### Lack of Ground Truth

Accepting the presence of biological and staining variability, there are still many problems related to the definition of appropriate ground truth for pathology specimens. Generally, within the pathology community, the diagnosis rendered from a histopathology section of an excisional biopsy is considered ground truth for the patient. In addition, many researchers use patient followup as either ground truth for the biopsy or as further

evidence of a correct diagnosis. While this may be the standard, there is overwhelming evidence of inter- and intra-observer variability in diagnoses [5, 6, 15–24]; this results in some well-founded concern about using another pathological diagnosis as ground truth for new studies in pathological diagnoses [25]. The use of various grading systems has sought to alleviate these problems and lend a more quantitative bent to the diagnosis process; however, the sheer number of grading systems [4–6, 17, 19, 21, 22, 26, 27] and the constant revision of such systems [4, 5] does not lend a consistent view of any cancerous disease. The use of clinical data such as patient followup and survival can alleviate these problems somewhat, although a true gold standard for pathology specimens is not available in the same sense that it may be for other datasets (e.g., remote sensing).

### Diagnoses of “Atypias”

While most pathologists can easily identify and categorize truly benign and frankly malignant specimens, there are many cases which fall somewhere in the continuum between benign and malignant. These are often given the diagnosis of “atypia.” For cytology specimens, the diagnosis of atypia usually means a more invasive followup procedure to acquire a tissue specimen. It is still unknown how exactly to approach treatment of atypical conditions, but furthermore, there is no standard for grading atypias. The problem of pathologist variability in diagnosis is increased greatly for atypical conditions. It is hoped that further study of atypical specimens will bring to light a more quantitative analysis, and allow for more specific treatment. While this vague area between benign and malignant conditions poses an image analysis challenge, it also poses very important questions about the premalignancy potential of benign conditions.

### Human Cognition and the Diagnosis Process

Humans’ concept of the world is inherently object-based, as opposed to the largely pixel-based representation of computer vision. As such, human experts describe and understand images in terms of such objects. For pathologists, diagnosis criteria are inevitably described using terms such as “nucleus” and “cell.” It is thus important to develop computer vision methods capable of such object-level analysis. This is important not just for the biomedical imaging fields, but for most computer vision applications.

#### 1.3.2 Commonly Used Performance Metrics

We will present in Section 1.3.3 a brief overview of some current image analysis research, particularly for cancer detection applications. As different authors may present their results using different metrics, we will first provide a description of each to facilitate better understanding and comparison of the descriptions to follow.

- **Accuracy:** refers to the probability of a correct decision,  $N_C/N$  where  $N_C$  is the number of correctly determined cases and  $N$  is the total number of cases.
- **Error:** refers to the probability of an incorrect decision,  $N_I/N$  where  $N_I$  is the number of incorrectly determined cases and  $N$  is the total number of cases. Note that  $Accuracy = 1 - Error$ .
- **Sensitivity/Detection Rate (DR):** refers to the probability that a positive case will be correctly decided,  $N_{TP}/(N_{TP} + N_{FN})$  where  $N_{TP}$  is the number of true positives and  $N_{FN}$  is the number of false negatives.
- **Specificity:** refers to the probability that a negative case will be correctly decided,  $N_{TN}/(N_{TN} + N_{FP})$  where  $N_{TN}$  is the number of true negatives and  $N_{FP}$  is the number of false positives.
- **False Alarm Rate (FAR):** refers to the probability that a case labeled as positive is actually negative,  $N_{FP}/(N_{TP} + N_{FP})$  where  $N_{FP}$  is the number of false positives and  $N_{TP}$  is the number of true positives; this is the metric that will be referred to in the next section. Alternatively, and possibly more commonly, false alarm rate can be defined as  $N_{FP}/(N_{TN} + N_{FP})$  where  $N_{FP}$  is the number of false positives and  $N_{TN}$  is the number of true negatives.
- **ROC curve, AUC:** A Receiver Operating Characteristic (ROC) curve allows a graphical representation of the tradeoff between sensitivity and specificity (sensitivity on the y-axis and 1-specificity on the x-axis). Equivalently, this may also be plotted as true positive rate versus false positive rate. A full explanation of ROC curves would be quite extensive, but we will note a few essentials. First, a diagonal line from  $(0,0)$  to  $(1,1)$  represents the results for a random guess. Second, the line from  $(0,0)$  directly to  $(0,1)$  and then to  $(1,1)$  (i.e., the left and top edges of the plot box) represents a perfect classifier. Thus, the closer an actual ROC curve is to this ideal, the better the classifier. Third, the Area Under the Curve (AUC) is often used as a single number to quantify a classifier; note that a perfect classifier has an AUC of 1.0 and the random guess an AUC of 0.5.

### 1.3.3 State-of-the-Art

#### General Histo/Cytopathology Image Analysis

In this section we present a brief overview of some recent work on general histo- and cyto-pathology image analysis. The work presented here (summarized also in Table 1.1) covers a range of imaging modalities and tissue/cancer types wherein we see a wide range of performance.

**Infrared vibrational and Fourier transform spectroscopy**<sup>1</sup> have been used for classification of prostate [29] and cervical [27,30] tissue, in which very specific spectral information (at the molecular-level) is used, but spatial attributes are generally not utilized. Using a pool of features hand-picked by spectroscopist analysis of pathologist-marked regions, Fernandez et al. [29] achieved accuracies of 90.1%-100.0% for classification of histologic classes from 250 tissue microarray cores from 40 patients, and 100% classification of adenocarcinoma versus normal tissue. Focusing on the glycogen depletion in dysplastic (abnormal) tissue, Shaw et al. [30] achieved accuracies of 60%-70% in separating normal from abnormal Pap smears. A back propagation neural network was used by Zhao et al. [27] along with feature sets derived from intensity statistics and the wavelet domain; pixel-level classification demonstrated a sensitivity of 95.2%, and nuclear-level classification accuracy of 97.6% for a dataset of 40 images.

Segmentation of nuclei in **fluorescence imagery**<sup>2</sup> is presented by Lin et al. [31] and Wählby et al. [32] for 2D imagery and 3D confocal imagery. Lin et al. [31] used a recursive, tree-based algorithm, and reported accuracy of 96.3% for 10 images of rodent brains. Wählby et al. [32] used a marker-based watershed transformation, region merging, and shape-based cluster separation; an accuracy of 91%-96% was achieved for a total of 6 2D images of cervical and prostatic carcinomas.

Luck et al. [33] describe segmentation for **in vivo confocal reflectance imagery**<sup>3</sup> of cervical tissue based on nuclear modeling, anisotropic median diffusion, Gauss-Markov random fields, and a Bayesian classifier. The authors reported a 90% match to hand-segmented nuclei with an average of 6 false positives per frame.

The **gross examination**<sup>4</sup> and manual feature extraction on 212 liver nodules, combined with correlation-based feature selection and several classification schemes (including decision trees, k-Nearest Neighbor (k-NN) classification, naive Bayes, and neural networks), resulted in a 95%-100% sensitivity and specificity for diagnosis as one of 3 conditions in Ciocchetta et al. [34].

Demir et al. [35] and Gunduz et al. [36] present a nuclei segmentation algorithm for H&E stained brain biopsies using the **La\*b\* color space** and a k-means algorithm. Features extracted are based on the concept of cell graphs [36] and augmented cell graphs [35],

---

<sup>1</sup>Infrared spectroscopy uses infrared light to excite vibrations in the molecular structure of a specimen. These are reflected as energies which give insight into the underlying molecular constituents. In the more common case of Fourier transform infrared spectroscopy, instead of an array of energies, an interferogram is produced; the Fourier transform of this interferogram produces the energies [28].

<sup>2</sup>Fluorescent dyes are attached to antibodies specific to some feature of interest (e.g., certain proteins) and imaged by exciting the fluorescence of the dyes with appropriate incident light. This method can very specifically target certain molecular attributes of a biological specimen. Also, specimens may be genetically modified to express fluorescent proteins, e.g., green fluorescent protein (GFP).

<sup>3</sup>In vivo imagery is imagery of living tissue in its natural environment, confocal imaging allows for imaging of different focal planes through the tissue, and reflectance imagery captures the reflected light from the specimen (rather than the transmitted light as through a thin slice specimen).

<sup>4</sup>Examination of the whole tumor on a macroscopic level.

including such features as degree and eccentricity commonly defined for graph structures. Classification of normal, inflamed, and cancerous tissue was performed by an artificial neural network, with accuracies of 96.9%-97.1% for 64 patients in [35] and 85.7%-94.0% for 12 patients in [36].

The **HSV color space** was used by Sammouda et al. [37] for segmentation of H&E stained lung nuclei, using a Hopfield neural network and maximum drawable circle algorithm, and nuclear radii as features. 92%-97% of cancerous nuclei were correctly detected in 16 images.

Roula et al. [38] used a **multispectral dataset** (33 bands) of H&E stained prostate tissue, extracted texture and mathematical morphology features, reduced dimensionality using principal components analysis (PCA), and classified using quadratic discriminant analysis. Classification error for 230 hand-picked nuclei representing 3 conditions was 5.1%. **GENIE** (an automated feature extraction system developed at Los Alamos National Laboratory) has been used with a multispectral dataset (31 bands) of Pap-stained urine cytology images as presented by Angeletti et al. [39]. Using a training set of 12 cases and validation sets of 17 and 8 cases, GENIE was able to discriminate between malignant and benign urothelial cells with a sensitivity of 85%-87% and specificity of 96%. Additionally, GENIE was able to correctly classify atypical urothelial cell clusters as benign or malignant (determined clinically by 1-year followup) with an AUC of 0.728.

**Narrowband red images** (610 nm, 621 nm) have been used for detection of laryngopharyngeal cancer [8], discrimination of cervical cancers and atypias [9], and separation of benign hyperplastic prostatic lesions from true prostatic carcinoma [10]. Additionally Brewer et al. [40] used the **red channel** from standard RGB light microscopy to classify epithelial and stromal (connective tissue) nuclei in ovarian tissue. In Brewer et al. [40], 7 features were hand selected from 93 karyometric (nuclear) features to discriminate cancerous and benign conditions, resulting in accuracy of 66%-78% for H&E stained sections from 20 patients. Zahniser et al. [9] used narrowband images at 621 nm (for nuclear material) and 497 nm (for cytoplasmic material) of Feulgen and Orange II stained cervical Pap smears and linear discriminant analysis at both the single-cell and cell-cluster level. Zahniser et al. [9] reported classification accuracy of 100% for normal (14 cases), 45% for benign change (11 cases), and 97% for abnormal (29 cases). Both Neheret al. [8] and Mairinger et al. [10] used the CytoSavant image analysis system from OncoMetrics, Inc. (Vancouver, BC, Canada) to extract 114 nuclear features (plus mean, maximum, minimum, and variance for each) from 610 nm narrowband images of Feulgen stained tissue, and used stepwise linear discriminant function analysis for classification; Neheret al. [8] reported sensitivity and specificity of 72.7% and 82.4% for 145 cases and Mairinger et al. [10] reported sensitivity and specificity of 92% and 95% for 240 cases.

Similarly, **narrowband green images** (565 nm) were used for analysis of Feulgen-stained lung tissue [20,41] and prostate tissue [41]. Weyn et al. [20] used 82 features (plus mean and standard deviation of each), including densitometry (optical density-related), morphometry, texture, and syntactic structure analysis (SSA) (related to spatial ar-

rangement) measures. For a dataset of 39 cases of malignant mesothelioma, 20 cases of pulmonary adenocarcinoma, and 7 cases of hyperplastic mesothelium, k-NN classification yielded accuracies of 83.9%-96.8% for discrimination of the three conditions, 79.5%-94.9% in typing malignant mesothelioma, and 60.0%-82.9% for prediction of prognosis for malignant mesothelioma [20]. Weyn et al. in [41], on the other hand, derived features from the imagery using Voronoi diagrams, Gabriel's graphs, and minimum spanning trees, all of which quantitate spatial arrangement of tissue constituents. k-NN classification yielded correct disease classification of 73.9% (51 cases), correct typing of malignant mesothelioma of 82.6% (44 cases), and correct grading of prostatic adenocarcinoma of 75.7% (38 cases).

Analysis of **immunostained imagery**<sup>5</sup> by Weyn et al. [22] used a CD31 immunostain (highlighting endothelial cells) and Hematoxylin counterstain to quantify prognosis based on vascular patterns in colorectal, cervical, and lung tissue. Extracted features include fractal analysis, vessel-derived (some manual), syntactic structure analysis, and clinical data (manual), as well as the mean, standard deviation, skewness, and kurtosis for each feature. Prognostic accuracies using a k-NN classification were 83.3% for cervical (78 images), 70.6% for colorectal (74 cases), and 86.4% for lung (27 images).

Direct “apples to apples” comparison of the research discussed above is difficult since each research publication specifies results in terms of different metrics. With this in mind, an overview of state-of-the-art for general cancer imagery analysis is presented in Table 1.1. While there is a wide range in performances with these different analysis methods, we see that it is possible to achieve accuracies in classification of cancerous conditions above 90%.

## Histo/Cytopathology Image Analysis for Breast Cancer

Recent research in the automated analysis of breast cancer imagery is presented below, in reverse chronological order, as well as summarized in Table 1.2. It should be noted that the caveats of Table 1.1 are also applicable here. A few things should be noted before we begin our synopsis of state-of-the-art in breast cancer imagery analysis. First, our focus in this section will be on analysis of imagery similar to ours; thus, mammography images, for example, will not be considered. Second, there is a wealth of breast cancer research on molecular and genetic mechanisms behind the development of breast cancer. We, however, will focus on research involving standard clinical pathology lab biopsy preparations.

Ballerini and Franzén [42] (2004) utilized light microscopy of breast cancer tissue with immunohistochemically stained epithelium and Feulgen-staining of the nuclei. This method used fuzzy c-means clustering and conditional dilation to segment nuclei, and

---

<sup>5</sup>Immunostains use antibodies to specifically target molecules of interest, similar to fluorescence imaging, but use standard dyes for viewing with light microscopy.

**Table 1.1.** Performance of general state-of-the-art histo/cytopathology image analysis.

<b>Ref.</b>	<b>Imaging Modality</b>	<b>Year</b>	<b>Tissue</b>	<b>Dataset</b>	<b>Performance</b>
[29]	Vibrational spectroscopy	2005	Prostate	250 cores, 40 patients	Accuracies: 100% cancer/normal, 90.1%-100% for histologic classes
[30]	Vibrational spectroscopy	2000	Cervical	?	Accuracy: 60%-70% normal/abnormal
[27]	Fourier transform spectroscopy	2004	Cervical	40 images	95.2% pixel-level sensitivity 97.6% nuclear-level sensitivity
[31]	Fluorescence	2005	Brain	10 images	96.3% correct nuclear segmentation
[32]	Fluorescence	2004	Cervical	6 images	91%-96% correct nuclear segmentation
[33]	in vivo confocal reflectance	2005	Cervical	?	90% correct nuclear segmentation
[34]	Gross examination	2003	Liver	212 nodules	95%-100% sensitivity and specificity
[35]	Light (H&E)	2005	Brain	64 patients	96.9%-97.1% normal/inflamed/cancer
[36]	Light (H&E)	2004	Brain	12 patients	85.7%-94.0% normal/inflamed/cancer
[37]	Light (H&E)	2005	Lung	16 images	92%-97% detection of cancerous nuclei
[38]	Light (H&E, 33 bands)	2003	Prostate	230 nuclei	5.1% classification error for 3 conditions
[39]	Light (Pap, 29 bands)	2005	Bladder	37 cases	85%-87% sensitivity, 96% specificity
[8]	Light (Feulgen, 610 nm)	2004	Oral	145 cases	72.7% sensitivity, 82.4% specificity
[9]	Light (Feulgen, Orange II, 610 nm, 497 nm)	1994	Cervical	54 cases	Classification accuracies: 100% normal, 45% benign change, 97% abnormal
[10]	Light (Feulgen, 610 nm)	1999	Prostate	240 cases	92% sensitivity, 95% specificity
[40]	Light (H&E, red channel)	2004	Ovarian	20 patients	Accuracy: 66%-78% cancerous/benign
[20]	Light (Feulgen, 565 nm)	1999	Lung	66 cases	Accuracies: 79.5%-94.9% typing, 60.0%-82.9% prognosis
[41]	Light (Feulgen, 565 nm)	1999	Lung, Prostate	51 cases, 44 cases	Accuracies: 82.6% typing, 75.7% grading
[22]	Light (CD31 immunostain, Hematoxylin)	2004	Cervical, Colorectal, Lung	78 images, 74 cases, 27 images	Prognostic accuracies: 83.3% cervical, 70.6% colorectal, 86.4% lung

a neural network for classification. Extracted features include granulometric moments, fractal analysis, and mathematical morphology. 20 cases, with 10 images per case, were analyzed with this method, resulting in 87%-93% correct classification of normal tissue, fibroadenosis (a benign condition), and ductal and lobular cancer.

Harvey et al. [43] (2003) used the GENIE automated feature extraction system for detection of cancerous nuclei in multispectral H&E stained histopathology images of breast tissue. Using a training set of 7 images and a test set of 8 images, GENIE attained an average detection rate of 82.3%-87.4% and average false alarm rate of 0.4%-15.8%.

Lee and Street [44] (2003) present a neural network-based method to automatically detect, segment, and classify breast cancer nuclei in gray-scale cytological images from fine needle aspirations (FNA) of the breast. Nuclear features include size, perimeter, smoothness, concavity, and 24 radii from each nucleus. Overall, 94.1% of nuclei were correctly delineated in a dataset of 140 images, and 94%-96% of nuclei were correctly classified as malignant.

Latson et al. [17] (2003) implemented an automated segmentation algorithm for epithelial cell nuclei based on the application of fuzzy c-means to the hue band (of HSV color space) followed by a marker-based watershed transform. Results for a dataset of 39 H&E histopathology images found 57.2%-71.6% correctly segmented nuclei, with a variation in performance for typical hyperplasia, atypical hyperplasia, cribriform ductal carcinoma in situ, and solid ductal carcinoma in situ. Clumps, poorly segmented individual nuclei, and missed nuclei were 4.5%-16.7%, 22.5%-26.3%, and 0.4%-1.4%, respectively.

van de Wouwer et al. [45] (2000) used green-filtered (565 nm) light microscopy images of Feulgen-stained breast tissue sections to extract features for k-NN classification of breast tissue. Features included densitometry, first- and second-order texture parameters, wavelets, and mathematical morphology. For a dataset of 20 normal and 63 invasive ductal carcinomas, 67.1% of nuclei and 100% of patients were classified correctly.

Herrera-Espíñeira et al. [46] (1998) used two different segmentation algorithms, one for non-overlapping nuclei (histogram-based threshold) and one for overlapping nuclei (edge detection and ellipse fitting); the choice in algorithms was decided by the human observer based on the image at hand. Nuclear features allowed 89.4%-91.5% average accuracy in discriminating benign (47 cases) from malignant (95 cases) for Pap-stained grayscale cytology imagery.

Weyn et al. [18] (1998) used the same imagery and similar analysis to van de Wouwer et al. [45]. Results in this study were 76.1% accuracy in benign versus malignant classification of images and 100% accuracy for patients. Also studied here was cancer grading, with 61.5% image accuracy and 78.5% patient accuracy.

Wang et al. [15] (1997) present a method for detection of breast cancer nuclei in light microscopy images of tissue immunostained for estrogen and progesterone receptors and counterstained with Hematoxylin. This method used receptive field theory, soft

thresholding, and lighting correction to segment nuclei; the classification of nuclei was based on the YUV color space and derived features (average and variance) as well as a local texture measure. For a dataset of 28 images, the authors achieved a sensitivity of 83%.

Anderson et al. [19] (1997) applied a knowledge-guided approach previously developed by Thompson et al. [47] for segmentation of cribriform gland tissue to segmentation and architectural discrimination of H&E stained ductal breast lesions. Features were computed at the glandular and lumen level. The dataset was composed of 215 images from 22 cases of ductal carcinoma in situ and 21 cases of ductal hyperplasia. Glandular features provided 63% correct image and 65% correct patient classification, lumen features provided 70% correct image and 72% correct patient classification, combined features provided 83% correct patient classification.

Overall, for the breast cancer image analysis, summarized in Table 1.2, there is a range in classification accuracy. In general, however, the accuracy increases as the classification progresses from individual nuclei to image-level to patient-level. In particular, for H&E imagery as we will be working with, the image-level classification accuracy is less than 90%. It is thus our goal to demonstrate image-level classification accuracies above 90%. To our knowledge, none of the methods in Table 1.2 are being adopted in a clinical setting. The performance required for eventual incorporation of an automated analysis method into standard clinical practice is something that will require close collaboration with pathologists.

## 1.4 Overview of the QOALA Framework

Here we present an overview of the Quantitative Object- and spatial Arrangement-Level Analysis (QOALA) framework for our hierarchical image analysis. The overarching goal of this research is to develop techniques for higher-level image analysis, i.e., object-level analysis. This is a problem of interest in many applications. We seek the development of methods to assist in the quantification of cancer in histo/cytopathology imagery; striving, however, to maintain a flexible framework for the analyses that can allow the use of such methods in a broad range of application areas. Further discussion will focus on histo/cytopathology applications, although it should be noted that any of these specific analyses could be replaced with other application-specific analyses within the same framework.

We seek to develop an application in which an expert (pathologist) can interactively provide expertise (in the form of training data) to facilitate the estimation of a desired classifier. A simple example of such a task is the discrimination of cancerous from non-cancerous cell nuclei. Of more interest to pathologists, however, are more complex tasks, such as distinguishing different grades of cancers, determining the prognosis of the patient, suggesting treatment options, quantifying premalignancy, etc. All of these tasks

**Table 1.2.** Performance of state-of-the-art histo/cytopathology image analysis for breast cancer.

Ref.	Imaging Modality	Year	Type	Conditions	Dataset	Performance
[42]	Light (immunostained epithelium, Feulgen)	2004	Histo	Normal, fibroadenosis, ductal and lobular cancer	20 cases 200 images	87%-93% class. (image)
[43]	Light (H&E, 29 bands)	2003	Histo	Benign, malignant	15 images	82.3%-87.4% DR 0.4%-15.8% FAR
[44]	Light (grayscale)	2003	Cyto	Benign, malignant	140 images	94.1% seg. (nuclei) 94%-96% class. (image)
[17]	Light (H&E)	2003	Histo	Typical hyperplasia, atypical hyperplasia, cribriform ductal carcinoma in situ, solid ductal carcinoma in situ	39 images	57.2%-71.6% seg. (nuclei)
[45]	Light (Feulgen, 565nm)	2000	Histo	Normal, invasive ductal carcinoma	83 patients	67.1% class. (nuclei) 100% class. (patient)
[46]	Light (Pap, grayscale)	1998	Cyto	Benign, malignant	142 cases	89.4%-91.5% class.
[18]	Light (Feulgen, 565 nm)	1998	Histo	Normal, invasive ductal carcinoma	83 patients	76.1% class. (nuclei) 100% class. (patient) 61.5% grading (nuclei) 78.5% grading (patient)
[15]	Light (immunostained hormone receptors, Feulgen)	1997	Histo	Benign, malignant	28 images	83% sensitivity
[19]	Light (H&E)	1997	Histo	Ductal carcinoma in situ, ductal hyperplasia	215 images, 43 cases	63%-70% class. (image) 65%-83% class. (patient)

are currently accomplished via the subjective opinion of the pathologist; we seek to provide a quantitative analysis for such tasks, something that is rather lacking in current pathology practice.

### 1.4.1 Related Work

Fundamentally, object-level analysis depends greatly on some underlying segmentation mechanism. It is the segmentation methodology that determines what constitutes an object. Commonly, an object is defined as a connected group of pixels satisfying some similarity criterion. In recent object-based analysis research, an object has been defined in terms of contour [48, 49]; homogeneous color [50, 51] or texture [51]; as an anomaly (in some sense) compared to the rest of the image [52]; location, size, and shape [53]; and topological and connectivity relations [54, 55]. Segmentation is, in general, a very difficult problem to “solve” in an adequate fashion; segmentation of medical imagery is no exception to this. The inherent variability of biological structures provides an additional complication, which can make any generalization of common features to a particular cytologic or histologic structure very difficult. Nevertheless, there is a wealth of research describing a wide variety of methods specifically for biomedical segmentation; some applicable research can be found in References [17, 31–33, 37, 46, 56–60], and an overview in Reference [61].

It is our goal in this research to work beyond the details of the pixel-level segmentation issues. While the initial segmentation of histologic structures will be ultimately important and worthy of appropriate analysis (see Chapter 4), we wish to use this imperfect segmentation and work to gain further information from the imagery. Recent object-level research has dealt with a wide range of issues, including object-based image similarity metrics [48, 49], object-based image retrieval [50, 53, 62], characterization of image content using (among other metrics) spatial relations [49, 51, 54, 55], and object detection and recognition tasks [52, 63, 64].

One similar approach to the one proposed can be found in [65], where an object-oriented approach is used to facilitate analysis (morphological, architectural, and statistical) of tissue sections. Kamalov et al. [65] propose the use of a “Slide-Unit-Cell” data structure to capture the hierarchy associated with analysis of a histopathology slide. Further classes of data (e.g., “Dot” and “Membrane”) allow the image processing to operate on several different objects [65]. While similar, the proposed approach will be more flexible in the definition of the various objects; essentially, there will be no limitation to a predefined set of objects and the analysis routines associated with those objects.

The Cellenger<sup>®</sup> system from Definiens AG (München, Germany), released commercially in April 2003, provides “automated multivariate image analysis for cell-based assays” [66]. This system is built on the previously developed and patented eCognition Server which provides access to databases and parallel computing services from a user’s

desktop [67]. While the Cellenger<sup>®</sup> system is described as automated, the initial task of setting up the image analysis is not automated, and requires knowledge and experience in image processing as well as programming skills. The Cellenger<sup>®</sup> system represents an image as several layers of information, comprising different sized objects [68], where objects are built iteratively from a previous level of objects [69]. Image analysis tasks are specified in a modular fashion using existing analysis tasks common to medical image analysis; this system was developed for use in drug development trials [66]. Most details of the Cellenger<sup>®</sup> system are proprietary to Definiens and are, therefore, not publicly available.

### 1.4.2 QOALA Framework

We propose and present an iterative and modular approach to object-level analysis, using pathologist expertise at each decision point. This procedure involves alternating steps of classification and segmentation in each iteration and can continue with expert input until the analysis has reached a satisfactory level (according to the expert). This flexibility allows the same framework to be used for a wide variety of analysis tasks; e.g., a user could use this same system to perform a segmentation of connective tissue, or further the analysis and locate connective tissue adjacent to malignant epithelial cells. To our knowledge, this interactive, modular, and iterative approach is a new contribution to the field of object-level analysis. While the Cellenger<sup>®</sup> system is also modular and iterative, our system is more flexible as we rely on an intuitive interaction process to determine the resulting analysis procedures; this takes the form of a simple image markup procedure whereby the expert uses a mouse or tablet stylus to designate areas within the image. Additionally, our system requires little (if no) image processing expertise of the user since the actual analysis algorithms take as input only the image markup.

For clarification of our process, consider an example for the first iteration. At this point, the pathologist may identify different histologic material (e.g., nuclear material, cytoplasm) as the training data and the system uses machine learning to develop a classifier for each specified category. The pathologist would have the opportunity to view the resulting classification, and modify the training data and retrain if the current classifiers are not satisfactory.

Once the *classification* for this step is accepted, the pathologist would provide input on the proper *segmentation* of the identified classes. This training data would provide information on which pixels should be agglomerated to form a proper histologic segmentation. As in the previous step, the pathologist will have the opportunity to view the resulting segmentation and modify training data if necessary.

We are making a point of distinguishing the *classification* from the *segmentation* in this framework; this separation is to better focus any optimization techniques. For this example, in the *classification* step we are interested in learning classifiers that can accu-

rately identify *all* pixels belonging to a certain histologic class, whereas in the *segmentation* step we are interested in the proper agglomeration of sets of these properly identified pixels to form appropriate segmented structures.

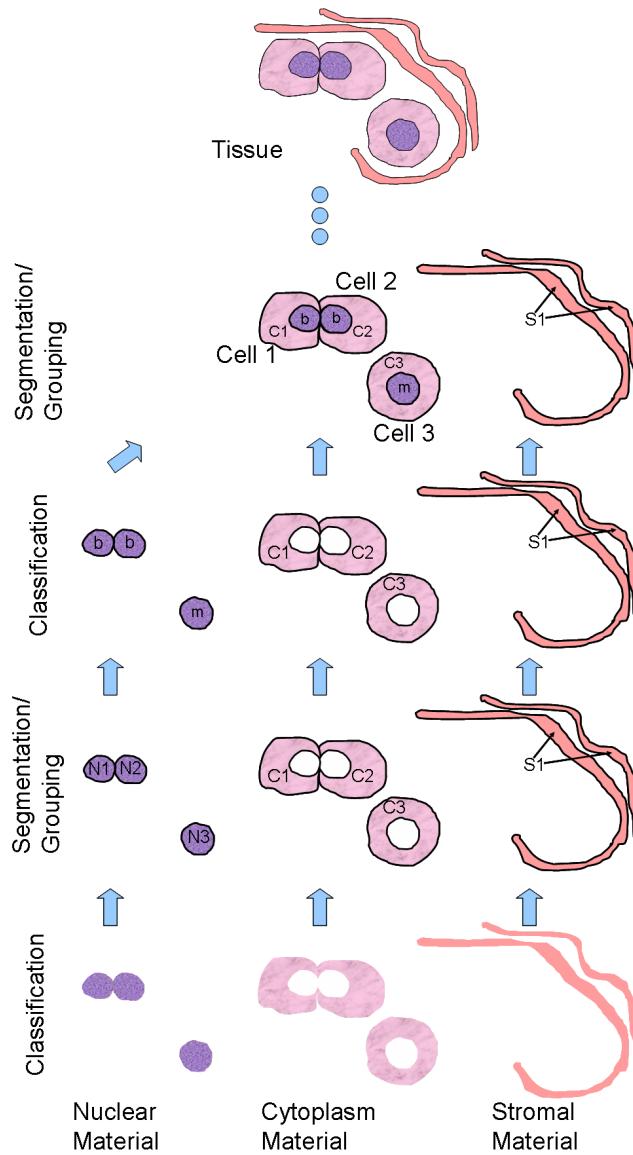
Subsequent iterations of the process would continue as alternating steps of classification and segmentation, where each new iteration yields a higher level classification and segmentation than the previous ones. In these further steps, the pathologist may, for example, classify certain nuclei as cancerous and group nuclear and cytoplasmic material into cells. This process will continue until the pathologist has achieved the level of analysis desired. This provides a flexible method of quantitative analysis, allowing the pathologist to interact with all steps of the processing, and tailor the analysis to a specific purpose in an easy and intuitive fashion.

An illustrative example of this two-step iterative process can be seen in Figure 1.2. Note that the proposed iterative process exemplified in Figure 1.2 does not incorporate any concept of feedback between the steps. While this is a very important consideration, and could ideally improve the overall process by allowing further analysis to modify lower-level results, it is beyond the scope of this work.

## 1.5 Outline of Dissertation and Research Contributions

We have thus far presented motivation for our research, an overview of some related pathology concepts, and a summary of current research in general cancer imagery analysis as well as research specific to breast cancer. We have also presented an overview of the QOALA framework for hierarchical image analysis. We will now briefly summarize the organization of this dissertation as well as the various research contributions. Further discussion of each research problem, including how our proposed research differs from other research, is included in subsequent chapters along with presentation of our results. We choose to focus our research efforts on the study of histology images, since we are interested in the quantification of multiple levels of information, including spatial arrangement of histologic entities. These methods, however, will be easily applicable to cytology imagery as well.

- **Chapter 2: Multispectral Analysis of Pixel-Level Nuclear Classification**  
Our datasets are relatively unique in histo- and cyto-pathology in that they are imaged multispectrally (usually with 29 bands covering the visible spectrum from 420 nm to 700 nm, though capability exists for other bands). While some researchers have found advantages to spectral imaging for some applications in medical image analysis [39, 70–73], the value of multispectral analysis for routine histo- and cytopathology preparations is still unknown. Throughout our proposed research we will strive to analyze the utility of the multispectral nature of our imagery. We present



**Figure 1.2.** Illustrative example of the two-step iterative classification-segmentation process. At the first level (bottom row), biologic materials are *classified* on a pixel-by-pixel basis. At the second level, the materials are *segmented* into their constituent objects, based on user feedback. Similarly, the third step is another level of *classification* distinguishing (e.g.) benign ("b") and malignant ("m") nuclei, and the fourth step is *segmentation* of "cell" objects. This process can continue until the user is satisfied with the level of abstraction; in this example, the process ends at the concept of "tissue" at the topmost level. Note also that the level of abstraction is controlled by user input at each level; this can be seen in the maintenance of the stromal elements as a single object.

in Chapter 2 analyses of our multispectral data for the pixel-level classification of nuclei.

### Research Contributions

- Quantitative and in-depth analysis of multispectral versus RGB imagery for pixel-level nuclear classification of H&E stained imagery, a common image analysis task in histopathology.

- **Chapter 3: Pixel-Level Classification of Cytoplasm and Stroma**

While the characteristics of cell nuclei are well established as useful for diagnostic purposes, it is expected that the characteristics of cytoplasm and stroma will be similarly useful. The discrimination between cytoplasm and stroma, however, is not commonly addressed in quantitative pathology studies, especially for standard H&E stained imagery. In Chapter 3 we present the development and analysis of a classifier for cytoplasm and stroma.

### Research Contributions

- Development of a pixel-level classifier for discrimination of cytoplasm and stroma.
- Analysis of features useful for the cytoplasm/stroma discrimination problem.

- **Chapter 4: Segmentation of Nuclei, Cytoplasm and Stroma**

The segmentation of cell nuclei on an object level is a very difficult problem. While there is much literature addressing this problem, we seek a solution that is feasible for our H&E stained imagery and that avoids too many underlying assumptions about the nuclear characteristics (e.g., parameterization as an ellipse). We also present the development of an object-level segmentation metric applicable to our nuclear segmentations, as well as other application areas.

### Research Contributions

- A *new* method to quantify segmentation performance.
- Comparison of several common segmentation methods on cell nuclei using the new metric.

- **Chapter 5: Feature Extraction and Selection**

In Chapter 5, we describe the various object-level and spatial-relation features that we extract from our various image objects. We provide a comprehensive list of features compiled from a variety of published research on histo- and cytological image analysis. We also present our results on the feature selection and classification performance using various categories of features.

### Research Contributions

- Comprehensive list of features for object-level and spatial-relation feature extraction.

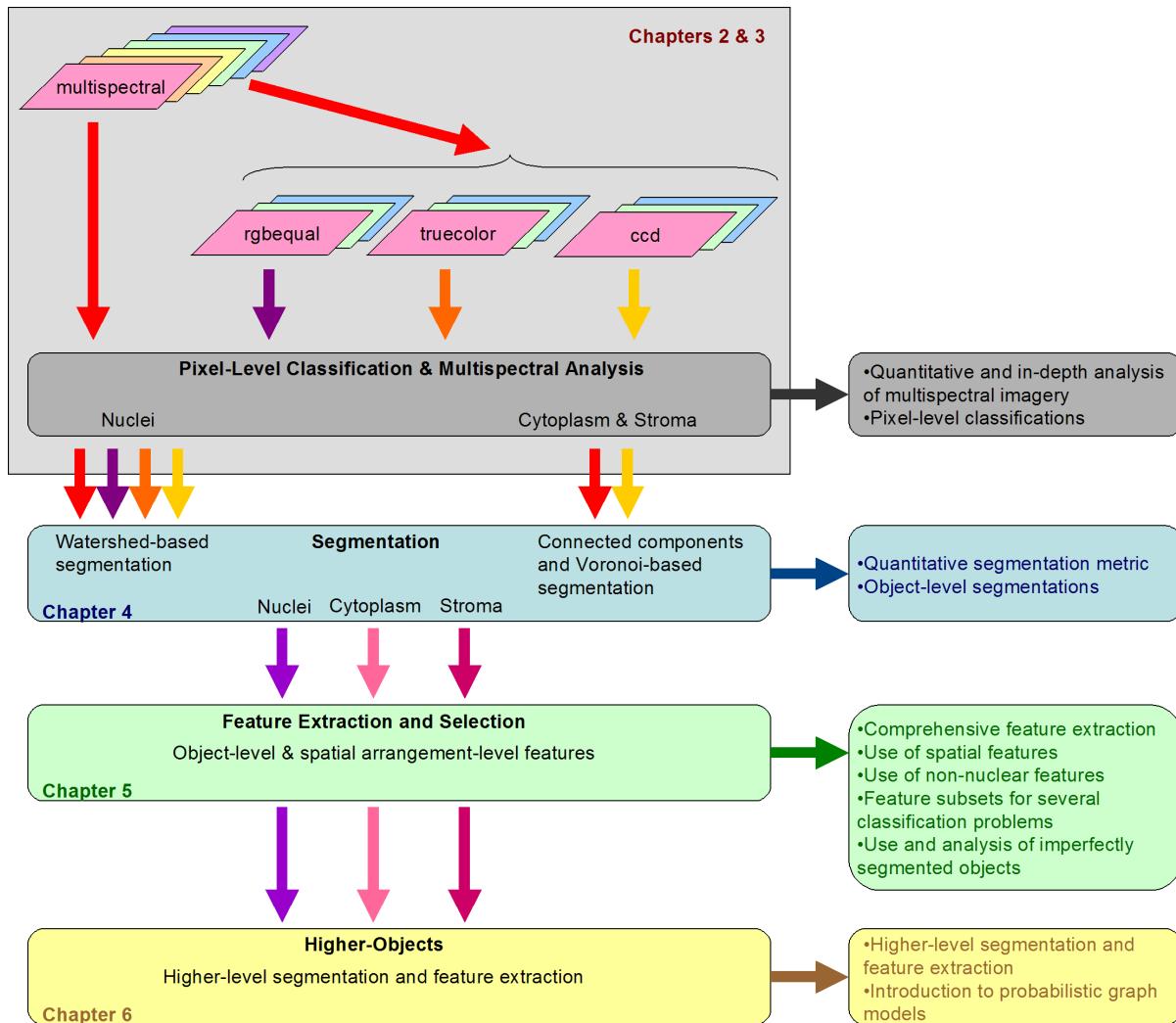
- Use of spatial arrangement features.
  - Use of non-nuclear features, namely cytoplasm and stroma features.
  - Feature subsets for several classification tasks.
  - Use of and analysis of performance for imperfectly segmented objects in a feature extraction, selection, and classification framework.
- **Chapter 6: Higher-Level Objects**

We use the Qualitative Spatial Reasoning (QSR) formulation Region Connection Calculus (RCC) as a means to segment, i.e., agglomerate, higher-level image objects. In Chapter 6 we use these higher-level image objects for classification and propose the use of these higher level objects for construction of a probabilistic image representation, which we call probabilistic graph models (PGMs).

#### Research Contributions

- Characterization of higher-level segmentation via RCC.
- Object- and image-level classification using higher-level objects.
- Introduction to probabilistic graph models.

A flowchart describing this dissertation research is shown in Figure 1.3.

**Figure 1.3.** Flowchart of dissertation research.

# Chapter 2

## Multispectral Analysis of Pixel-Level Nuclear Classification

We present analyses of our multispectral data designed to test the hypothesis that the additional spectral bands contain more information useful for classification than the three standard bands of RGB microscopy imagery. Results indicate only slight performance differences using multispectral imagery as opposed to derived RGB imagery. These performance differences are not, however, statistically significant in many cases. These conclusions hold for classification using all available image bands as well as using single image bands, indicating that the multispectral bands do not contain any more useful spectral information than do the RGB bands for this nuclear classification task. Preliminary results from this chapter are presented in [74] and [75].

### 2.1 Motivation

The use of multispectral imaging capabilities is relatively new to the field of cyto- and histo-pathology, particularly for transmitted brightfield microscopy.<sup>1</sup> Preliminary research and development of spectral imaging hardware for spectral brightfield microscopy occurred in the late 1990's [73, 76], and use of this technology for cancer research has occurred within the last few years [38, 39, 70, 71]. These publications have begun to explore the use of potential additional information contained in the spectral data (29-33 wavelengths in the visible spectrum, from 400 nm to 720 nm, spaced 10 nm apart), in particular for multiply stained ( $> 2$  stains) specimens. The added benefit of multispectral imaging for standard H&E histopathology imagery, however, is still largely

---

<sup>1</sup>Transmitted brightfield microscopy is the microscopy setup commonly encountered in high-school biology labs. A broad spectrum light source is directed underneath a specimen, and the transmitted light is viewed and/or captured with a camera setup.

unknown [76], although some promising results were presented in [38]. The working hypothesis in many cases is that the additional spectral bands contain more information than standard 3-band RGB images for image classification and analysis tasks.

## 2.2 Related Work

While the use of multispectral light microscopy is new to cyto/histopathology, many researchers have used single or dual narrow-band filters to enhance imagery for particular stains, most using a red filter (or the red channel of RGB imagery) for enhancement of Hematoxylin or Feulgen staining (both stains are blue) [8–10, 19, 40], and some using a green filter for enhancement of Feulgen staining [18, 20, 41, 45]. Other techniques, such as Fourier transform spectroscopy [72, 77] and vibrational spectroscopy [29, 30] have been around longer and have also used multispectral imaging capabilities.

There have been comparisons of spectral unmixing algorithms (to separate constituent dyes) between RGB imagery and multispectral brightfield imagery [70] and Fourier transform spectral imagery [72]. Both of these studies demonstrate the advantage of multispectral data for spectral unmixing, which in turn can be used for further processing tasks, e.g., segmentation. Recently, a study was published describing segmentation of cervical cells in Pap smears using the same multispectral imaging technology (VariSpec<sup>TM</sup>) as our data (refer to Section 1.2.2) [78]. No quantitative results were presented, however, nor any substantiation behind the claims that spectral data are better suited for the segmentation task.

We present analyses of our multispectral data designed to test the hypothesis that the additional spectral bands contain more information useful for classification than the three standard bands of RGB microscopy imagery. Section 2.3 presents methods for computation of RGB imagery from multispectral imagery. Description of the classifiers and performance metric for classification appear in Section 2.4. Results and analysis of the nuclear classifications are presented in Section 2.5. A summary of our research and directions for future research are presented in Sections 2.6 and 2.7, respectively.

## 2.3 Computation of Derived RGB Images

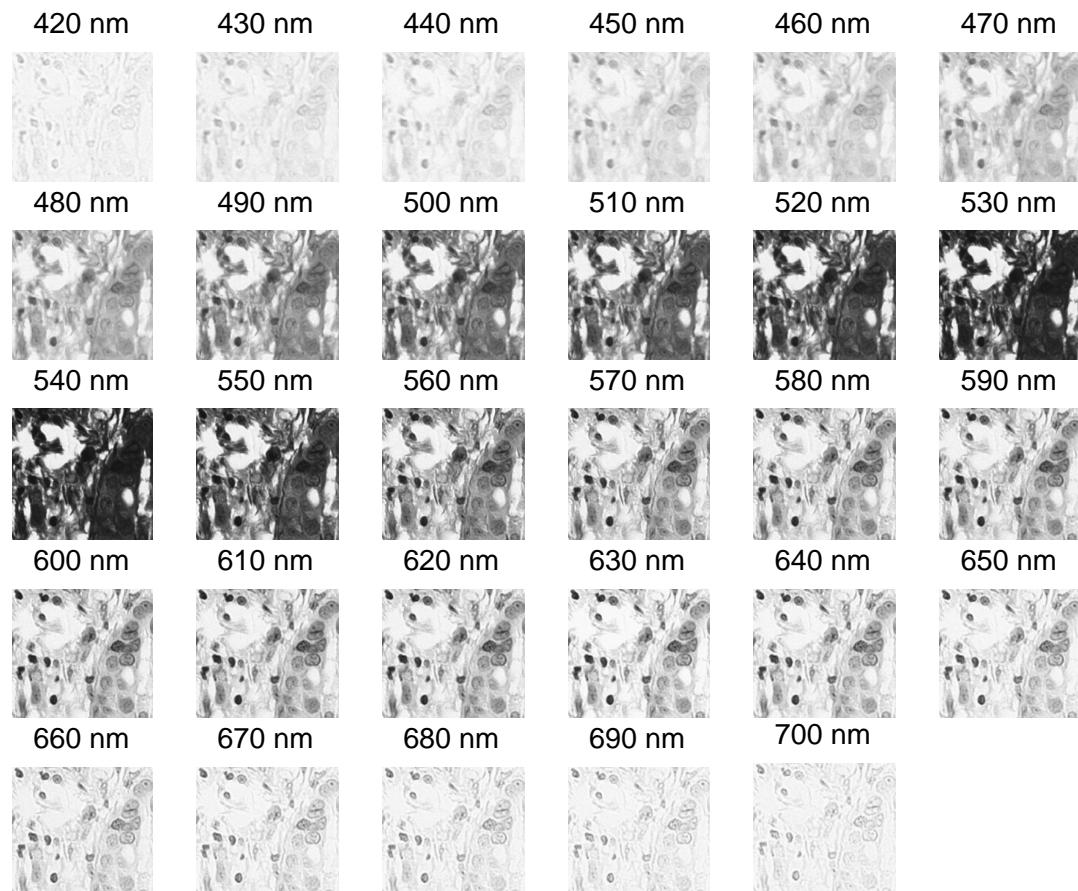
The dataset used for these multispectral analyses consists of 58 H&E histopathology images of breast tissue from the Yale Tissue Microarray Facility (<http://tissuearray.org/facility/inventory/list.php>). These data were captured from 5 microarrays (ytma10, 12, 49, and 55), with (6, 6, 34, and 6) images captured per array, respectively; in total there are 26 malignant images, and 32 benign (including 6 normal from ytma55). These 58 images are not microarray images in the general sense since they are single histopathology images as might be obtained from standard clinical biopsy specimens.

The multispectral images have 29 bands, spaced 10 nm apart, ranging within the visible spectrum from 420 to 700 nm. Each band is represented in an image stack as an 8 bit,  $768 \times 896$  grayscale image; refer to Figure 2.1 for an example image stack.

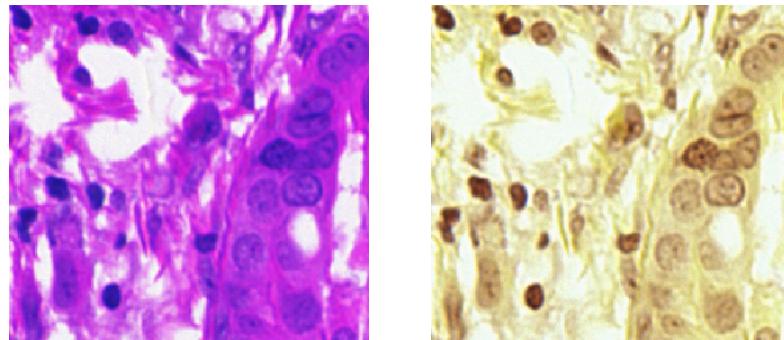
One could foresee many methods for the derivation of RGB imagery from multispectral. Single representative wavelengths may be chosen for red (R), green (G), and blue (B); this can result in a qualitatively wide variety of visualizations, two of which are shown in Figure 2.2 (a) and (b). This approach, however, limits any further analysis to the three narrow bands chosen. Another approach is to find some allocation for all the multispectral bands to R, G, and B. We use here:

1. **rgbequal:** Created by (approximately) equally allocating the 29 bands to R, G, and B, similar to the approach in [72], reflecting a rough approximation of the three spectral ranges associated with the three colors red, green, and blue, albeit with some ambiguity in allocation of intermediate colors (e.g., yellow). We define the blue band of such images as the average of the first 9 multispectral bands (420-500 nm), the green band as the average of the next 10 bands (510-600 nm), and the red band as the average of the last 10 bands (610-700 nm). An example rgbequal image is shown in Figure 2.2 (c).
2. **truecolor:** Created by converting the illumination wavelength for each band into the constituent RGB values as perceived by humans, then averaging the contribution to R, G, and B for each band. This method uses the MatlabCentral function *spectrumRGB*. The transfer functions for conversion of wavelength values to RGB values by *spectrumRGB* are shown in Figure 2.3. Note the second lobe in the red channel response that occurs in the blue wavelengths; this is due to the human perception of violet toward that end of the visible spectrum. The MatlabCentral file exchange can be found online at <http://www.mathworks.com/matlabcentral/fileexchange>. An example truecolor image is shown in Figure 2.2 (d).
3. **ccd:** A modification of truecolor imagery to better match the spectral response of common 3-CCD color cameras used in microscopy setups for biomedical research. This method also uses the *spectrumRGB* function, but removes the second lobe in the red response; thus, ccd imagery will differ from truecolor imagery only in the red channel. An example ccd image is shown in Figure 2.2 (e).

Some researchers have shown the advantages of other color spaces besides RGB, in particular, YUV [15], HSV [17, 37], and HSI [15, 80]. In all of these applications, the new color space is computed from an image originally specified in RGB. We have chosen to compare performance of our multispectral data to that of RGB imagery, since that provides the most straightforward derivation of 3-band imagery from multispectral and requires no further processing of the multispectral data itself.

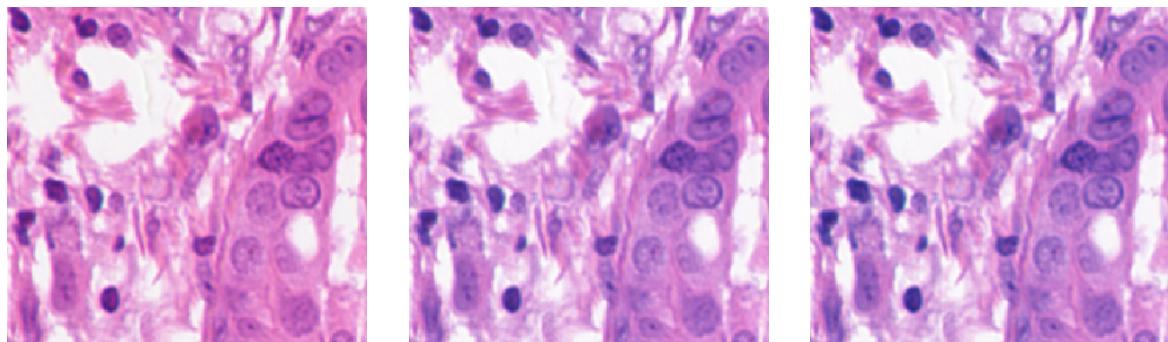


**Figure 2.1.** Example multispectral image stack, with each band displayed as an 8-bit grayscale image. Only a portion of the entire histology image is displayed here to allow better appreciation of detail. Note the good contrast for nuclei in the red portion of the spectrum (e.g., 570-630 nm) and the lack of inter-tissue contrast in the green portion of the spectrum (e.g., 520-550 nm).



(a) Bands chosen to represent spectral peaks of human cones (580, 540, and 450 nm) [79].

(b) Bands chosen arbitrarily (660, 630, and 560 nm).

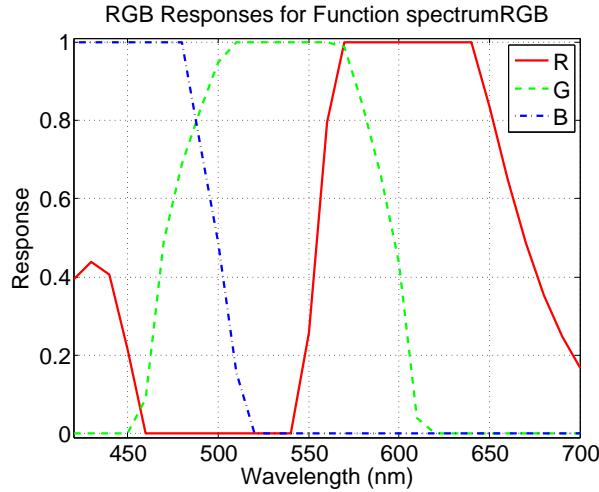


(c) Bands allocated equally and averaged, “rgbequal.”

(d) Bands allocated with MatlabCentral function *spectrumRGB*, “truecolor.”

(e) Bands allocated to approximate spectral responses of common 3-CCD color cameras, “ccd.”

**Figure 2.2.** RGB representations of the multispectral stack in Figure 2.1.



**Figure 2.3.** RGB responses for the MatlabCentral function *spectrumRGB*, used in the computation of truecolor and ccd RGB imagery. Note the second lobe in the red response due to the perception of violet at the smaller wavelengths; this second lobe is eliminated for computation of ccd imagery.

## 2.4 Pixel-Level Classification

In this section we describe the classifiers used for pixel-level classification of nuclei, and present the metric used to quantify classification performance.

### 2.4.1 Classifiers

We use six classifiers for our study of nuclear classification. These classifiers were chosen based on their established performance and use for multispectral data, sparsity of parameters to optimize, computational efficiency, and the use of (primarily) spectral information. The use of primarily spectral information is important in these analyses since the basic hypothesis in question deals with the spectral information content of our imagery.<sup>2</sup> The exceptions to these characteristics are noted in the classifier descriptions to follow. Notation in the following text will use bold-faced letters and symbols to denote vectors. Additionally, image pixels and class means are considered as vectors,

$$\mathbf{y} = [b_1, b_2, \dots, b_B]^\top \quad (2.1)$$

---

<sup>2</sup>Since texture is an important characteristic of cell nuclei, it is reasonable to speculate about the use of a classification method that uses texture information in its formulation. The interested reader is referred to Appendix D.

where  $b_i$ ,  $i = 1, \dots, B$ , is the value of the pixel in the  $i$ th band, and  $B$  is the total number of image bands.

**Maximum Likelihood (ML)** calculates the likelihood of class membership for an observation and assigns the observation to the class with the highest likelihood [81]. That is, a pixel is assigned the label of the class that it is most likely to be a member of. Likelihood is defined probabilistically, using the estimated joint probability density or mass function. The likelihood  $L$  as a function of the current observation  $\mathbf{y}$  is defined as:

$$L(\mathbf{y}) = f_{\mathbf{y}}(\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n | \mathbf{y}) \quad (2.2)$$

where  $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n$  are  $n$  random variables corresponding to  $n$  classes. The ML estimate,  $\hat{\mathbf{y}}$ , of  $\mathbf{y}$  is thus defined as

$$\hat{\mathbf{y}} = \max L(\mathbf{y}) \quad (2.3)$$

We are interested here, however, only in the class label  $z$  of  $\hat{\mathbf{y}}$ ,

$$z = C_i \quad \text{if} \quad \hat{\mathbf{y}} \in C_i \quad (2.4)$$

Mean and covariance is computed for each class, assuming Gaussianity, which results in a quadratic discriminant boundary.

**Minimum Euclidean Distance (MED)** finds the Euclidean distances between a current observation and the class means. The observation is assigned the class label of the closest class mean, i.e., the minimum Euclidean distance. The MED classification of observation  $\mathbf{y}$  is thus defined as

$$z = C_i \quad \text{if} \quad |\mathbf{y} - \mathbf{m}_i| < |\mathbf{y} - \mathbf{m}_j|, \quad i = 1, \dots, n, \quad i \neq j \quad (2.5)$$

where  $\mathbf{m}_i$  is the mean of the  $i$ th class for  $n$  total classes, and  $|\cdot|$  is the Euclidean distance metric.

**Spectral Angle Mapper (SAM)** finds the angle between a current observation and the class means. The observation is assigned the class label for which this angle is minimum. In this case,

$$z = C_i \quad \text{if} \quad \theta_i = \angle(\mathbf{y}, \mathbf{m}_i) < \theta_j = \angle(\mathbf{y}, \mathbf{m}_j), \quad i = 1, \dots, n, \quad i \neq j \quad (2.6)$$

Using the definition of the dot product of two vectors,  $\mathbf{a}$  and  $\mathbf{b}$ ,

$$\mathbf{a} \cdot \mathbf{b} = |\mathbf{a}| |\mathbf{b}| \cos(\theta) \quad (2.7)$$

Equation (2.6) can be rewritten as

$$z = C_i \quad \text{if} \quad \arccos\left(\frac{\mathbf{y} \cdot \mathbf{m}_i}{|\mathbf{y}| |\mathbf{m}_i|}\right) < \arccos\left(\frac{\mathbf{y} \cdot \mathbf{m}_j}{|\mathbf{y}| |\mathbf{m}_j|}\right), \quad i = 1, \dots, n, \quad i \neq j \quad (2.8)$$

**Fisher Linear Discriminant Analysis (FLDA)** constructs a decision hyperplane based on a linear combination of the input data planes. More specifically, it projects the multi-dimensional data to one dimension, maximizes a function representing the difference between the projected class means, and normalizes by the within-class scatter along a direction perpendicular to the decision hyperplane [82]. Further discussion here will be limited to the two-class case (following the explanation of Bishop in [82]), since we are considering here a two class problem. An input observation  $\mathbf{x}$  (of  $B$  dimensions) is projected to one dimension via the mapping

$$y = \mathbf{w}^\top \mathbf{x} \quad (2.9)$$

To implement a criterion to maximize class separation, we could choose the weight vector  $\mathbf{w}$  to maximize the function

$$\mathbf{w}^\top (\mathbf{m}_2 - \mathbf{m}_1) \quad (2.10)$$

subject to  $\mathbf{w}$  being unit length. To normalize by the within-class scatter, the within-class covariance is used:

$$s_k^2 = \sum_{n \in C_k} (y^n - \mathbf{w}^\top \mathbf{m}_k)^2, \quad k = 1, 2 \quad (2.11)$$

where  $y^n$  is the  $n$ -th element of class  $C_k$ . The Fisher criterion is thus defined as

$$J(\mathbf{w}) = \frac{(\mathbf{w}^\top \mathbf{m}_2 - \mathbf{w}^\top \mathbf{m}_1)^2}{s_1^2 + s_2^2} \quad (2.12)$$

Substituting previous equations into Equation (2.12), differentiating with respect to  $\mathbf{w}$ , and simplifying, it can be shown that

$$\mathbf{w} \propto \mathbf{S}_W^{-1} (\mathbf{m}_2 - \mathbf{m}_1) \quad (2.13)$$

where  $\mathbf{S}_W$  is the total within class covariance, in this case

$$\mathbf{S}_W = \sum_{n \in C_1} (\mathbf{x}^n - \mathbf{m}_1)(\mathbf{x}^n - \mathbf{m}_1)^\top + \sum_{n \in C_2} (\mathbf{x}^n - \mathbf{m}_2)(\mathbf{x}^n - \mathbf{m}_2)^\top \quad (2.14)$$

The projected data may now be used to construct a discriminant function  $d$  and threshold  $y_0$ <sup>3</sup> such that

$$z = \begin{cases} C_1 & \text{if } d(\mathbf{x}) \geq y_0, \\ C_2 & \text{else.} \end{cases} \quad (2.15)$$

The FLDA classifier is equivalent to the ML formulation if there are equal covariance

---

<sup>3</sup>The computation of a threshold is generally formulated as a decision boundary problem to minimize, e.g., the probability of misclassification. Additionally, since  $\mathbf{w}^\top \mathbf{x}$  is the sum of random variables, invoking the Central Limit Theorem, the class conditional density/mass functions can be defined using normal distributions [82].

matrices for the two classes; this results in a linear discriminant boundary.

An **automated feature extraction (AFE)** tool called **GENIE** (GENetic Imagery Exploitation) is based on evolutionary computation and is designed to explore the entire feature space of spectral and spatial operators for multispectral data, and evolve a solution best fit for the classification task. More practically speaking, GENIE selects a subset from the available data planes and develops an algorithm, consisting of spectral and spatial operators, to extract various feature planes which are subsequently fed to a standard classifier backend. GENIE selects an initial set of algorithms consisting of randomly selected operators and randomly selected data planes as input. Throughout the evolution process, only appropriate algorithms with appropriate data input will survive, as quantified by the classification performance of the algorithms. Classification performance is defined as an equal tradeoff between detection rate and false alarm rate. For this work, GENIE uses a simple Fisher linear discriminant backend, although the capability exists for other backend classifiers. For more information on GENIE, see Appendix E and Reference [83].

**Support Vector Machines (SVMs)** construct a linear hyperplane that maximizes the margin<sup>4</sup> between classes. This is a nonlinear optimization problem: for a two class problem, find the parameters  $\mathbf{w}$  and  $w_0$  of the hyperplane

$$\mathbf{w}^\top \mathbf{x} + w_0 = 0 \quad (2.16)$$

minimizing the quantity

$$J(\mathbf{w}, w_0, \boldsymbol{\xi}) = \frac{1}{2} \|\mathbf{w}\|^2 + C \sum_{i=1}^N \xi_i \quad (2.17)$$

subject to

$$y_i [\mathbf{w}^\top \mathbf{x}_i + w_0] \geq 1 - \xi_i, \quad \xi_i \geq 0, \quad i = 1, 2, \dots, N \quad (2.18)$$

where  $\boldsymbol{\xi}$  is a vector of “slack variables” representing the three categories of training vectors: correctly classified training outside of the class separation band, correctly classified training inside of the class separation band, and incorrectly classified training [84];  $C$  is a positive constant that weights the relative contribution of the two terms in  $J$ ;  $\mathbf{x}_i$  is the  $i$ -th training vector; and  $y_i$  is the class label ( $\pm 1$ ). In the case of nonlinear SVMs, the data is first mapped to a higher dimensional space, where a linear hyperplane is computed to separate the classes, using a kernel function which defines the inner product operation in the higher dimensional space [84]. We have implemented an SVM using SVM<sup>light</sup> [85], with linear (**LSVM**),

$$K(\mathbf{x}, \mathbf{z}) = \mathbf{x} \cdot \mathbf{z} \quad (2.19)$$

---

<sup>4</sup>The minimum separation (i.e., distance) between samples from different classes.

and quadratic (**NLSVM**),

$$K(\mathbf{x}, \mathbf{z}) = (s\mathbf{x}^\top \mathbf{z} + c)^2 \quad (2.20)$$

kernels. For the nonlinear SVM, the kernel parameters ( $s$  and  $c$  for quadratic) must be explicitly optimized for the training data; this is the only classifier used in this study which requires optimization of parameters.<sup>5</sup>

## 2.4.2 Performance Metric

We choose a general metric of classification performance that equally penalizes both types of classification errors: 1) true (nuclei) pixels incorrectly labeled as false (non-nuclei) and 2) false pixels incorrectly labeled as true. In particular, the performance metric is defined as

$$P = 0.5(R_d + (1 - R_f)) \quad (2.21)$$

where  $R_d$  is the fraction of true pixels classified correctly (detection rate),  $R_f$  is the fraction of false pixels classified incorrectly (false alarm rate), and the factor of 0.5 scales the metric to the range [0, 1]. Note that a perfect segmentation will yield a performance score of 1 (100%), while a score of 0.5 (50%) can be obtained by a trivial solution of all pixels labeled true (or false). We will present our results in terms of this metric  $P$ , as well as the individual metrics comprising  $P$ : detection rate (DR) and false alarm rate (FAR).

The performance metric in Equation (2.21) is valid for cases where ground truth is defined on a pixel-by-pixel basis for some region(s) of the image; the best estimate of performance is achieved when ground truth is defined for the entire image. As a compromise between the necessity of comprehensive ground truth for proper quantification of classification accuracy, and the tedious and time-consuming aspect of human delineation of such ground truth, we have marked a  $200 \times 200$  pixel window in each of the 58 histology images.<sup>6</sup> These windows were chosen to best encompass the range of visual features in the imagery, according to subjective evaluation by the human viewer. We feel that this approach will not be any more biased toward certain image features than would a random approach; in many cases, a randomly located  $200 \times 200$  pixel window could encompass little, if any, cellular material. This window is used to determine classification performance for each image, as well as provide training data for classifier optimization.

---

<sup>5</sup>Strictly speaking, the AFE tool GENIE has many free parameters (e.g., number of feature planes) that can be optimized. For the purpose of this study, however, we use the default parameters. Refer to Reference [83] for more description of the parameters associated with GENIE.

<sup>6</sup>The imagery and ground truth markups are available from <http://flour.ece.ucsb.edu:8080/bisquick/>.

## 2.5 Pixel-Level Classification Results

In this section we demonstrate the nuclear classification performance of our imagery using several well known classifiers. We use multispectral images, each of the three derived RGB images, as well as individual bands from each of the image types (multispectral, rgbequal, truecolor, and ccd). Nuclear classification is an important task since many of the characteristics of cancer are expressed in the cell nuclei. Indeed, much of cytology analysis is based solely on nuclear characteristics. Additionally, it is easy to define a reasonably precise pixel-level ground truth for nuclei which lends this task to a thorough examination of classification accuracy. We discuss the pixel-level classification of cytoplasm and stroma in Chapter 3.

We present classification performance using all available image bands of each image type in Section 2.5.1, followed by classification performance of individual image bands in Section 2.5.2, and a discussion of the differences in performance using all available bands versus individual bands in Section 2.5.3. We conclude our discussion of results by presenting the image bands used preferentially in the AFE and FLDA classifiers (Section 2.5.4), analysis of the entropic content of individual image bands (Section 2.5.5), and a principal components analysis of the multispectral imagery (Section 2.5.6).

### 2.5.1 Classification Using All Available Image Bands

We split our dataset in (approximately) half<sup>7</sup> to create a training set of images and a test set of images; half of the benign and malignant subsets were randomly assigned to the training or test set to allow for even representation of benign and malignant characteristics in both sets. The same random half of the data was provided to each classifier. Applying the six aforementioned classifiers<sup>8</sup> to each complete image stack (i.e., all available bands of multispectral, rgbequal, truecolor, and ccd), and averaging over the images contained in the training (in-sample) and test (out-of-sample) sets, we achieve the results shown quantitatively in Table 2.1 and graphically in Figure 2.4. Since the AFE tool GENIE is stochastic, averages for ten independent runs are presented. Since the optimization of the quadratic kernel for NLSVM is very time consuming, results presented here are for a randomly selected 10% of the training data.

In Table 2.1, the best in-sample (IS) and out-of-sample (OS) performance scores are in bold-face. The best performance occurs for either the multispectral or ccd imagery, with ML and AFE performing better with ccd; MED, FLDA, and both SVMs performing better with multispectral, and SAM split between the two for IS and OS. It is important

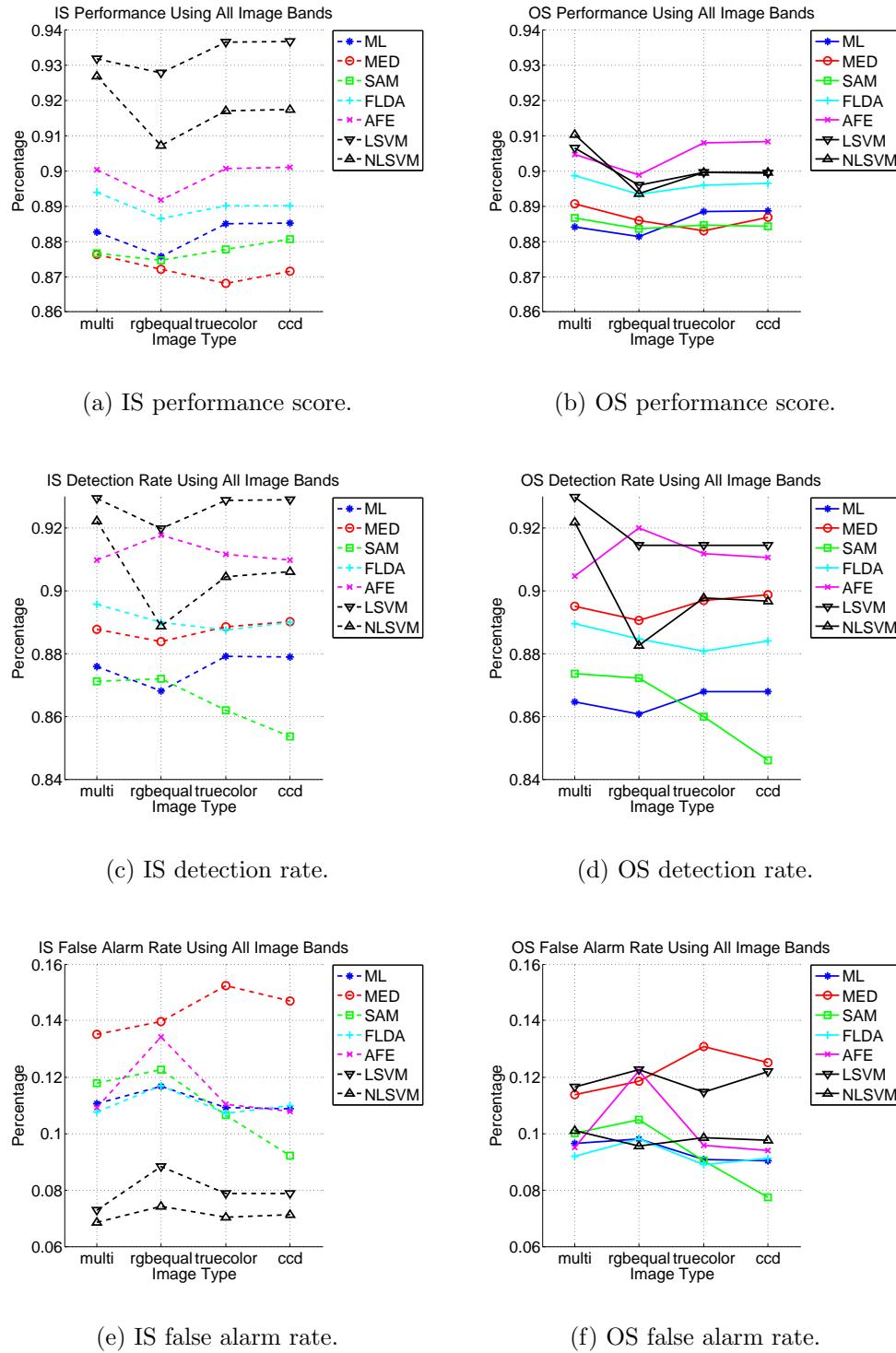
---

<sup>7</sup>One of the benign images was left out of the training set, i.e., the training set actually contains 28 images and the test set 30 images rather than an even split of 29 images each. All results take this into account.

<sup>8</sup>Many thanks to Zhiqiang Bi for performing and optimizing the SVM classifications presented here.

**Table 2.1.** Detection rate (DR), false alarm rate (FAR), and performance (P) using all available image bands. Results are presented as in-sample (IS)/out-of-sample (OS) and as percentages. Bold rates correspond to the best rates over all image types for each classifier.

Classifier	Image Type	DR	FAR	P
ML	multispectral	0.876/0.865	0.111/0.097	0.883/0.884
	rgbequal	0.868/0.861	0.117/0.098	0.876/0.881
	truecolor	<b>0.879/0.868</b>	<b>0.109/0.091</b>	0.885/0.889
	ccd	<b>0.879/0.868</b>	<b>0.109/0.090</b>	<b>0.885/0.889</b>
MED	multispectral	0.888/0.895	<b>0.135/0.114</b>	<b>0.876/0.891</b>
	rgbequal	0.884/0.891	0.140/0.120	0.872/0.886
	truecolor	0.889/0.897	0.152/0.131	0.868/0.883
	ccd	<b>0.890/0.899</b>	0.147/0.125	0.872/0.887
SAM	multispectral	<b>0.871/0.874</b>	0.118/0.100	<b>0.877/88.7</b>
	rgbequal	<b>0.872/0.872</b>	0.123/0.105	0.875/0.884
	truecolor	0.862/0.860	0.107/0.090	0.878/0.885
	ccd	0.854/0.846	<b>0.092/0.077</b>	<b>0.881/0.884</b>
FLDA	multispectral	<b>0.896/0.890</b>	0.108/0.092	<b>0.894/0.899</b>
	rgbequal	0.890/0.885	0.117/0.098	0.887/0.893
	truecolor	0.888/0.881	<b>0.107/0.089</b>	0.890/0.896
	ccd	0.890/0.884	0.110/0.091	0.890/0.897
AFE	multispectral	0.908/0.905	0.109/0.095	0.900/0.905
	rgbequal	<b>0.918/0.920</b>	0.134/0.122	0.892/0.899
	truecolor	0.912/0.912	0.110/0.096	<b>0.901/0.908</b>
	ccd	0.910/0.911	<b>0.108/0.094</b>	<b>0.901/0.908</b>
LSVM	multispectral	<b>0.929/0.922</b>	<b>0.073/0.101</b>	<b>0.932/0.911</b>
	rgbequal	0.920/0.883	<b>0.088/0.096</b>	0.928/0.894
	truecolor	<b>0.929/0.898</b>	0.079/0.099	<b>0.937/0.900</b>
	ccd	<b>0.929/0.897</b>	0.079/0.098	<b>0.937/0.900</b>
NLSVM	multispectral	<b>0.922/0.930</b>	<b>0.069/0.117</b>	<b>0.927/0.901</b>
	rgbequal	0.888/0.915	0.074/0.123	0.907/0.896
	truecolor	0.905/0.915	<b>0.070/0.115</b>	0.917/0.900
	ccd	0.906/0.915	0.071/0.122	0.918/0.900



**Figure 2.4.** Average performance and detection/false alarm rates using all available image bands for in-sample (IS) and out-of-sample (OS).

**Table 2.2.** Performance of the Minimum Euclidean Distance (MED) classifier using all image bands for different randomly selected training sets. Set 1 is the training set used for all the previously reported results. To facilitate easy observation of the better performing set (in-sample (IS) or out-of-sample (OS)), the better performance is presented in bold face.

Training Set	IS $P$	OS $P$
<b>Set 1</b>	0.876	<b>0.891</b>
<b>Set 2</b>	0.871	<b>0.897</b>
<b>Set 3</b>	<b>0.884</b>	0.883
<b>Set 4</b>	<b>0.890</b>	0.876
<b>Set 5</b>	0.878	<b>0.887</b>
<b>Set 6</b>	<b>0.884</b>	0.879
<b>Set 7</b>	<b>0.892</b>	0.874
<b>Set 8</b>	<b>0.884</b>	0.882
<b>Set 9</b>	0.879	<b>0.886</b>
<b>Set 10</b>	0.876	<b>0.890</b>

to note, however, that these increases in performance are only 0.004 IS and 0.004 OS when averaged over all classifiers.

It is surprising that the OS performance is generally higher (0.0076 increase in performance) than the IS performance for all classifiers but the SVMs, since performance generally decreases for unseen data. Looking at the performance of individual images in the test and training set, we note that there are several images in both the training and test sets with markedly poor performance; it appears that the IS set has slightly poorer performing images than the OS set. As a further test of this conjecture, we retrained a classifier (in this case, MED) with 9 additional randomly chosen training sets (10 training sets total including the original set used for all previous results). The IS and OS performance on multispectral imagery for these different training sets are shown in Table 2.2. It appears from Table 2.2 that the choice of images for the training and test sets affects the relative performance between the two, i.e., with other randomly chosen training and test sets, we see that we have an IS performance better than OS performance about half of the time.

From the results in Table 2.1 and Figure 2.4, we see that we can achieve a performance  $P$  (Equation (2.21)) ranging from 0.876 to 0.937 for IS and from 0.887 to 0.911 for OS, depending on the particular classifier. Results are also presented for DR and FAR in Table 2.1 and Figure 2.4 as reference. FAR and DR are commonly used metrics for classification performance, and were defined in Section 1.3.2. Future observations and analysis of results will most likely involve a relative weighting assigned to DR and FAR, in terms of performance required by pathologists.

**Table 2.3.** Wilcoxon paired signed-rank test p-values for performances of multispectral versus RGB imagery, presented as in-sample (IS)/out-of-sample (OS). Starred (\*) entries correspond to statistical significance at the p-value of 0.05. Using the Bonferroni correction for these 21 tests, we look for the corrected p-value of  $0.05/21 \approx 0.0024$ , which are presented as bold entries in the table. Note that all bold entries also satisfy the less conservative p-value of 0.05 although they are not explicitly starred.

Classifier	Image–multi vs.		
	rgbequal	truecolor	ccd
<b>ML</b>	4.7e-1/5.4e-1	1.5e-1/1.0e-1	1.5e-1/8.2e-2
<b>MED</b>	<b>1.3e-4/5.3e-5</b>	<b>8.2e-4/1.4e-4</b>	<b>1.0e-3/8.3e-4</b>
<b>SAM</b>	3.1e-2*/5.7e-3*	1.6e-1/7.3e-1	5.8e-2/8.3e-1
<b>FLDA</b>	8.3e-3*/6.5e-2	4.7e-2*/7.5e-2	3.4e-2*/1.2e-1
<b>AFE</b>	<b>3.8e-4/3.0e-3*</b>	5.8e-1/1.1e-1	8.0e-1/2.9e-2*
<b>LSVM</b>	1.8e-1/ <b>1.2e-3</b>	8.8e-2/6.3e-1	5.6e-2/4.3e-1
<b>NLSVM</b>	<b>4.2e-6/6.9e-5</b>	<b>6.4e-4/4.7e-3*</b>	<b>9.6e-4/6.0e-3*</b>

A paired Wilcoxon signed-rank test<sup>9</sup> is used to determine the statistical significance of these differences in performance; results are shown in Table 2.3 where only half of these differences are statistically significant for a p-value of 0.05 (the starred entries in Table 2.3). Only a few performance differences satisfy the corrected p-value of  $0.05/21 \approx 0.0024$  (bold entries), according to the Bonferroni correction.<sup>10</sup>

We have shown in this section, using a pairwise Wilcoxon signed rank test, that only a few performance differences between multispectral and RGB imagery are actually statistically significant. Taking the performance differences from Table 2.1 for all those shown to be statistically significant in Table 2.3, the expected performance increase can be determined for all statistically significant differences. In the following, bold differences correspond to significance at the Bonferroni corrected p-value and the remainder for significance at the uncorrected p-value of 0.05:

- **IS: 0.004, 0.008, and 0.004** increase in favor of multispectral over rgbequal, true-

---

<sup>9</sup>The Wilcoxon signed-rank test is an alternative to the common Student’s t-test for situations in which the underlying data distribution cannot be assumed to be normal and/or the classes cannot be assumed to have the same variance. The paired Wilcoxon signed-rank test is a two-sided test used to compare the distributions of two related measurements, and specifically tests for a zero median value of the pairwise sample differences. The p-value is the probability that the samples are symmetrically distributed about a single central (i.e., median) point [86].

<sup>10</sup>The Bonferroni correction compensates for the possibility that some number of statistical tests will randomly appear to be significant. For a p-value of  $0.05 = 1/20$ , it is expected that one out of every twenty tests will, by chance, appear to be significant. The Bonferroni correction, which is conservative, specifies that the corrected p-value should be  $p_c = p/n$ , where  $p$  is the original p-value, and  $n$  is the number of tests being performed over the same data [87].

color, and ccd, respectively, for MED; 0.002 in favor of multispectral over rgbequal for SAM; 0.007, 0.004 and 0.004 in favor of multispectral over rgbequal, truecolor, and ccd, respectively, for FLDA; **0.008** in favor of multispectral over rgbequal for AFE; and **0.020**, **0.010**, and **0.009** in favor of multispectral over rgbequal, truecolor, and ccd, respectively, for NLSVM.

- **OS:** **0.005**, **0.008**, and **0.004** increase in favor of multispectral imagery over rgbequal, truecolor, and ccd, respectively, for MED; 0.003 in favor of multispectral over truecolor for SAM; 0.006 in favor of multispectral over rgbequal and 0.003 in favor of ccd over multispectral for AFE; **0.011** in favor of multispectral over rgbequal for LSVM; and **0.017**, 0.011, and 0.011 in favor of multispectral over rgbequal, truecolor, and ccd, respectively, for NLSVM.

### 2.5.2 Classification Using Single Image Bands

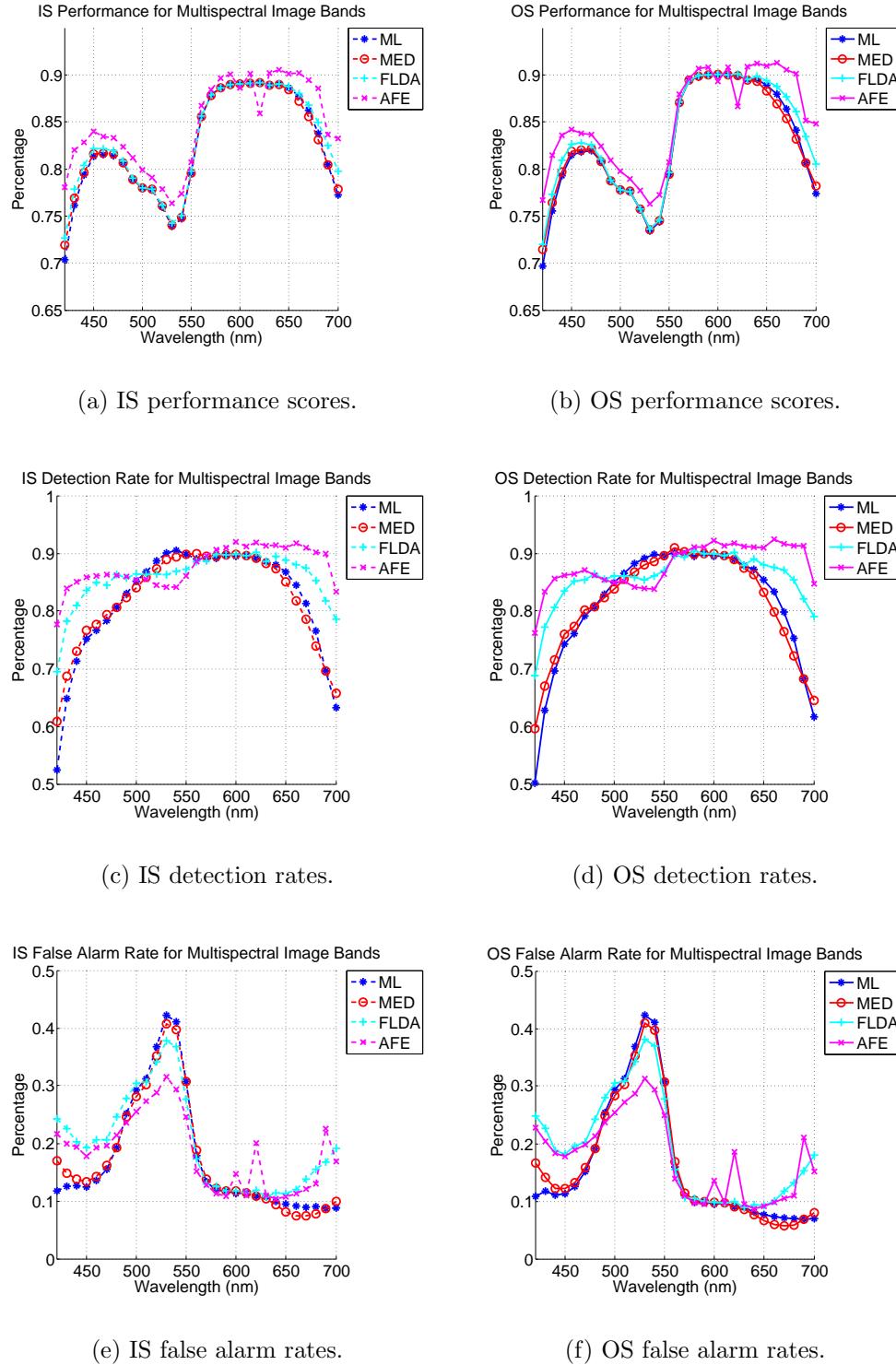
To gain more understanding of the relative contributions of specific image bands, the ML, MED, FLDA, and AFE classifiers are applied to each individual image band for each image type. The SAM classifier is excluded here since it will fail on one-band images,<sup>11</sup> and the SVMs are excluded for computational reasons (it would be prohibitively computationally intensive to optimize kernel parameters for each image band). Performance scores for classification using single multispectral bands are shown in Figure 2.5 (a) and (b) where the best performance scores occur in the red portion of the spectrum, with poorer performance in the green portion and at the extremes of the spectrum. Empirical observations of typical multispectral imagery (refer to Figure 2.1) reveal a good contrast between nuclei and non-nuclei pixels in the red bands, decent contrast in the blue bands, and very little contrast in the green bands and extreme ends of the spectrum. These empirical observations correspond nicely with the graphs in Figure 2.5 (a) and (b).

Looking at the DR and FAR for single multispectral bands, shown in Figure 2.5 (c)-(f), we note that the DR for the bands is more consistent across the spectrum of image bands, with a rolloff at the extremes of the spectrum. The FAR on the other hand, displays much more variation over the image bands; thus it seems that FAR has the major effect on overall performance score.

In these classification runs, the AFE tool tends to outperform the other classifiers, although it does not have as clear an advantage as when using all image bands. Additionally, AFE has lower DR in the green portion of the spectrum, but also lower FAR. AFE is the only classifier that has degraded performance for the bands at 600 and 620

---

<sup>11</sup>Referring to Equation (2.7), in the case of one-element vectors,  $\mathbf{a} = |\mathbf{a}|$ , thus  $\mathbf{a} \cdot \mathbf{b} = |\mathbf{a}||\mathbf{b}| \Rightarrow \cos(\theta) = 1 \Rightarrow \theta = 0$ . Thus, the SAM classifier will assign a class label of ‘1’ to the observation since the spectral angle is apparently 0. This reduces the classification problem to the trivial solution mentioned previously: assigning a single label to all pixels, resulting in a performance score of 0.5.



**Figure 2.5.** Performance and detection/false alarm rates on single multispectral bands for in-sample (IS) and out-of-sample (OS).

nm; this is due to the poor performance of one of the 10 independent runs for these particular image bands. The last thing we note is the better performance of the AFE classifier in the blue portion of the spectrum compared with the other classifiers. This is most likely due to the ability of AFE to extract local spatial features to assist in the more difficult classification of blue bands. Using purely spectral information in the red bands (i.e., ML, MED, and FLDA), however, performs nearly as well as AFE.

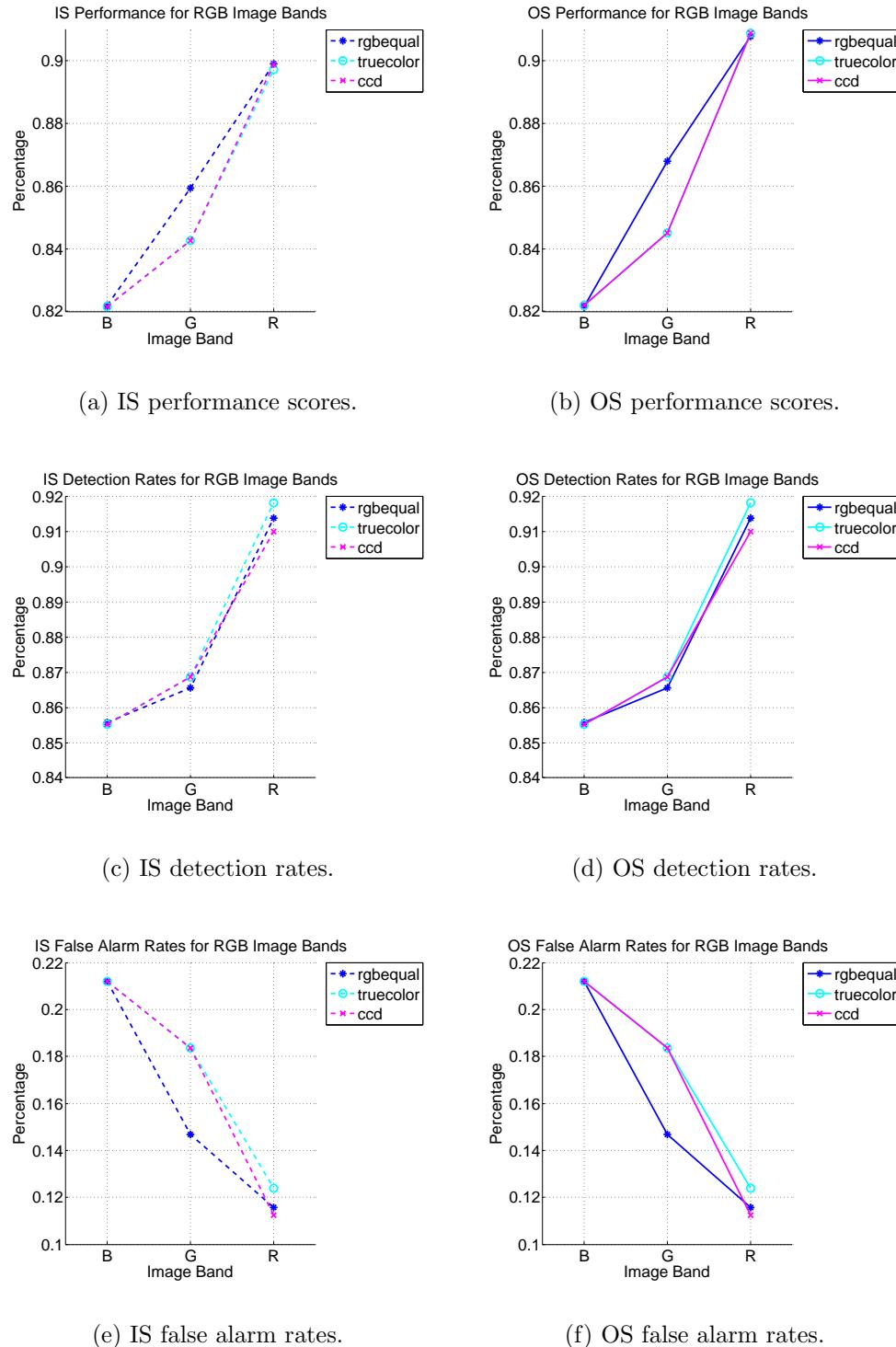
We present now the classification results for single RGB channels. Plotted in Figure 2.6 are the performance, DR, and FAR of the AFE classifier on single image bands of rgbequal, truecolor, and ccd imagery. The AFE classifier is plotted here since it consistently yielded the highest performance scores; Appendix F contains plots for the other three classifiers. It should be noted that while the green and blue channels of ccd imagery are identical to those of truecolor imagery, all three ccd channels are plotted for completeness. Similar to the multispectral single band plots, the red channels yield the best performance, as well as the highest DR and lowest FAR; this is most likely due to the good contrast for nuclei in the red channels (refer to Figure 2.1). The most significant differences between the RGB image types occurs in the green channels.

The distinction in performance of red channels between the RGB image types is not large; refer to Table 2.4. Here the single best performing multispectral band yields a performance increase of 0.005 IS and 0.006 OS, similar to the performance increase using all multispectral bands versus all RGB bands. As reference, Table 2.4 also displays the IS and OS performance using all image bands.

It is interesting that the poorest performing RGB channel is the blue, whereas with multispectral data, the green portions of the spectrum are the poorest performers. While it may seem contradictory that in RGB imagery the green channel outperforms the blue channel when the opposite is true in multispectral imagery, it is important to remember how the multispectral bands are allocated to each of the RGB bands. Consider, for example, the allocation of bands in rgbequal imagery: the bands from 510 nm to 600 nm are averaged to yield the green channel. Referring to Figure 2.5 (a) these bands have a large variation in performance. Thus, to obtain the green channel, we are averaging multispectral bands, several of which have relatively good performance. A similar situation occurs with the truecolor and ccd imagery, albeit with a weighting applied to the various “green” bands.

We find the analysis of performance on single image bands satisfactory from an intuitive standpoint. Since the nuclei are stained with the blue-colored Hematoxylin which will block red light, the red portions of the spectrum have the best contrast and perform best for this nuclear classification task. While green light is also blocked by the Hematoxylin, so also is it blocked by the Eosin, rendering the green portion of the spectrum less informative for the task at hand.

The distinction in performance of red channels between the RGB image types is not large; we do note, however, that the single best performing multispectral band yields a



**Figure 2.6.** Performance and detection/false alarm rates on single RGB bands with the AFE classifier for in-sample (IS) and out-of-sample (OS).

**Table 2.4.** Maximum performance using all available image bands and single image bands presented as in-sample (IS)/out-of-sample (OS). ‘P’ is the performance using all image bands, ‘Max P’ is the maximum performance of a single image band, and ‘Max P Band’ is the band yielding the best performance.

Classifier	Imagery	All Bands	Single Bands	
		P	Max P	Max P Band
ML	<b>multispectral</b>	0.883/0.884	0.892/0.900	620/590 nm
	<b>rgbequal</b>	0.876/0.881	0.883/0.888	R/R
	<b>truecolor</b>	0.885/0.889	0.889/0.898	R/R
	<b>ccd</b>	0.885/0.889	0.889/0.898	R/R
MED	<b>multispectral</b>	0.876/0.891	0.892/0.900	620/600 nm
	<b>rgbequal</b>	0.872/0.886	0.880/0.882	R/R
	<b>truecolor</b>	0.868/0.883	0.888/0.898	R/R
	<b>ccd</b>	0.872/0.887	0.889/0.898	R/R
FLDA	<b>multispectral</b>	0.894/0.899	0.892/0.901	620/620 nm
	<b>rgbequal</b>	0.887/0.893	0.884/0.893	R/R
	<b>truecolor</b>	0.890/0.896	0.889/0.899	R/R
	<b>ccd</b>	0.890/0.897	0.889/0.900	R/R
AFE	<b>multispectral</b>	0.900/0.905	0.905/0.913	640/660 nm
	<b>rgbequal</b>	0.892/0.899	0.899/0.908	R/R
	<b>truecolor</b>	0.901/0.908	0.897/0.909	R/R
	<b>ccd</b>	0.901/0.908	0.892/0.901	R/R

performance increase of 0.006 as compared to the single best RGB bands, averaged over all 4 classifiers. This performance increase is consistently in favor of single multispectral image bands, but is not generally statistically significant (refer to Table 2.5), especially for the more complex (and physically realistic) allocations of the truecolor and ccd imagery.

We have shown in this section that performance differences between single multispectral image bands and single RGB image bands are not statistically significant. This indicates that the individual multispectral image bands are not yielding any more useful spectral information than are the individual RGB image bands for classification of nuclei pixels in H&E stained imagery.

### 2.5.3 Single Image Bands versus All Image Bands

One curious result from the previous section is that use of single image bands tend to outperform classification using all image bands; this results in a performance increase of 0.007 IS and 0.007 OS (most easily seen in Table 2.4). While these differences in performance are not large, the consistency of the bias toward single image bands is

**Table 2.5.** Wilcoxon paired signed-rank test p-values for performances of the best multispectral band versus the red channel of RGB imagery, presented as in-sample (IS)/out-of-sample (OS). Starred (\*) entries correspond to statistical significance at the p-value of 0.05. Using the Bonferroni correction for these 12 tests, we look for the corrected p-value of  $0.05/12 \approx 0.0042$ , presented as bold entries in the table. Note that all bold entries also satisfy the less conservative p-value of 0.05 although they are not explicitly starred.

Classifier	Multi Band	Band		
		rgbequal R	truecolor R	ccd R
ML	620/590 nm	2.9e-2*/3.2e-2*	4.0e-2*/3.1e-1	3.1e-2*/3.4e-1
MED	620/600 nm	6.9e-2*/2.2e-2*	1.0e-1/2.5e-1	1.0e-1/1.7e-1
FLDA	620/620 nm	6.3e-3*/ <b>1.7e-3</b>	2.5e-1/2.5e-1	1.2e-1/3.6e-1
AFE	640/660 nm	<b>2.3e-4</b> /9.4e-2	<b>1.8e-3</b> /4.0e-1	<b>1.7e-3</b> /4.7e-1

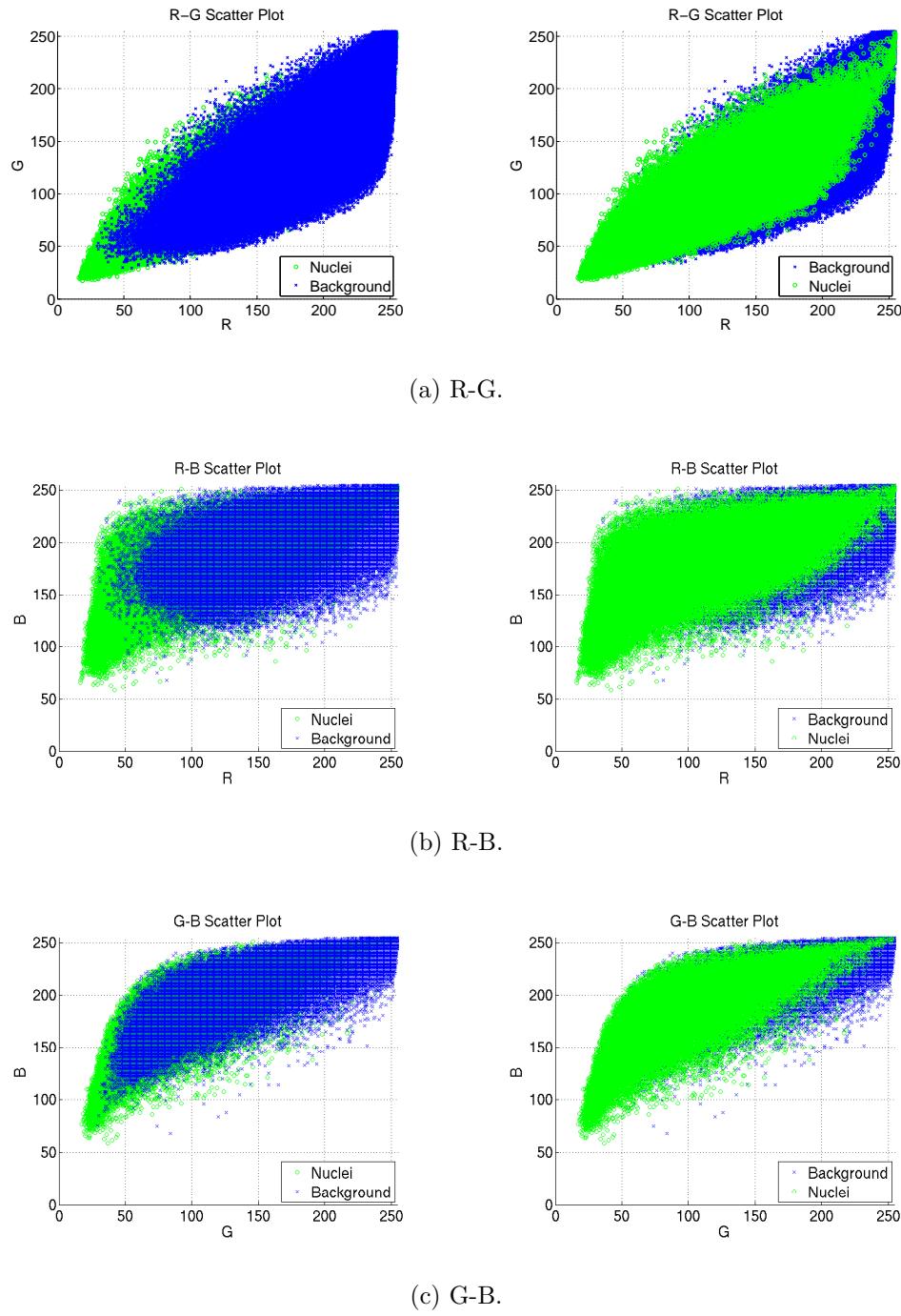
**Table 2.6.** Wilcoxon paired signed-rank test p-values for performances of best single image bands versus all image bands, presented as in-sample (IS)/out-of-sample (OS). Starred (\*) entries correspond to statistical significance at the p-value of 0.05. Using the Bonferroni correction for these 16 tests, we look for the corrected p-value of  $0.05/16 \approx 0.0031$ , presented as bold entries in the table. Note that all bold entries also satisfy the less conservative p-value of 0.05 although they are not explicitly starred.

Classifier	Image			
	multi	rgbequal	truecolor	ccd
ML	<b>1.0e-3/2.4e-4</b>	<b>2.9e-4</b> /7.3e-3*	4.0e-1/2.0e-2*	3.6e-1/2.3e-2*
MED	3.3e-3*/5.0e-2*	4.3e-2*/8.8e-1	<b>1.4e-3</b> /1.1e-2*	<b>1.4e-3</b> /2.2e-2*
FLDA	9.2e-2/8.3e-1	3.8e-2*/9.1e-1	2.6e-1/7.3e-1	2.7e-1/6.6e-1
AFE	3.4e-2*/2.2e-1	<b>3.1e-3/1.5e-4</b>	5.6e-2/9.3e-1	3.6e-2*/7.2e-1

striking. Additionally, this performance difference may be statistically significant in as many as half of the cases (refer to Table 2.6).

To attempt to explain this result, we look at the distribution of nuclei and background pixels in the form of scatter plots (Figure 2.7). Since all the classifiers considered use a linear (or approximately linear<sup>12</sup>) decision hyperplane, they may have a slightly easier time correctly discriminating a 1-D distribution (e.g. just the red channel). Referring to the 2-D scatter plots in Figure 2.7, the distributions are distinctly non-Gaussian (an assumption made by several of the classifiers), and it appears that a curved decision boundary may improve performance. It is our conjecture that the addition of more

<sup>12</sup>While the ML classifier results in a quadratic boundary, we see in Figure 2.7 that the two classes have approximately equal scatter. This would indicate approximately equal covariance matrices which will yield an approximately linear boundary.



**Figure 2.7.** Scatter plots of nuclei versus non-nuclei in ccd imagery of the whole dataset. Nuclei points are green o's and non-nuclei points are blue x's. Each scatter plot is displayed both with blue and green on top to allow better appreciation of the extent of each class.

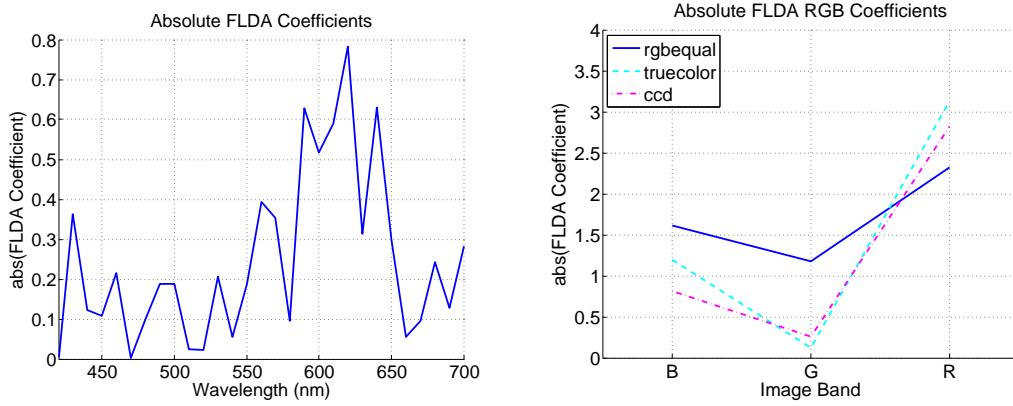
image bands may decrease performance slightly, by adding additional dimensions for which a linear hyperplane cannot adequately compensate. We would also like to note here that although the scatter plots in Figure 2.7 appear to have significant inter-class overlap, it is the density of these points that allows for as good a classification as we achieve with simple linear classifiers.

As a test of our conjecture, we return to our nonlinear SVM results from the previous section (refer to Table 2.1). There is an average performance of 0.927 IS and 0.901 OS, as compared to the best non-SVM performance of 0.901 IS and 0.908 OS from the AFE classifier. This would seem to indicate the superiority of a curved (quadratic, in this case) decision boundary for the higher dimensional classification problems, at least for IS data. While we also note in Table 2.1 that the linear SVM (LSVM) outperforms the quadratic SVM (NLSVM), we must keep in mind that the NSLVM kernel was trained for only 10% of the training data, resulting in a less than optimal solution as compared to the LSVM which had access to 100% of the training data.

#### 2.5.4 Analysis of FLDA Coefficients and Bands Chosen in AFE Solutions

We expect that the single image bands which yield the best performance should also be the bands used most often by the classifiers. A direct examination of this is possible with the FLDA and AFE classifiers. For FLDA, image bands are weighted and summed; the higher the *absolute value* of the coefficient, the more important the image band. A plot of these coefficients for multispectral and RGB imagery is shown in Figure 2.8. For the AFE classifier, more important image bands should be chosen more often in solutions; a plot of the average number of times an image band is chosen in an AFE solution is shown in Figure 2.9, where the 10 independent runs have been averaged. Once again, in both the FLDA and AFE classifier, there is a preference for the red portion of the spectrum.

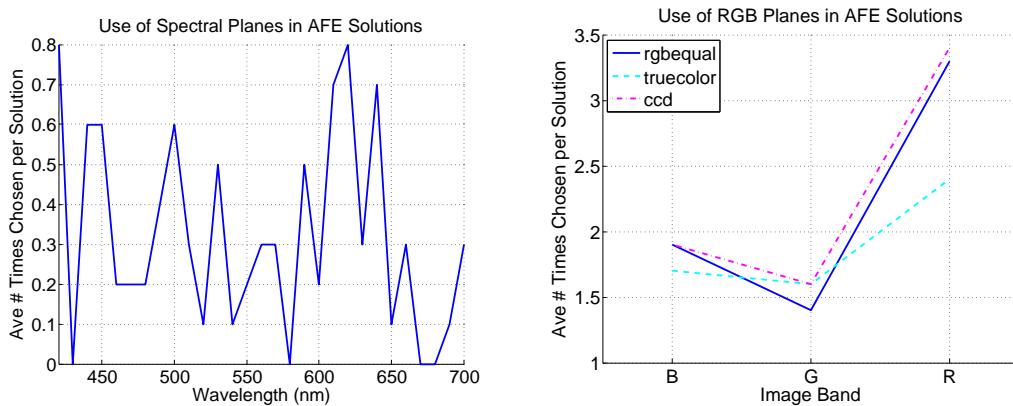
We note also that with RGB imagery (Figures 2.8 (b) and 2.9 (b)), the FLDA classifier weights the red channel the most, followed by the blue, and finally green channels. Similarly, the AFE classifier chooses the red channel most often, followed in turn by blue and green. Comparing the multispectral plots for the AFE and FLDA classifiers (Figures 2.8 (a) and 2.9 (a)), there are striking similarities in the relative use/weighting of bands, particularly in the red portion of the spectrum (i.e., 580-650 nm). The more prevalent use of green and blue bands in the AFE classifier, compared to FLDA, may be due to the classifier's ability to extract local spatial features, making those bands more useful beyond the raw spectral attributes used by the FLDA classifier. Additionally, some of these differences may average out if more AFE runs were considered. (In particular, the AFE classifier displays an odd preference for 420 nm which has very little contrast (refer to Figure 2.1)). Overall, considering the disparate nature of these two classifiers, we find it interesting that they both display similar preferences for particular image bands.



(a) Absolute value of FLDA coefficients for multispectral imagery.

(b) Absolute value of FLDA coefficients for RGB imagery.

**Figure 2.8.** Absolute value of Fisher Linear Discriminant Analysis (FLDA) coefficients. Larger absolute value indicates greater weight (importance) of the image band to the overall classification performance.



(a) Multispectral bands chosen in AFE solutions.

(b) RGB bands chosen in AFE solutions.

**Figure 2.9.** Spectral bands chosen in Automated Feature Extraction (AFE) solutions, averaged over 10 independent runs.

We use the analysis in this section as a complement to the analysis of performance on single image bands. Specifically, we have shown that image bands that yielded better performances are also the image bands chosen preferentially in both the FLDA and AFE classifiers. While it may be more qualitatively satisfying if the plots of Figures 2.8 (a) and 2.9 (a) would bear more resemblance to those of Figure 2.5 (a) and (b), it is important to remember that these two analyses are very distinct from one another. In the case of Figure 2.5 (a) and (b), we are limiting the classifiers to a single image band, and optimizing the performance, whereas for Figures 2.8 (a) and 2.9 (a) we are providing the classifiers with a choice of all available image bands and optimizing performance. As a more intuitive example, for the FLDA classifier, even if a specific image band X performs well when used alone, this same image band X may not yield as much information as, say, the linear combination of bands Y and Z. We have shown, therefore, in this analysis, a classifier preference for image bands that have also been shown to yield better performance when used singly in classification.

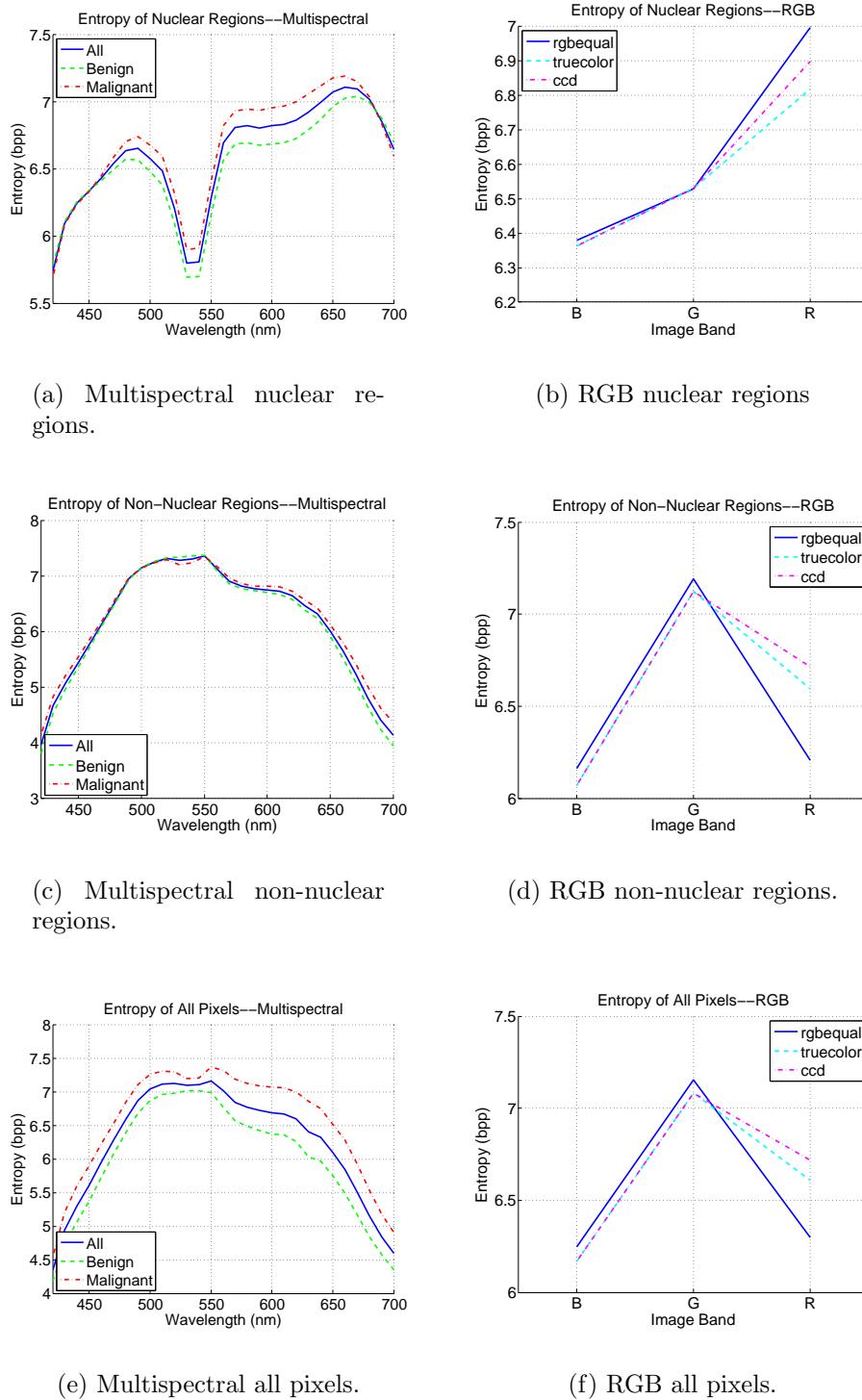
### 2.5.5 Entropic Analysis of Image Bands

In order to quantify the source of the slight performance differences between the various image types, we look at the entropy of regions in our imagery. For this, entropy is defined as

$$H = - \sum_{i=1}^N P(a_i) \log_2 P(a_i), \quad (2.22)$$

where  $N$  is the total number of grayscale levels,  $a_i$  is the  $i$ -th level, and  $P(a_i)$  is the probability of the  $i$ -th level (generally estimated from the image histogram) [79].

Entropies of our imagery are plotted in Figure 2.10 and summarized in Table 2.7. From these results, we make several observations. First, the highest entropy occurs in the red portion of the spectrum, corroborating well with the best performing image bands being in the red portion of the spectrum. Second, malignant images have a higher entropy than the benign images; this is not surprising since malignant nuclei tend to have more pronounced chromatin texture. There is, however, significant overlap between entropies of individual images and the average malignant or benign entropies. Third, our multispectral imagery has slightly higher nuclear entropy than any of the derived RGB imagery. This could lead to the erroneous conclusion that multispectral imagery should therefore have the best performance; entropy is, however, only one particular measure of information content. The information that human experts use in their segmentation of nuclei involves higher level concepts beyond pixel-level entropy.



**Figure 2.10.** Entropy of images/image regions. For clarity we plot only the entropy of the total dataset for the RGB plots.

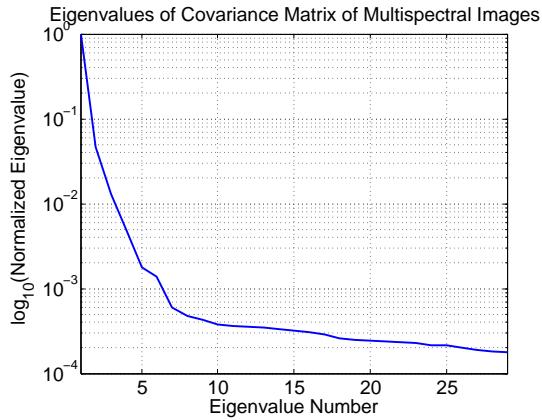
**Table 2.7.** Entropy of image regions.

Region	Imagery	Max Entropy	Max Entropy Band
Nuclei	<b>multispectral</b>	7.11 bpp	660 nm
	<b>rgbequal</b>	7.00 bpp	R
	<b>truecolor</b>	6.82 bpp	R
	<b>ccd</b>	6.90 bpp	R
Non-Nuclei	<b>multispectral</b>	7.35 bpp	550 nm
	<b>rgbequal</b>	7.19 bpp	G
	<b>truecolor</b>	7.12 bpp	G
	<b>ccd</b>	7.12 bpp	G
All Pixels	<b>multispectral</b>	7.16 bpp	550 nm
	<b>rgbequal</b>	7.15 bpp	G
	<b>truecolor</b>	7.08 bpp	G
	<b>ccd</b>	7.08 bpp	G

### 2.5.6 Principal Components Analysis of Multispectral Images

We use Principal Components Analysis (PCA) [88] as a dimensionality reduction method to see how many “important” bands actually exist within our multispectral imagery. Input to the PCA algorithm is the  $(768 \cdot 896) \times 29$  matrix where rows correspond to a single image pixel and columns are the pixel values for each of the 29 multispectral image bands. The average sorted eigenvalues of the covariance matrix of this input are plotted in Figure 2.11, where the eigenvalues for each image are normalized such that the largest eigenvalue has unit value. There was very little difference in this plot for malignant and benign subsets of our data, so the eigenvalue plot is averaged over our entire dataset. It was somewhat surprising that there was only one dominant eigenvalue, with the second ranked eigenvalue at approximately one-tenth the value of the dominant one. Given that there are two stains in our histopathology imagery, we expected that there would be two dominant eigenvalues. Figure 2.12 shows the projection of an example image onto the first three eigenvectors. The first projection seems to highlight nuclear regions (i.e., the Hematoxylin), the second projection seems to highlight the connective tissue and cytoplasm (i.e., the Eosin), and the third and subsequent projections do not have any obvious correlation with the tissue stains.

We have thus found that PCA indicates the presence of two dominant eigenvalues, if the principal components responsible for 97% of the variation in the data are considered. This indicates the presence of only two information-bearing projections of the imagery for this nuclear classification task, providing insight into the approximately equivalent performance of the RGB imagery and multispectral. We have also shown that these two informative projections demonstrate a direct relationship to the two image stains. Interestingly, the first component is responsible for 93% of the total variation; this component



**Figure 2.11.** Logarithmic plot of the eigenvalues of multispectral imagery, from Principal Components Analysis (PCA). Eigenvalues for each image are normalized so that the largest eigenvalue has unit value.

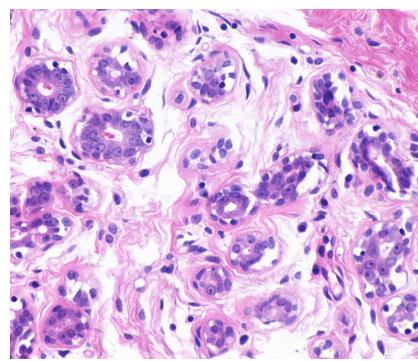
is generally correlated with Hematoxylin, but is sometimes correlated instead with what appears to be a grayscale version (i.e., intensity) of the original image; the reasons for this are not immediately clear. The possibility that other image bands may contain important diagnostic information for further analysis (i.e., analysis beyond a simple pixel-level nuclear classification) is still an open question [76].

## 2.6 Summary

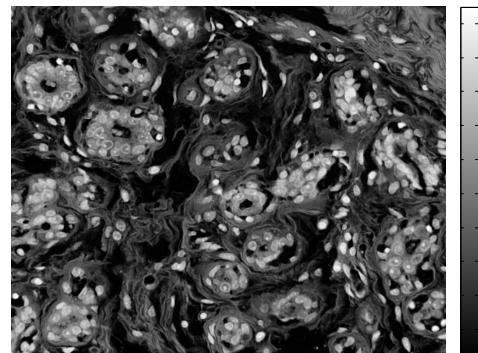
We have shown in this chapter a demonstration of performance for different image types and different classifiers in a nuclear classification task. First, results indicate only slight performance differences using multispectral imagery as opposed to derived RGB imagery. These performance differences are not, however, statistically significant in many cases, especially in the OS set and for the physically realistic ccd and truecolor imagery. These conclusions hold for both classification using all available image bands as well as using single image bands, indicating that the multispectral bands do not contain any more specific spectral information than do the RGB bands for this nuclear classification task.

Second, we have also shown that the single image bands with the best performance are the image bands chosen more often/weighted more heavily by the AFE and FLDA classifiers. Analysis of the entropy of image bands similarly reveals a preference for the red portion of the spectrum for nuclei.

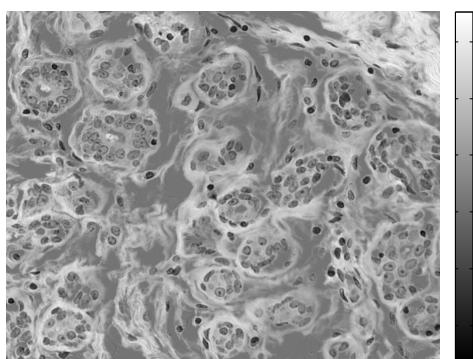
Finally, we have shown through the use of PCA as a dimensionality reduction method, that only two bands are carrying 97% of the variation in our image data, and appear



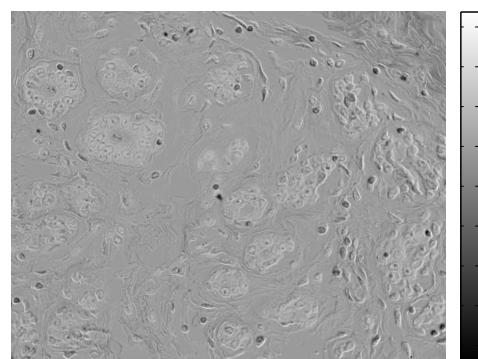
(a) ccd image.



(b) Projection on first eigenvector.



(c) Projection on second eigenvector.



(d) Projection on third eigenvector.

**Figure 2.12.** Projection of an example image onto the first three Principal Components Analysis (PCA) eigenvectors.

to be correlated with the two image stains. This result provides some insight into the roughly equivalent performance of RGB imagery to multispectral.

While the results presented here are intriguing, they are by no means complete, since we are considering only a single pixel-level classification task. This is the first thorough quantitative and comparative analysis between multispectral and RGB imagery for a common classification task.

## 2.7 Future Directions

As this work on pixel-level classification of histological images is of integral importance to many applications, there are several distinct directions that future research may take. First, we have considered only a single low-level classification task for nuclei. Further research could involve pixel-level classification of other histologic entities (some work will be presented in Chapter 3), molecularly specific pixel-level classification [76] (to identify subtle differences in the spectra associated with different dye-protein interactions), as well as the object-level classification of histologic entities. We will touch on the latter area in Chapter 5, but our main focus will be the development of object-level analysis and not an analysis of the different imagery for such an analysis task.

Second, while we have presented analysis for the RGB color space, there may be other color spaces more suitable for this and other classification tasks. As previously mentioned, other research has demonstrated the utility of YUV [15], HSV [17, 37], and HSI [15, 80]. Additionally, the conversion of the imagery to optical density (OD), which provides a linear relationship between image intensity and staining density, may yield a useful image space to explore the virtues of multispectral and standard RGB imagery. Similarly, dimensionality reduction techniques such as PCA and Independent Components Analysis (ICA) [89] may also provide an informative image space.

Finally, a very important research direction is the incorporation of feedback to the system. In this case, one could “fine-tune” the results of the pixel-level classification based on the results of a further higher-level analysis. In particular, the results of higher-level analysis can be incorporated into the performance metric for the pixel-level classification. This avenue of research is broad and applicable to a variety of problems, especially as more research investigates higher-level image analysis. As it stands, there seems to be a bit of a rift between the “standard” low-level image analysis techniques and the newly developed higher-level techniques. Methods to help bridge this gap are sorely needed.

# Chapter 3

## Pixel-Level Classification of Cytoplasm and Stroma

In our imagery, cytoplasm and stroma are not easily distinguished based solely on spectral information, especially when the inter-image variability in staining is taken into account. Thus, we find a pixel-level classification for the combined cytoplasm/stroma class and then discuss methods of discrimination between cytoplasm and stroma using texture and other features. We present a method to distinguish between tissue and background, whereby we can simply subtract the previous nuclei classification results from the new tissue class to yield a combined cytoplasm and stroma class. This cytoplasm and stroma class can be further processed to discriminate between the two classes. We find our best overall average performance (average of performance for classification of cytoplasm and stroma classes) to be approximately 0.86.

### 3.1 Motivation

While the characteristics of cell nuclei are well established as useful for diagnostic purposes, it is expected that the characteristics of cytoplasm and stroma will be similarly useful. The qualities of cytoplasm and stroma in breast cancer do not appear to have been specifically addressed in the literature, but stroma in pancreatic cancer has been noted as displaying distinctive growth patterns [90, 91]. It is likely that some characteristics of the cytoplasm and stroma are affecting the pathologist's perception of many histopathology slides [92, 93]. The discrimination between cytoplasm and stroma, however, is not commonly addressed in quantitative pathology studies, especially for standard H&E stained imagery (Sims et al. in [90] and [91] use sirius red and light-green and still require manual classification of color clusters). We seek to develop a method to automatically and robustly distinguish cytoplasm from stroma in our pathology imagery. This will allow a full object-level description to be developed for both entities in addition

to the commonly quantified nuclear features.

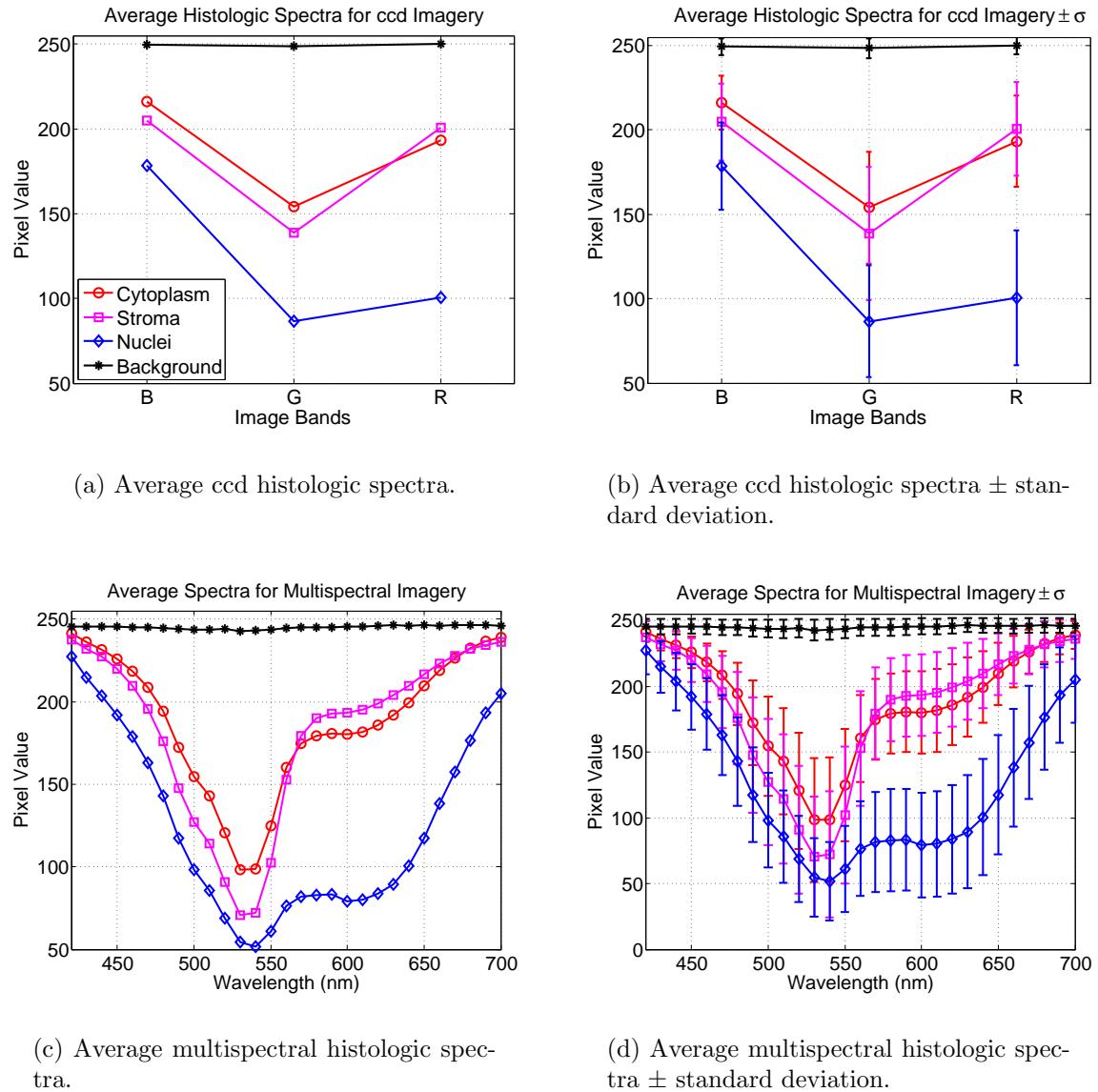
Of ultimate interest is the separate classification of cytoplasm and stroma pixels. This proves to be a difficult (if not impossible) task using solely spectral information, especially considering the wide variability in the spectral appearance of cytoplasm (from light purple to vivid pink); refer to Figure 3.1. We are thus more interested in the 2-class problem, distinguishing between tissue and background, whereby we can simply subtract the previous nuclei classification results from the new tissue class to yield a combined cytoplasm and stroma class. This cytoplasm and stroma class can be further processed to discriminate between the two classes. In general, stromal structures display a sinuous and fibrous structure with a strong parallel direction between the various fibers. Cytoplasm, on the other hand, has a more homogeneous and granular texture; as an example, refer to Figure 3.2.

## 3.2 Related Work

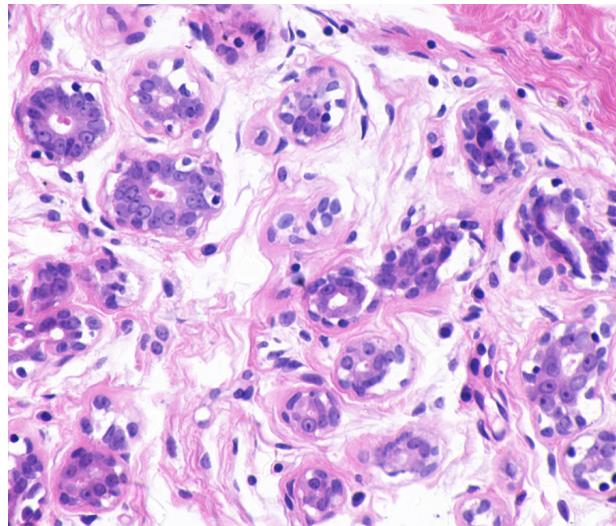
As previously mentioned, publications on classification/discrimination of cytoplasm and stroma for H&E imagery are quite sparse. Masood et al. [94] use the Nuance system from CRi (Woburn, MA; the same company that manufactures the VariSpec™ LCTF) to obtain multispectral imagery, and rely on the fundamental assumption that the various histological classes are spectrally distinct. It also appears that the segmented classes in [94] must be manually assigned to their respective histologic classes. No quantitative results are presented for the segmentation accuracy of this method.

Teverovskiy et al. [95] and Tabesh et al. [96] use a rule-based approach to segment several histologic entities including white space, red blood cells, nuclei (stromal, apoptotic, and epithelial), stroma, lumen, cytoplasm, and nucleoli. This approach relies on several measurements of color information as well as some shape-based metrics; it also relies on a large number of empirically determined thresholds. Image “objects” are determined by a region growing algorithm using color and shape regularity. These objects are merged to form the various histologic classes based on spectral and morphologic features. The quantitative results presented in [95] and [96] focus on Gleason grading scores resulting from the automated analysis system; no quantitative results are presented for segmentation accuracy of histologic entities.

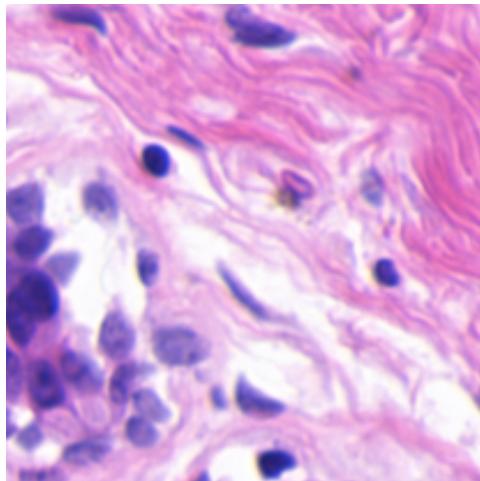
In our imagery, cytoplasm and stroma are not easily distinguished based solely on spectral information, especially when the inter-image variability in staining is taken into account. In this chapter we present the development and analysis of a classifier for cytoplasm and stroma. Section 3.3 presents the pixel-level classification for the combined cytoplasm/stroma class and Section 3.4 discusses methods of discrimination between cytoplasm and stroma. We summarize our research in Section 3.5 and conclude with possible directions for future research in Section 3.6.



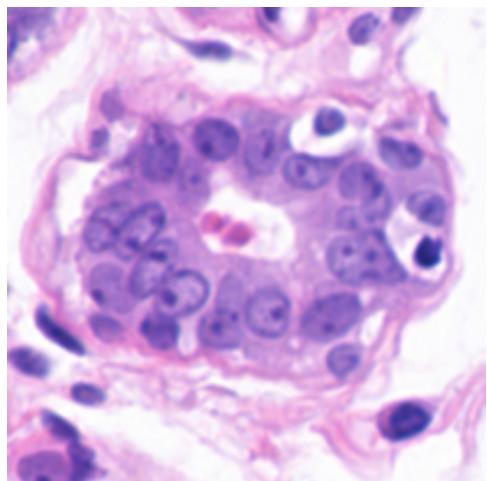
**Figure 3.1.** Multispectral and ccd average histologic spectra are plotted in (a) and (c) to allow better appreciation of the spectral shape, and average spectra  $\pm$  the standard deviation are plotted in (b) and (d) to illustrate the overlap of the cytoplasm and stroma classes across the spectrum. Background and nuclei classes are shown as reference; note that these classes do not have the same overlap as do the cytoplasm and stroma classes.



(a) Example ccd image.



(b) Stroma can range in color and intensity, but generally displays a pink color and a sinuous, fibrous texture.



(c) Cytoplasm is generally granular to homogeneous in texture and is most apparent here on the inner edge of this ductal structure.

**Figure 3.2.** Example of stroma and cytoplasm. For context, the entire image is shown in (a), while detail of stroma and cytoplasm regions are shown in (b) and (c).

### 3.3 Pixel-Level Classification

Since cytoplasm and stroma are both stained predominantly with Eosin, the problem of the classification of the combined cytoplasm and stroma class may be solved with solely the use of spectral information. We turn our attention to simple methods for the classification of tissue versus background. One simple method would be to find a threshold for the green channel which could separate out the tissue portions of the image. While this is certainly a valid approach, we have reservations about the use of a threshold-based method with imagery that is subject to such wide variation in staining intensity. This leads us to approach the problem as an unsupervised classification.

#### 3.3.1 k-means Clustering and Region Merging

As opposed to the supervised classification techniques used in Chapter 2, unsupervised techniques do not rely on a training markup but rather attempt to cluster the image pixels according to some similarity criterion defined *a priori*. The k-means clustering algorithm is chosen for pixel-level classification of cytoplasm and stroma versus nuclei and background. k-means is a widely known and used algorithm which clusters data into  $k$  clusters while minimizing the intracluster variance [97]:

$$V = \sum_{k'=1}^k \sum_{\mathbf{x} \in C_{k'}} \|\mathbf{x} - \mathbf{m}_{k'}\|^2 \quad (3.1)$$

where  $C_{k'}$  is the  $k'$ -th cluster. The number of clusters  $k$  must be specified by the user and is generally based on the desired number of output classes or knowledge of the image constituents.

Noting that there are three spectrally distinct classes, background (white), nuclei (blue-purple), and cytoplasm/stroma (pink-red), we choose  $k = 3$ . k-means clusters are assigned to the background  $C'_b$ , nuclei  $C'_n$ , and cytoplasm/stroma  $C'_{cs}$  as follows:  $C'_b$  is the class with lightest average green value,  $C'_n$  is the class with the darkest average green value, and  $C'_{cs}$  is the remaining k-means class. This rule-based approach is based on the observation that the nuclei tend to be the darkest staining histologic entities in the image.

While it may seem odd to classify cytoplasm and stroma pixels as the absence of nuclei and background pixels, it is the background and nuclei pixels that are the easiest to quantify. Indeed, we have thorough pixel-level markups of nuclei used in Chapter 2. These ground truths can be used as a measure for the classification ability of the k-means classifier. This will lend a quantitative metric to the classification of the k-means algorithm beyond the qualitative analysis of the cytoplasm/stroma classification.

Although we expect three main spectrally distinct classes, the merging of a larger

number of classes may prove more accurate. The rule-based approach presented above is modified for use with  $k > 3$ , where the average values of the red, green, and blue channels ( $R(k)$ ,  $G(k)$ , and  $B(k)$ , respectively) have been calculated ahead of time:

1. Find the cluster with the lightest average gray value in the green band. The green band is used since this band encompasses the wavelengths of light that are blocked by both Hematoxylin and Eosin. Thus, the intensity of the green band gives an indication of the intensity of total stain. This cluster with the lightest average green value is assigned to the background. Call this cluster  $C'_b$ .
2. Find the cluster with the darkest average gray value in the green band. It is expected that the darkest regions of the image will correspond to the high-density chromatin areas of the nuclei. Call this cluster  $C'_n$ .
3. Iterate through the following steps
  - (a) Find the most commonly occurring spatially adjacent label to  $C'_n$ .
  - (b) Defining  $k_{sa}$  as the label of this most common adjacent cluster, if

$$B(k_{sa}) > R(k_{sa}) + \text{mean}(B - R), \quad (3.2)$$

then assign this cluster to nuclei ( $C'_n = C'_n \cup C_{k_{sa}}$ ), and proceed to step (c). Otherwise, assign this and all remaining clusters to cytoplasm/stroma ( $C'_{cs} = C_{k_{sa}} \cup (C'_b \cup C'_n)^C$ ) and terminate the iterative process. In the preceding equations  $\cup$  is the union and  $(\cdot)^C$  is the set complement operator.

- (c) (Degenerate case) If  $C'_n \cup C'_b \equiv \bigcup C_k$ , terminate the iterations. Otherwise, repeat step (a).

### 3.3.2 Classification Results

We quantify the performance of this k-means merging algorithm by computing the classification performance on just the nuclei using the same ground truth used in Chapter 2. This allows a comparison of the nuclei classification capabilities of the k-means merging algorithm as compared to the supervised learning methods that were extensively studied in Chapter 2. Only the ccd RGB imagery is used for this quantification. The k-means merging method presented above would work as is for the rgbequal and truecolor imagery, and could be easily modified to work with the multispectral imagery. We choose to focus our efforts on the ccd imagery, however, since we have shown in the previous section the approximately equivalent performance of spectral classification algorithms between the different imagery. Choice of the ccd imagery is also motivated by positive feedback from our pathology collaborators as to the aesthetic equivalence of our ccd imagery to standard H&E stained imagery in histology textbooks [93]. The three

metrics presented in Chapter 2, namely detection rate (DR), false alarm rate (FAR), and performance (P) are used; refer to Section 2.4.2 and Equation 2.21 therein. Results for  $k = 3, \dots, 15$  are presented in Table 3.1. For easy reference, Table 3.2 includes the OS rates for ccd imagery from the six classifiers of Section 2.4.1.

Referring to Table 3.1, there are fairly consistent results across the different values of  $k$ , with a slight performance advantage (on average) for  $k = 3$ . Since the larger values of  $k$  require longer computation time to converge,  $k = 3$  is chosen as the optimal value both in terms of computational efficiency and overall performance. Table 3.1 also shows the performance for classification of background pixels, compiled using a sparser markup indicating slide background pixels. There is a very good performance for this class across all values of  $k$ . We were expecting a degraded performance from the k-means algorithm as compared to the classifiers used in Section 2.4.1, shown in Table 3.2, but instead found comparable performance. As such, both the k-means classification of the combined cytoplasm/stroma class as well as nuclei could be used for further analysis. The discrimination between cytoplasm and stroma will be discussed in subsequent sections.

## 3.4 Discrimination between Stroma and Cytoplasm

Now that we have demonstrated pixel-level classification of both nuclei and slide background, leaving the combined cytoplasm and stroma class remaining, we focus our attention on the discrimination between cytoplasm and stroma.

### 3.4.1 Gabor Feature-Based Discrimination

A wavelet-based texture measure is chosen as a powerful but compact representation of texture (i.e., high energy spatial information). In general, stromal structures display a sinuous and fibrous structure with a strong parallel direction between the various fibers. As an example, refer to Figure 3.2. Due to the strong orientation component of the stromal texture, the Gabor wavelet is chosen as a basis for a texture discrimination between cytoplasm and stroma.

Gabor functions are Gaussians modulated by complex sinusoids, and have been proposed as an apt representation of human visual receptor fields [98, 99]. The two-dimensional Gabor functions in the time- and frequency-domain are implemented according to [100, 101] (in which the redundancy between the various Gabor filters is minimized), and a modification of the code from [100] is used,<sup>1</sup> where the Gabor filtering is performed in the frequency domain to avoid the computationally intensive convolutions required in the time domain. The mean and standard deviation of the resulting filtered

---

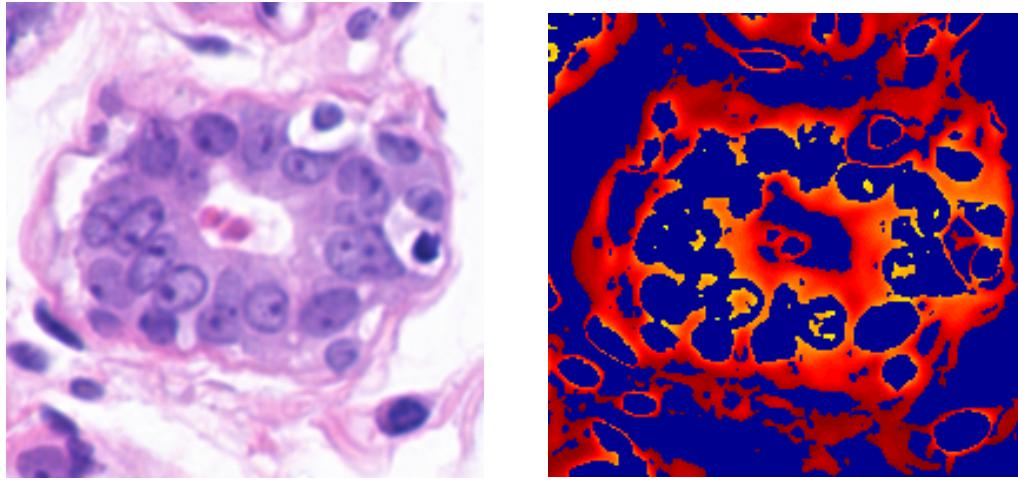
<sup>1</sup>Available at <http://vision.ece.ucsb.edu/texture/software/>

**Table 3.1.** Detection rate (DR), false alarm rate (FAR), and performance (P) for k-means classification. Bold rates correspond to the best rates. Performance for nuclei is compiled using the comprehensive ground truth of Chapter 2, while performance for the background class is compiled using a sparser ground truth indicating regions of slide background.

k	Nuclei			Background		
	DR	FAR	P	DR	FAR	P
<b>3</b>	0.903	<b>0.103</b>	<b>0.900</b>	0.982	0.040	0.971
<b>4</b>	0.900	0.194	0.852	<b>0.997</b>	0.016	0.990
<b>5</b>	0.941	0.163	0.889	0.995	0.009	0.993
<b>6</b>	0.909	0.123	0.893	0.994	0.005	<b>0.995</b>
<b>7</b>	0.925	0.143	0.891	0.993	0.004	0.994
<b>8</b>	0.932	0.143	0.894	0.990	0.003	0.994
<b>9</b>	0.949	0.171	0.889	0.983	0.002	0.990
<b>10</b>	<b>0.954</b>	0.167	0.834	0.978	0.002	0.988
<b>11</b>	0.949	0.162	0.893	0.968	0.002	0.983
<b>12</b>	0.939	0.148	0.896	0.964	<b>0.001</b>	0.098
<b>13</b>	0.944	0.163	0.890	0.951	<b>0.001</b>	0.975
<b>14</b>	0.948	0.164	0.892	0.945	<b>0.001</b>	0.972
<b>15</b>	0.931	0.155	0.888	0.940	<b>0.001</b>	0.970

**Table 3.2.** Out-of-sample detection rate (DR), false alarm rate (FAR), and performance (P) for Section 2.4.1 classifiers. Bold rates correspond to the best rates.

Classifier	DR	FAR	P
<b>ML</b>	0.868	0.090	0.889
<b>MED</b>	0.899	0.125	0.887
<b>SAM</b>	0.846	<b>0.077</b>	0.884
<b>FLDA</b>	0.884	0.091	0.897
<b>AFE</b>	0.911	0.094	<b>0.908</b>
<b>LSVM</b>	0.897	0.098	0.901
<b>NLSVM</b>	<b>0.915</b>	0.122	0.900



(a) Reference ccd image, displaying both stroma and cytoplasm.

(b) Gabor output (mean of the 12 features).

**Figure 3.3.** Gabor output. The nuclei and background have been masked out in the Gabor output in (b), leaving only the combined cytoplasm and stroma class. Note that stromal areas display a larger value in the Gabor output (red areas in (b)), and the cytoplasm areas display smaller values (orange and yellow in (b)).

output are commonly used features for texture characterization [100, 101]; we use six orientations and one scale, resulting in twelve features.

Since we expect that stromal regions will have stronger Gabor output (refer to Figure 3.3, we would like to determine a robust threshold for the Gabor output. We avoid any absolute threshold due to the staining intensity variations in our imagery. Defining a relative threshold as the point at which the rate of change of the thresholded images is maximum, this should correspond to the threshold at which the major cytoplasm regions are eliminated. Results using these thresholds are shown in Table 3.3, where there are poor DRs for both cytoplasm and stroma and an overall performance (averaged cyto-

**Table 3.3.** Detection rate (DR), false alarm rate (FAR), and performance (P) for Gabor cytoplasm/stroma discrimination, using a sparse ground truth markup of cytoplasm and stroma regions for compilation of results. Total performance is the average of cytoplasm and stroma performance.

	Cytoplasm			Stroma			Total P
	DR	FAR	P	DR	FAR	P	
Gabor	0.442	0.071	0.685	0.555	0.069	0.743	0.714

plasm and stroma performance) of 0.714. Since there are twelve results from the Gabor filtering, the mean over these feature planes is presented. The mean is used rather than the maximum since we are interested in regions with a strong response in any orientation, rather than just in the six specific orientations of the Gabor filters.

Empirical observations of the resulting segmentations suggest that the addition of the red channel information may improve the segmentation performance. This leads to the question of whether there are other features that could be of use in a feature selection framework.

### 3.4.2 Nuclear Proximity, Gabor, and Spectral Feature-Based Discrimination

Cytoplasm and stroma are distinguished by the human expert based on proximity to the nucleus, texture, and sometimes color information (some stroma is dark pink) [93]. As such, we choose to define three categories of features related to color, nuclear proximity, and texture for use in a standard classification framework. We will discuss each of these feature categories in turn, followed by the feature selection and classification aspects.

#### Proximity to Nuclei

Proximity to nuclei is one of the most telling characteristics of cytoplasm besides its general textural difference from stroma. The concept of “proximity” is more of a high-level architectural concept, but we seek to model this in a simple geometric fashion. While there are some heuristics about the extent of cytoplasm with respect to nuclear size, we will simply compute the Euclidean distance to the nearest nucleus, and allow the classification method to determine an appropriate threshold based on our training data. Thus, for each cytoplasm/stroma pixel, we have a single distance feature to the nearest nucleus perimeter:

$$d_{nn} = \min_{N_p} |N_p - CS_i| \quad (3.3)$$

where  $N_p$  is the set of nuclear perimeter pixels,  $CS_i$  is the current cytoplasm/stroma pixel, and  $|\cdot|$  is the 2-D Euclidean distance metric.

#### Gabor Texture Features

For texture features, we use a vector consisting of twelve values per pixel, where for each orientation and scale both the mean and standard deviation of the filtered output is computed. Additionally, the average over the mean values and the standard deviation

values are computed, resulting in a length-14 vector:

$$\mathbf{g} = [m_{11}, \sigma_{11}, \dots, m_{SK}, \sigma_{SK}, \bar{m}, \bar{\sigma}]^\top \quad (3.4)$$

## Spectral Features

While the ccd imagery was used for the k-means classification of the combined cytoplasm and stroma class, the multispectral information for this class can be substituted. While there is significant spectral overlap between the cytoplasm and stroma classes (refer again to Figure 3.1), the use of spectral information, in combination with our other feature categories, may improve the discrimination between cytoplasm and stroma. Thus we have a length-32 feature vector for spectral information, where the first 29 indices correspond to the 29 bands of the multispectral imagery, and the following three are the (R, G, B) values of the ccd imagery:

$$\boldsymbol{\lambda} = [\lambda_{420}, \lambda_{430}, \dots, \lambda_{700}, \lambda_R, \lambda_G, \lambda_B]^\top \quad (3.5)$$

Combining the proximity, texture, and spectral features, we have a length-47 feature vector,

$$\mathbf{f} = [d_{nn} \ \mathbf{g}^\top \ \boldsymbol{\lambda}^\top]^\top \quad (3.6)$$

for each pixel in the combined cytoplasm and stroma class.

## Feature Selection Methods

Feature selection is a means to select the relevant and important features from a large set of features, many of which may be redundant, irrelevant, or not particularly useful (and possibly detrimental) to the classification performance. This is an increasingly important area of research now that automated quantitative image analysis techniques are becoming more mainstream, as it is common practice to extract as many features as possible for future analysis. While humans have innate abilities to process and understand imagery, they do not tend to excel at explaining *how* they reach their decisions. As such, large feature sets are generated in the hopes that some subset of features incorporates the information the human expert is using for analysis.

Well-known feature selection methods include the sequential search methods, namely sequential forward selection (SFS) [102] and sequential backward selection (SBS) [103]. SFS works by sequentially adding the feature that most improves the classification performance; similarly, SBS begins with the entire feature set and sequentially removes the feature that most improves the classification performance.<sup>2</sup> Both SFS and SBS suffer

---

<sup>2</sup>It may seem counter-intuitive that removal of a feature could improve the classification performance. As previously mentioned, however, not all features are similarly useful for classification. Removal of

from the “nesting effect” whereby features that are selected (SFS) or discarded (SBS) cannot be revisited in a later step and are thus suboptimal [104]. Pudil et al. [104] proposed the use of floating search methods, sequential floating forward search (SFFS) and sequential floating backward search (SFBS), in which previously selected/discard features can be re-evaluated at later steps. While these methods still cannot guarantee optimality of the selected feature subset, they have been shown to perform very well compared to other feature selection methods and are, furthermore, much more computationally efficient [104].

More recent feature selection research has focused on such methods as boosting [105–107] and grafting [108, 109]. SFFS, however, has been shown to have comparable performance to grafting for several problem sets [110], and has been shown to outperform many other feature selection methods [111]. SBFS has similar performance to SFFS in many cases, but tends to suffer from higher computational complexity in the initial steps due to the larger feature set with which it begins.

## Results

We choose SFFS as our feature selection method, Fisher linear discriminant analysis (FLDA) as the classifier, and use the PRTools package from Delft University [112], which includes a wide variety of feature selection and classification methods. The choice of the FLDA classifier is motivated by the good performance of this classifier for the nuclei classification task of Chapter 2. Additionally, preliminary tests with a quadratic classifier yielded poorer results for this cytoplasm/stroma discrimination. The total training set contains 190,140 samples (139,047 stroma and 51,093 cytoplasm) and the test set contains 185,929 samples (131,958 stroma and 53,971 cytoplasm). Since these datasets are very large and would be computationally expensive to process, a random subset of 1000 samples is selected for the training set. Since the training data are unbalanced, 500 samples are selected from each stroma and cytoplasm examples for each set. Using the unbalanced data tends to result in algorithms that classify the majority of the image as stroma, thus achieving a good score for the majority of the samples. By balancing the training data, we hope to achieve a better performance on both cytoplasm and stroma. To test the performance of the classification algorithm, we test over both the entire training and test sets.

The performance of the SFFS feature selection is shown in Table 3.4. In addition to feature selection over the 47-D feature set, SFFS is also applied over smaller feature subsets: the 35-D subset excluding the orientation specific Gabor features and the 6-D subset excluding both the orientation specific Gabor features and the multispectral features.<sup>3</sup> The 47-D feature set can achieve a performance of 0.80 IS and 0.77 OS.

---

features that may be redundant, irrelevant, or contradictory will improve the classification performance.

<sup>3</sup>Thus the 6-D feature subset consists of the nuclear proximity feature, the mean and standard

**Table 3.4.** Detection rates, false alarm rates, and performance for Sequential Floating Forward Selection (SFFS) cytoplasm/stroma discrimination, using the sparse ground truth markup of cytoplasm and stroma for compilation of results. Results are presented as in-sample (IS)/ out-of-sample (OS). Bold rates correspond to the best rates and total performance is the average of cytoplasm and stroma performance.

<b>Feature Set</b>	<b>Cytoplasm</b>		<b>Stroma</b>		<b>Total P</b>
	<b>DR</b>	<b>FAR</b>	<b>DR</b>	<b>FAR</b>	
47-D	0.73/0.64	0.14/0.10	0.86/0.90	0.27/0.36	0.80/0.77
35-D	0.61/0.53	0.09/0.07	0.91/0.93	0.39/0.47	0.76/0.73
6-D	0.11/0.10	<b>0.03/0.02</b>	<b>0.97/0.98</b>	0.89/0.90	0.54/0.54

The use of the 35-D feature set was motivated by the desire for none of the features to be orientation specific, as this will degrade the generalizability of the algorithm. The use of the 6-D feature set was intended to explore the utility of the multispectral information for this cytoplasm/stroma discrimination task. We expected a slight increase in performance using the 35-D feature set as opposed to the 47-D subset, but instead we find a slight decrease in performance for this dataset. There is an even further decrease in performance for the 6-D feature set, indicating that there may be some subtle spectral information within the multispectral bands that is useful for the cytoplasm/stroma discrimination. Application to other H&E stained histopathology imagery may see an advantage to the orientation invariant 35-D and 6-D feature sets.

### 3.4.3 Genie Pro Feature-Based Discrimination

We also use Genie Pro for discrimination of cytoplasm and stroma. Genie Pro computes features using a pool of predefined image operators and the selection of relevant features is based on evolutionary computation. Thus the selection of features is “evolved” based on the performance on training data. For more information on Genie Pro refer to Appendix E. For the definition of ground truth, the nuclei and background classes are defined as an unknown and the class that Genie Pro assigns to these regions is considered irrelevant. Rather, it is the discrimination between cytoplasm and stroma that is of interest here. Genie Pro was allowed to train for 100,000 iterations (approximately a day on an Intel Core Duo 2.33 GHz tablet PC with 2 GB of RAM), and the performance was analyzed. This high number of iterations allows Genie Pro ample time to refine its algorithm to achieve the best possible classification performance. Results are shown in the first row of Table 3.5. Here we see the best performance so far, with 0.89 IS and 0.86 OS.

---

deviation of the average Gabor response, and the three ccd RGB channels.

**Table 3.5.** Genie Pro cytoplasm-stroma results, presented as in-sample (IS)/out-of-sample (OS). Bold rates correspond to the best rates. Original refers to the original training data; masked refers to the original training data applied to a masked image, where the combined cytoplasm/stroma class is masked from the background and nuclei; and ccd refers to the training data applied to the ccd RGB image.

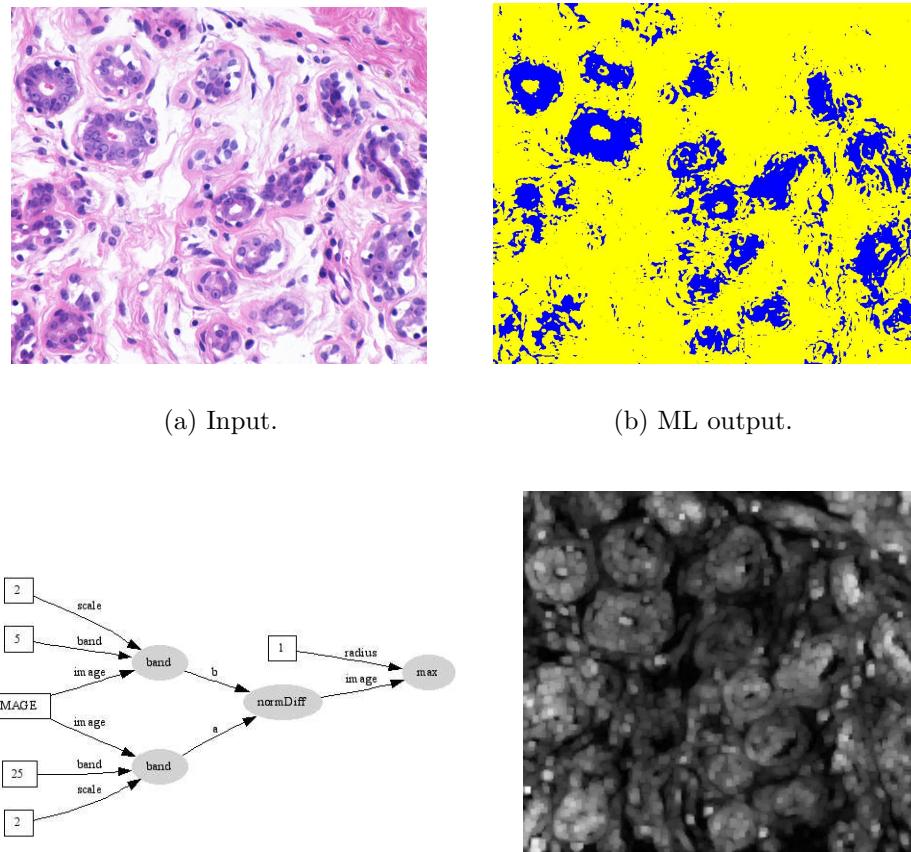
Training	Cytoplasm		Stroma		Total P
	DR	FAR	DR	FAR	
Original	<b>0.97/0.92</b>	0.23/0.21	0.83/0.77	<b>0.01/0.06</b>	<b>0.89/0.86</b>
Masked	0.77/0.59	0.31/0.26	0.83/0.82	0.11/0.15	0.80/0.75
ccd	0.78/0.61	<b>0.15/0.22</b>	<b>0.91/0.84</b>	0.12/0.16	0.85/0.77

We also applied this 2-class markup (incorporating only cytoplasm and stroma as described above) to a masked image; this masked image contains the combined cytoplasm/stroma class as output from the k-means rule-based algorithm of Section 3.3.1. The use of this training data was motivated by the hypothesis that providing Genie Pro with an unambiguous unknown class (i.e., an obviously unique background to the cytoplasm/stroma class that we are trying to discriminate) would provide easier means to focus on the cytoplasm/stroma discrimination problem at hand. These results are shown in the second row of Table 3.5, where we see poorer performance.

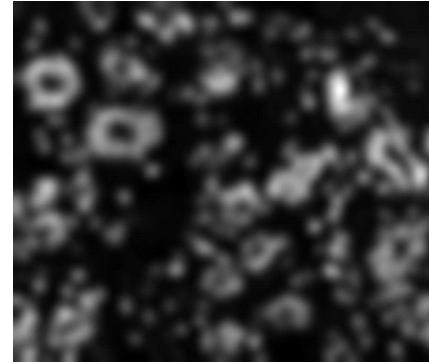
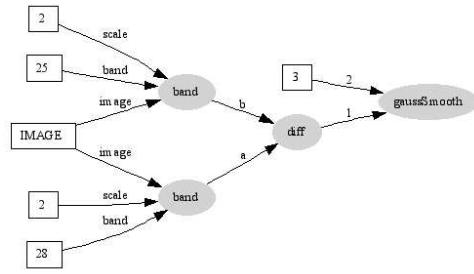
One advantage of Genie Pro is the ability to visualize the resulting algorithm. This can allow insight into the features that Genie Pro uses in the classification process. The algorithm resulting from the original training data is shown in Figure 3.4. Within Figure 3.4, the various attributes correspond to the feature planes that Genie Pro has extracted from the imagery. Both the graphical representation of the algorithm for feature extraction as well as the output for one of our dataset images are shown. The final output is determined from the attribute planes using a ML classification. In actuality, there are two steps in the feature extraction in Genie Pro: the first is the spectral and textural feature extraction which is shown in Figure 3.4 and the second is a morphological post-processing designed to mitigate clutter in the resulting ML output.

We note in Figure 3.4 that three of the five Genie Pro attributes are using the difference of spectrally close bands; in two cases two green bands (attributes 3 and 4), and in one case two red bands (attribute 1). This may indicate some extra information contained in the multispectral data useful for the discrimination of cytoplasm and stroma. Attribute 0 uses two spectrally distinct bands and attribute 2 uses a single band.

Genie Pro was run on ccd imagery using the same image markup to look at the attributes extracted when there are not spectrally similar bands available. Results for this training run are shown in the last row of Table 3.5, and the attributes in Figure 3.5. Table 3.5 shows that the ccd classification does not perform as well as the multispectral, but better than the masked multispectral.

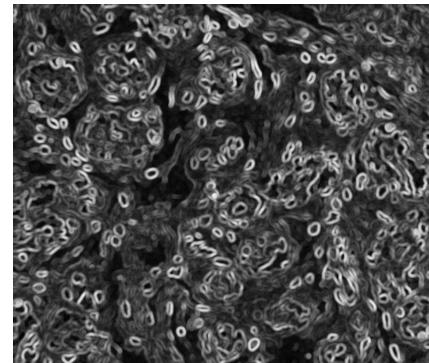
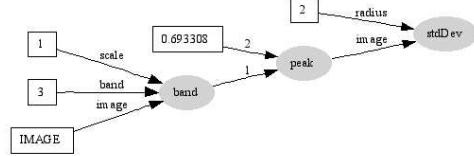


**Figure 3.4.** Genie Pro algorithm using multispectral imagery as input. (continued on next 2 pages)



(e) Attribute 1. Absolute difference of two red bands (700 and 670 nm), i.e.,  $|a - b|$ , for which the pixel values of nuclei vary more than for cytoplasm and stroma; this results in an image of cell nuclei, which is subsequently smoothed with a Gaussian filter.

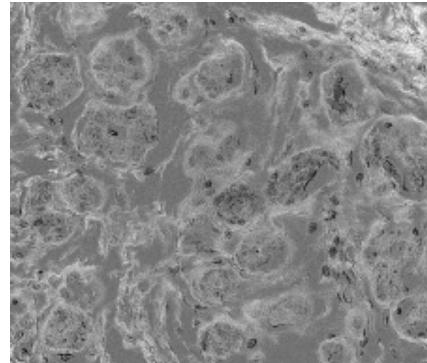
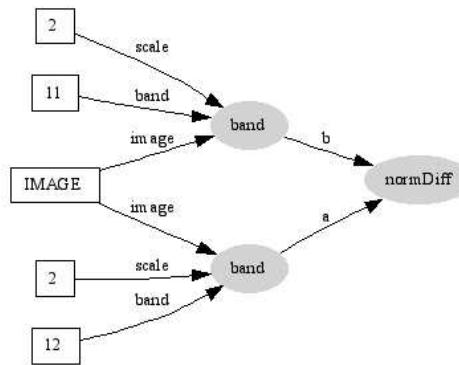
(f) Output 1. Gaussian smoothed nuclei.



(g) Attribute 2. Blue band at 470 nm is clipped at max and min values and then Gaussian smoothed (“peak”), and the standard deviation of a disc-shaped neighborhood is computed for each pixel.

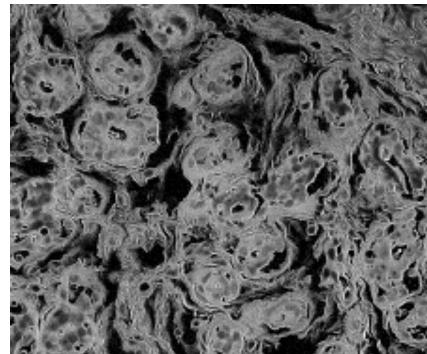
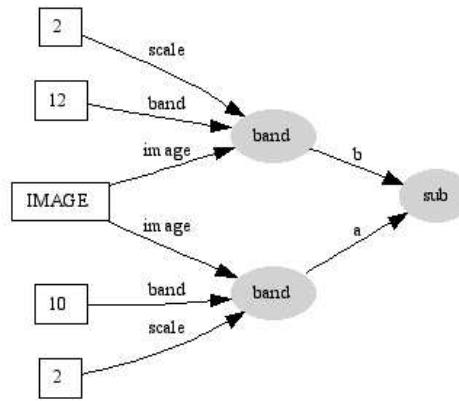
(h) Output 2. Edges of tissue and nuclei regions.

**Figure 3.4.** (Cont.) Genie Pro algorithm, multispectral imagery. (continued on next page)



(i) Attribute 3. The normalized difference ( $\frac{a-b}{a+b}$ ) of two spectrally adjacent green bands  $a$  (540 nm) and  $b$  (530 nm).

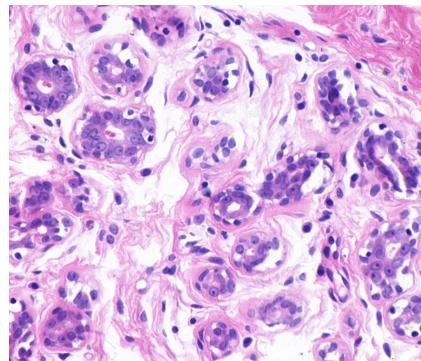
(j) Output 3. Positive values appear as lighter gray, negative values as darker gray. Appears to highlight stroma as lighter colored than any other histologic class.



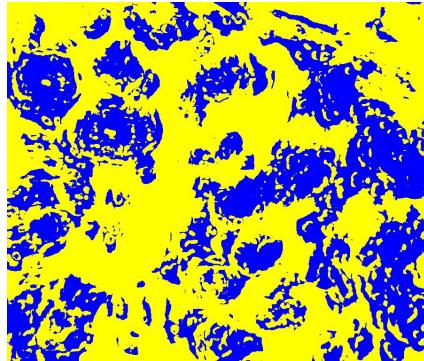
(k) Attribute 4. The subtraction of green bands  $a$  (520 nm) and  $b$  (540 nm), where the output is zero if  $a < b$ .

(l) Output 4. Highlights areas of tissue by the subtraction of a more spectrally transparent band from one that is less so.

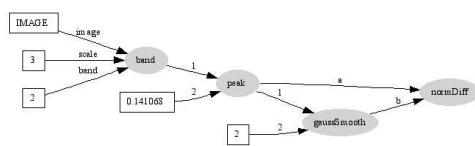
**Figure 3.4.** (Cont.) Genie Pro algorithm, multispectral imagery.



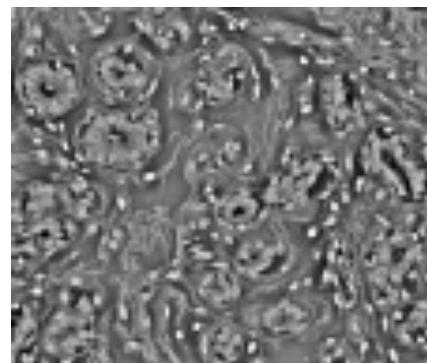
(a) Input.



(b) ML output.

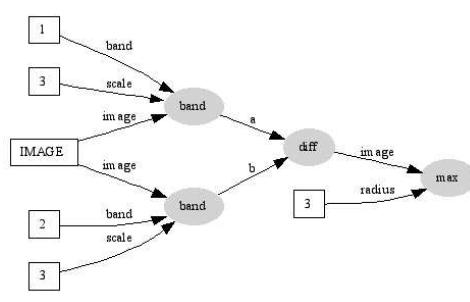


(c) Attribute 0. Computes the normalized difference ( $\frac{a-b}{a+b}$ ) of the clipped and Gaussian smoothed (“peak”) version of the blue channel ( $b$ ) with a further Gaussian smoothing ( $a$ ).

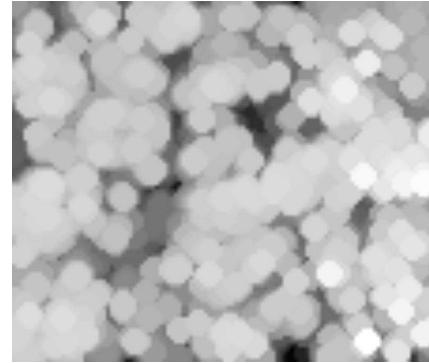


(d) Output 0. Appears to enhance tissue regions, with some nuclei particularly prominent.

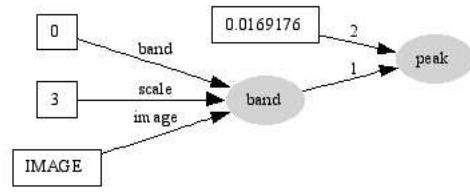
**Figure 3.5.** Genie Pro algorithm using ccd imagery as input. (continued on next 2 pages)



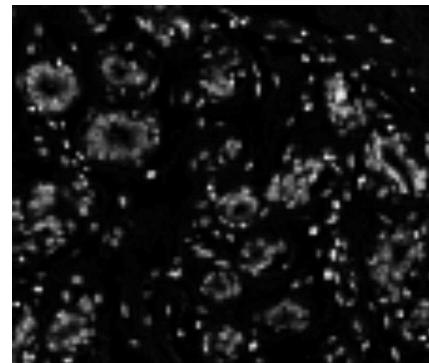
(e) Attribute 1. Takes the absolute difference of the green and blue channels and dilates the output.



(f) Output 1. The absolute difference highlights areas of tissue, with nuclei appearing brightest. The morphological dilation expands these areas, again providing the brightest values at and around nuclei.

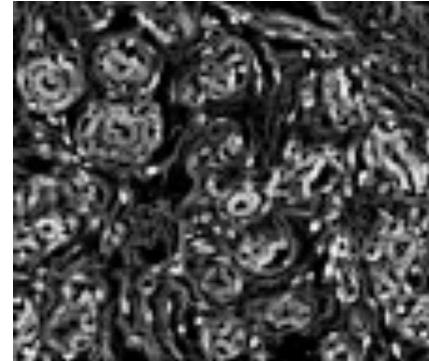
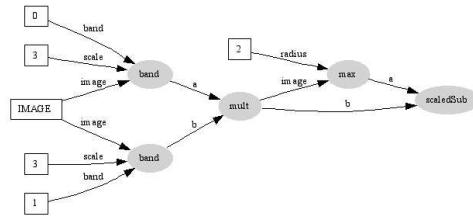


(g) Attribute 2. Clips and Gaussian smooths the red channel, enhancing nuclei.



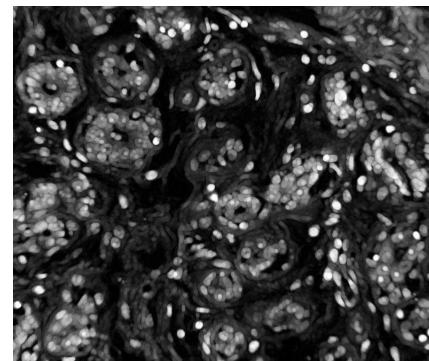
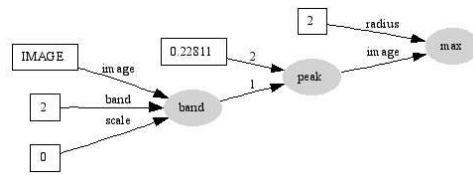
(h) Output 2. Nuclei.

**Figure 3.5.** (Cont.) Genie Pro algorithm, ccd imagery. (continued on next page)



(i) Attribute 3. Multiplies the red and green channels, followed by the scaled subtraction (scaled to be greater than 0) of the multiplication and a dilated version of the multiplication.

(j) Output 3. The multiplication of the red channel (with dark nuclei) and the green channel (with dark tissue) results in an image with dark nuclei. The subtraction of the dilated multiplication image from the multiplication image enhances edges of these dark regions. Thus, we have edges of nuclei (most prominent) and edges of other tissue.



(k) Attribute 4. Clips and Gaussian smooths the blue channel, followed by a morphological dilation.

(l) Output 4. Dilated nuclei.

**Figure 3.5. (Cont.)** Genie Pro algorithm, ccd imagery.

**Table 3.6.** Genie Pro single attribute plane results for the multispectral solution, presented as in-sample (IS)/out-of-sample (OS). The results from the original solution using all five attribute planes is included here for easy reference.

Attribute	Cytoplasm		Stroma		Total P
	DR	FAR	DR	FAR	
<b>0</b>	0.88/0.91	0.58/0.59	0.42/0.36	0.08/0.04	0.66/0.66
<b>1</b>	0.81/0.80	<b>0.41/0.32</b>	<b>0.69/0.72</b>	0.17/0.08	<b>0.78/0.73</b>
<b>2</b>	0.93/0.92	0.67/0.67	0.19/0.16	0.20/0.21	0.55/0.56
<b>3</b>	<b>1.00/1.00</b>	0.70/0.70	0.05/0.04	<b>0.00/0.00</b>	0.59/0.59
<b>4</b>	0.83/0.85	0.64/0.65	0.34/0.30	0.22/0.21	0.57/0.58
<b>All</b>	0.97/0.92	0.23/0.21	0.83/0.77	0.01/0.06	0.89/0.86

Most of the ccd attributes seem to be related to the nuclei, with the possible exception of attribute 3 which appears to be enhancing edges of tissue. Multispectral attribute 1 appears to be related to nuclei as well. Necessarily, all the masked attributes are some derived feature of solely the combined cytoplasm/stroma class. We hypothesize that Genie Pro is using the nuclei information to help discriminate cytoplasm from stroma.

To test the hypothesis that Genie Pro is using nuclei information to help discriminate cytoplasm from stroma, the multispectral and ccd solutions were modified to provide access to only one of the attribute planes for classification. While it is well established that a combination of features can be more discriminating than single features, this will give some idea of the relative usefulness of the individual attribute planes.

The classification performance for each attribute plane for the multispectral Genie Pro solution is shown in Table 3.6 and for the ccd Genie Pro solution in Table 3.7. The best single attribute for the multispectral solution is attribute 1, which appears to be a smoothed version of the nuclei. Similarly, in the ccd solution, attributes 0 and 1 are the best performing single features and both appear to be related to nuclei; the worst performing attribute is 3, which does not have as direct a relation to nuclei.

For a complete analysis of Genie Pro's ability to discriminate between cytoplasm and stroma, we would need to develop several independent solutions for each of the training cases (multispectral, ccd, and masked). We have shown, however, a proof of concept that Genie Pro is making use of nuclei information to help with the discrimination of cytoplasm and stroma. We have also shown that these attributes related to the nuclei are the best performing single feature planes within the solutions. Based on our analysis here, we will use the Genie Pro solutions for further analysis of cytoplasm and stroma regions.

**Table 3.7.** Genie Pro single attribute plane results for the ccd solution, presented as in-sample (IS)/out-of-sample (OS). The results from the original solution using all five attribute planes is included here for easy reference.

Attribute	Cytoplasm		Stroma		Total P
	DR	FAR	DR	FAR	
0	0.56/0.55	<b>0.38/0.40</b>	<b>0.87/0.84</b>	0.18/0.20	0.70/ <b>0.72</b>
1	<b>0.80/0.84</b>	0.49/0.45	0.61/0.64	<b>0.15/0.11</b>	<b>0.73</b> /0.69
2	0.40/0.30	0.41/0.41	0.79/0.86	0.28/0.26	0.62/0.63
3	0.39/0.35	0.68/0.66	0.71/0.71	0.29/0.31	0.52/0.53
4	0.57/0.63	0.54/0.58	0.62/0.61	0.22/0.19	0.62/0.61
All	0.97/0.92	0.23/0.21	0.83/0.77	0.01/0.06	0.89/0.86

## 3.5 Summary

We have presented in this chapter our methods for classification and discrimination of cytoplasm and stroma. We find our best overall average performance (OS) to be approximately 0.86. The discrimination between cytoplasm and stroma is a difficult problem, and one that has only recently begun to be addressed in the literature. Many of these methods assume a spectral difference between the two histologic classes which we did not find to be evident in our data (refer again to Figure 3.1).

The method of [95] and [96] uses some morphologic information in addition to spectral information for the classification of histologic entities, however, no quantitative results for the accuracy of such a segmentation is provided. From the example image segmentation provided in [96], we note similarities to our classification. First, we note a similar qualitative performance on nuclear classification. While the papers [95] and [96] discuss a region merging approach to separate clumped nuclei, the example image does not have any specification of the individual nuclear boundaries. Thus we will not be able to make any comparisons between our work in Chapter 4 and Tabesh et al. in [96]. The pixel-level nuclear classifications, however, seem to be on par with ours. Second, we note misclassification for both cytoplasm and stroma, generally confused with each other. This indicates the inherent difficulty in distinguishing the two classes. This may also indicate that it is less important to accurately distinguish the two classes as it is to characterize the features of the combined cytoplasm-stroma class as a whole. While it is ultimately of interest to characterize the cytoplasm and stroma classes individually, we may be able to provide valuable higher-level image information without such a distinction.

Even with the misclassified cytoplasm and stroma, Tabesh et al. demonstrate an average accuracy of 96.7% for cancer diagnosis and 81.0% for distinction of low and high Gleason grades [96]. This would seem to indicate the utility of such an imperfect segmentation for higher level analysis. Indeed, this is one of our interests in pursuing

our hierarchical image analysis method: we would like to investigate the extent to which higher-level image analysis can build upon imperfect lower-level analysis.

### 3.6 Future Directions

It would be of great use to utilize a dataset for which a more comprehensive ground truth could be developed for cytoplasm and stroma, e.g., serial sections classified with vibrational spectroscopy [29] or stained with other more specific stains. This would allow for a more comprehensive analysis of the performance of various cytoplasm and stroma classification methods.

While we have presented analysis of this problem using an intuitively satisfying feature set, there may well be other features better suited for this classification problem. Indeed, some of these well-suited features may stem from a higher-level conceptual analysis of the imagery, leading back to the need for feedback from higher levels of analysis. Given our intriguing results indicating the utility of nuclei for the discrimination of cytoplasm and stroma, it would be interesting to further investigate the potential use of other nuclear proximity features, e.g., a simple dilation of the nuclei pixels.

Additionally, further investigation of multiple independent Genie Pro algorithms for cytoplasm and stroma discrimination could be of use either directly for the discrimination process or for further insight into relevant features.



# Chapter 4

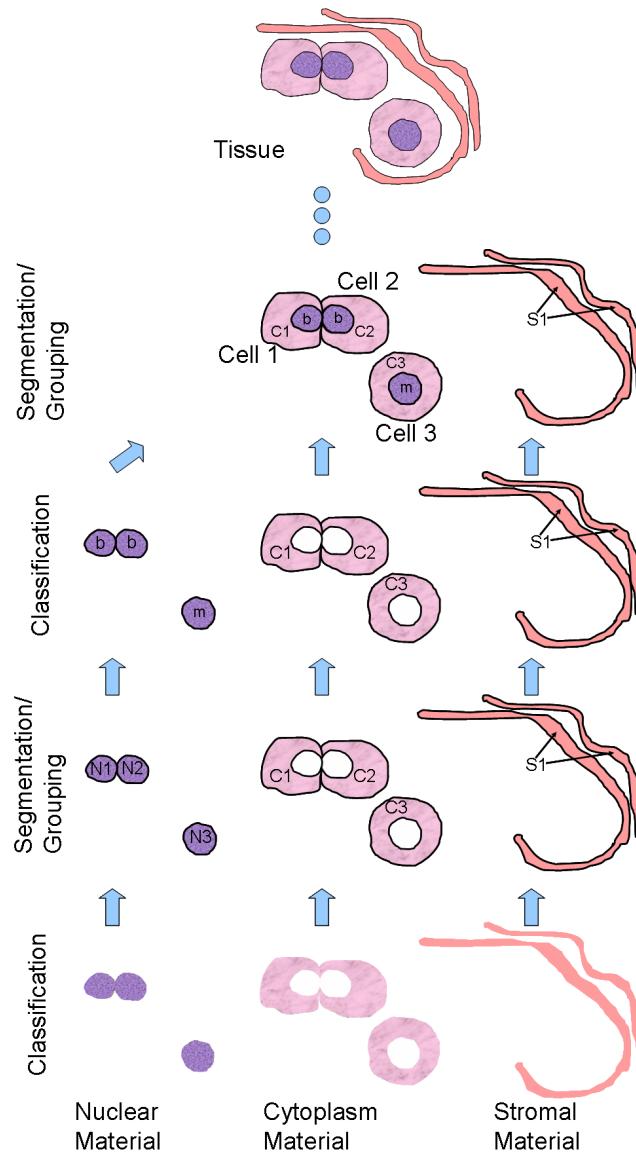
## Nuclei Segmentation: Methods and Evaluation Metrics

We present here our research on the first-level segmentation (the first segmentation row of Figure 4.1) for cell nuclei, cytoplasm, and stroma. This work takes the pixel-level classification output of Chapters 2 and 3 and seeks to delineate individual objects by assigning the pixels to specific histologic objects. We develop a new object-level metric for segmentation evaluation and show its correspondence to qualitative observations of general segmentation characteristics. This metric is also used to compare several methods for delineation of cell nuclei, and to illustrate the dependence of this higher-level segmentation on the accuracy of the underlying pixel-level classification. We continue with analysis of multispectral versus RGB imagery, and show that for nuclear segmentation, the RGB imagery performs better than the multispectral. Additionally, we briefly discuss our methods of segmentation for cytoplasm and stroma.

### 4.1 Motivation

We seek domain-specific first-level segmentations (delineation of objects) for our histopathology imagery to serve as a foundation for our hierarchical object-level analysis. The segmentation of cell nuclei on an object level is a very difficult problem. While there is much literature addressing this problem, we seek a solution that is feasible for our H&E stained imagery and that avoids too many underlying assumptions about the nuclear characteristics (e.g., parameterization as an ellipse). We also present the development of an object-level segmentation metric applicable to our nuclear segmentations, and possibly other application areas.

As noted in Chapter 3, the discrimination of cytoplasm and stroma is not commonly addressed in the literature. Also not addressed is the segmentation of cytoplasm and



**Figure 4.1.** Repeated Figure 1.2: Illustrative example of the two-step iterative classification-segmentation process. At the first level (bottom two rows), biologic materials are first *classified* on a pixel-by-pixel basis and then *segmented* into their constituent objects, based on user feedback. Similarly, the second level is a *classification* distinguishing (e.g.) benign (“b”) and malignant (“m”) nuclei, and *segmentation* of “cell” objects. This process can continue until the user is satisfied with the level of abstraction; in this example, the process ends at the concept of “tissue” at the topmost level. Note also that the level of abstraction is controlled by user input at each level; this can be seen in the maintenance of the stromal elements as a single object.

stroma into discrete objects. Some authors (e.g., Gil et al. [113]) mention the possible use of segmentation of further histologic entities such as cytoplasm, but consider the solution to be infeasible in the near future. In this chapter, we discuss the methods we use for the segmentation of cytoplasm and stroma, which will be used to extract object level features for both histologic entities.

The work presented here begins the development of object-level analysis of our imagery, i.e., the first segmentation of the framework of Figure 4.1.

## 4.2 Related Work

### 4.2.1 Object-Level Analysis

As mentioned in Section 1.4, an image object is generally defined as a connected group of pixels satisfying some similarity criterion. In recent object-based analysis research, an object has been defined in terms of contour [48, 49]; homogeneous color [50, 51] or texture [51]; as an anomaly (in some sense) compared to the rest of the image [52]; location, size, and shape [53]; and topological and connectivity relations [54, 55]. Some relevant research on segmentations of biological entities can be found in References [17, 31–33, 37, 46, 56–60], and an overview in Reference [61].

### 4.2.2 Segmentation Evaluation

The subject of objective and quantitative evaluation of segmentation performance has received less attention than has the development of various segmentation algorithms themselves. The evaluation of segmentation algorithms is often presented as a qualitative overview of several representative segmented images. Many researchers in the fields of computer vision and image analysis have noted the lack of objective and quantitative metrics for segmentation quality, most often attributed to the lack of a well-defined definition of image segmentation [114–122].

In Zhang’s commonly cited survey of segmentation evaluation methods [121], he discriminates between three main categories of segmentation evaluation metrics: analytical metrics, empirical goodness metrics, and empirical discrepancy metrics. Analytical metrics look at the segmentation algorithm formulations and analyze their properties. Empirical goodness metrics define a “good” segmentation in terms of properties that humans intuitively associate with a good segmentation, e.g., smoothness of boundaries or homogeneity of color. Empirical discrepancy metrics rely on the use of a ground truth image and seek to provide a measure of the difference between the segmentation output and the ground truth segmentation. We focus on empirical discrepancy methods, as we wish to focus on methods that can be easily adopted into a machine learning framework.

We avoid metrics of segmentation performance that rely on point-to-point or region-to-region correspondences (e.g., [115, 123–125]) as this is a sufficiently difficult problem that it is an area of research in its own right. While there is much research in the use of multiple ground truths, often manually defined by multiple human experts, we stick to the case of one ground truth assumed to be the gold standard. Research using one reference ground truth include:

- Fenster and Chiu [126] and Udupa et al. [117] call for the simultaneous use of accuracy, precision, and efficiency to rate segmentation algorithms. They propose that the use of only one of these metrics will fail to completely characterize the segmentation performance.
- Yasnoff et al. [120] proposes the use of two measures for segmentation accuracy: percentage of pixels misclassified and the pixel distance error. Pixel distance error is defined as the Euclidean distance from a misclassified pixel to the closest pixel of the misclassified class. This is aggregated for all misclassified pixels.
- Everingham et al. [127] propose the use of Pareto front, originally used in political economics, as a method of aggregating multiple fitness metrics into one fitness function. They develop a method for ROC-like analysis of these aggregate fitness functions which provides more quantification of the range of segmentation performance under different circumstances.
- Hoover et al. [118] propose the use of five possible classification scenarios for the quantification of segmentation accuracy of laser range images. The five scenarios are: 1) correct detection, 2) oversegmentation, 3) undersegmentation, 4) missed, and 5) noise.
- Zhang and Gerbrands [122] propose a very different approach to quantifying segmentation accuracy. Rather than focusing on pixel-level quantities such as percentage of pixels misclassified, they propose the use of object-level metrics as a measure of the accuracy. The particular object-level metrics used for quantification of the segmentation accuracy are the measurements that are ultimately desired for a further image analysis beyond segmentation; this method has been named Ultimate Measurement Accuracy (UMA).

### 4.2.3 Nuclear Segmentation

Many publications on nuclear segmentation provide only a few examples of the segmentation results, and do not present any quantitative results gathered from a dataset. Some papers address the quantitative analysis of the segmentation by having the expert pathologist review the automated segmentation and quantifying the number and/or percent of “correctly” segmented nuclei (e.g., see [17, 128]). While this begins to address the

issue of quantitative segmentation evaluation, it does not provide an automated means to compute the performance metrics.

Additionally, many nuclear segmentation algorithms presented in the literature make use of immunostaining for more specific location of cellular structures, or the Feulgen stain which is specific to DNA, or even the physical extraction of nuclei prior to image acquisition (see [10]). The use of cytology imagery is more often addressed since this provides a less complicated image structure with more instances of isolated cells and/or well-delineated cell clusters, e.g., [44, 46, 59, 129].

The use of standard H&E histology imagery for nuclear-based quantitative analysis is rarer and has met with mixed success:

- Demir et al. [35] forgo the delineation of individual nuclei and use the cell clusters as nodes for a graph-based analysis of H&E stained brain biopsies.
- Glotsos et al. [128] cluster nuclear pixels based on texture features using an SVM and use an active contour model based on gradient vector flow to delineate the nuclei for H&E stained astrocytomas and bladder cancer. Accuracies are reported to be  $\sim 95\%$  with  $\sim 2\text{-}3\%$  oversegmentation and undersegmentation. While these results are very promising, their imagery appears to have better separated nuclei than ours.
- Latson et al. [17] use a fuzzy c-means clustering of the hue band of H&E stained breast biopsy, followed by morphological cleanup and a marker-based watershed on the distance transform of the nuclear classification. Markers are defined as the regional minima of the distance transform. Results are 57.2%–71.6% correctly segmented nuclei, 4.5%–16.7% clumped nuclei, 22.5%–26.3% “badly segmented” nuclei, and 0.4%–1.4% missed nuclei.

#### 4.2.4 Segmentation of Cytoplasm and Stroma

The segmentation of cytoplasm and stroma is an area of research that does not seem to have been addressed much in the literature. No references were found regarding the segmentation of stroma into discrete components. Very few references were found regarding the segmentation of cytoplasm. In particular, Canzonieri et al. [130], Spillman et al. [21], and Zhao et al. [27] mention the use of the nuclear to cytoplasm area ratio as a discriminating feature of cancerous versus normal tissue. These references, however, do not consider the segmentation of cytoplasm into discrete components, but rather compute the ratio over the entire image or subimages.

Zahniser et al. [9] use an isodensity contour tracing algorithm to segment both the Feulgen stained nuclei (using a 621 nm red image) and the Orange II stained cytoplasm (using a 497 nm green image) in cervical cytology specimens. The information in the

cytoplasm segmentation is used to determine whether the image object is a single cell or a cell cluster for subsequent feature extraction, and also for computation of the nuclear to cytoplasmic area ratio.

Honda [131] shows that a Voronoi tessellation closely approximates cellular patterns in two dimensions, particularly for single sheets of epithelial cells.

Jones et al. [132] present a modification of the Voronoi segmentation, in which they provide a tradeoff between the true Voronoi segmentation and the edge locations in the gradient image. This method allows for actual cell boundaries to dominate when they are present, and allow the Voronoi approximation to dominate when there are no edge cues available. This method was applied to both synthetic images, and images of drosophila (fruit fly) cells fluorescently stained for DNA and actin (a cytoskeletal protein). 92% of the automatically determined boundaries were within 5 pixels of the manually segmented boundaries, for cells approximately 25 pixels in diameter.

We present here our research on the first-level segmentation (the second row of Figure 1.2) for cell nuclei, cytoplasm, and stroma. We first describe the object-level metric for segmentation accuracy that we have developed as well as the specification of object-level ground truth for cell nuclei in Section 4.3. We then present results for two categories of nuclear segmentation (Sections 4.4 and 4.5), including analysis of the segmentation accuracy and efficiency. We analyze the variability of these segmentation methods over our entire dataset and present the expected maximum segmentation performance given an ideal pixel-level classification (Section 4.6). The segmentation of cytoplasm and stroma is briefly discussed in Section 4.7, followed by a summary (Section 4.9) and future work (Section 4.8).

## 4.3 Segmentation Metric

The following metric<sup>1</sup> was defined with the segmentation of cell nuclei, i.e., roughly circular or elliptical objects, in mind. For the segmentation of cell nuclei, we wish to penalize not only the size of regions missed and extraneous regions, but also the shape of those same regions. Additionally included are terms to penalize over- and under-segmentation. We introduce the quadrant sum as a method of quantifying deviation in shape from the ground truth by comparing the mass across two orthogonal axes through the object's center of mass. While this section will focus on elliptical objects, use of the quadrant sum for arbitrarily shaped objects will be shown in Section 4.3.3. The work presented in this section along with preliminary results from the rest of the chapter was published by the author in [133].

---

<sup>1</sup>We are using the term metric in the general sense, not in the strict mathematical definition of a function which satisfies non-negativity, positive-definitiveness, and the triangle inequality.

### 4.3.1 Definition

We define our segmentation metric as:

$$P = \frac{1}{N_D} \sum_{i=1}^{N_D} \max \left( 0, \left[ 1 - \alpha_1 \frac{SR - 1}{\delta_{SR}} - \alpha_2 \frac{1}{1.75} \left( \frac{PM}{GT} + \frac{2QS_{PM}}{GT} \right) \right. \right. \\ \left. \left. - \alpha_3 \frac{1}{1.75} \left( \frac{EP}{GT} + \frac{2QS_{EP}}{GT} \right) \right] \right] \cdot \left( 1 - \alpha_4 \frac{N - N_D}{N} \right) - \alpha_5 \frac{ER}{N \cdot \delta_{ER}} \right) \quad (4.1)$$

where

$$0 \leq \alpha_i \leq 1, \quad i = 1, \dots, 5.$$

Taking each additive term in the equation above, we will define the variables. First we would like to point out that for the sake of clarity and brevity, Equation (4.1) does not include the necessary clipping functions to assure that each term is less than 1. We will discuss the need for these clipping functions and explicitly display them in the discussions of individual terms to follow.

In Equation (4.1),  $N$  is the number of ground truth nuclei defined in the user markup and  $N_D$  is the number of nuclei detected by the segmentation algorithm; thus the summation averages scores for individual nuclei. We penalize for each nucleus detected (a segmented region overlapping with a ground truth region):

#### 1. The number of segmented regions:

$$\text{term}_1 = \alpha_1 \min \left( 1, \frac{SR - 1}{\delta_{SR}} \right) \quad (4.2)$$

$SR$  is defined as the number of segmented regions overlapping the current ground truth nucleus, and  $\delta_{SR}$  as the upper limit for number of segmented regions. For a perfect segmentation there would be only one segmented region per ground truth region and  $\delta_{SR} = 1$  would be an intuitive value for evaluation of very good segmentations; we leave this as a parameter, however, to allow for comparison of poorer segmentations with more tendency to oversegment. The minimum function is used to clip this term to a maximum value of 1 for any number of segmented regions greater than  $\delta_{SR}$ . Overall, the weight  $\alpha_1$  can be thought of as the penalty for an oversegmented nucleus.

#### 2. The size and shape of the region of pixels missed:

$$\text{term}_2 = \alpha_2 \min \left( 1, \frac{1}{1.75} \cdot \left( \frac{PM}{GT} + \min \left( 1, \frac{2 \cdot QS_{PM}}{GT} \right) \right) \right) \quad (4.3)$$

$PM$  is defined as the number of pixels missed; that is, the pixels belonging to the ground truth markup of the nucleus, but missed by the segmentation algorithm.

$GT$  is the number of pixels in the ground truth markup which we use to normalize our size metric  $PM$ . Thus,  $\frac{PM}{GT}$  quantifies the size of the region of missed pixels. This is similar, though not identical, to the percentage of misclassified pixels used in [120] since we are separately dealing with the pixels missed and the extra pixels (i.e., two different types of misclassification).

We also look at the spatial distribution of the missed pixels, since we wish to penalize certain spatial distributions more than others. For example, a distribution of missed pixels in an annulus about the centroid of the nucleus will affect the shape other higher-level metrics far less than a distribution of missed pixels encompassing half of the nucleus. Note that this is a different approach than a simple pixel distance error as in [120] and tends towards an appreciation of the ultimate goal of this image analysis as in [122]. We take the “quadrant sum” of the pixels missed,  $QS_{PM}$  as follows:

$$QS_{PM} = \|r_1 + r_3 - r_2 - r_4\| + \|r_1 + r_2 - r_3 - r_4\| \quad (4.4)$$

where  $r_i$  are the number of pixels in the respective quadrants 1 – 4, and the quadrants are defined in a counter-clockwise direction. Looking at the polar coordinate representation of the  $PM$  region, the number of pixels in each of the four quadrants can be determined:

$$\begin{aligned} r_1 &= \sum \|e^{j\theta_{PM}}\|, \quad \text{for } 0 < \theta_{PM} < \frac{\pi}{2} \\ r_2 &= \sum \|e^{j\theta_{PM}}\|, \quad \text{for } \frac{\pi}{2} < \theta_{PM} < \pi \\ r_3 &= \sum \|e^{j\theta_{PM}}\|, \quad \text{for } -\frac{\pi}{2} < \theta_{PM} < -\pi \\ r_4 &= \sum \|e^{j\theta_{PM}}\|, \quad \text{for } 0 < \theta_{PM} < -\frac{\pi}{2} \end{aligned} \quad (4.5)$$

Thus,  $QS_{PM}$  is a measure of symmetry about the x- and y-axes of the region, where the origin is at the ground truth centroid.  $QS_{PM}$  is normalized by  $\frac{GT}{2}$ , since the maximum value this sum can (theoretically) take is half of the area. Due to the discrete nature of the regions, however, it is possible that  $QS_{PM}$  may slightly exceed  $\frac{GT}{2}$ ; to compensate for this, we take the minimum of 1 and  $\frac{2 \cdot QS_{PM}}{GT}$ . While this metric does normalize by the total size of the region, it is not duplicating the size metric  $\frac{PM}{GT}$ , since it is calculating here a measure of asymmetry according to the region size. While this is a simple and easy to compute metric, there is no reason why another shape metric could not be substituted, with appropriate attention to the inclusion of the size metric.

Overall,  $\alpha_2$  can be thought of as the penalty for regions of pixels missed, penalizing both size and shape. More details of the performance of this QS term is explained in Figure 4.2 for circular and elliptical regions, including the motivation for the normalization factor of 1.75. Additionally, the interested reader is referred to Ap-

pendix G where examples of actual  $PM$  regions are ranked according to term 2 for one of the dataset images.

### 3. The size and shape of the region of excess pixels:

$$\text{term}_3 = \alpha_3 \min \left( 1, \frac{1}{1.75} \cdot \left( \min \left( 1, \frac{EP}{GT} \right) + \min \left( 1, \frac{2 \cdot QS_{EP}}{GT} \right) \right) \right) \quad (4.6)$$

Similar to term 2,  $EP$  is defined as the number of excess pixels; that is, the pixels segmented as part of the nuclear region that do not correspond with the ground truth markup. Since it is possible in the case of a severely undersegmented nucleus that  $\frac{EP}{GT} > 1$ , we take the minimum of 1 or the ratio.

Analogously to term 2, the quadrant sum of the excess pixels,  $QS_{EP}$ , is calculated:

$$QS_{EP} = \|r_1 + r_3 - r_2 - r_4\| + \|r_1 + r_2 - r_3 - r_4\| \quad (4.7)$$

where, for this quantity, the polar representation of the  $EP$  region is used:

$$\begin{aligned} r_1 &= \sum \|e^{j\theta_{EP}}\|, \quad \text{for } 0 < \theta_{EP} < \frac{\pi}{2} \\ r_2 &= \sum \|e^{j\theta_{EP}}\|, \quad \text{for } \frac{\pi}{2} < \theta_{EP} < \pi \\ r_3 &= \sum \|e^{j\theta_{EP}}\|, \quad \text{for } -\frac{\pi}{2} < \theta_{EP} < -\pi \\ r_4 &= \sum \|e^{j\theta_{EP}}\|, \quad \text{for } 0 < \theta_{EP} < -\frac{\pi}{2} \end{aligned} \quad (4.8)$$

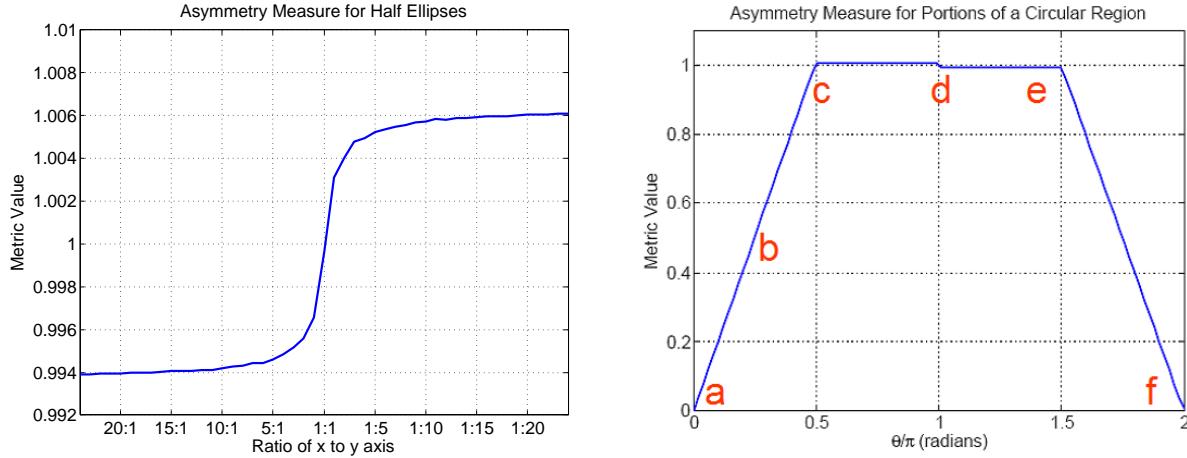
The same normalization factor  $\frac{GT}{2}$  is used for  $QS_{PM}$ . It is possible, however, that this term can be much larger than 1, in particular for an adjacent nucleus segmented as part of the current nuclear region; as such, we again use the minimum function, taking the minimum of 1 and  $\frac{2 \cdot QS_{EP}}{GT}$ . Again the sum of the two factors is normalized by 1.75.  $\alpha_3$  is thus the penalty for size and shape of excess pixel regions, and is related to the degree of undersegmentation of the nucleus. Appendix G contains examples of actual  $EP$  regions ranked according to term 3 for one of the dataset images.

Averaging these three terms provides a measure of the segmentation performance on all detected nuclei. We wish also, however, to weight this average by the general detection rate of nuclei. Thus, the average of the previous three terms is scaled by:

### 4. The fraction of nuclei detected:

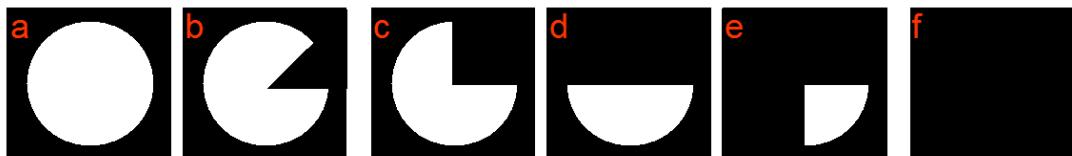
$$\text{term}_4 = 1 - \alpha_4 \frac{N - N_D}{N} \quad (4.9)$$

This term with  $\alpha_4 = 1$  would simply be the detection rate; the fraction of missed



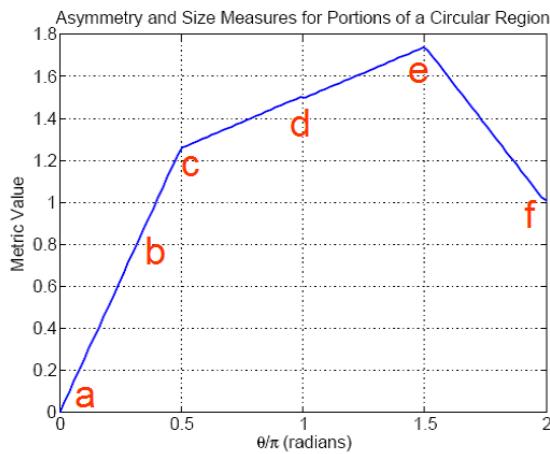
(a) Effect on the QS metric of ellipticity and orientation of missed pixels. The region missed is below the x-axis: ellipses plotted to the left of circular are missing half of their area along the major axis and to the right of circular, half their area along the minor axis. Note the possibility for the metric to be slightly larger than 1.

(b) Effect on the QS metric of the portion of a circular region of pixels missed. The maximum value for this metric occurs at (and around)  $\theta = \pi$ , when half of the region is missed. The metric tapers off to zero for small and large angles; this illustrates the need for a separate size metric, since this metric is scoring only the asymmetry.

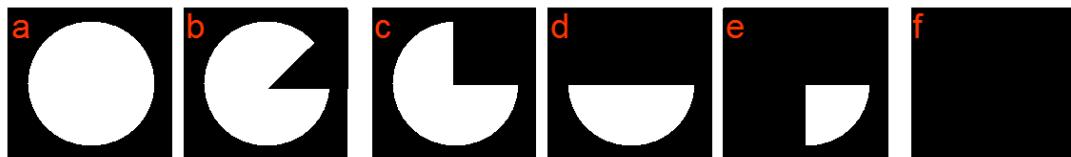


(c) Example wedges removed from the discrete circular region in (b).

**Figure 4.2.** Performance of the Quadrant Sum (QS) metric for pixels missed in discrete elliptical and circular regions. The QS metric in these plots has been normalized by  $\frac{GT}{2}$ , and the size metric by  $GT$ , where  $GT$  is the number of ground truth pixels. (continued on next page)



(d) Additive effect of the QS and size metrics. The combination of these two metrics yields the desired penalty. Note the maximum value of  $\sim 1.75$ .



(e) Example wedges removed from the discrete circular region in (d).

**Figure 4.2.** (Cont.) Performance of the Quadrant Sum (QS) metric for pixels missed in discrete elliptical and circular regions. The QS metric in these plots has been normalized by  $\frac{GT}{2}$ , and the size metric by  $GT$ , where  $GT$  is the number of ground truth pixels.

pixels  $\frac{N-N_D}{N}$  is scaled by weight  $\alpha_4$  and then subtracted from 1 to get a modified fraction of detected nuclei. In many cases it may be sufficient and desirable to set  $\alpha_4 = 1$ , but it is left as a parameter since in the segmentation of nuclei, the accuracy of the nuclei that are segmented may be of more importance than the actual detection rate. This harkens back to the theory of UMA [122], wherein it is the accuracy of further image analysis that determine the accuracy of the underlying segmentation.

Finally we wish to penalize over the whole region of ground truth:

##### 5. The number of extra segmented regions:

$$\text{term}_5 = \alpha_5 \min \left( 1, \frac{ER}{N \cdot \delta_{ER}} \right) \quad (4.10)$$

While terms 1-3 are defined for a single nucleus, and term 4 scales the average of terms 1-3, this term looks at the excess segmented regions that have no correspondence to a ground truth nucleus. For this term,  $ER$  is defined as the number of excess segmented regions and  $\delta_{ER}$  as the fraction of total ground truth nuclei that we will allow as excess regions; in general  $\delta_{ER} = 1$ .  $\alpha_5$  is, therefore, the penalty for excess segmented regions, similar to the concept of noise in [118].

Overall, the choice of  $\alpha_i$  reflects a weighting of the relative importance of the various penalties. Similarly, the choice of  $\delta_{SR}$  and  $\delta_{ER}$  reflects a choice in the latitude given to certain errors in segmentation. A reasonable choice for default parameters would be  $\alpha = [0.5 \ 0.5 \ 0.5 \ 1 \ 0.5]$ ,  $\delta_{SR} = 1$ , and  $\delta_{ER} = 1$ , reflecting an equal penalty for under- and over-segmentation errors ( $\alpha_1$ ,  $\alpha_2$ , and  $\alpha_3$ ), a direct weighting by the detection rate ( $\alpha_4$ ), equal importance given to the correct detection and segmentation of cell nuclei and the avoidance of erroneously detected and segmented nuclei ( $\alpha_5$ ), one segmented region allowed per nucleus ( $\delta_{SR}$ ), and weighting of the erroneously segmented regions proportional to the total number of cell nuclei ( $\delta_{ER}$ ). It is important to note, however, that while the choice of these parameters will effect the absolute values of the metric terms, a direct comparison of segmentation performance for different algorithms may be achieved with any reasonable parameter choice.

#### 4.3.2 Metric Variation versus Segmentation Quality

We apply the segmentation metric (Equation (4.1)) to the watershed transform of the complemented Euclidean distance transform (WSCDT) of a thresholded red channel for an example ccd image. The threshold is varied over the entire range of gray scale values it can assume,  $[0, 255]$ , and all pixels less than the threshold value are retained. The use of the red channel is motivated by the high contrast for nuclei present in this channel.

While the WSCDT is a common approach to watershed segmentation, for the sake of clarity, we will describe in more detail the individual steps.

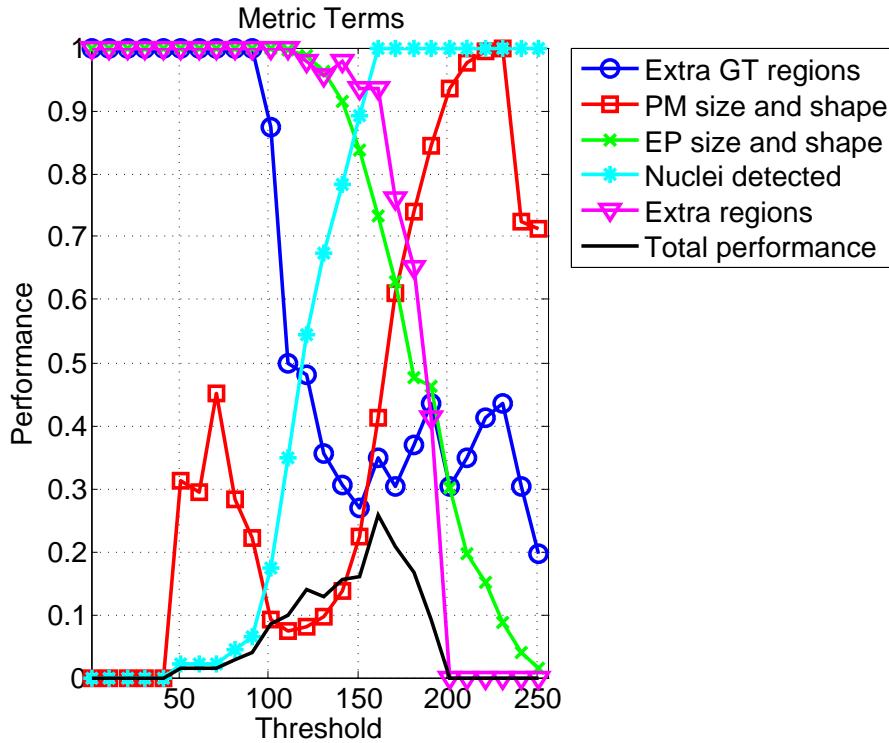
1. Compute the negative of the Euclidean distance transform on the complemented binary image, setting the distance of all background pixels in the binary image to a depth of  $-\infty$ .
2. Compute the watershed transform on the resulting distance transform.

Thus by varying the threshold and computing the segmentation metric (Equation (4.1)) of the WSCDT segmentation, we will get a sense of the expected variation in our metric for a range of segmentation possibilities. These possibilities include the two extremes whereby either all or none of the pixels have been classified as nuclei. The performance of the individual metric terms as well as the overall performance is displayed in Figure 4.3. It is important to note that in this figure the *performance* of the individual terms are plotted rather than the terms themselves; thus the subtraction of each term from a value of 1 is plotted.

Figure 4.3 illustrates that the performance is zero for both extremes of the threshold classification. Observation of individual terms shows expected trends, namely that:

- Term 1 (extra GT regions) decreases in performance as the threshold increases. As the regions thresholded as nuclei become larger with more complicated boundaries, the distance transform has multiple minima per connected component in the binary thresholded image. This results in a tendency to oversegment.
- Term 2 (pixels missed) increases in performance as more pixels are attributed to nuclei. The dip in performance at high thresholds is due to an assumption that the largest watershed region is the background; this is generally valid for reasonable classifications, but becomes invalid as nearly the entire image is classified as foreground.
- Term 3 (extra pixels) decreases in performance as nuclei tend to merge in the binary thresholded image.
- Term 4 (nuclei detected) increases in performance as more pixels are attributed to nuclei.
- Term 5 (extra regions) decreases in performance as more extraneous regions are thresholded as nuclei. The performance of this term returns to 1 for a threshold of 256, since there are no longer any extraneous regions; this is not apparent in Figure 4.3 since we have downsampled the plot for less clutter.

We have presented here a general segmentation metric computed on an object level. This metric uses simple quantities that are easy to compute using the segmentation



**Figure 4.3.** Metric variation versus segmentation quality for an example ccd image. The red channel was thresholded, retaining all pixels less than the threshold, and was then segmented with the Watershed on the Complemented Distance Transform (WSCDT) method. It should be noted that all terms plotted here are performance, i.e., one minus the penalty, where the penalties are the terms previously discussed in relation to the segmentation metric. The terms are denoted by a brief description in the legend, but they are also plotted in numerical order, i.e., blue circles are term 1, red squares are term 2, and so forth. We have downsampled the plot for less clutter.

and ground truth images, namely the regions of pixels segmented that do not belong to a ground truth region, and the regions of pixels not segmented that do belong to a ground truth region. We have also shown the variation in this metric for a variety of segmentations using a simple watershed-based segmentation (WSCDT). We will first make a few remarks about the ground truth markup process, as well as the application of this metric to non-elliptically shaped objects and a comparison to other metrics. We will then present segmentation results on cell nuclei using our newly defined metric.

### 4.3.3 Application to Non-Elliptically Shaped Objects

We would like to briefly discuss the applicability of the QS metric to non-elliptically shaped objects; in this section the focus will be on the use of the PM QS metric, but the arguments are identical for the EP case. The use of the centroid of the ground truth object is what allows this metric to work for irregularly shaped objects. For a planar object with uniform density, the mass (number of pixels in this case) will be equal across any arbitrary line through the center of mass (equivalent to the centroid in the uniform density case). By defining orthogonal axes through the centroid, we can eliminate the chance of the arbitrary line corresponding to a reflectional symmetry of the region of pixels missed. An example of the application of the PM QS metric to a hand silhouette is shown in Figure 4.4, where we see the utility of the QS metric in quantifying the deviation in shape of the hand, rather than simply the size of the deviation in segmentation.

### 4.3.4 Comparison to Other Segmentation Metrics

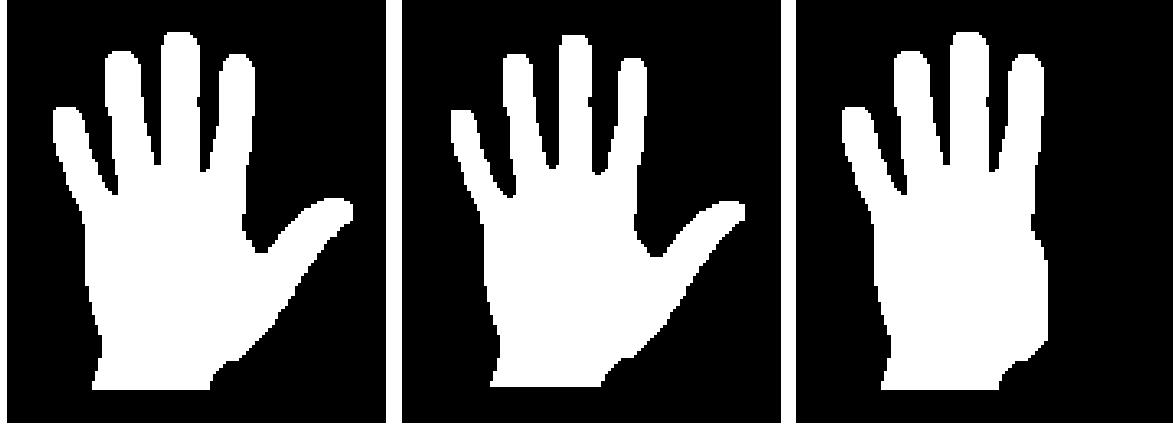
The Vinet measure (as coined by [124]) was originally developed for region-based stereo matching [125]. Its main purpose is to establish a region-to-region correspondence between images by looking for the maximum region overlap. A dissimilarity measure is then defined using the cardinality of the overlap. Thus, the Vinet measure can be thought of as a measure of the pixels missed and extra pixels, i.e., the pixels that are not contained in the intersection of the two objects. Figure 4.4 shows the size metric for the hand silhouette. This size metric, however, does not capture any difference in shape of the objects, as does the QS metric.

The Hausdorff distance defines the deviation between two sets of points as the largest distance between any two corresponding points [115, 134]. Defined formally, given two sets of points  $A = \{a_1, \dots, a_m\}$  and  $B = \{b_1, \dots, b_n\}$ , the Hausdorff distance is:

$$H(A, B) = \max(h(A, B), h(B, A)) \quad (4.11)$$

where

$$h(A, B) = \max_{a \in A} \min_{b \in B} \|a - b\| \quad (4.12)$$



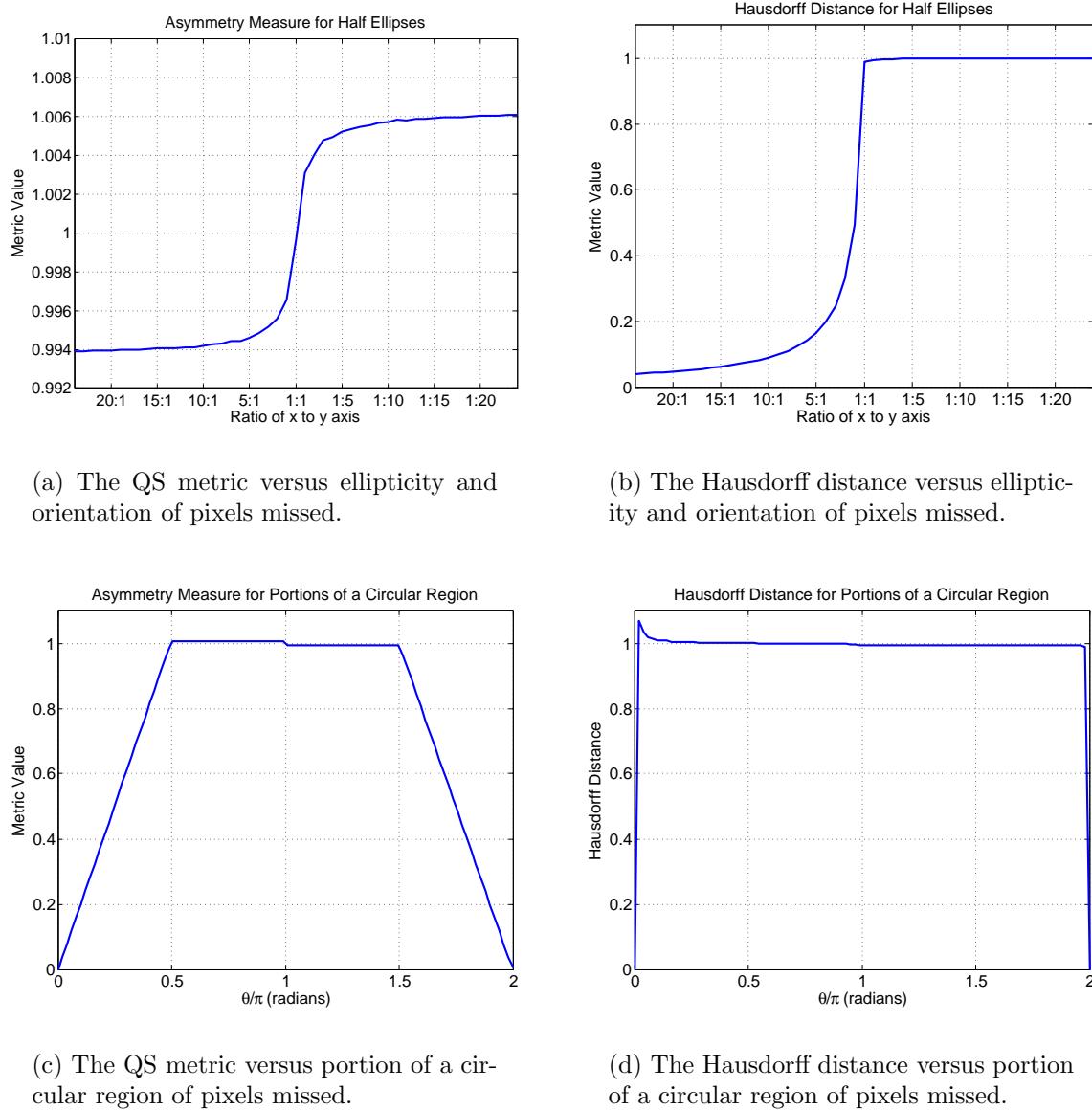
(a) Original hand silhouette with  $GT = 5270$  object pixels.

(b) Erosion by 1 pixel; total of  $PM = 524$  pixels eroded (missed).  $\frac{2 \cdot QS}{GT} = 0.188$ ,  $\frac{PM}{GT} = 0.099$ ,  $term_2 = 0.107$ .

(c) Thumb removed; total of  $PM = 524$  pixels missed.  $\frac{2 \cdot QS}{GT} = 0.397$ ,  $\frac{PM}{GT} = 0.099$ ,  $term_2 = 0.227$ .

**Figure 4.4.** Application of the QS and size metrics to an example silhouette and “segmentations.” Qualitatively, the segmentation in (b) retains more resemblance to the original silhouette in (a) than does the segmentation in (c), where the entire thumb is missed. A size metric alone (as the one used in [124]) would rank the two results in (b) and (c) as equally good segmentations, while the use of the QS metric penalizes the change in shape of (c). Note that in (b) the addition of the shape metric does not change the value of the original size-based metric by much (0.8%).

This is commonly used for boundary matching in, e.g., object detection and image retrieval applications; more recently it has been used for face recognition [135] and hand recognition [136]. It is unclear whether this metric can easily be applied to objects consisting of multiple boundaries and/or objects with holes. Additionally, in the application to segmentation of cell nuclei, we are more interested in the object itself rather than just the boundary. Figure 4.5 shows the Hausdorff distance for elliptical and circular regions, where, for reference, the same plots for the QS metric are included (from Figure 4.2). It appears that, while the Hausdorff distance may be well suited for boundary matching problems, it is not well suited for applications in which a measure of shape deviation is of importance. Additionally, it is not clear how to properly normalize the Hausdorff distance.



**Figure 4.5.** Comparison of the Quadrant Sum (QS) metric and Hausdorff distance for pixels missed in discrete elliptical and circular regions. The QS metric in these plots has been normalized by  $\frac{GT}{2}$ , where  $GT$  is the number of ground truth pixels, and the Hausdorff distance has been normalized by the largest diameter of the object. The Hausdorff distance provides a poor measure of the deviation in shape, and is not easily normalized.

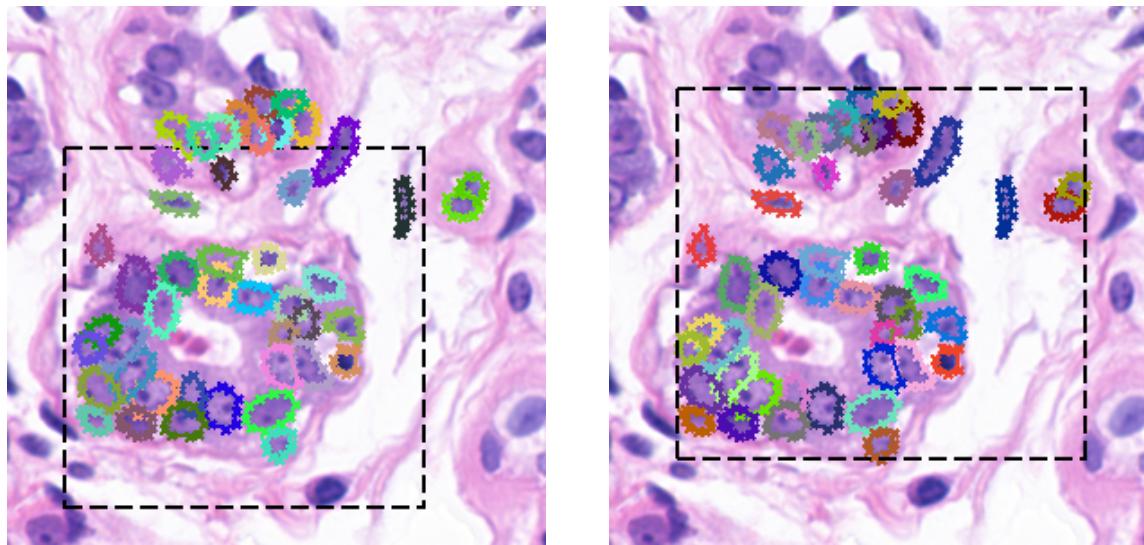
### 4.3.5 Ground Truth Image Markup

We have developed an interactive system for user delineation of object-level ground truth. Within this system, the user designates objects by delineating the perimeter; objects may overlap, and can be defined with as many or as few points as desired. These points are used to create an object mask containing the pixels within the designated perimeter. The user uses a truth window, a rectangular box, to help focus the image markup to a particular image region. There is no reason why this could not be extended to include image markups involving several truth windows, but we concentrate here on the case of one truth window to better correspond with the pixel-level image markups we use in Chapter 2.

While it is easy to specify a pixel-level markup within a designated truth window, such a specification becomes more complicated with an object-level markup. Consider the fact that in a pixel-level markup, an object that spans the truth window boundary can be marked up to the boundary without losing any important information for the overall classification. In an object-level markup, however, the actual extent and border of the object is of utmost importance. Moreover, if objects are marked within a rough concept of a truth window, there is the possibility that the truth window will contain parts of objects that have not been delineated by the user. This will lead to erroneously low performance since the segmentation metric will assume that these regions were incorrectly segmented.

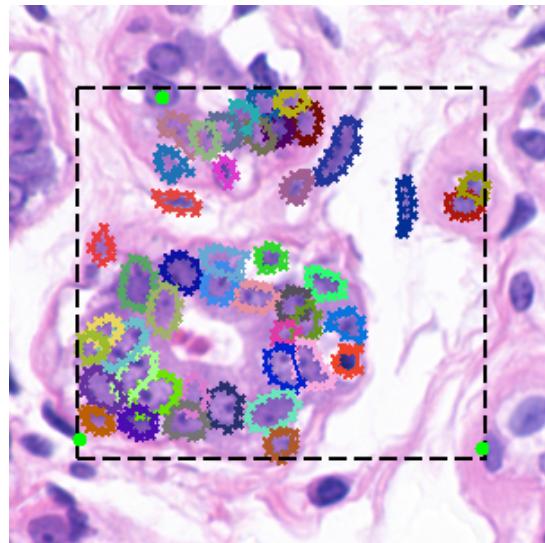
To help alleviate this problem, after the user indicates that the delineation of objects within the chosen truth window is complete (see Figure 4.6 (a)) the truth window is recomputed as the minimum bounding rectangle of the object markups (Figure 4.6 (b)). Using this new truth window, the user is asked to mark a minimum of one point for each unmarked object that is either completely or partially enclosed by the new truth window. These points should ideally be the centroid for unmarked regions entirely enclosed within the truth window and a point within the object *and* the truth window for objects partially contained (Figure 4.6 (c)). This information is used in a connected-components analysis to determine if extra segmented regions are associated with an object that has not been delineated in the ground truth markup. This allows for some disambiguation of the truly extra segmented regions from those segmented regions that can not be accurately accounted for in the ground truth markup.

We store as the ground truth file 1) the values of the perimeter points (which are generally not specified according to the discrete image grid, but a continuous one), 2) the object mask, 3) a mask of the entire image markup (computed by the Boolean OR of the various object markups), and 4) the points designating non-delineated objects. While it is possible to store only the perimeter and non-delineated object points (and thereby save disk space), the pixel masks are also stored to speed future computation. For the dataset of 58 images, defining object-level ground truth in a  $\sim 200 \times 200$  window, with an average of 50 marked and 9 unmarked objects per truth window, the ground truth files use  $\approx 8$  MB of disk space, with an average individual file requiring about 138 kB.



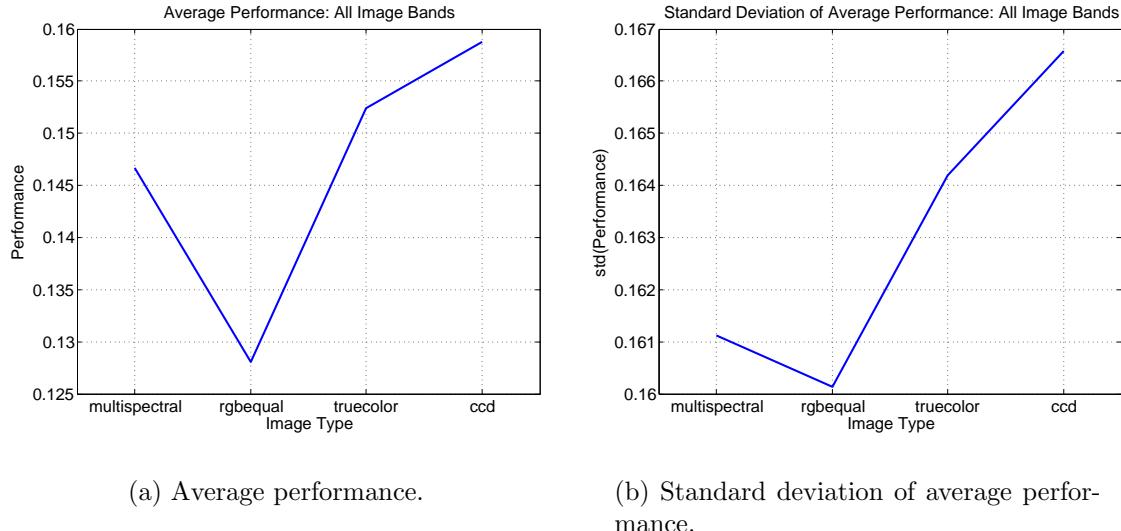
(a) Object delineation within the originally specified truth window. Each object perimeter is assigned a random color for display purposes.

(b) Recomputed truth window, the minimum bounding rectangle of the marked objects.



(c) Specification of unmarked objects within the recomputed truth window, displayed here as green dots.

**Figure 4.6.** Illustration of the object-level image markup process.



**Figure 4.7.** Watershed on the Complemented Distance Transform (WSCDT) segmentation performance using all image bands, averaged over the five pixel-level classifiers used in Chapter 2 (ML, MED, SAM, FLDA, and AFE).

## 4.4 Watershed-Based Segmentation of Nuclei

We investigate here various watershed-based segmentation methods for extraction of cell nuclei. Default weights (refer to Section 4.3.1) of  $\alpha = [0.5 \ 0.5 \ 0.5 \ 1 \ 0.5]$ ,  $\delta_{SR} = 1$ , and  $\delta_{ER} = 1$  are assigned.

### 4.4.1 Watershed on the Complemented Distance Transform (WSCDT)

The segmentation metric (Equation (4.1)) is applied to the watershed transform of the complemented Euclidean distance transform (WSCDT) of a pixel-level classification. As the binary image, we use the output of the pixel-level classifiers discussed in Chapter 2. In further discussion, no distinction will be made between the five classifiers; instead average segmentation performance of this algorithm and performance differences over the image types and image bands will be presented. For consideration of Wilcoxon p-values, which require paired comparisons, the classifiers will be individually considered.

Figure 4.7 shows the segmentation performance using all image bands for WSCDT. Note that the ccd images perform best overall, but also have the highest standard deviation in performance. Multispectral and rgbequal imagery have lower performance and lower standard deviation. The standard deviation in performance versus image type

**Table 4.1.** Wilcoxon paired signed-rank test p-values for performances of multispectral versus RGB imagery for Watershed on the Complemented Distance Transform (WSCDT) nuclear segmentation. Starred (\*) entries correspond to statistical significance at the p-value of 0.05. Using the Bonferroni correction for these 15 tests, we look for the corrected p-value of  $0.05/15 \approx 0.0033$ , which are presented as bold entries in the table. Note that all bold entries also satisfy the less conservative p-value of 0.05 although they are not explicitly starred.

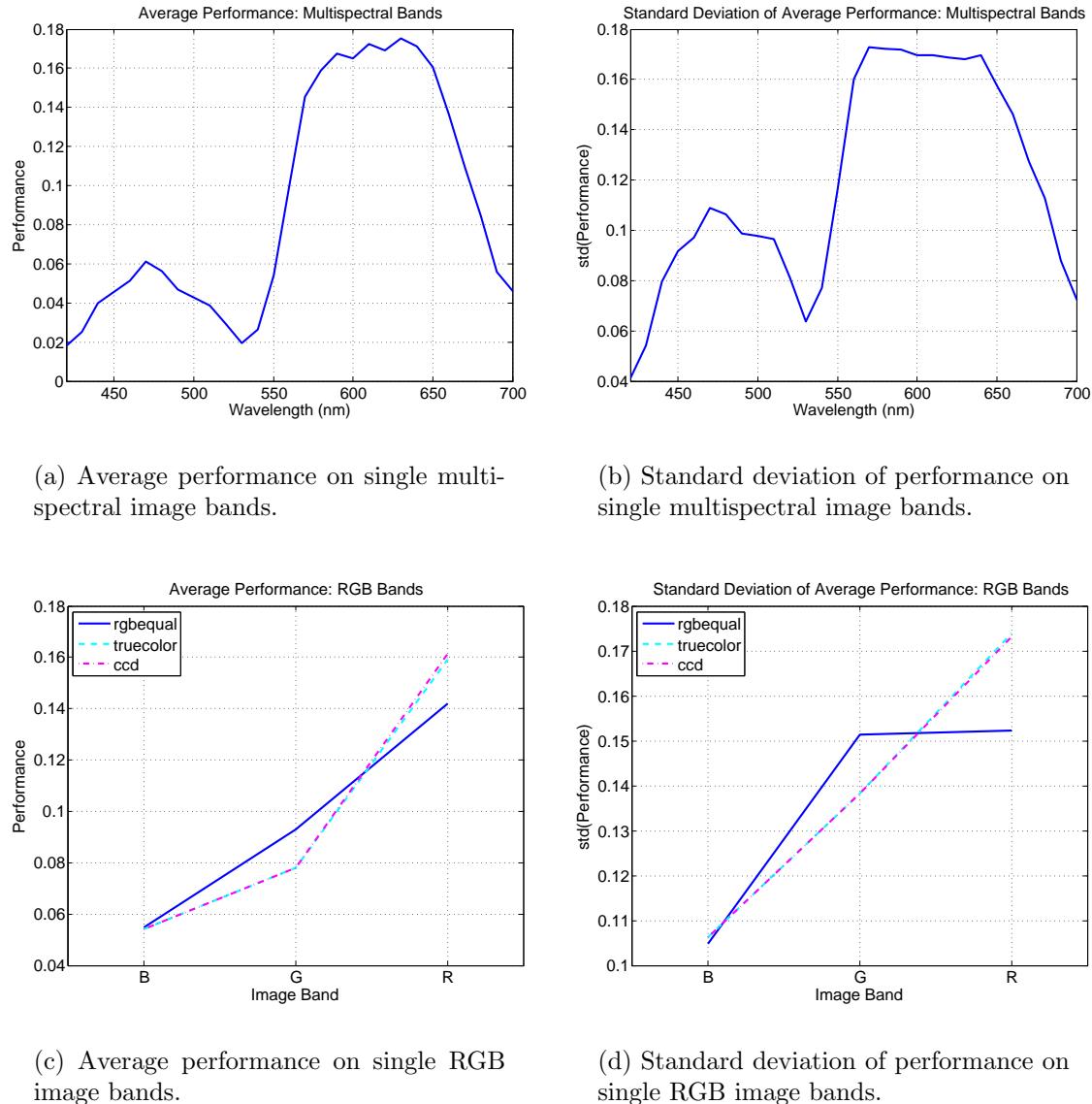
Classifier	Image–multi vs.		
	rgbequal	truecolor	ccd
<b>ML</b>	6.7e-2	<b>7.7e-6</b>	<b>4.1e-6</b>
<b>MED</b>	5.9e-2	6.7e-3*	3.6e-3*
<b>SAM</b>	<b>1.6e-5</b>	3.6e-1	5.1e-2
<b>FLDA</b>	<b>3.0e-4</b>	6.3e-3*	<b>1.7e-3</b>
<b>AFE</b>	<b>3.7e-7</b>	7.0e-1	1.1e-2*

follows the same trend as the average performance versus image type. There is a performance increase of  $\sim 0.01$  for ccd imagery versus multispectral, slightly larger than the performance differences seen in the pixel level classifiers (Chapter 2). It is important to note also that all scores are very poor (within the range of 0.125-0.160). Additionally, the standard deviation in scores is nearly the same as (and slightly larger than) the average performance, indicating that there is a wide spread in performance over the individual images. Table 4.1 shows the Wilcoxon p-values for each of the five classifiers, comparing the WSCDT performance for multispectral imagery versus RGB. It appears that about half of these differences are statistically significant, indicating that while multispectral outperforms rgbequal imagery for WSCDT nuclear segmentation, truecolor and ccd imagery outperform the multispectral. Remember that the multispectral versus RGB imagery distinction comes straight from the application of pixel-level classifiers to nuclei from Chapter 2.

The performance on single image bands for multispectral and RGB imagery is shown in Figure 4.8. Not surprisingly, for multispectral bands (Figure 4.8(a) and (b)) we note the same trend in performance that as in Chapter 2 (refer to Figure 2.5). Similarly, the RGB plots appear very similar to those in Chapter 2 (Figure 2.6). Again, the performance is poor and the standard deviation in performance is on the same order as the actual performance.

#### 4.4.2 Watershed on H-minima Transforms (WSHmin)

The segmentation metric (Equation (4.1)) is now applied to the watershed transform of an h-minima transform of the complemented distance transform. This has the practical



**Figure 4.8.** Watershed on the Complemented Distance Transform (WSCDT) segmentation performance on single image bands, averaged over the four classifiers used for single band analysis in Chapter 2 (ML, MED, FLDA, and AFE).

effect of disregarding minima that are relatively shallow. More precisely, we:

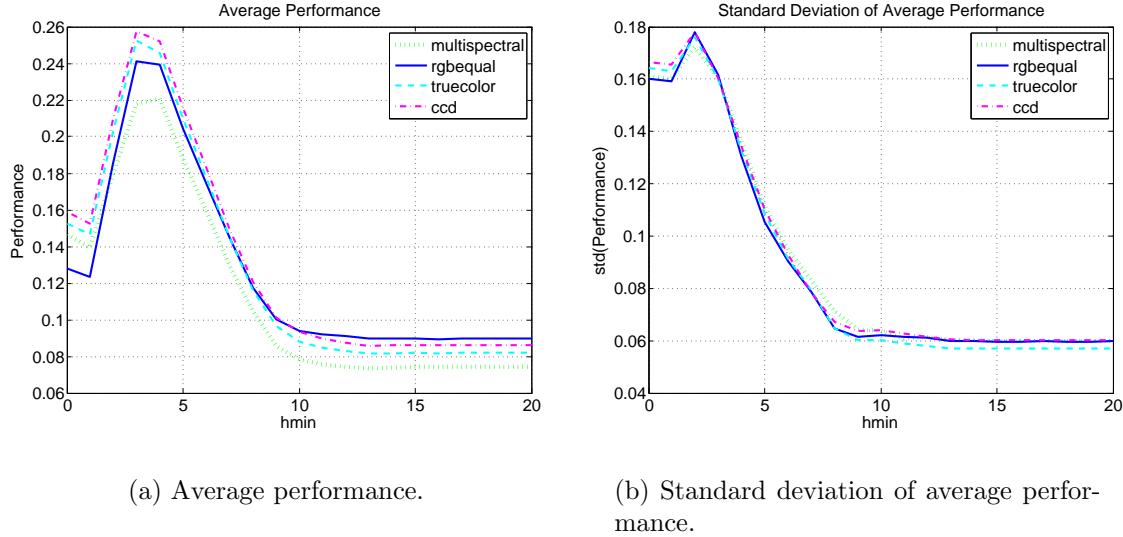
1. Compute the Euclidean distance transform of the complemented binary image.
2. Suppress all minima less than some value  $h_{min}$ .
3. Take the negative of the h-minima transform of the distance transform, and set the distance of all background pixels in the binary image to a depth of  $-\infty$ .
4. Compute the watershed transform on the resulting h-minima transform of the distance transform.

Again, the binary images used as input to this algorithm are the outputs of the various pixel-level classifiers used in Chapter 2. For analysis here, the focus is on the performance using all image bands, since the computation of multiple h-minima transforms increases the computation time required. The AFE classifier is excluded here, since accurate performance results requires an averaging of the 10 independent classifier runs. For an  $h_{min}$  range of [1, 20], as shown in Figure 4.9, the four image types (multispectral, rgbequal, truecolor, and ccd), and the four pixel-level classifiers (ML, MED, SAM, and FLDA), 18,560 separate h-minima and watershed transforms must be computed. Including the AFE classifier would increase this to 30,160; including all image bands would require 607,840 transforms.

In Figure 4.9, performance at first increases to a maximum at an  $h_{min}$  value of 3 or 4, after which the performance decreases and levels out. Multispectral imagery has the poorest performance for all values of  $h_{min}$  except 0 and 1. Again, however, these performances are not particularly good, and the standard deviation is on par with the actual performance. Table 4.2 shows the Wilcoxon p-values for each of the four classifiers, comparing the WSHmin performance for multispectral imagery versus RGB for  $h_{min} = 4$ . It appears most of these differences are statistically significant, indicating that all three types of RGB imagery outperform the multispectral.

Figure 4.10 plots the value of the various segmentation metric terms as they vary with  $h_{min}$ . The behavior of the individual metric terms is consistent with intuition: as  $h_{min}$  increases, the suppression of minima results in fewer total segmented regions and the performance of terms 1 (extra regions per ground truth region) and 5 (extra segmented regions) increases, while the performance of terms 2 (pixels missed), 3 (extra pixels), and 5 (nuclei detected) decreases. The brief increase in term 3 is due to smaller extraneous watershed regions being suppressed by the h-minima transform; after a certain point, however, the h-minima transform creates few enough minima that nuclei regions are merged by the watershed transform, resulting in the subsequent decrease in performance of term 3.

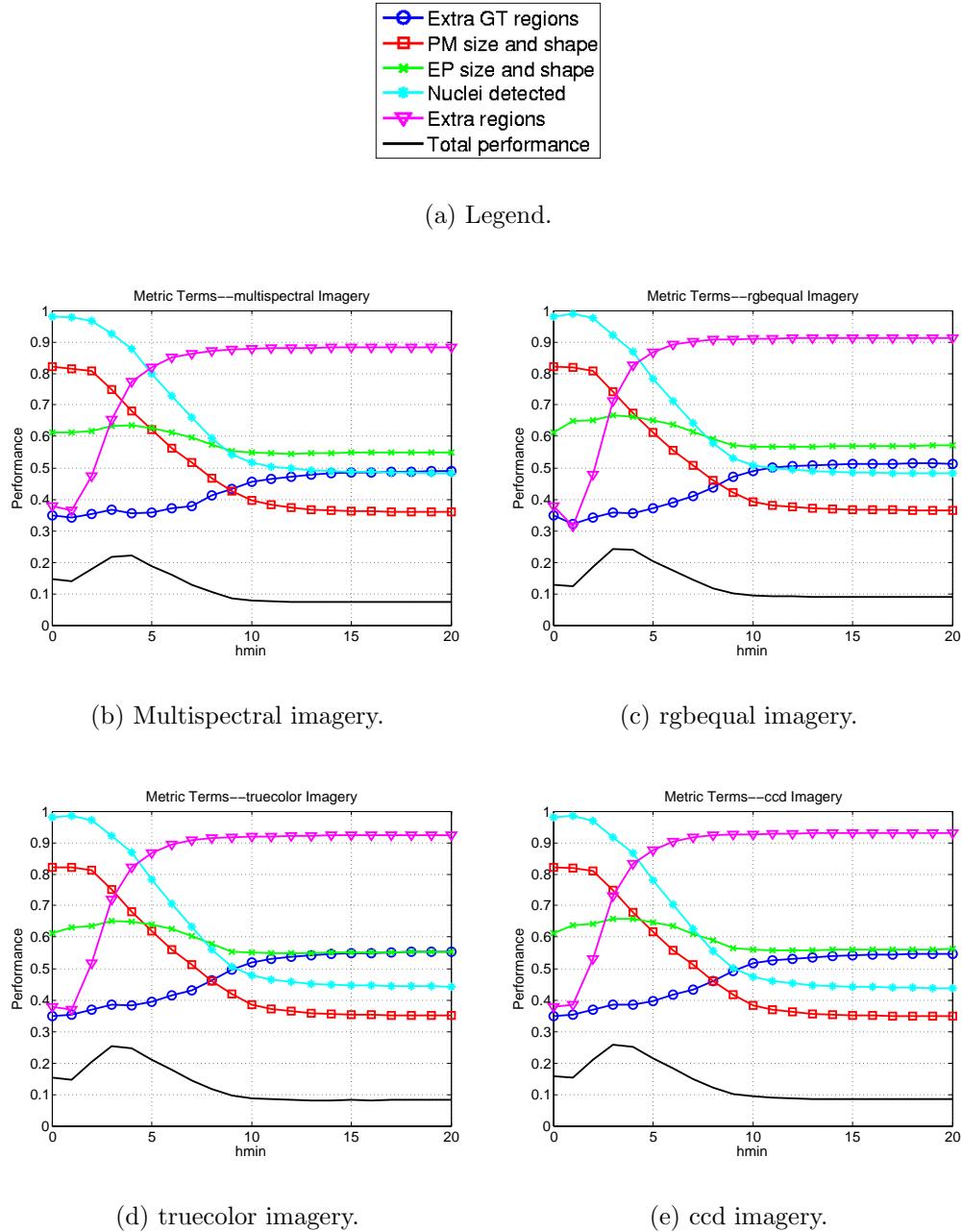
Figure 4.11 shows example segmentation results for different values of  $h_{min}$  along with the corresponding performance. This illustrates the fact that the initial increase in



**Figure 4.9.** Watershed on H-minima Transforms (WSHmin) segmentation performance for different values of  $h_{min}$ .  $h_{min} = 0$  (i.e., Watershed on the Complemented Distance Transform (WSCDT)) is included in these plots for easy reference. These results are averaged over the ML, MED, SAM, and FLDA pixel-level classifiers from Chapter 2.

**Table 4.2.** Wilcoxon paired signed-rank test p-values for performances of multispectral versus RGB imagery for Watershed on H-minima Transforms (WSHmin) nuclear segmentation,  $h_{min} = 4$ . Starred (\*) entries correspond to statistical significance at the p-value of 0.05. Using the Bonferroni correction for these 12 tests, we look for the corrected p-value of  $0.05/12 \approx 0.0042$ , which are presented as bold entries in the table. Note that all bold entries also satisfy the less conservative p-value of 0.05 although they are not explicitly starred.

Classifier	Image–multi vs.		
	rgbequal	truecolor	ccd
<b>ML</b>	<b>3.4e-6</b>	<b>6.7e-9</b>	<b>2.8e-9</b>
<b>MED</b>	2.0e-2*	<b>1.1e-4</b>	2.7e-2*
<b>SAM</b>	1.1e-2*	<b>3.4e-3</b>	3.1e-2*
<b>FLDA</b>	7.5e-1	4.7e-1	3.0e-1



**Figure 4.10.** Value of Watershed on H-minima Transforms (WSHmin) performance metric terms versus  $h_{min}$  for different imagery. All terms plotted here are performance, i.e., one minus the penalty, where the penalties are the terms previously discussed in relation to the segmentation metric. The terms are denoted by a brief description in the charts, but they are also plotted in numerical order, i.e., blue circles are term 1, red squares are term 2, and so forth.

performance with increase in  $h_{min}$  is most likely due to the decrease in oversegmentation, although we do lose some well segmented nuclei in the process. Similarly, we can see that subsequent decrease in performance as  $h_{min}$  further increases is most likely due to the further loss of objects.

#### 4.4.3 Watershed Using Granulometry-Based Markers (WSGran)

Granulometries are a common technique to elicit information about the size distribution of objects in an image. Granulometries are calculated by applying successively larger structuring elements in a morphological image operation and analyzing the residue image. More specifically, we use the morphological opening of a binary image with a disk-shaped structuring element (SE). The radius of the SE is increased by 1 pixel each iteration and the residue of the image (the sum of the pixels) is calculated. Iteration terminates when the image residue is zero, i.e., the structuring element is large enough to completely remove all non-zero pixels. Thus, there is a set of structuring elements  $GSE_k$  which are used to calculate the morphological opening of a binary image  $I \circ GSE_k$ . Defining  $\Psi_k$  as the area (number of pixels) in the  $k$ -th residue image, the function

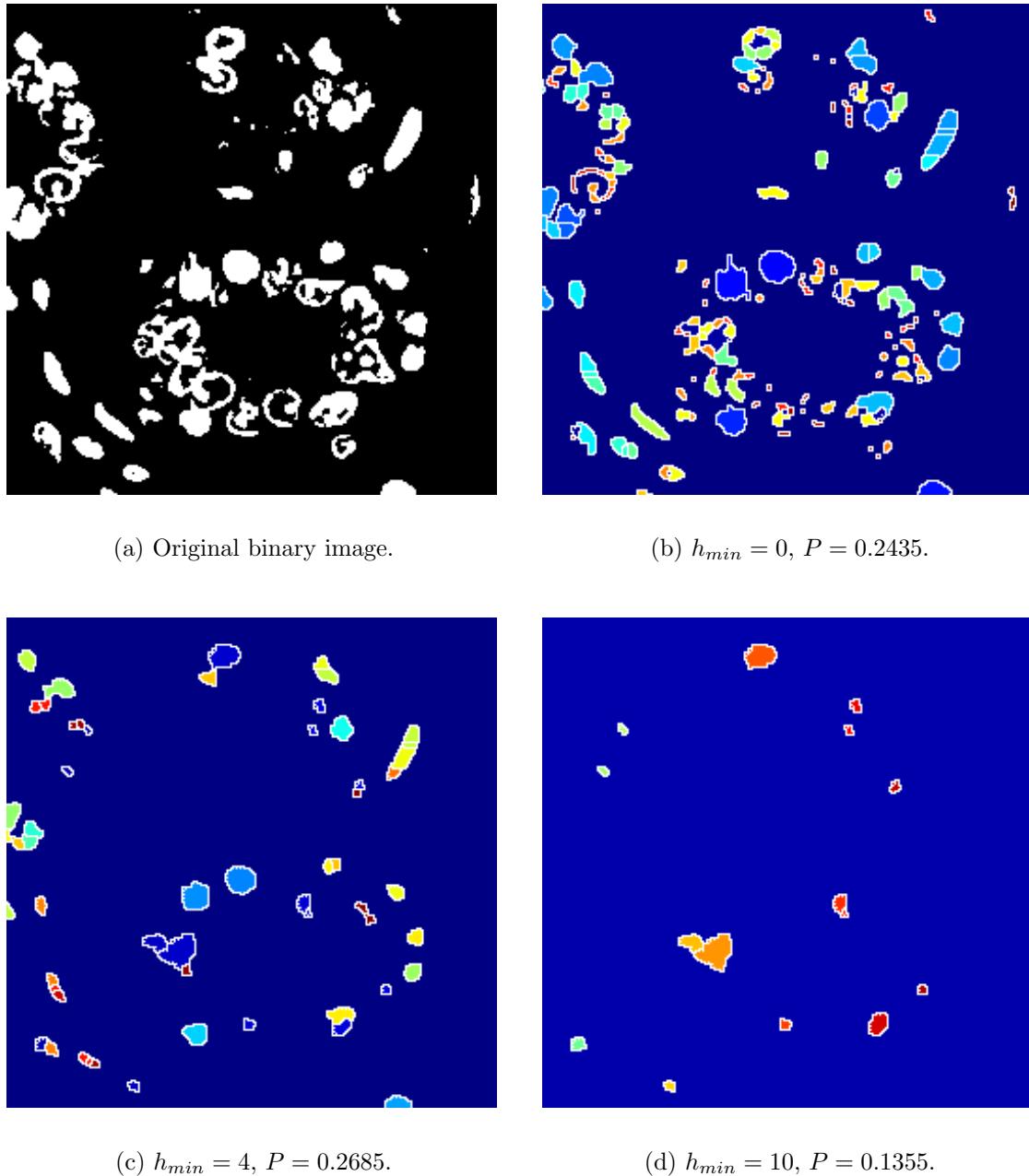
$$\phi(k) = 1 - \frac{\Psi_k}{\Psi_0} \quad (4.13)$$

is used to calculate the size distribution of image  $I$  [137, 138]. Looking at the first derivative (the element-wise subtraction) of  $\phi(k)$  will yield a local maximum for a structuring element with approximate size of a large number of objects in the original binary image. The derivative  $\phi'(k)$  is often called the pattern spectrum of image  $I$ .

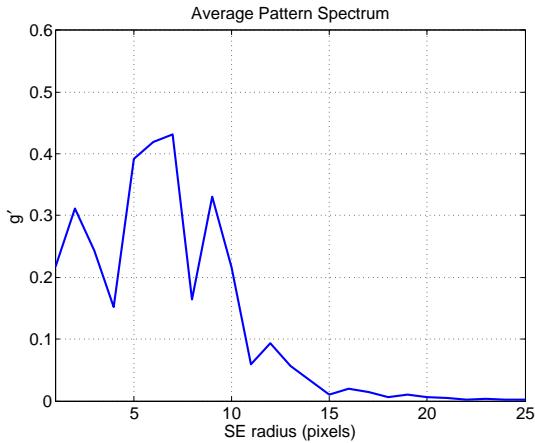
The average pattern spectra of the dataset, as well as the benign and malignant subsets of the dataset are shown in Figure 4.12. Significant local maxima occur at SE radii of 2, 5-7, and 9, with less significant minima at radii of 12, 16, and 19. These pattern spectra are averaged over all five pixel-level classifiers using all image bands.

We expected a more significant difference between the pattern spectra of benign and malignant imagery, namely that the larger nuclei inherent in malignant imagery would manifest itself as larger objects in the granulometry plots. Instead, it appears that the pattern spectra of benign and malignant imagery are quite similar, with the main differences being the relative height of the local maxima. Oddly enough, it is the malignant imagery that has more objects of radius 2 and 5, and the benign more at radius 7. We hypothesize that this effect is due to the fragmented pixel-level segmentation of malignant nuclei due to the more prominent chromatin texture.

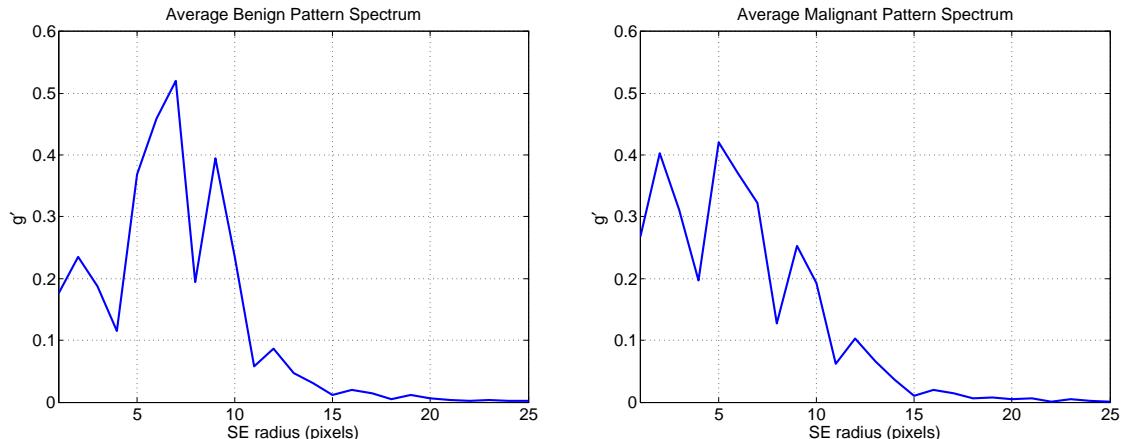
Figure 4.13 shows the pattern spectrum of a single benign image and single malignant image (output of the FLDA classifier), which are used as example images to support our hypothesis about the reason for differing malignant and benign pattern spectra. In



**Figure 4.11.** Example Watershed on H-minima Transforms (WSHmin) segmentations where each color indicates a different segmented region. Note that as  $h_{min}$  increases, the oversegmentation is diminished, but many valid nuclei are completely missed.



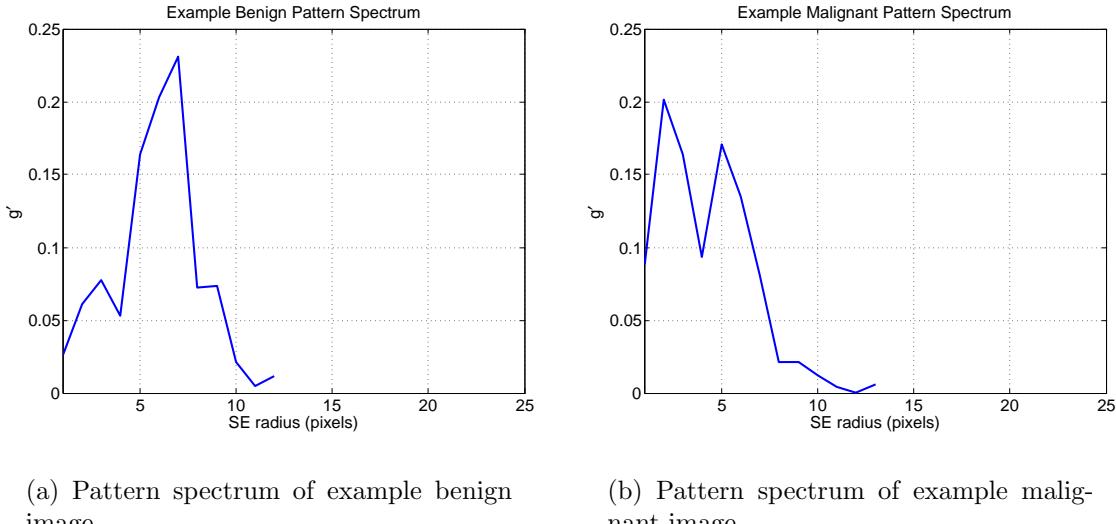
(a) Average pattern spectrum over all dataset images.



(b) Average pattern spectrum over benign images.

(c) Average pattern spectrum over malignant images.

**Figure 4.12.** Pattern spectra of dataset images, displayed on the same scale.



**Figure 4.13.** Pattern spectra of example images. The x-axis (structuring element (SE) radius) is scaled between 1 and 25 to allow better comparison to the plots in Figure 4.12, even though these particular pattern spectra do not contain non-zero values past a radius of 12 or 13.

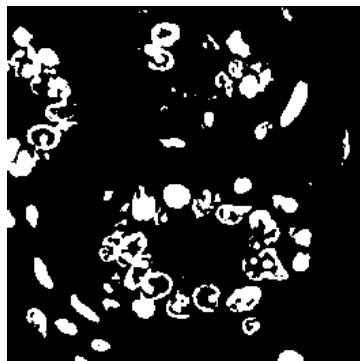
Figure 4.13 the two images chosen as examples display the same trends in local maxima as the average plots in Figure 4.12.

Figure 4.14 shows the image residues for SE radii of 2, 5, 6, 7, and 9 (the local maxima) for these example images. Note that in the original binary images (Figure 4.14 (a) and (d)) the malignant image contains many small regions within several nuclei, due to the chromatin distribution. Thus, the residue after a SE of radius 2 contains significantly fewer objects than the corresponding residue of the benign image; this corresponds with the more significant maxima in the pattern spectrum of the malignant imagery at a SE radius of 2. A similar effect can be seen at radius 5, whereas the benign residues have a larger relative difference for radii of value 6, 7, and 9. These observations correspond nicely with the pattern spectra plots in Figure 4.13.

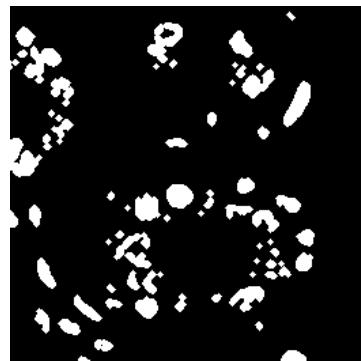
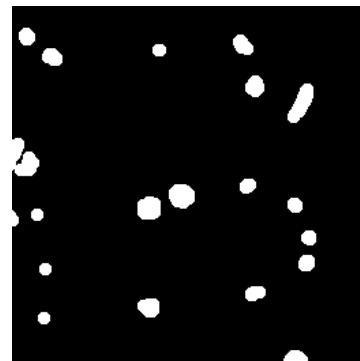
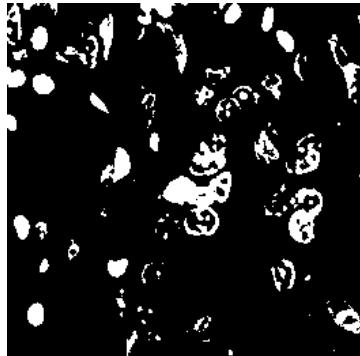
While it was initially surprising to find such a similarity between the pattern spectra of benign and malignant images, this may ultimately prove helpful in the definition of nuclei markers for a marker-based watershed (or other segmentation) algorithm. We compute watershed transforms with minima of the input image imposed according to foreground and background markers:

- Foreground markers: We use image residues from structuring elements of varying radii:

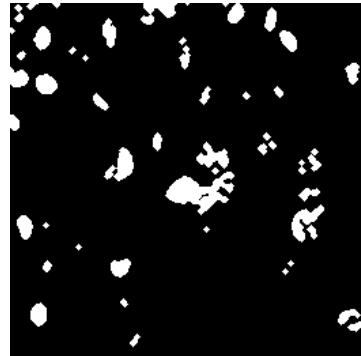
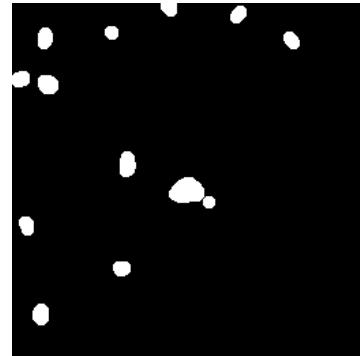
$$F = I \circ GSE \quad (4.14)$$



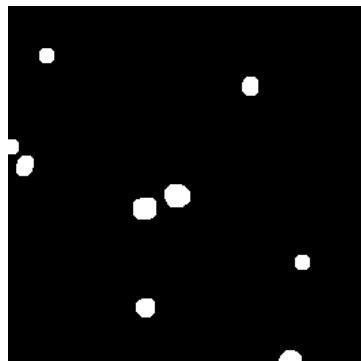
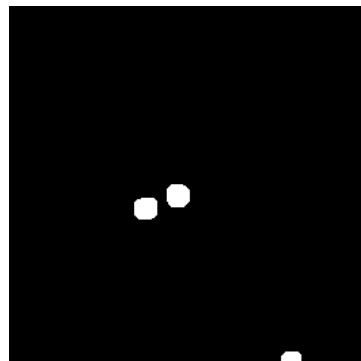
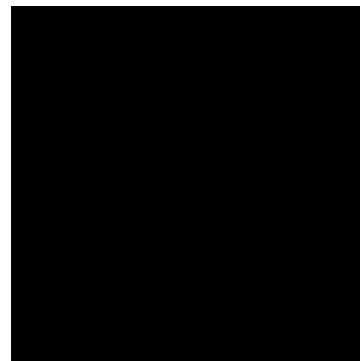
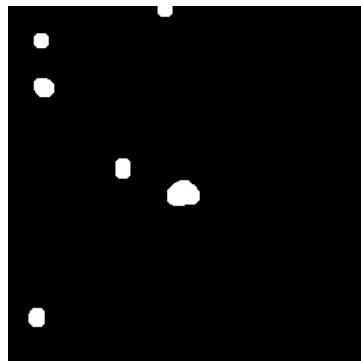
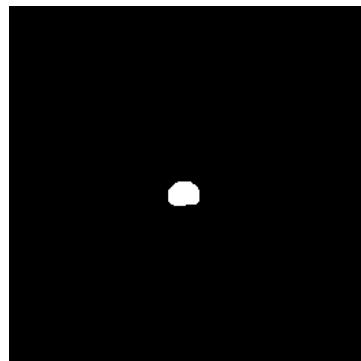
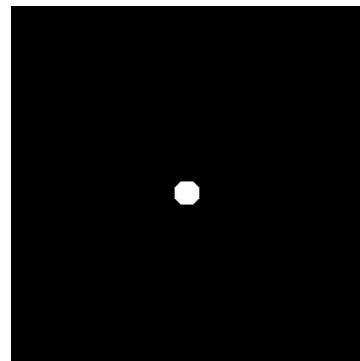
(a) Original binary benign image.

(b) Benign residue  $r = 2$ .(c) Benign residue  $r = 5$ .

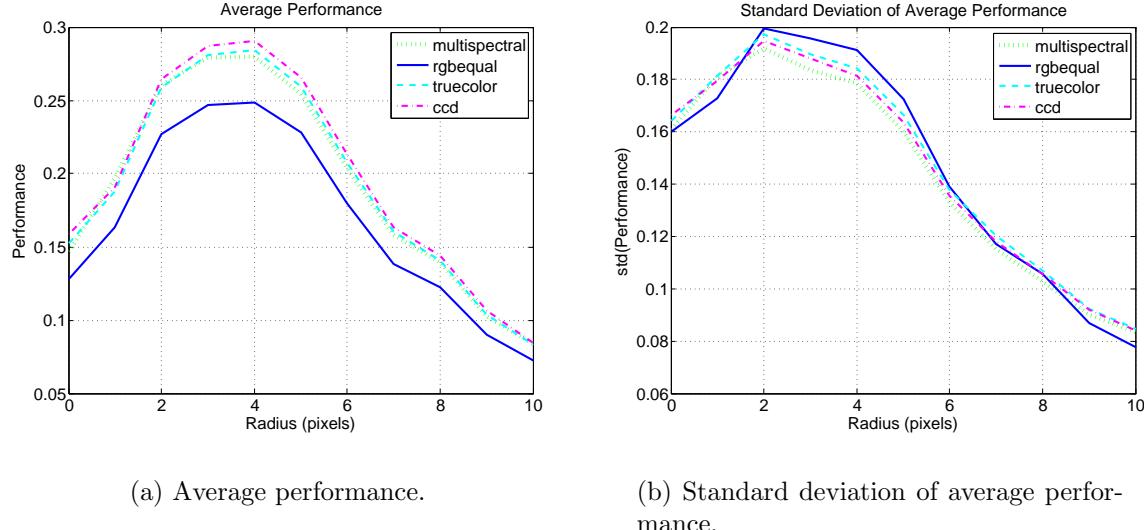
(d) Original binary malignant image.

(e) Malignant residue  $r = 2$ .(f) Malignant residue  $r = 5$ .

**Figure 4.14.** Example benign and malignant residue images for structuring elements of radii  $r$  at local maxima. (continued on next page)

(g) Benign residue  $r = 6$ .(h) Benign residue  $r = 7$ .(i) Benign residue  $r = 9$ .(j) Malignant residue  $r = 6$ .(k) Malignant residue  $r = 7$ .(l) Malignant residue  $r = 9$ .

**Figure 4.14.** (Cont.) Example benign and malignant residue images for structuring elements of radii  $r$  at local maxima.



**Figure 4.15.** Watershed using Granulometry-based markers (WSGran) segmentation performance for different values of structuring element radius.  $r = 0$  (i.e., Watershed on the Complemented Distance Transform (WSCDT)) is included in these plots for easy reference. These results are averaged over the five pixel-level classifiers used in Chapter 2 (ML, MED, SAM, FLDA, and AFE).

where  $I$  is the original binary image,  $\circ$  is the image opening operator, and  $GSE$  is the structuring element.  $GSE$  is a discrete circle of radius  $r$ , where  $r$  is chosen as the approximate size of most image objects, according to the pattern spectrum. These markers serve as an approximation of the center of the binary objects.

- Background markers: We use the erosion of the complement of the original binary image:

$$B = \tilde{I} \ominus SE \quad (4.15)$$

where  $\tilde{I}$  is the binary complement of  $I$ ,  $\ominus$  is the erosion operator, and  $SE$  in this case is a fixed size discrete circle of radius 3. These markers impose a minima in all parts of the background of the binary image. The erosion operator creates a separation between the background markers and the object boundaries.

The markers can be thought of as seeds for the watershed transform.

Figure 4.15 shows the performance of the WSGran method versus the SE radius used to determine the foreground markers. There is a general increase in performance as the SE radius increases, up to a radius of 4 pixels, followed by a decrease in performance. Also, the standard deviation in performance is once again similar to the average performance. Similar to the WSCDT and WSHmin performances, multispectral imagery seems to

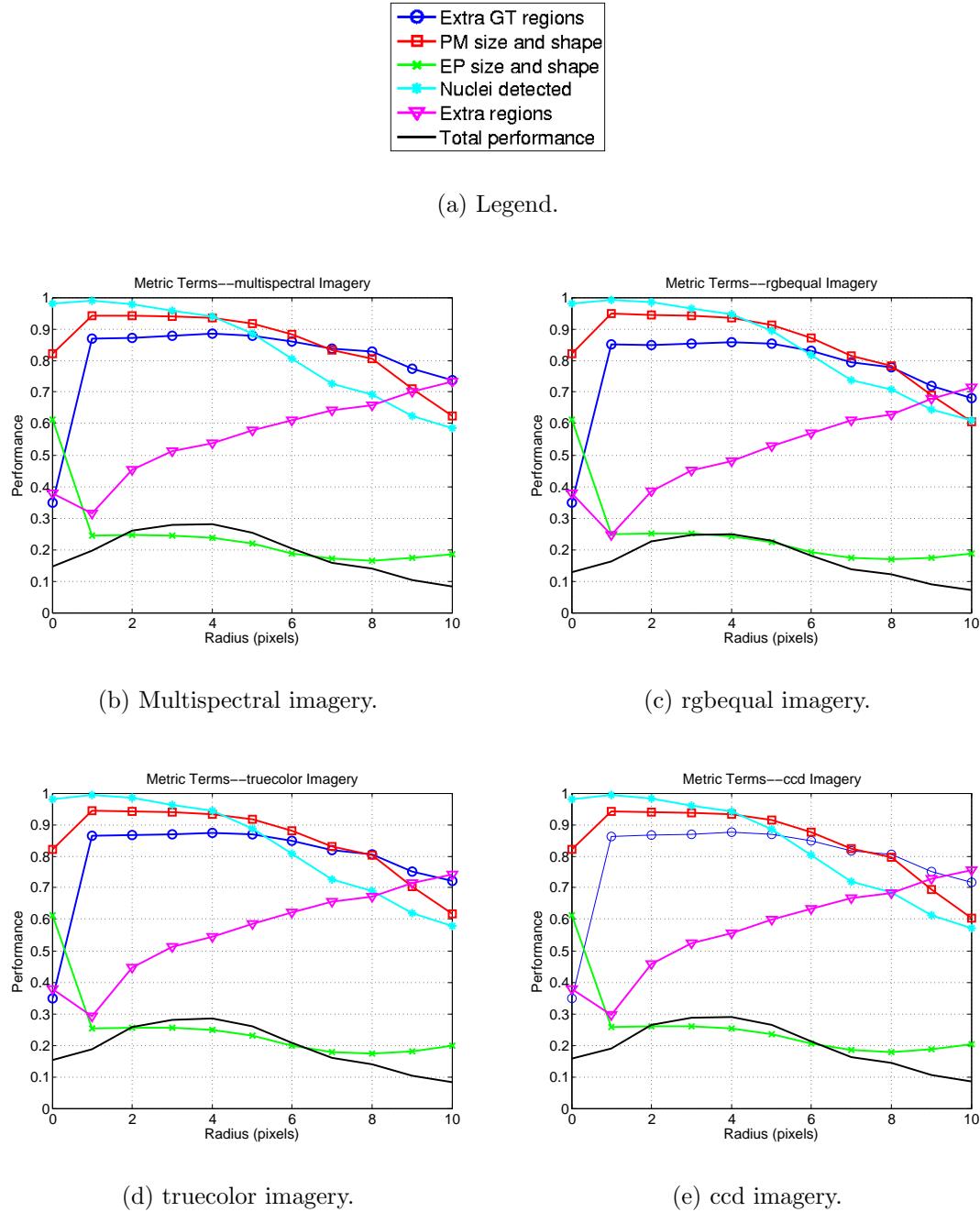
**Table 4.3.** Wilcoxon paired signed-rank test p-values for performances of multispectral versus RGB imagery for Watershed using Granulometry-based markers (WSGran) nuclear segmentation,  $r = 4$ . Starred (\*) entries correspond to statistical significance at the p-value of 0.05. Using the Bonferroni correction for these 15 tests, we look for the corrected p-value of  $0.05/15 \approx 0.0033$ , which are presented as bold entries in the table. Note that all bold entries also satisfy the less conservative p-value of 0.05 although they are not explicitly starred.

Classifier	Image–multi vs.		
	rgbequal	truecolor	ccd
<b>ML</b>	5.9e-1	<b>6.1e-6</b>	<b>3.1e-6</b>
<b>MED</b>	6.4e-2	2.7e-1	5.0e-1
<b>SAM</b>	<b>7.3e-7</b>	3.8e-3*	6.7e-1
<b>FLDA</b>	<b>1.9e-6</b>	<b>1.0e-5</b>	<b>3.3e-6</b>
<b>AFE</b>	<b>2.8e-6</b>	<b>1.1e-3</b>	<b>7.0e-7</b>

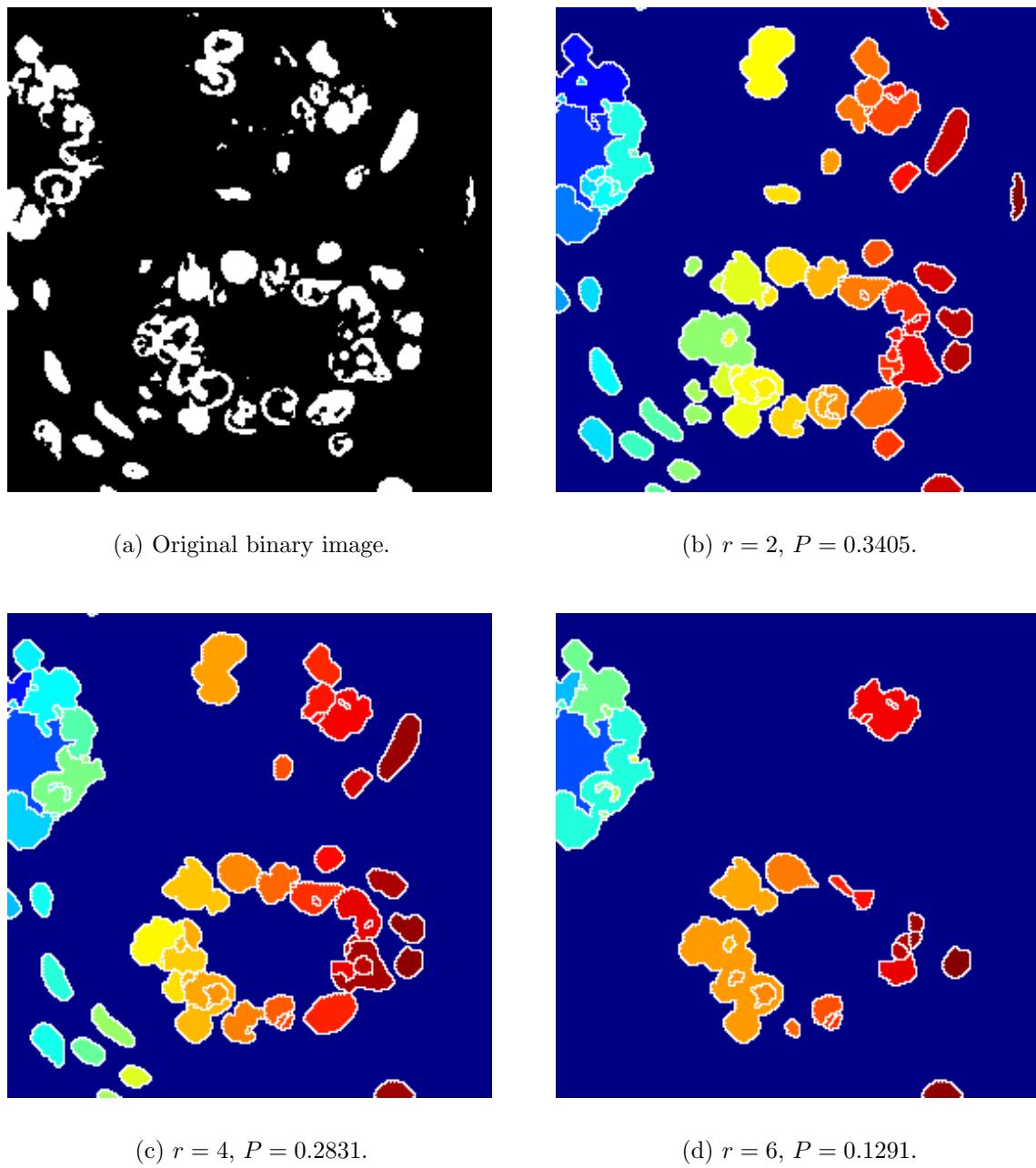
perform the worst. Wilcoxon p-values for multispectral versus the RGB imagery are shown in Table 4.3 where it appears that many of these differences in performance are statistically significant.

Figure 4.16 shows the performance of individual metric terms for different SE radii. With the introduction of granulometry-derived markers (radius 1), there is a sudden jump in the performance of term 1 (extra segmented regions per ground truth region) and a sudden decrease in performance of term 3 (extra pixels). This indicates an overall tendency towards undersegmentation of the granulometry-based marker method. For SE radii of 1-4, there is a general increase in the performance of terms 1 (extra segmented regions per ground truth region) and 5 (excess segmented regions), relatively flat performance for terms 2 (pixels missed) and 3 (extra pixels), and a decrease in term 4 (nuclei detected). This corresponds with the intuition that as the SE size is increased, there will be less markers and thus less watershed regions (increasing the performance of terms 1 and 5), while the overall pixels included in each ground truth nucleus should remain relatively constant. As the SE radius increases beyond 4 pixels, the performance of all terms begins to decrease (with the exception of term 5). Again, this corresponds nicely with intuition that as the SE is increased beyond the average size of objects in the image, we will begin to lose entire objects from the marker selection and watershed segmentation. Refer again to Figure 4.14 for illustration of the effects of SE size.

Figure 4.17 shows example segmentations for different SE radii as markers. We note qualitatively here that the WSGran segmentations yield fewer small regions compared to the WSHmin segmentations. This can be seen in the many merged nuclei in Figure 4.17 compared to Figure 4.11. To this end, note the relatively poor performance of term 3 (extra pixels) for all values of SE radius (refer to Figure 4.16). Conversely, WSHmin has a relatively poor performance for term 1 (extra regions per GT region), indicating that



**Figure 4.16.** Value of Watershed using Granulometry-based markers (WSGran) performance metric terms versus structuring element radius  $r$  for different imagery. It should be noted that all terms plotted here are performance, i.e., one minus the penalty, where the penalties are the terms previously discussed in relation to the segmentation metric.



**Figure 4.17.** Example Watershed using Granulometry-based markers (WSGran) segmentations, where each color indicates a different segmented region.

WSHmin has a tendency to oversegment. The performance of other terms is comparable between the two methods.

#### 4.4.4 Blobdetector-Based Segmentation (WSBlob)

For this method we use an *a priori* assumption about the shape of cell nuclei, namely that they are roughly circular in shape and approximately the same diameter. This method is motivated by the cell nuclei detector presented by Byun et al. in [139], which we will refer to as the “blobdetector.”

Byun et al. [139] use an inverted Laplacian of Gaussian (LoG) filter for detection of blobs in fluorescent confocal retinal imagery. For use in our brightfield imagery, we use a non-inverted LoG filter in the same blobdetector framework of [139].<sup>2</sup> The LoG filter is shown in Figure 4.18 for reference; the filter size is chosen to be the median diameter of the blobs of interest. The blobdetector framework also implements a local-maxima search over the LoG filtered image, using another parameter characterizing the minimum distance between blob centers (generally assumed to be approximately half the filter width). We note a few relevant characteristics of the blobdetector method:

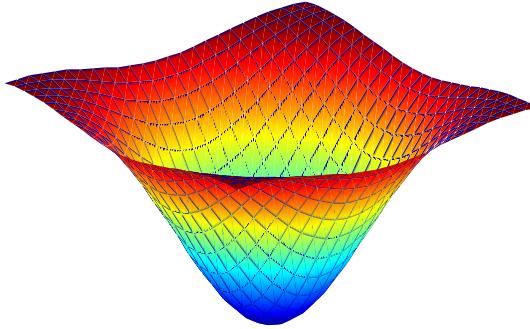
1. The use of a circular model for cell nuclei is motivated by the desire for rotation invariance of the filtered output, and is shown in [139] to provide both quantitatively and qualitatively good results for roughly elliptical objects.
2. While one can fine tune the filter size and object separation parameters, the method does not appear to be overly sensitive to choice of these parameters.
3. The method was applied to a range of image types and was shown to have good performance for cell nuclei detection, as measured by average error in nuclei counting.

This method, however was developed for the task of nuclei detection and counting applications; it does not address the issue of nuclear delineation. We are interested, however, in using the detection capabilities of this method as a seed for a subsequent watershed segmentation. Thus, although some *a priori* assumption about nuclei shape is used for detection of the rough location of nuclei, the watershed transform is used to delineate the shape of individual nuclei. This method (WSBlob) proceeds in a similar fashion to other marker-based watershed methods previously discussed:

- Detect nuclei using the red channel of the ccd imagery and use these locations as foreground markers for the watershed transform. A filter size of 25 pixels in diameter and an inter-blob distance of 12 was empirically chosen.

---

<sup>2</sup>Code available at <http://www.bioimage.ucsb.edu/software.html>



**Figure 4.18.** Laplacian of Gaussian used as a model of cell nuclei.

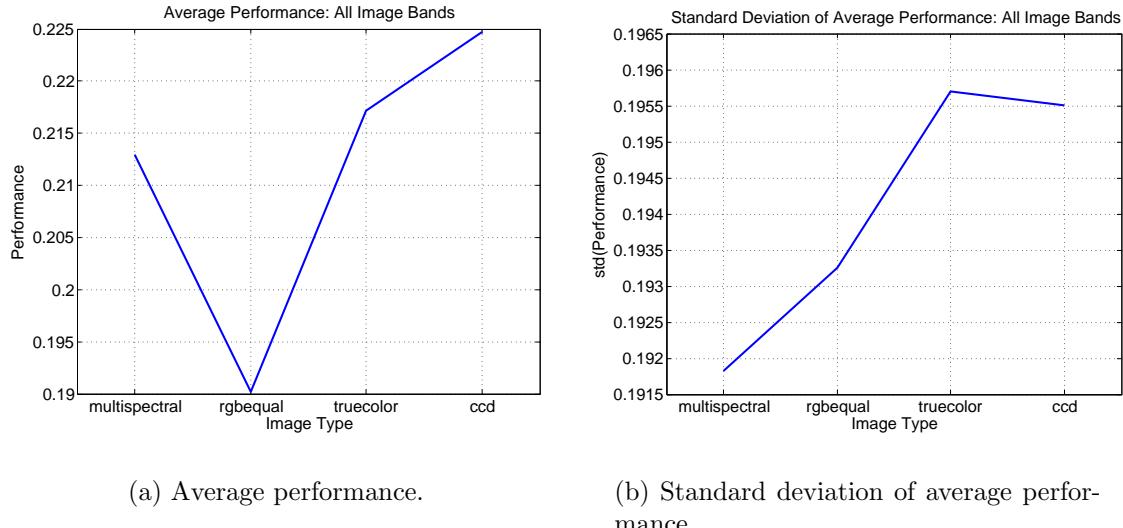
**Table 4.4.** Wilcoxon paired signed-rank test p-values for performances of multispectral versus RGB imagery for blobdetector-based (WSBlob) nuclear segmentation. Starred (\*) entries correspond to statistical significance at the p-value of 0.05. Using the Bonferroni correction for these 15 tests, we look for the corrected p-value of  $0.05/15 \approx 0.0033$ , which are presented as bold entries in the table. Note that all bold entries also satisfy the less conservative p-value of 0.05 although they are not explicitly starred.

<b>Classifier</b>	<b>Image</b> -multi vs.		
	rgbequal	truecolor	ccd
<b>ML</b>	3.5e-2*	<b>5.2e-7</b>	<b>6.3e-7</b>
<b>MED</b>	2.6e-2*	6.9e-2	8.8e-1
<b>SAM</b>	<b>2.0e-5</b>	4.5e-1	<b>6.1e-4</b>
<b>FLDA</b>	<b>3.3e-6</b>	<b>7.5e-5</b>	<b>1.7e-5</b>
<b>AFE</b>	<b>5.6e-6</b>	5.5e-2	<b>8.8e-4</b>

- Use the eroded complement of the binary nuclei classification as background markers.

By using the shape-based information from the blobdetector method as foreground markers and the spectral information from the pixel-level classification to develop the background markers and the distance transform, the false nuclei detections of the blobdetector method can be ignored if they fall outside of the spectrally classified nuclei regions. This makes the assumption that the pixel-level classification of nuclei more accurately determines regions of nuclear material, while the blobdetector more accurately locates individual nuclei.

Average results for the WSBlob method are shown in Figure 4.19 which illustrates a performance between that of the WSCDT and other methods. There is again a large standard deviation in performance. Wilcoxon p-values for multispectral versus RGB



**Figure 4.19.** Blobdetector-based (WSBlob) segmentation performance using all image bands, averaged over the five pixel-level classifiers used in Chapter 2 (ML, MED, SAM, FLDA, and AFE).

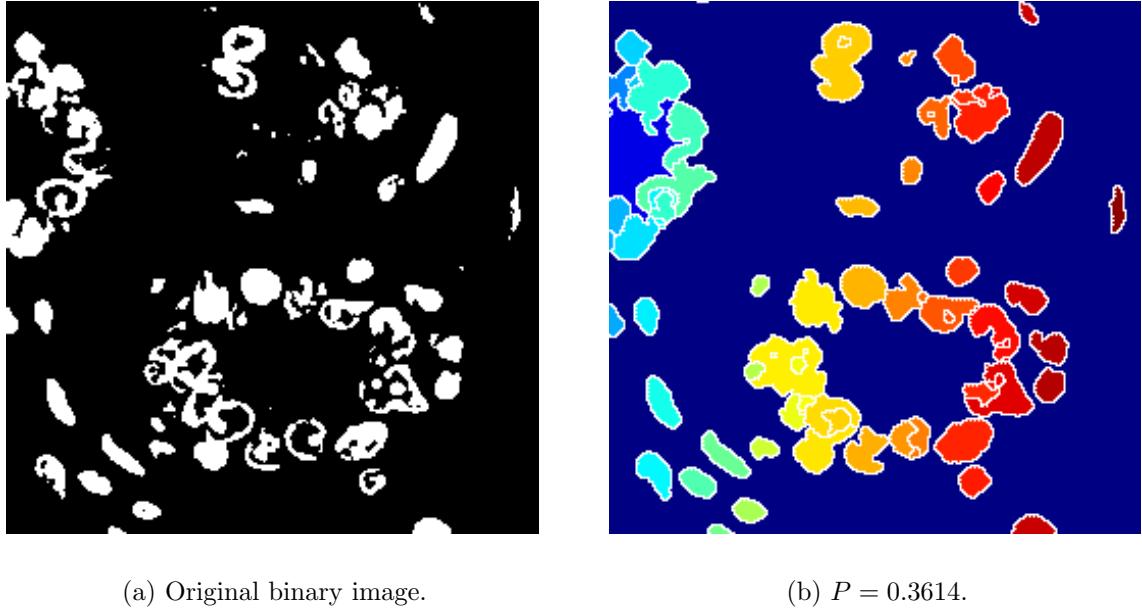
imagery is shown in Table 4.4. An example WSBlob segmentation is shown in Figure 4.20.

## 4.5 Concavity-Based Segmentation of Nuclei

Motivated by observations that shape is a large factor in humans' ability to properly discriminate individual nuclei, we turn our attention to a shape-based segmentation method, using concavities as the basis of possible segmentation lines. We use the method presented by Kumar et al. in [140]. We will first provide an overview of the concavity-based method before describing our results.

### 4.5.1 Overview of the Concavity-Based Segmentation Algorithm (Kumar)

The concavity-based segmentation method of Kumar et al. [140] is the culmination of previous work by Kumar and other researchers, namely [141–144]. It uses a rule-based approach for the segmentation of binary objects, beginning with a measure of concavity depth, adopted from a study of several measures of concavity in [145]. In this method, the segmentation of one binary blob into multiple constituent objects is accomplished with the use of a “split line.” This split line represents the boundary between the constituent



**Figure 4.20.** Example blobdetector-based (WSBlob) segmentation, where each color indicates a different segmented region.

objects. A split line may occur between two object concavities, or between a single concavity and the opposite object boundary. Binary objects are split recursively until no further valid split lines are found.

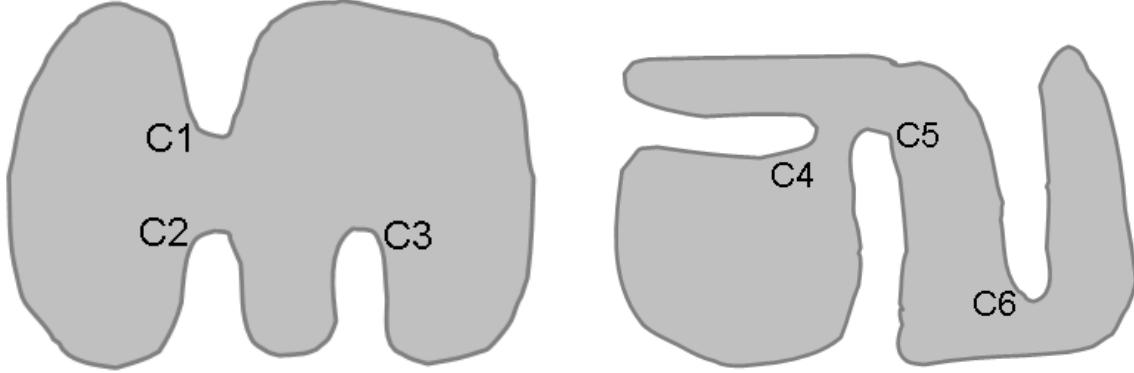
There are six features used to determine the presence of valid split lines [140]:

- **Concavity depth (CD):** Denoting the chords of the convex hull of a binary object as  $K_i$  and the corresponding boundary arcs as  $B_i$  (the boundary pixels of the object between the endpoints of  $K_i$ ), the concavity pixel  $C_i \in B_i$  is the pixel with the largest perpendicular distance  $CD_i$  between a boundary pixel  $b \in B_i$  and the chord  $K_i$ .
- **Saliency (SA):** Using intuition that a “good” split line will have large concavities at each endpoint as well as a small length, the saliency metric is defined as:

$$SA_{ij} = \frac{\min(CD_i, CD_j)}{\min(CD_i, CD_j) + d(C_i, C_j)} \quad (4.16)$$

where  $d(\cdot)$  is the Euclidean distance function.

- **Concavity-concavity alignment (CC):** This metric represents the relative alignment of the concavities with respect to each other. Denoting the unit vector connecting the concavity pixel  $C_i$  with the midpoint of chord  $K_i$  as  $\mathbf{v}_i$ , and similarly



(a) Oppositely aligned concavities C1 and C2 define a more intuitive split line ( $CC_{12} \approx 0$ ,  $CL_{12} \approx 0$ ), while parallel concavities C2 and C3 do not ( $CC_{23} \approx \pi$ ,  $CL_{23} \approx \pi/2$ ). Concavities C1 and C3 form a split line with  $CC_{13} \approx 0$  and  $CL_{13} \approx \pi/4$ .

(b) The split line defined by C4 and C5 ( $CC_{45} \approx \pi/2$ ,  $CL_{45} \approx \pi/2$ ) is a reasonable scenario; the split line between C5 and C6 ( $CC_{56} \approx 0$ ,  $CL_{56} \approx 3\pi/2$ ) is a possible scenario, but less likely.

**Figure 4.21.** Illustrative examples of Concavity-Concavity (CC) and Concavity-Line (CL) alignment metrics.

the unit vector associated with  $C_j$  as  $\mathbf{v}_j$ , we denote the alignment of the concavities as the radial value

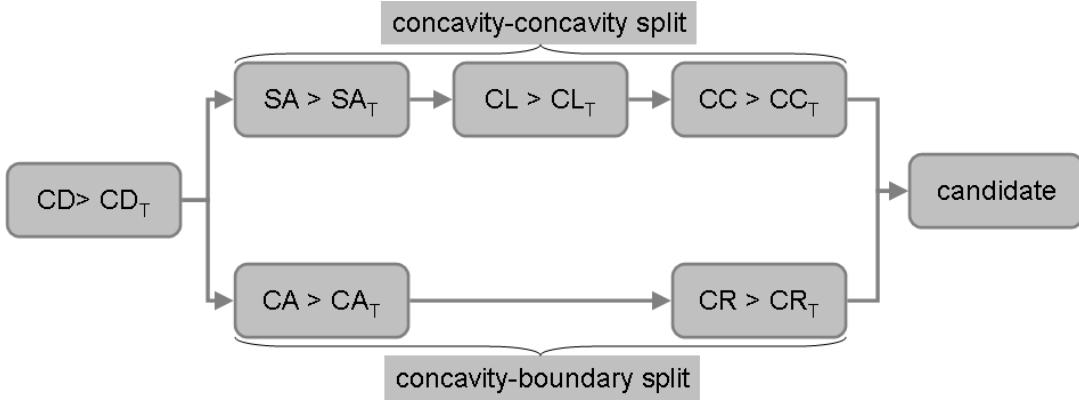
$$CC_{ij} = \cos^{-1}(-\mathbf{v}_i \cdot \mathbf{v}_j) = \pi - \cos^{-1}(\mathbf{v}_i \cdot \mathbf{v}_j) \quad (4.17)$$

$0 \leq CC_{ij} \leq \pi$ , with smaller values indicating a better concavity-concavity alignment. This can be further understood by the intuitive consideration that an ideal split line would occur for two concavities with oppositely aligned  $\mathbf{v}_i$  and  $\mathbf{v}_j$ , meaning that  $-\mathbf{v}_i \cdot \mathbf{v}_j = 1$  and thus  $CC_{ij} = 0$ . Conversely, a poor split line would occur between two parallel concavities, in which case  $-\mathbf{v}_i \cdot \mathbf{v}_j = -1$  and  $CC_{ij} = \pi$ . An illustration of this can be seen in Figure 4.21 (a).

- **Concavity-line alignment (CL):** This metric measures the alignment of the candidate split line with respect to the alignment of both endpoint concavities. Again using the unit vectors  $\mathbf{v}_i$  and  $\mathbf{v}_j$  as defined above, as well as the unit vector  $\mathbf{u}_{ij}$  defining the direction of the candidate split line, the concavity-line alignment is defined as the radial value

$$CL_{ij} = \max(\cos^{-1}(\mathbf{v}_i \cdot \mathbf{u}_{ij}), \cos^{-1}(\mathbf{v}_j \cdot (-\mathbf{u}_{ij}))) \quad (4.18)$$

$0 \leq CL_{ij} \leq \pi$ , again with smaller values indicating a better alignment. In practice,



**Figure 4.22.** Flowchart of the candidate split line selection process.

however, the maximum value of  $CL_{ij}$  is expected to be not much larger than  $\pi/2$ ; refer to Figure 4.21 (b) for illustration.

- **Concavity angle (CA):** Denoting  $\mathbf{w}_{i1}$  and  $\mathbf{w}_{i2}$  as the unit vectors between concavity pixel  $C_i$  and the endpoints of chord  $K_i$ , the concavity angle is defined as the angle between these two unit vectors:

$$CA_i = \cos^{-1}(\mathbf{w}_{i1} \cdot \mathbf{w}_{i2}) \quad (4.19)$$

A small concavity angle indicates a sharp concavity which is intuitively the best choice to begin a split line.

- **Concavity ratio (CR):** The relative depth of the concavity pixel  $C_i$  is computed by normalizing by the largest concavity depth ( $CD_m$ ) in the object:

$$CR_i = \frac{CD_i}{CD_m} \quad (4.20)$$

Once concavities exceeding a concavity depth threshold  $CD_T$  are found,  $SA$ ,  $CC$ , and  $CL$  are used to determine valid candidates for concavity-concavity split lines.  $CA$  and  $CR$  are used to determine valid candidate concavity-boundary split lines, in the case that there is either no valid concavity-concavity split line, or only one concavity exceeding the threshold  $CD_T$  exists. The algorithm for finding candidate split lines is summarized in the flowchart in Figure 4.22.

Choice of the size thresholds  $CD_T$ ,  $SA_T$ ,  $CL_T$ ,  $CC_T$ ,  $CA_T$ , and  $CR_T$  were determined empirically in [140]. We determine our thresholds based on training data, which will be discussed in a later section.

For cases where a choice among multiple candidate split lines must be decided, Kumar et al. introduce another metric  $\chi$  which is reduced to a linear decision boundary between

concavity depths  $C_i$  and  $C_j$  and the distance between them,  $d(C_i, C_j)$  [140]. Since this introduces two more free parameters (the slope and y-intercept of the decision line), in addition to the six thresholds, we instead propose the following metric to determine the “best” split line.

For concavity-concavity split lines, choose the split line candidate that maximizes

$$D_{cc} = \frac{CD_1 + CD_2}{2CD_T} + \frac{SA}{SA_T} + \left(1 - \frac{CL}{CL_T}\right) + \left(1 - \frac{CC}{CC_T}\right) \quad (4.21)$$

where  $CD_1$  and  $CD_2$  are the concavity depths of the two endpoints, and the remaining terms are as defined previously. For concavity-boundary splits, maximize

$$D_{cb} = \frac{CD}{CD_T} + \left(1 - \frac{CA}{CA_T}\right) + \frac{CR}{CR_T} \quad (4.22)$$

Since these measures are computed for valid split lines, all the various parameters have satisfied the corresponding thresholds. Thus normalizing by the threshold values will yield a value in the range  $[0, 1]$ . For the parameters where a smaller value is better (i.e., the parameters that must satisfy a ‘<’ relationship with the threshold), the normalized value is subtracted from 1. This allows us to directly choose the largest  $D_{cc}$  or  $D_{cb}$  from the candidate split lines.

#### 4.5.2 Ground Truth Markup

We extend the interactive system described in Section 4.3.5 to include the ability to delineate ground truth markup for use in concavity-based segmentation. After the object-level delineation of Section 4.3.5, the user is presented with both the original image with an overlay of the individual object perimeters (for reference) as well as with a new window which will display individual connected components of the binary total markup (the Boolean OR of the individual object markups).

The total markup image is used to elicit user input on appropriate split lines for the data in question. The user is sequentially presented with individual connected components of the total markup, with instructions to define appropriate split lines. For reference and disambiguation, the user may still access the original color/grayscale image with an overlay of the object perimeters. As new split lines are specified, they are assigned consecutive numbers and added to the total markup image. This new image will be saved as ground truth for determination of the necessary concavity-based segmentation thresholds. A screen capture illustrating this markup process is shown in Figure 4.23.

### 4.5.3 Choice of Thresholds

We choose appropriate thresholds for the six concavity-based segmentation parameters using the ground truth markup described in the previous section. From this concavity markup, we can easily locate valid split lines and the concavity(ies) associated with them. These will yield the distributions of the various parameters for valid split lines. In order to develop a threshold to distinguish valid from invalid split line candidates, we must also compute the distributions of the six parameters for non split lines.

To this end, for each connected component in the concavity markup, CD, CR, and CA are computed for each concavity in the object that is not associated with a split line in the markup. Then SA, CL, and CC are computed for the split line between the current concavity and all other concavities in the object. Thus the distribution of all six parameters is obtained for a variety of invalid split lines.

For an object with  $n$  concavities (equivalently,  $n$  convex hull chords),  $c$  valid concavity-concavity splits, and  $b$  valid concavity-boundary splits, there will be  $2c + b$  representative points in the distribution of valid CD values;  $c$  points in the distribution of valid SA, CL, and CC values; and  $b$  points in the distribution of valid CA and CR values. For the distribution of invalid points, there will be  $n - 2c - b$  for CD, CA, and CR and  $O(n^3)$ <sup>3</sup> for SA, CL, and CC.

Since all split lines must satisfy a particular concavity depth, all points not satisfying  $CD_T$  are eliminated prior to determining SA, CL, and CC (or CA and CR). No further points are eliminated during the threshold calculations since the subsequent parameters are considered equal in merit, and the order in which they are tested is arbitrary.

The threshold is chosen by exhaustive search over the range of parameter values, and choosing the threshold that maximizes the performance:

$$P = w_1 N_{C1} - N_{I2} \quad (4.23)$$

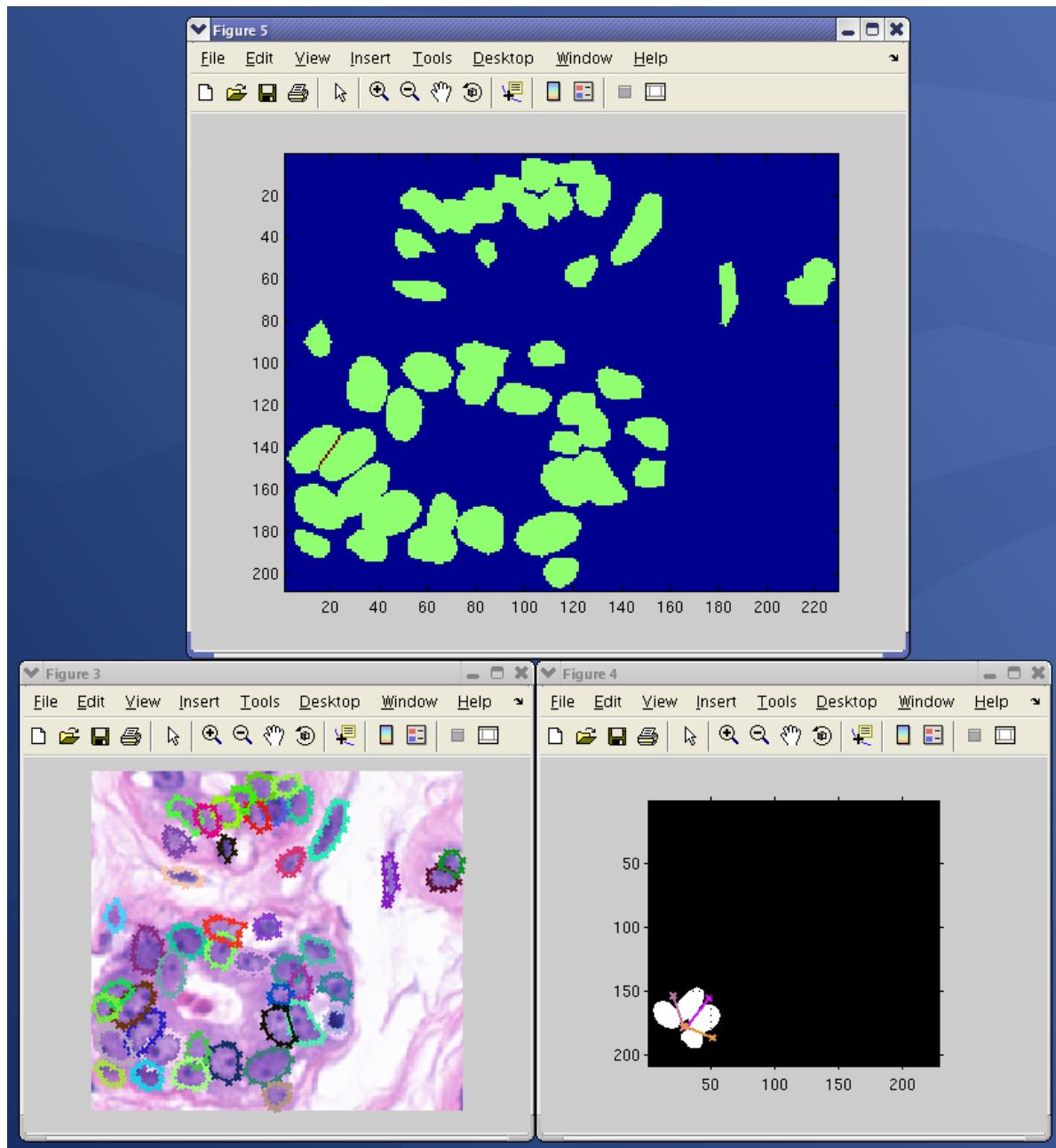
where  $N_{C1}$  is the number of correctly classified class 1 samples, and  $N_{I2}$  is the number of incorrectly classified class 2 samples. The weight  $w_1$  is chosen such that

$$w_1 N_1 = N_2 \quad (4.24)$$

where  $N_1$  and  $N_2$  are the total number of class 1 and class 2 samples, respectively; this takes care of unbalanced training. Note that the performance in Equation (4.23)

---

<sup>3</sup>The exact number of SA, CL, and CC values can be shown to be  $\frac{1}{3} [n^3 - 3(2c + b)n^2 + (3(2c + b) - 1)n + (2c + b)((2c + b)^2 + 1)]$ . This result is obtained from the observation that there will be  $(n - 2c - b)(n - 2c - b - 1) + (n - 2c - b - 1)(n - 2c - b - 2) + \dots + (2)(1)$  invalid split lines. Letting  $m = n - 2c - b - 1$ , this sum can be represented as  $\sum_{k=1}^m (k + 1)k$ , for which the closed form solution  $\frac{1}{3}m(m + 1)(m + 2)$  exists. Substitution and algebraic manipulation yields the equation above.



**Figure 4.23.** Illustration of the Kumar ground truth markup process. Clockwise from the top window: the total markup image with user specified split lines, the current connected component where the user marks the split lines, and the original image with object perimeters for reference.

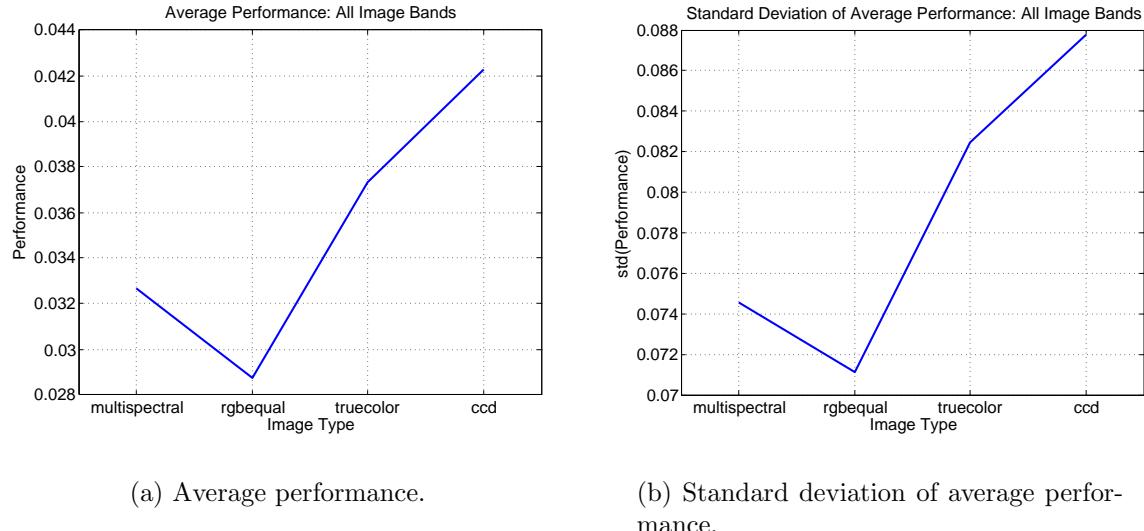
**Table 4.5.** In-sample thresholds for concavity-based segmentation. The Inequality column is a reminder of the direction of the decision for each parameter. The four thresholds presented ( $t_{IS}$ ,  $t_{benign}$ ,  $t_{malignant}$ ,  $t_{kumar}$ ) are the thresholds determined using all in-sample images, just benign in-sample images, just malignant in-sample images, and, as reference, the parameters used in the paper by Kumar et al. [140], respectively.

Parameter	Inequality	$t_{IS}$	$t_{benign}$	$t_{malignant}$	$t_{kumar}$
<b>CD</b>	>	0.95	0.95	1.08	3.00
<b>SA</b>	>	0.18	0.24	0.18	0.12
<b>CL</b>	<	111.60°	109.62°	111.60°	70.00°
<b>CC</b>	<	145.80°	145.80°	142.20°	105.00°
<b>CA</b>	<	120.12°	120.12°	118.90°	90.00°
<b>CR</b>	>	0.09	0.09	0.11	0.17

is used only to find the threshold for which the term is maximized, not for any further comparison; accordingly, we do not concern ourselves with any normalization factors for  $P$ .

Using the procedures outlined above, we obtain the thresholds shown in Table 4.5. These were obtained using the in-sample images as training; these are the same in-sample images used for training in Chapter 2. As a point of reference, Table 4.5 also displays the thresholds determined using benign and malignant in-sample images, as well as the parameters chosen empirically in [140]. A few comments are in order:

- First, while the thresholds from [140] ( $t_{kumar}$ ) are presented for comparison, these thresholds were chosen for a dataset consisting of spore specimens, not nuclei in histopathology imagery.
- Second, the concavity depth is relatively small (approximately a pixel); this would seem to indicate that most concavities in the images are considered indicative of a split line.
- Third, for concavity-concavity split lines, we note that the threshold for concavity-concavity alignment ( $CC_T$ ) is relatively large, only 35° less than the maximum value of 180°, indicating that this parameter is not as discriminating as other parameters. Additionally, the value of  $CL_T$  is not much larger than 90°, as expected per previous discussion of the CL metric.
- Fourth, for concavity-boundary split lines, the threshold for CR is very small, while CA is a middling value. It appears from these data, that concavity-boundary splits are (not surprisingly) more difficult to determine. One should keep in mind, however, that a concavity-boundary split candidate is considered only after all concavity-concavity split candidates have been eliminated.



(a) Average performance.

(b) Standard deviation of average performance.

**Figure 4.24.** Concavity-based segmentation (Kumar) performance.

- Lastly, we note the very similar threshold results between the benign and malignant in-sample images, indicating that these parameters are general between the two classes.

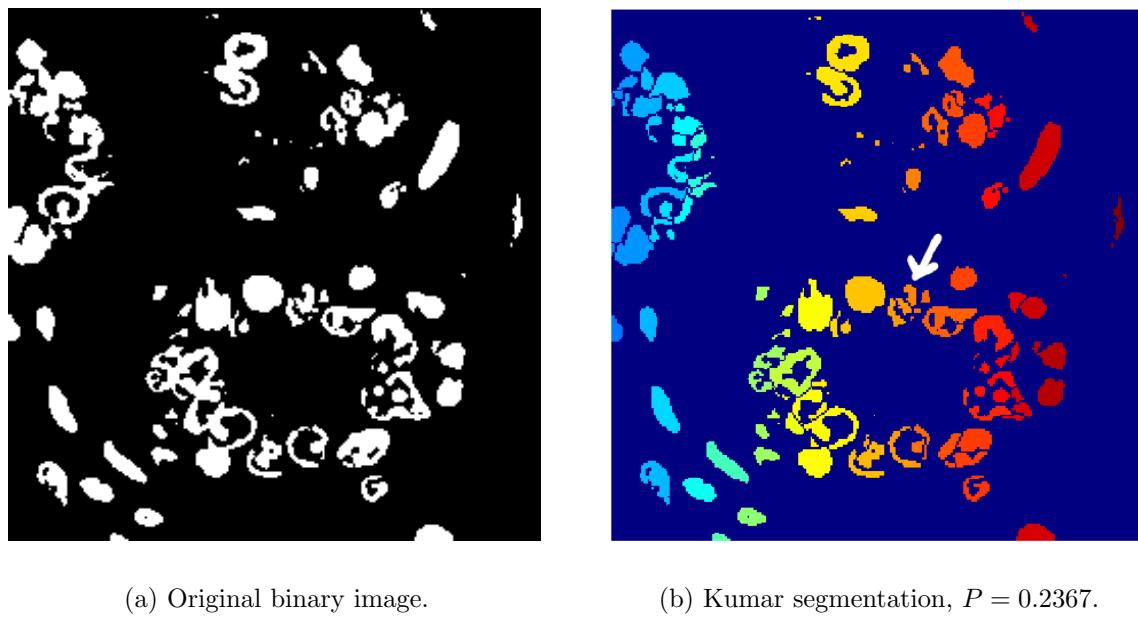
#### 4.5.4 Results

We discuss here the results using the concavity-based segmentation method of Kumar et al. [140] on binary pixel-level classifications. We will subsequently refer to these segmentations as the Kumar method. Average performance of the Kumar method is presented in Figure 4.24 for the different image types, as well as the Wilcoxon p-values in Table 4.6. This method has very poor performance, especially in comparison to the previously presented methods. Referring to the example segmentation shown in Figure 4.25, we conclude that this decrease in performance is due to three main factors:

1. Decreased performance of the GT term (term 1): When presented with binary nuclei with some pixels missing (due to imperfections in the classification process), the Kumar method tends to split these nuclei into multiple sections, due to the presence of many artificial concavities. As an example in Figure 4.25, look for the nucleus indicated by the arrow which has been split into multiple parts.
2. Decreased performance of the PM term (term 2): The WSCDT, WSHmin, and WSGran methods presented in the previous section all rely on the watershed transform of the Euclidean distance transform of the complemented binary image for

**Table 4.6.** Wilcoxon paired signed-rank test p-values for performances of multispectral versus RGB imagery for Kumar nuclear segmentation. Starred entries correspond to statistical significance at the p-value of 0.05. Using the Bonferroni correction for these 15 tests, we look for the corrected p-value of  $0.05/15 \approx 0.0033$ , which are presented as bold entries in the table. Note that all bold entries also satisfy the less conservative p-value of 0.05 although they are not explicitly starred.

Classifier	Image—multi vs.		
	rgbequal	truecolor	ccd
ML	7.8e-3*	<b>4.9e-4</b>	<b>4.9e-4</b>
MED	5.2e-1	8.8e-1	1.7e-1
SAM	<b>2.4e-4</b>	2.6e-1	9.1e-1
FLDA	<b>2.9e-3</b>	1.5e-1	1.1e-1
AFE	6.8e-3*	5.6e-3*	<b>1.7e-6</b>



**Figure 4.25.** Example Kumar segmentation, where each color indicates a different segmented region. Note that this method tends to oversegment, particularly when the underlying pixel-level segmentation is imperfect; refer to the nucleus indicated by the arrow.

the underlying segmentation. This results in an “interpolation” of sorts in regards to regions of missing pixels close to the object centroid. Thus, the Kumar segmentation is penalized for missing these pixels; this is most easily seen by comparing the WSGran and Kumar segmentation outputs (compare Figure 4.17 (b) with Figure 4.25 (b)).

3. Decreased performance of the ER term (term 5): The WSCDT, WSHmin, and WSGran methods all include some inherent filtering of small regions. In the case of WSGran this filtering is directly related to the use of granulometries to determine markers. For WSHmin this filtering of small objects is a side effect of the suppression of shallow minima; small objects will have small Euclidean distances which results in shallow minima. The WSCDT eliminates small regions due to a subtlety in the discrete nature of the algorithm as implemented in Matlab: since each watershed line must be an 8-connected line, any region with too few pixels for both a watershed line and an interior basin will be discounted as part of the background (or the interior of the larger enclosing basin). It is this term that appears to be the largest detriment to the overall performance of the Kumar algorithm (judged by looking at the average difference in the terms between Kumar and WSCDT for the same images).

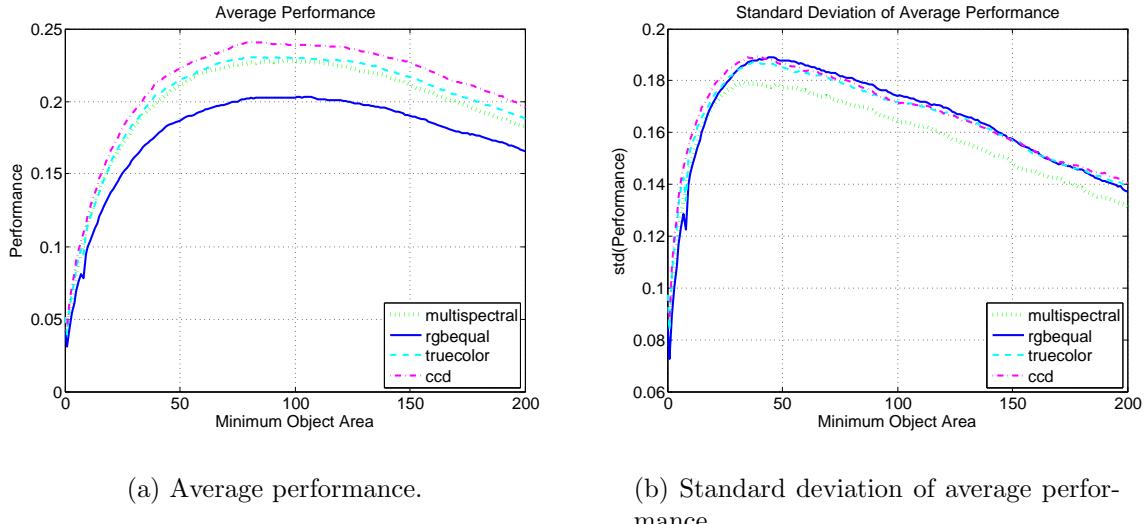
It is these observations that motivate the use of the watershed transform and area-based filtering techniques using the output of the Kumar segmentation as markers for individual nuclei.

#### 4.5.5 Concavity-Based Segmentation as Markers for Watershed Segmentation (WSCM)

Now we use the results of the concavity-based segmentation discussed in the previous section as markers for a subsequent watershed segmentation. This is motivated by the observation that while the concavity-based segmentation is designed to determine the best places to split a clump of objects, the watershed transform is better suited to determine the shape of the resulting splits. For this segmentation, we use the centroids of the regions found by the Kumar method as markers for the foreground (nuclei) objects. An erosion of the background is used as the background markers. Additionally, prior to the watershed transform, all binary regions with an area less than some predetermined threshold are removed.

The plots of average performance of WSCM with respect to the minimum object area are shown in Figure 4.26. The removal of small areas does increase the performance of the WSCM method up to a minimum object area of approximately 100 pixels, after which the performance begins to slowly degrade.

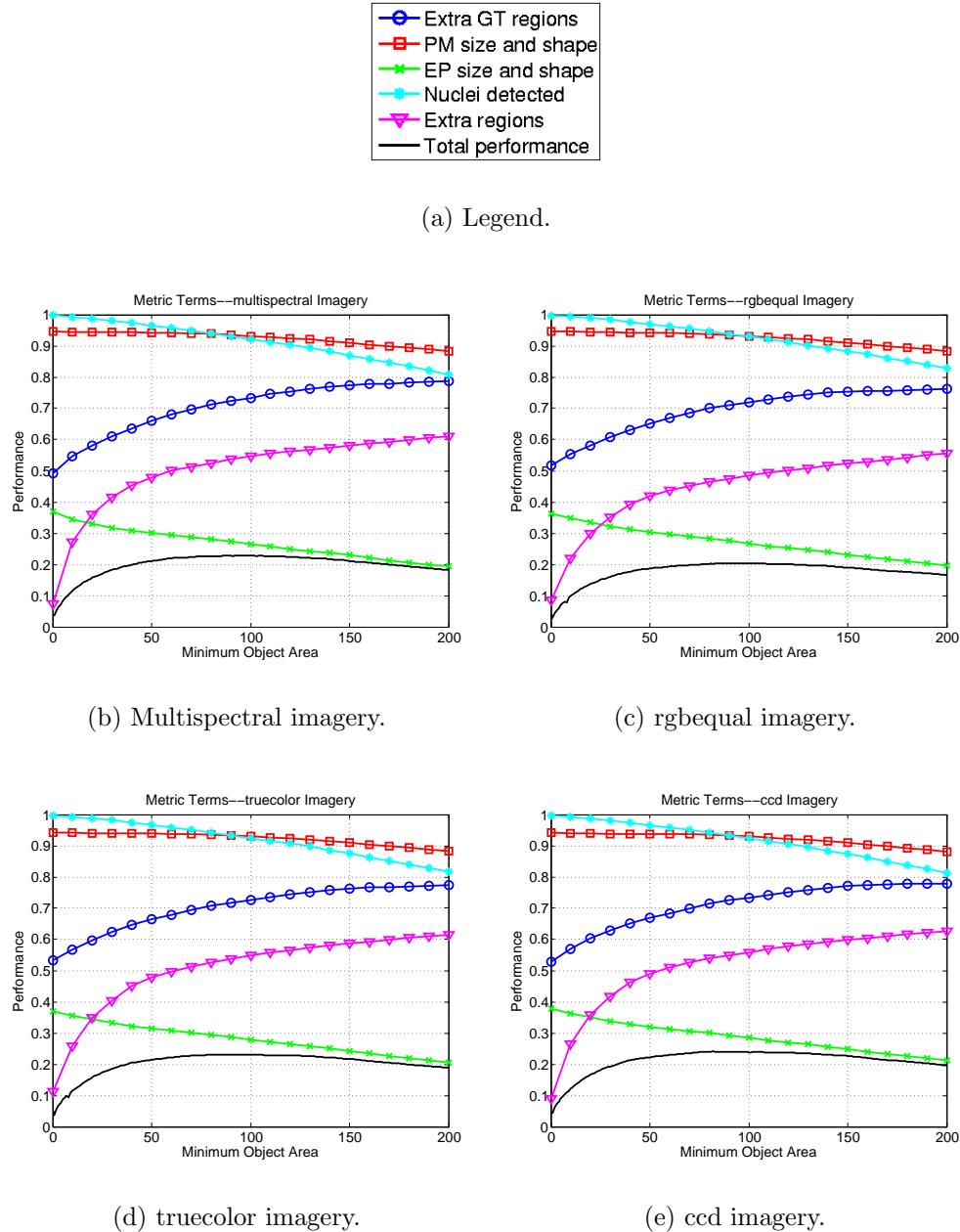
Figure 4.27 shows the performance of individual metric terms for different minimum



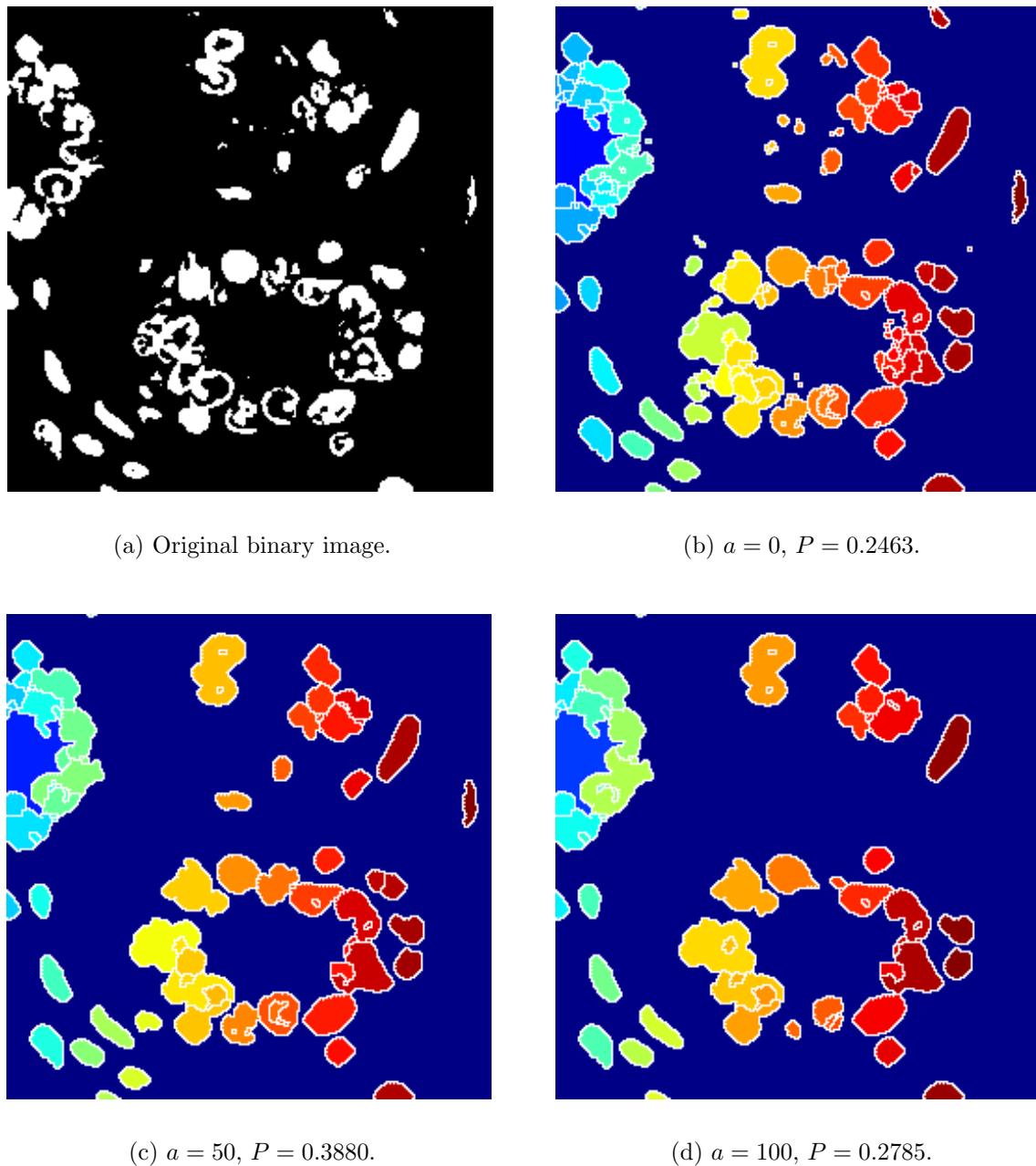
**Figure 4.26.** Concavity-Based Segmentation as Markers for Watershed Segmentation (WSCM) segmentation performance versus minimum object area  $a$ .

object areas. The increase in overall performance as minimum object area increases is due to the increase in performance of terms 1 (extra segmented regions per ground truth region) and 5 (excess segmented regions), where term 5 appears to have the largest effect on performance increase. Simultaneously, terms 2, 3, and 4 (pixels missed, extra pixels, and nuclei detected, respectively) slowly decrease. As terms 1 and 5 level off, the decrease in terms 2, 3, and 4 result in a gradual decrease of the overall performance. This analysis corresponds nicely with our analysis of the major contributing factors to the poorer performance of the Kumar method with respect to the other watershed-based methods.

Figure 4.28 shows example segmentations for different minimum object areas. Qualitatively the overall tendency towards undersegmentation (similar to that seen with WS-Gran) is relatively constant with respect to the minimum object area. This can also be seen in Figure 4.27 in the low performance of term 3 (extra pixels), and the slow decrease in this term with respect to minimum object area. Comparing the WS-Gran outputs (Figure 4.17) to the WSCM outputs, there is a qualitative similarity; indeed, comparison of the individual metric terms for WS-Gran (Figure 4.16) yields similar values and trends to the two segmentations. As reference, the Wilcoxon p-values for the WSCM method, minimum object area  $a = 105$ , are shown in Table 4.7.



**Figure 4.27.** Value of Concavity-Based Segmentation as Markers for Watershed Segmentation (WSCM) performance metric terms versus minimum object area  $a$  for different imagery. It should be noted that all terms plotted here are performance, i.e., one minus the penalty, where the penalties are the terms previously discussed in relation to the segmentation metric. The terms are denoted by a brief description in the charts, but as reference, they are plotted in numerical order, i.e., blue circles are term 1, red squares are term 2, and so forth.



**Figure 4.28.** Example Concavity-Based Segmentation as Markers for Watershed Segmentation (WSCM) segmentations, where each color indicates a different segmented region.

**Table 4.7.** Wilcoxon paired signed-rank test p-values for performances of multispectral versus RGB imagery for Concavity-Based Segmentation as Markers for Watershed Segmentation (WSCM) nuclear segmentation,  $a = 105$ . Starred entries correspond to statistical significance at the p-value of 0.05. Using the Bonferroni correction for these 15 tests, we look for the corrected p-value of  $0.05/15 \approx 0.0033$ , which are presented as bold entries in the table. Note that all bold entries also satisfy the less conservative p-value of 0.05 although they are not explicitly starred.

Classifier	Image—multi vs.		
	rgbequal	truecolor	ccd
<b>ML</b>	4.2e-2*	<b>2.2e-5</b>	<b>5.3e-6</b>
<b>MED</b>	3.7e-3*	1.6e-2*	1.4e-1
<b>SAM</b>	<b>3.2e-6</b>	2.3e-2*	9.3e-2
<b>FLDA</b>	<b>7.9e-7</b>	<b>4.4e-4</b>	<b>5.3e-5</b>
<b>AFE</b>	<b>1.1e-4</b>	<b>1.4e-3</b>	<b>8.0e-7</b>

**Table 4.8.** Average computation time per image, in seconds, on an Intel® Xeon™ 3.6 GHz CPU.

Method	Time (s)
<b>WSCDT</b>	1.3
<b>WSHmin</b>	1.7
<b>WSGran</b>	2.7
<b>WSBlob</b>	64.9
<b>Kumar</b>	330.9
<b>WSCM</b>	354.7

#### 4.5.6 A Note on Computational Efficiency

One of the disadvantages of the concavity-based method is the computation time. Table 4.8 lists the average computation time per image for the six methods thus far discussed (WSCDT, WSHmin, WSGran, WSBlob, Kumar, WSCM). The concavity-based methods require approximately 5 times more computation time than even the most computationally intensive (WSBlob) of the other methods, and approximately 123 times more computation than WSGran, WSHmin, and WSCDT.

The WSCDT, WSHmin, and WSGran methods utilize existing MATLAB functions, and require only one pass through the image for the computations. The increase in computation time for the WSBlob method is due to the local maxima search of the blob-detector method. The concavity-based method of Kumar et al. [140], on the other hand, loops through the connected components of the binary image and iteratively calculates the split lines. As a connected component is split, the method will loop through each

**Table 4.9.** Ideal maximum average nuclear segmentation performance ( $P$ )  $\pm$  standard deviation. The total markup from the ground truth is used as the pixel-level classification. The Parameter(s) columns specify the image type and/or segmentation parameter value for which the maximum average performance occurs.

Method	Ideal		Actual	
	P	Parameter(s)	P	Parameter(s)
<b>WSCDT</b>	$0.64 \pm 0.11$	N/A	$0.18 \pm 0.17$	GENIE, ccd
<b>WSHmin</b>	$0.69 \pm 0.10$	$h_{min} = 1$	$0.27 \pm 0.15$	SAM, ccd, $h_{min} = 3$
<b>WSGran</b>	$0.68 \pm 0.10$	$r = 2$	$0.34 \pm 0.17$	GENIE, ccd, $r = 4$
<b>Kumar</b>	$0.71 \pm 0.12$	N/A	$0.07 \pm 0.12$	GENIE, ccd
<b>WSCM</b>	$0.64 \pm 0.12$	N/A	$0.29 \pm 0.17$	GENIE, ccd, $a = 103$
<b>WSGT</b>	$0.69 \pm 0.12$	N/A	N/A	N/A

of these new connected components and compute the split lines. Thus the Kumar and WSCM methods are (not quite linearly) dependent on the number of connected components in the image, with an average of 0.26 and 0.28 seconds computation time per connected component, respectively. The watershed methods, on the other hand, have a very uniform computation time across the various images. Due to the unsatisfactory computation time for the concavity-based methods, and the lack of increase in performance using these methods, we choose to stick with the WSGran and WSHmin segmentations for future research.

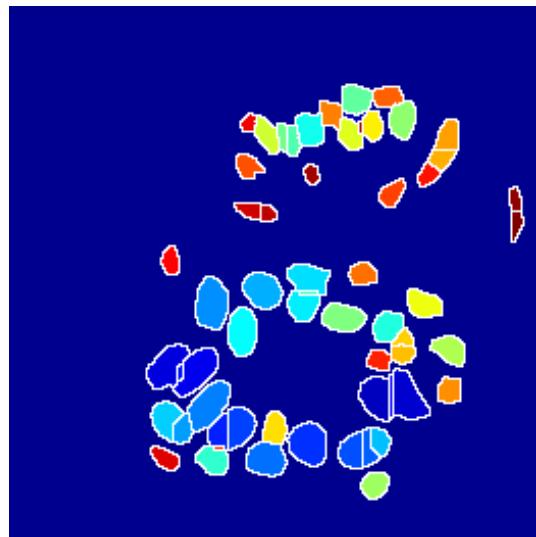
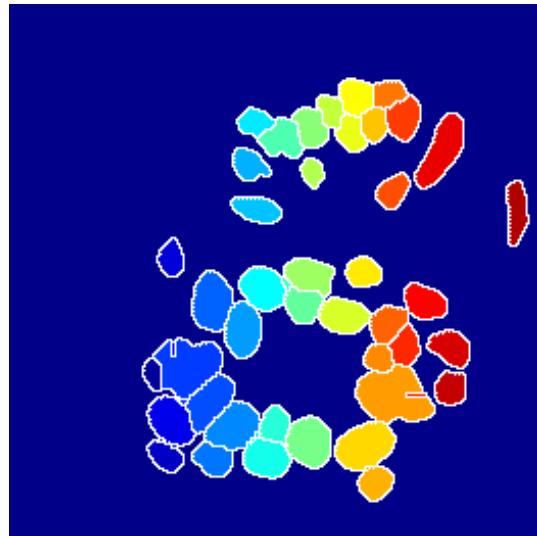
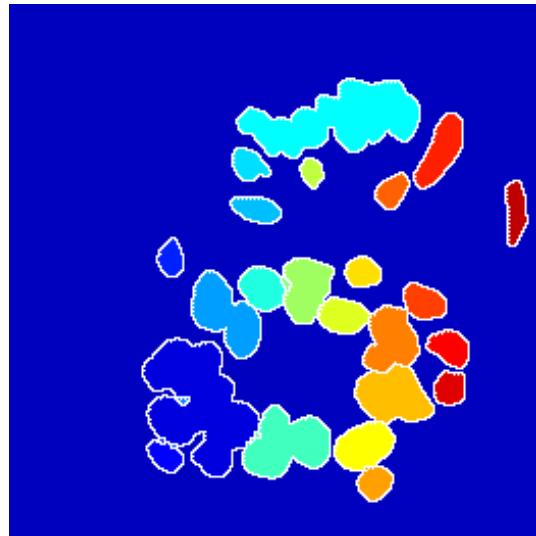
## 4.6 Ideal Maximum Segmentation Performance of Nuclei

We are interested in how the imperfect pixel-level nuclear classifications impact the nuclear segmentations presented in this chapter. To this end, we applied each of our segmentation methods to the entire markup mask stored in the ground truth file. Table 4.9 displays the maximum average performance and standard deviation for an ideal pixel-level input, as well as the maximum performances and standard deviations for the actual segmentations (on real nuclei classifier outputs). Example outputs for each of these ideal cases is shown in Figure 4.29 (WSHmin is not included since it had identical performance to WSCDT for  $h_{min} = 1$ ).

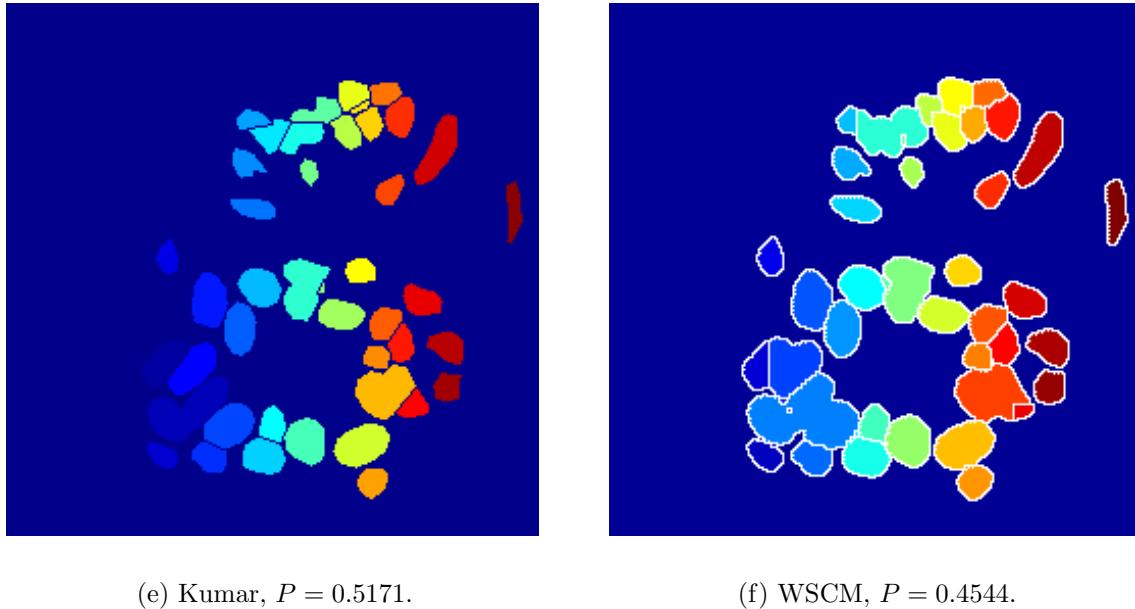
There is surprisingly good performance using the standard WSCDT method on the ground truth image mask. This indicates that the boundaries of correctly defined cell nuclei are smooth enough for a reasonable watershed-based segmentation (see Figure 4.29 (b)). This is further indicated by the performance of the watershed using the ground truth nuclei centers as markers, denoted by WSGT in Table 4.9 and Figure 4.29 (c). The



(a) Original ground truth image.

(b) WSCDT,  $P = 0.4769$ .(c) WSGT,  $P = 0.4150$ .(d) WSGran,  $r = 2$ ,  $P = 0.5872$ .

**Figure 4.29.** Example ideal segmentations, where each color indicates a different segmented region. (continued on next page)



**Figure 4.29.** (Cont.) Example ideal segmentations, where each color indicates a different segmented region.

WSHmin method performs equally well, for small values of  $h_{min}$ , but the performance tapers off quickly as  $h_{min}$  increases beyond a value of 1. The WSGran performance is slightly less than that of WSCDT, and is maximum for a granulometric SE radius of  $r = 2$  (see Figure 4.29 (d)). The Kumar method performed the best for these ideal input images, validating the theory behind that algorithm; an example is shown in Figure 4.29 (e). The WSCM method was applied only for  $a = 0$  since the reasoning behind suppressing small areas had to do with imperfections in the underlying nuclear classifications. The WSCM method has a lower performance than the Kumar method for an ideal binary input (see Figure 4.29 (f)). This indicates that the watershed transform may not be the ideal algorithm for use in segmentation of clumped nuclei; a higher-level shape-based algorithm such as the Kumar method may be more applicable.

We now compare the ideal and actual segmentations based on their performance and the performance of the individual metric terms, summarized in Table 4.10. The major contribution to decreased actual performance for WSCDT appears to be due to term 5, indicating that the actual WSCDT segmentations have more extra regions not associated with ground truth nuclei. The detection rate for nuclei (term 4) is almost the same as in the actual segmentations as in the ideal; thus the degraded performance of the actual segmentations have to do with errors in proper delineation of the individual nuclei, along with the errors associated with extra regions. Similar analysis hold for the comparison of actual to ideal WSHmin segmentations, although the actual WSHmin segmentations

**Table 4.10.** Ideal maximum average nuclear segmentation term performance. The total markup from the ground truth is used as the pixel-level classification.

Method	P	term <sub>1</sub>	term <sub>2</sub>	term <sub>3</sub>	term <sub>4</sub>	term <sub>5</sub>
WSCDT–Ideal	0.64	0.55	0.97	0.77	1.00	0.98
WSCDT–Actual	0.18	0.39	0.85	0.60	0.99	0.43
WSHmin–Ideal	0.63	0.54	0.97	0.77	1.00	0.98
WSHmin–Actual	0.27	0.38	0.72	0.69	0.89	0.79
WSGran–Ideal	0.68	0.99	1.00	0.48	1.00	0.89
WSGran–Actual	0.34	0.90	0.95	0.24	0.96	0.63
Kumar–Ideal	0.71	0.75	0.97	0.69	1.00	1.00
Kumar–Actual	0.07	0.51	0.83	0.41	1.00	0.19
WSCM–Ideal	0.64	0.78	1.00	0.61	1.00	0.89
WSCM–Actual	0.29	0.75	0.94	0.28	0.93	0.64
WSGT	0.63	0.74	1.00	0.64	1.00	0.90

have much better performance for term 5 (the suppression of shallow minima eliminates many of the extra regions). Actual WSGran segmentations have significantly decreased performance for terms 3 and 5 compared to the ideal WSGran segmentations. This indicates that the actual WSGran segmentations are more undersegmented than the ideal, although the ideal are quite undersegmented to begin with (refer to Figure 4.29 (c)). Actual Kumar and WSCM segmentations similarly have lower performance for terms 3 and 5, and Kumar additionally has decreased performance for term 1 (extra watershed regions per ground truth region). As reference, the term performances are also presented for the WSGT example. There is very similar performance to that of the WSCDT, with slightly better performance on term 1 and slightly worse performance on term 3, indicating a tendency towards undersegmentation in WSGT rather than oversegmentation in WSCDT.

A consistent theme throughout is the degradation in performance of terms 3 and 5. This implies that the pixel-level nuclear classifications provide two difficulties not present (or as prevalent) in the ground truth. First, the pixel-level classifications present clumps of nuclei with fewer clues (e.g., concavities) as to the correct split lines. Qualitatively, in many images, there are subtle spectral changes between adjacent nuclei that indicate the presence of a cell membrane and/or cytoplasm. This region is often only on the order of one pixel wide for closely clustered nuclei, and is most likely misclassified as nuclear material by the pixel-level classifiers. Also, many clustered nuclei display only shape-related cues as to the existence of multiple nuclei. Second, the pixel-level classifications yield the presence of extra regions not associated with ground truth nuclei. These are most likely the false alarm components of the classifications.

## 4.7 Segmentation of Cytoplasm and Stroma

### 4.7.1 Voronoi-Based Segmentation of Cytoplasm

Per the work of Honda [131] and Jones et al. [132], we use the centroids of the nuclei as the markers of a Voronoi-based segmentation for cytoplasm. Since we are only interested in the cytoplasm contained within each Voronoi cell, we mask the cytoplasm pixels contained within each Voronoi cell prior to any feature extraction of the cytoplasm objects. The pixel-level classification of cytoplasm is obtained using the cytoplasm classifier from Chapter 3. An example Voronoi-based segmentation of cytoplasm can be seen in Figure 4.30 (b).

### 4.7.2 Connected Components-Based Segmentation of Stroma

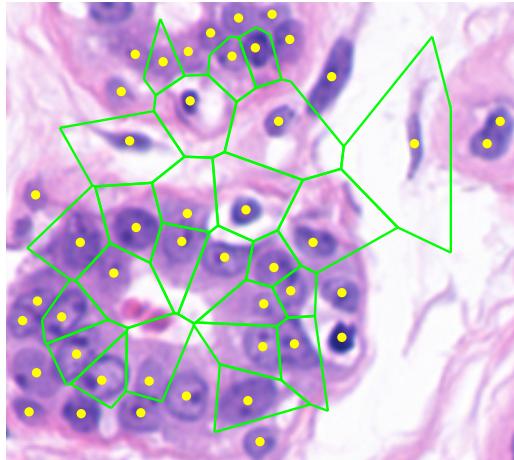
There do not seem to exist any pathology-based rules regarding the discrete nature of stromal components of tissue. That is, the stroma is generally regarded as a single entity, while the spatial distribution especially with respect to cells and other similar metrics may be considered important features. The pixel-level classification of stroma is obtained using the stroma classifier from Chapter 3. Then, we treat each connected component of stroma as an object. An example of such a stromal segmentation can be seen in Figure 4.30 (c).

### 4.7.3 Voronoi-Based Segmentation of Cytoplasm and Stroma

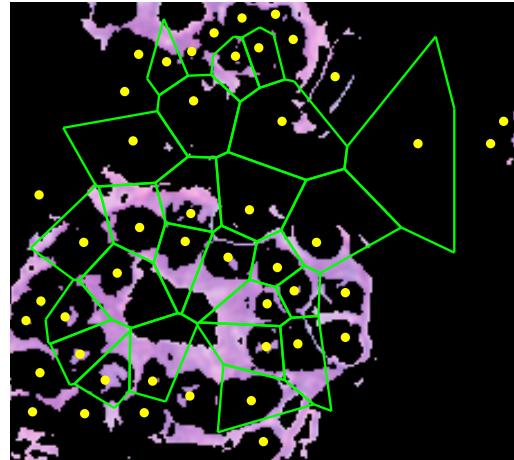
As discussed in Chapter 3, the characterization of features of the combined cytoplasm-stroma class may provide useful information for further analysis. Thus, we also use the Voronoi segmentation for the combined cytoplasm and stroma class. We are, therefore, associating both cytoplasm and stroma pixels within the Voronoi cell of a nucleus as “belonging” to that biological cell. An example of this segmentation can be seen in Figure 4.30 (d).

## 4.8 Summary

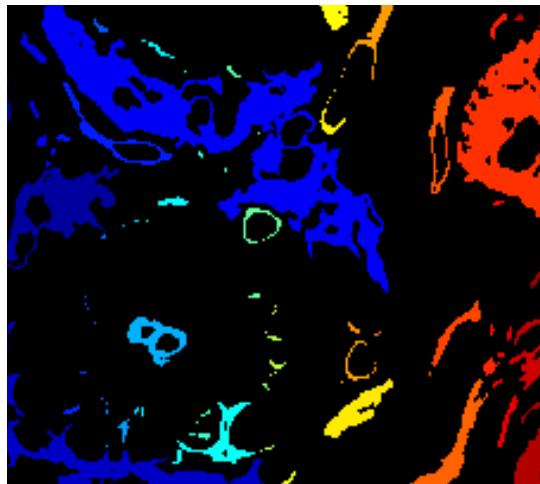
In this chapter, we have presented our segmentation metric for cell nuclei as well as the application of that metric to several different nuclei segmentations. We have also briefly presented methods for segmentation of cytoplasm and stroma.



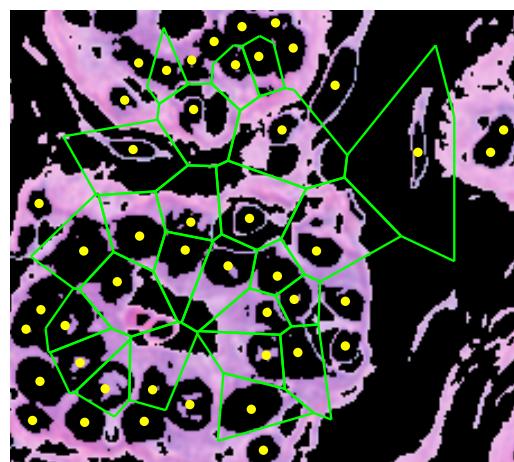
(a) Original image with GT nuclei centroids used to develop the Voronoi tessellation.



(b) Voronoi segmentation of the cytoplasm class. Each nucleus can be associated with the cytoplasm and features of the cytoplasm within the Voronoi cell.



(c) 8-connected component segmentation of the stroma class, where each color indicates a different segmented region.



(d) Voronoi segmentation of the combined cytoplasm and stroma class.

**Figure 4.30.** Example cytoplasm and stroma segmentations.

### 4.8.1 Segmentation Evaluation

We have shown this segmentation metric and its constituent terms to correspond well with the qualitative observations of segmentation accuracy, including the general tendency of an algorithm to over- or under-segment an image. This metric also allows for a direct quantitative comparison between the outputs of different segmentation algorithms. While the metric defines a single performance, we have shown the usefulness of observing the performance of the individual metric terms.

We have also discussed a new method for specification of ground truth for this object-level segmentation problem. This involves not only the delineation of cell nuclei within an approximate truth window, but also the marking of non-delineated objects within the truth window. This allows us to focus our segmentation evaluation on only those objects that were delineated by the user.

In comparison to other work in segmentation evaluation, our metric does not require the computation of region or boundary correspondences. While this metric was originally conceived and developed to match human intuition about a “good” nuclear segmentation, the metric has many similarities to the 5-scenario metric of Hoover et al. [118].

### 4.8.2 Nuclear Segmentation

We have presented several methods for segmentation of nuclei given the pixel-level classification of nuclear material. We used the watershed transform as the basis for most of our segmentations as this method is known to provide a good match to object shape. The other segmentation method that we presented used the concavities of the object boundary to search for split lines. We constructed training data and conducted an exhaustive search for the best threshold parameters for this rule-based method, as well as defining a new criterion for the “best” split line among several possibilities. We also developed a method that combines the advantages of both the concavity-based and watershed-based methods.

All of our nuclei segmentations have relatively poor performance, and a very high standard deviation, indicating the complete failure of decent segmentation for many images. We have shown, however, that these poor performances are due in large part to errors in the underlying pixel-level nuclear classification. While there is obviously room for improvement in the nuclei segmentation, we are more interested here in a proof-of-concept of our object-level classification and segmentation framework. We will, therefore, use the best of our segmentation outputs as input to the further steps in our scheme, and address some of the shortcomings of the first-level segmentations.

Additionally, we have continued with analysis of multispectral versus RGB imagery, showing that for nuclear segmentation RGB imagery performs better. The statistical significance of this performance difference was verified using Wilcoxon signed rank tests.

Before moving on to the future directions for this research, we would like to note the difficulty of this nuclear segmentation problem. While many researchers have addressed it to some extent, the results for H&E stained histopathology imagery have generally been unsatisfactory, especially for automated systems that do not invoke user input for defining regions of clumped nuclei (e.g., [17]).

## 4.9 Future Directions

### 4.9.1 Segmentation Evaluation

It would be valuable to perform a human studies experiment to correlate our segmentation metric with human perception of the segmentation accuracy, similar to the analysis in [120]. Our development of the metric was based on human intuition regarding “good” nuclear segmentations, but it would be good to statistically quantify the similarities between the metric rankings and human rankings. This is particularly important in our application domain, as clinical diagnosis is based on the perception of this imagery by the human pathologist.

Similarly, it may be important to quantify the precision/variability of our metric. This could entail a comprehensive analysis of the metric performance for a variety of (possibly subtly) different segmentations. It is desired that the variation in our metric would be directly proportional to the variation in segmentation output. We have already presented an analysis of the variation of our segmentations and the metric output over our image set, which is an equally important quantification of variability.

Another direction for future research involves the possible use of this metric in other application domains. While the specific terms were developed with the segmentation of cell nuclei in mind, it is reasonable to expect that similar metrics would be directly applicable to another problem, e.g., landcover classification for satellite imagery.

The approach to segmentation evaluation in [127] is very intriguing, particularly the ability to perform ROC-like analysis over large dimensions of metrics. The incorporation of the five terms of our segmentation metric may benefit from a more formal aggregation and analysis. This would also allow the easier addition of further metrics, such as computational efficiency.

The concept of Ultimate Measurement Accuracy, presented in [122], is quite synergistic with our proposed research focus. For each additional level in our analysis, we look back to the previous steps and try to quantify and understand the effects that errors in these initial analyses may have on the current higher-level analysis.

### 4.9.2 Nuclear Segmentation

While our nuclear segmentation results were not stellar, we have shown that a great deal of the degradation in performance is associated with an imperfect pixel-level nuclear classification. This highlights the need for feedback in our hierarchical analysis, whereby it is the accuracy of the higher-level analysis that is of ultimate importance; this is very similar to the theory behind the UMA method for segmentation evaluation [122].

Further work on nuclear segmentation will most likely be very domain specific. While we have avoided any segmentations that rely on assumptions of the underlying object shape, e.g., ellipse fitting, such algorithms could be used to compute markers for subsequent segmentation. The use of concavity-based markers for a watershed segmentation actually decreased the performance, so care should be taken when applying this approach.

The best of the nuclear segmentation methods for ideal input is the Kumar concavity-based algorithm. The poor performance of actual imagery with this algorithm is, in part, due to the presence of many artifactual concavities and holes from imperfect pixel-level classification. It would be interesting to quantify any differences in the characteristics of “true” concavities and the artifactual ones and similarly for holes. A better discrimination of “true” concavities may allow for the actual Kumar performance to approach that of the ideal.

Most of the segmentation methods presented have a tendency to undersegment, even for ideal pixel-level input. There may be ways to combine the results of an undersegmented and oversegmented output that could provide a nice separation of individual nuclei. This would be a region merging algorithm, taking into account *a priori* knowledge in the form of the undersegmented output as well as other information (e.g., concavities, shape).



# Chapter 5

## Feature Extraction and Selection

The extraction and use of relevant image features for automated analysis of medical imagery, particularly cancer imagery, is a topic of great interest. Research on useful features for cancer classification and diagnosis has often been approached by the definition of features flagged by clinicians as particularly important features for the diagnosis process. The vast majority of these features are nuclear features. We propose the use of a comprehensive set of both object-level and spatial-relation features for the classification and characterization of histopathology imagery. We apply feature extraction and selection to nuclei, cytoplasm, and stromal entities within the imagery and show very good classification using even solely cytoplasm and stromal information. Additionally, we explore the use of imperfectly segmented objects within this classification framework and show that this feature extraction and selection approach is capable of utilizing such imperfect objects and yielding classification performances close to that of perfectly segmented objects. The use of an automated feature extraction and selection scheme for classification of imperfectly segmented nuclei has not been studied previously.

### 5.1 Motivation

Feature selection is a means to select the relevant and important features from a large set of features, many of which may be redundant, irrelevant, or not particularly useful (and possibly detrimental) to the classification performance. This is an increasingly important area of research now that automated quantitative image analysis techniques are becoming more mainstream, as it is common practice to extract as many features as possible for future analysis. While humans have innate abilities to process and understand imagery, they do not tend to excel at explaining *how* they reach their decisions. As such, large feature sets are generated in the hopes that some subset of features incorporates the information the human expert is using for analysis.

## 5.2 Related Work

The extraction and use of relevant image features for automated analysis of medical imagery, particularly cancer imagery, is a topic of great interest. Research on useful features for cancer classification and diagnosis has often been approached by the definition of features flagged by clinicians as particularly important features for the diagnosis process. Starting in the late 1990's, this research has proceeded on two fronts: the recognition of important image features that are discretized manually by the expert (e.g., "degree of hyperplasia") [21, 130, 146–148], and features extracted in an automated fashion [15, 16, 44, 46, 149, 150]. These works consider a few (tens of) features and analyze the utility of those features and subsets of features for a specific analysis such as detection or typing of a cancerous condition.

At the same time, effort was put into the use of a more comprehensive feature set with automated feature selection methods to narrow down the important features. While there were definite biological motivations behind the extraction of these features, the analysis of biological significance of the feature subsets is more often left to a later analysis as compared to the methods above. Notable research here includes the CytoSavant system (OncoMetrics, Inc.; Vancouver, BC, Canada) proposed in [151] and [152], used in [8] and [153]; the work at the University of Antwerp, Belgium [18, 20, 45]; and the work presented by Brewer et al. in [40]. Additionally, Rodenacker and Bengtsson compiled a list of features for use in cytometry analysis [154].

The vast majority of these features are nuclear features. No research has explicitly addressed the use of cytoplasm, stromal, or other histologic features in an automated system, although several researchers have hinted at the possible benefit of such analysis [90–93]. Additionally, while many of these methods rely on semi-automated nuclear segmentation (assuring a reasonable level of accuracy in nuclear delineation), some require the manual delineation of nuclei, e.g., [16, 18, 20, 45, 150, 153]. The use of an automated feature extraction and selection scheme for classification of imperfectly segmented nuclei has not been studied.

The use of spatial-relation features for quantifying cellular arrangement was proposed in the early 1990's [155–157], but didn't find application to clinical imagery until recently [20, 22, 35, 41]. In particular, Weyn et al. in [20] and [41] use Voronoi tessellations, Gabriel's graphs, and minimum spanning trees to quantify the spatial arrangement of cell nuclei in lung cancer. Weyn et al. in [22] use similar features to quantify the spatial arrangement of vessels in several different types of cancers. Demir et al. in [35] form a weighted acyclical graph using cell nuclei clusters as the nodes and Euclidean distance between nodes to weight the graph edges. Overall, however, the use of spatial arrangement of histologic entities has been relatively untouched, especially in comparison to the wealth of research on nuclear features that has occurred during the same timeframe.

We propose the use of a comprehensive set of both object-level and spatial-relation

features for the classification and characterization of histopathology imagery. It is our hypothesis that a feature extraction and selection process may be performed on imperfectly segmented histologic entities and still yield similar classification performance as compared to feature extraction and selection performed on well-segmented entities. In this chapter we define all the object-level features extracted from the imagery in Section 5.3 and the spatial-relation features in Section 5.4, including a discussion of the implementation of region-based spatial-relation features (Section 5.5). Feature selection is performed and results are presented for object- and image-level classification in Section 5.6. Discussion of the best feature subsets is presented in Section 5.7. A summary and future work appears in Sections 5.8 and 5.9.

## 5.3 Object-Level Feature Definitions

The object-level features can be categorized as belonging to one of four categories: size and shape, radiometric and densitometric, texture, and chromatin-specific. While the radiometric and densitometric, texture, and chromatin-specific features could be considered low-level features (i.e., generally extracted from a local neighborhood), they are included here since we are extracting each of them from an object region. The size and shape metrics are true object-level metrics; this distinction will become important in Chapter 6. In the feature descriptions to follow, fractal dimension (FD) related features are discussed separately, although each of the FD features can be attributed to one of the four previously mentioned categories. A summary of the object-level features is listed in Table 5.1; the feature definitions (Sections 5.3.1–5.3.5) may be skipped if the details are not of interest to the reader. In the following, features denoted by the superscript  $\dagger$  are features included as is from the original Matlab *regionprops* function.

### 5.3.1 Size and Shape Features

- **Area $^\dagger$ :** The total number of pixels in the object [8, 16, 18, 20, 21, 27, 31, 44–46, 65, 130, 146–149, 151, 152, 154].

$$Area = \sum_n \sum_m \Omega(n, m) \quad (5.1)$$

where  $\Omega(n, m)$  is the object mask consisting of ones within the object and zeros elsewhere.

- **Elliptical Features:** Computed for the ellipse with the same second moments as the object.

- **Major and Minor Axis Length<sup>†</sup>:** [18, 20, 27, 45, 46, 154]

$$majlen = 2\sqrt{2} \cdot \sqrt{m_{xx} + m_{yy} + \sqrt{(m_{xx} - m_{yy})^2 + 4m_{xy}^2}} \quad (5.2)$$

$$minlen = 2\sqrt{2} \cdot \sqrt{m_{xx} + m_{yy} - \sqrt{(m_{xx} - m_{yy})^2 + 4m_{xy}^2}} \quad (5.3)$$

where

$$\begin{aligned} m_{xx} &= \frac{1}{M} \sum_{i=1}^M (x_i - \bar{x})^2 + \frac{1}{12} \\ m_{yy} &= \frac{1}{M} \sum_{i=1}^M (y_i - \bar{y})^2 + \frac{1}{12} \\ m_{xy} &= \frac{1}{M} \sum_{i=1}^M (x_i - \bar{x})(y_i - \bar{y}) \end{aligned} \quad (5.4)$$

and  $\bar{x}$  and  $\bar{y}$  are the x- and y-centroids, respectively, of the object:

$$\bar{x} = \frac{\sum_n \sum_m n\Omega(n, m)}{Area} \quad (5.5)$$

$$\bar{y} = \frac{\sum_n \sum_m m\Omega(n, m)}{Area} \quad (5.6)$$

and  $M$  is the total number of pixels in the object.

- **Eccentricity<sup>†</sup>:** A measure of how much a conic section deviates from circular [8, 16, 46, 65, 130, 149, 152, 154]. A circle has eccentricity 0, a line (i.e., a degenerate ellipse) an eccentricity of 1, and an ellipse between 0 and 1.

$$ecc = \frac{2\sqrt{\left(\frac{majlen}{2}\right)^2 - \left(\frac{minlen}{2}\right)^2}}{majlen} \quad (5.7)$$

- **Orientation<sup>†</sup>:** Defined as an angle between the major axis of the ellipse and the x-axis of the original image coordinates [8, 152].

$$\theta_O = \begin{cases} \tan^{-1} \left( \frac{m_{yy} - m_{xx} + \sqrt{(m_{yy} - m_{xx})^2 + 4m_{xy}^2}}{2m_{xy}} \right) & \text{if } m_{yy} > m_{xx}, \\ \tan^{-1} \left( \frac{2m_{xy}}{m_{xx} - m_{yy} + \sqrt{(m_{xx} - m_{yy})^2 + 4m_{xy}^2}} \right) & \text{else.} \end{cases} \quad (5.8)$$

- **EllipticalDeviation:** A measure of how much the image object deviates from the ellipse with the same second moments [154]. This measure takes into

account the pixels included in the ellipse and not in the object, as well as those pixels included in the object but not in the ellipse. This metric is normalized by the total object area. Defining  $E(n, m)$  as the ellipse mask, elliptical deviation is defined as

$$EllDev = \frac{\sum_{(n,m) \in \Omega, \notin E} \Omega(n, m) + \sum_{(n,m) \in E, \notin \Omega} E(n, m)}{Area} \quad (5.9)$$

- **Convex Hull Features:** These features rely on the computation of the convex hull of the object mask.

- **ConvexArea<sup>†</sup>:** The area of the convex hull of the object [154].

$$ConvexArea = \sum_n \sum_m convhull(\Omega(n, m)) \quad (5.10)$$

- **ConvexDeficiency:** The pixels within the convex hull that are not within the object [154].

$$ConvexDeficiency = \frac{ConvexArea - Area}{Area} \quad (5.11)$$

- **Solidity<sup>†</sup>:** The fraction of pixels within the convex hull that are within the object.

$$Solidity = \frac{Area}{ConvexArea} \quad (5.12)$$

- **Filled Image Features:** For computation of these features, attention is paid to both the object regions as well as the holes within the object.

- **FilledArea<sup>†</sup>:** The area of the filled object.

$$FilledArea = \sum_n \sum_m fill(\Omega(n, m)) \quad (5.13)$$

- **EulerNumber:**

$$EulerNumber = N_{regions} - N_{holes} \quad (5.14)$$

where  $N_{regions}$  is the number of regions contained within the object, and  $N_{holes}$  is the number of holes within the object.

- **Bounding Box Features:** These features are related to the bounding box of the object, which is defined as the smallest rectangle that completely contains the object, where the sides of the rectangle are aligned with the image axes.

- **Extent<sup>†</sup>:** The fraction of pixels within the bounding box that are also within

the object.

$$\text{Extent} = \frac{\text{Area}}{N \cdot M} \quad (5.15)$$

where  $N$  and  $M$  are the width and height of the bounding box.

- **AspectRatio:** The ratio of the height to width of the bounding box:

$$\text{AspectRatio} = \frac{M}{N} \quad (5.16)$$

- **Boundary Features:** These features are related to the boundary of the object.

- **Perimeter<sup>†</sup>:** The distance around the boundary of the object, where boundary pixels are 8-connected [18, 20, 44–46, 154].

$$\text{Perimeter} = \sum_{n=1}^N \sqrt{(x(n+1) - x(n))^2 + (y(n+1) - y(n))^2} \quad (5.17)$$

where  $x$  and  $y$  are the x- and y-coordinates of the  $N$  boundary pixels, and by definition  $x(N+1) = x(1)$  and  $y(N+1) = y(1)$  (to take care of the distance between the last and first boundary pixels).

- **Radii:** The vector of distances between the boundary pixels and the object centroid [16, 18, 20, 37, 44, 45, 65, 130, 152, 154].

$$\text{Radii} = \sqrt{(\mathbf{x} - \bar{\mathbf{x}})^2 + (\mathbf{y} - \bar{\mathbf{y}})^2} \quad (5.18)$$

where  $\mathbf{x}$  and  $\mathbf{y}$  are the vectors of the x- and y-coordinates of the boundary pixels and  $\bar{x}$  and  $\bar{y}$  are the x- and y-centroids of the object.

- **PerimeterFFT:** Characterizes the variations in the object boundary by treating the object radii as a one dimensional signal [8, 16, 65, 150, 152, 154]. The Fast Fourier Transform (FFT) is used to measure energies at various frequencies:

$$\text{PerimeterFFT} = \sum_{n=0}^{N-1} x(n) e^{-j2\pi nk/N}, \quad k = 0, 1, \dots, N-1 \quad (5.19)$$

where we choose  $N = 1024$ . As handled by the Matlab *fft* code,  $x(n)$  is zero-padded if less than 1024 samples and truncated if more. Energies are

computed as:

$$PerimeterFFTLow = \sum_{n=0}^{340} |\text{PerimeterFFT}(n)|^2 \quad (5.20)$$

$$PerimeterFFTMed = \sum_{n=341}^{683} |\text{PerimeterFFT}(n)|^2 \quad (5.21)$$

$$PerimeterFFTHi = \sum_{n=684}^{1023} |\text{PerimeterFFT}(n)|^2 \quad (5.22)$$

$$PerimeterFFTh = \sum_{n=32 \cdot h}^{32 \cdot (h+1) - 1} |\text{PerimeterFFT}(n)|^2 \quad (5.23)$$

where  $h = 0, 1, \dots, 32$ .

- **PerimeterCurvature:** The angle between line segments connecting subsequent boundary points [154].

$$\theta_{curve} = \arctan \left( \frac{y(n+1) - y(n)}{x(n+1) - x(n)} \right), \quad n = 1, \dots, N \quad (5.24)$$

where  $x$  and  $y$  are the x- and y-coordinates of the  $N$  boundary pixels, and by definition  $x(N+1) = x(1)$  and  $y(N+1) = y(1)$  (to take care of the curvature between the last and first boundary pixels).

- **BendingEnergy:** A measure analogous to the physical energy required to bend a rod. This metric is calculated as the sum of the difference in curvature between subsequent boundary points [31, 150, 154].

$$BendingEnergy = \sum_{n=1}^N (\theta_{curve}(n+1) - \theta_{curve}(n)) \quad (5.25)$$

where  $n = 1, \dots, N$  and, similar to the extension of  $x$  and  $y$  in Equation (5.24),  $\theta_{curve}(N+1) = \theta_{curve}(1)$ .

- **Other Shape Features:** Provide measures of the shape of the object, similar to some previously defined parameters (e.g., Eccentricity). These features, however, are not dependent on either the second-moment ellipse or the bounding box.

- **EquivDiameter<sup>†</sup>:** The diameter of the circle with the same area as the object [18, 20, 45].

$$EquivDiameter = 2 \sqrt{\frac{\pi}{Area}} \quad (5.26)$$

- **Sphericity:** Ratio of the radii of the largest circle fully inscribed in the object to that of the smallest circle which fully circumscribes the object. This is also equivalent to the ratio of the smallest to largest radii of the object, measured from the centroid [8, 16, 18, 20, 44–46, 65, 130, 152].

$$Sphericity = \frac{\min(\mathbf{Radii})}{\max(\mathbf{Radii})} \quad (5.27)$$

- **Compactness:** A measure of the roundness of the object, giving the minimum value of 1 for circles [8, 16, 18, 20, 31, 45, 65, 130, 150, 152, 154]:

$$Compactness = \frac{4\pi Area}{Perimeter^2} \quad (5.28)$$

- **InertiaShape:** Another measure of roundness, again giving 1 for circles [8, 65, 152]:

$$InertiaShape = \frac{2\pi \sum_n \sum_m \mathbf{Radii}(n, m)^2}{Area^2} \quad (5.29)$$

where in this case **Radii** is extended to include the radii of all object pixels, not just the boundary pixels. Note that this feature is a normalized moment of inertia for the object.

- **Centers of Mass:** The deviation of the center of mass from the object centroid, calculated for both the grayscale intensity image and the normalized optical density image.

- **GCenterMass:**

$$GCenterMass = \frac{\sqrt{(x_{GCM} - \bar{x})^2 + (y_{GCM} - \bar{y})^2}}{\text{mean}(\mathbf{Radii})} \quad (5.30)$$

where the center of mass  $(x_{GCM}, y_{GCM})$  of the grayscale intensity image  $A_I$  is calculated as

$$x_{GCM} = \frac{1}{N} \sum_n \sum_m n A_I(n, m) \quad (5.31)$$

$$y_{GCM} = \frac{1}{N} \sum_n \sum_m m A_I(n, m) \quad (5.32)$$

where  $N$  is the number of pixels in the  $A_I$  image.

- **ODCenterMass:**

$$ODCenterMass = \frac{\sqrt{(x_{ODCM} - \bar{x})^2 + (y_{ODCM} - \bar{y})^2}}{\text{mean}(\mathbf{Radii})} \quad (5.33)$$

where, similarly, the center of mass  $(x_{ODCM}, y_{ODCM})$  of the optical density image  $OD$  is calculated as in Equation (5.32).

- **Reflection Symmetry Features:** This is a metric of the reflection symmetry of the object [18, 45, 46, 158], based on the quadrant sum as presented in Chapter 4, specifically Equations (4.4) and (4.5). The quadrant sum is computed for the object mask ( $BWReflSymmetry$ ), the grayscale intensity image ( $GReflSymmetry$ ), and the normalized optical density image ( $ODReflSymmetry$ ).

### 5.3.2 Radiometric and Densitometric Features

- **ImageBands:** The values of the image bands for each object pixel.

$$\text{ImageBands} = [A(n, m, 1), \dots, A(n, m, B)], \quad \forall (n, m) \in \Omega \quad (5.34)$$

where  $A$  is the original image,  $B$  is the number of image bands in  $A$ , and  $\Omega$  is the object mask, as in (5.1). Thus **ImageBands** is an  $Area \times B$  matrix with each row corresponding to an object pixel and each column corresponding to an image band.

- **Intensity:** Calculated as the mean value for each object pixel, averaged across the image bands [8, 152, 154]:

$$\text{Intensity}(k) = \frac{1}{B} \sum_{b=1}^B \text{ImageBands}(k, b), \quad k = 1, \dots, Area \quad (5.35)$$

- **ODV:** The values of the optical density<sup>1</sup> image for each image band, as well as the intensity image: [8, 18, 20, 45, 65, 151, 152]

$$\text{ODV} = [\text{OD}_1, \text{OD}_2, \dots, \text{OD}_B, \text{OD}_I] \quad (5.36)$$

where each row of **ODV** is an object pixel, and each column is the optical density value for an image band. That is, **OD<sub>\*</sub>** are column vectors:

$$\text{OD}_b = -\log_{10} \left( \frac{\text{ImageBands}(k, b)}{A_{mx}} \right), \quad k = 1, \dots, Area, \quad b = 1, \dots, B \quad (5.37)$$

$$\text{OD}_I = -\log_{10} \left( \frac{\text{Intensity}(k)}{A_{mx}} \right), \quad k = 1, \dots, Area \quad (5.38)$$

where  $A_{mx}$  is the maximum value of  $A$  (i.e., the transmission value for a blank slide, usually 255 for 8-bit images), and  $B$  is the number of image bands.

---

<sup>1</sup>For transmission microscopy of stained tissue, optical density (OD) provides a linear relationship between image intensity and staining density. Refer to Appendix H for more explanation and an example.

- **IOD:** The Integrated Optical Density (IOD) is the summation of the optical density values of the object [8, 9, 18, 20, 45, 65, 151, 152]. For the case of nuclei, this can provide a measure of the amount of DNA present within each nucleus. In this implementation, there is a vector of values corresponding to each of the columns in **ODV**.

$$\mathbf{IOD}(b) = \sum_{k=1}^{Area} \mathbf{ODV}(k, b), \quad b = 1, \dots, B + 1 \quad (5.39)$$

- **MOD:** The Mean Optical Density (MOD) is the IOD normalized by the nuclear area [8, 18, 20, 45, 151, 152].

$$\mathbf{MOD}(b) = \frac{\mathbf{IOD}(b)}{Area} \quad (5.40)$$

- **Hue:** The hue value for each object pixel [9, 79].

$$\theta = \arccos \left( 0.5 \frac{2R - G - B}{\sqrt{(R - G)^2 + (R - B)(G - B)}} \right) \quad (5.41)$$

$$\mathbf{Hue}(k) = \begin{cases} \theta(k) & \text{if } B < G, \\ 2\pi - \theta(k) & \text{else.} \end{cases} \quad (5.42)$$

where  $R$ ,  $G$ , and  $B$  are the values of the red, green, and blue channel , respectively, of  $A(n, m)$ ,  $(n, m) \in \Omega$ . For images with less than three channels, the **Hue** parameter is not calculated, and for images with more than three channels (i.e., multispectral), the image is first converted to an RGB image via the ‘ccd’ transform described in Chapter 2.

### 5.3.3 Texture Features

- **Co-occurrence Matrix Features:** These metrics are computed from a gray-level co-occurrence matrix (GLCM), which tallies the number of occurrences of gray-level  $i$  in a specific spatial relation with gray-level  $j$ . Common spatial relations are horizontally and vertically adjacent pixels at a distance of some  $k$  pixels. Since we are not interested in the orientation of texture, the horizontal and vertical matrices are accumulated as recommended in [154]. The co-occurrence matrix is calculated with the Matlab function *graycomatrix* for the intensity image discretized to 64 levels (to reduce computational intensity), and for distances of  $k = 1, \dots, 5$  pixels. Defining the GLCM as  $C(i, j)$ ,  $i, j = 0, \dots, 63$ , the following metrics are used, where the metrics denoted by  $\dagger$  were used as included in the Matlab *graycoprops* function:

- **GLCMIInertia<sup>‡</sup>:** [8, 18, 20, 45, 65, 151, 152, 154]

$$GLCMIInertia = \sum_i \sum_j (i - j)^2 C(i, j) \quad (5.43)$$

- **GLCMEnergy<sup>‡</sup>:** [8, 18, 20, 45, 65, 151, 152, 154]

$$GLCMEnergy = \sum_i \sum_j C^2(i, j) \quad (5.44)$$

- **GLCMEEntropy:** [8, 18, 20, 45, 65, 151, 152, 154]

$$GLCMEEntropy = \sum_i \sum_j C(i, j) \log_2(C(i, j)) \quad (5.45)$$

- **GLCMHomogeneity:** [8, 18, 20, 45, 65, 151, 152]

$$GLCMHomogeneity = \sum_i \sum_j \frac{1}{1 + (i - j)^2} C(i, j) \quad (5.46)$$

- **GLCMMaxProb:** [18, 20, 45]

$$GLCMMaxProb = \max_{i,j} C(i, j) \quad (5.47)$$

- **GLCMClusterShade:** [8, 18, 20, 45, 65, 151, 152, 154]

$$GLCMClusterShade = \sum_i \sum_j (k - M_x + j - M_y)^3 C(i, j) \quad (5.48)$$

$$M_x = \sum_i \sum_j i C(i, j) \quad (5.49)$$

$$M_y = \sum_i \sum_j j C(i, j) \quad (5.50)$$

- **GLCMClusterProminence:** [8, 18, 20, 45, 65, 151, 152, 154]

$$GLCMClusterProminence = \sum_i \sum_j (k - M_x + j - M_y)^4 C(i, j) \quad (5.51)$$

- **GLCMCorrelation<sup>‡</sup>:** [8, 18, 20, 45, 65, 151, 152, 154, 159]

$$GLCMCorrelation = \frac{\sum_i \sum_j (i - M_x)(j - M_y)C(i, j)}{S_x \cdot S_y} \quad (5.52)$$

$$S_x = \sqrt{\sum_i (i - M_x)^2 C(i, j)} \quad (5.53)$$

$$S_y = \sqrt{\sum_j (j - M_y)^2 C(i, j)} \quad (5.54)$$

The same features are also calculated using a co-occurrence matrix of the optical density image, discretized to 64 levels and for distances of  $k = 1, \dots, 5$  pixels. These metrics are computed via the same equations, replacing the co-occurrence matrix of the grayscale intensity image with the co-occurrence matrix of the optical density image, yielding the features *ODCMInertia*, *ODCMEnergy*, *ODCMEntropy*, *OCDMHomogeneity*, *ODCMMaxProb*, *ODCMClusterShade*, *ODCMClusterProminence*, and *ODCMCorrelation*.

- **Run-Length Texture Features:** These features are based on the gray level run length (GLRL) matrix  $G(i, j|\theta)$ , which quantifies the number of runs of length  $j$ , gray level  $i$ , in direction  $\theta$ . From this 2-dimensional matrix, several features can be computed. Typically, the image is discretized to a smaller number of gray levels since the method is sensitive to noise; in this implementation the grayscale intensity image is discretized to 16 levels. We have implemented the method of [160] which requires the computation of two vectors rather than the full 2-dimensional GLRL matrix. Specifically, the run-length distribution  $r(j|\theta)$  and the gray-level distribution  $g(i|\theta)$  are defined as

$$r(j|\theta) = \sum_i G(i, j|\theta) \quad (5.55)$$

$$g(i|\theta) = \sum_j G(i, j|\theta) \quad (5.56)$$

and the total number of runs in the image  $S$  is

$$S = \sum_i \sum_j G(i, j|\theta) = \sum_i g(i|\theta) = \sum_j r(j|\theta) \quad (5.57)$$

The following features are extracted from the GLRL matrix  $G(i, j|\theta)$  for  $\theta = 0, \frac{\pi}{4}, \frac{\pi}{2}, \frac{3\pi}{4}$ .

- **SRE:** The short runs emphasis [20].

$$SRE = \frac{1}{S} \sum_j \frac{r(j|\theta)}{j^2} \quad (5.58)$$

- **LRE:** The long runs emphasis [20].

$$LRE = \frac{1}{S} \sum_j r(j|\theta) j^2 \quad (5.59)$$

- **GLN:** The gray-level nonuniformity [20].

$$GLN = \frac{1}{S} \sum_i g(i|\theta)^2 \quad (5.60)$$

- **RLN:** The run-length nonuniformity [20].

$$RLN = \frac{1}{S} \sum_j r(j|\theta)^2 \quad (5.61)$$

- **RP:** The runs percentage [20].

$$RP = \frac{1}{Area} \sum_j r(j|\theta) \quad (5.62)$$

- **LGRE:** The low gray-level runs emphasis [160].

$$LGRE = \frac{1}{S} \sum_i \frac{g(i|\theta)}{i^2} \quad (5.63)$$

- **HGRE:** The high gray-level runs emphasis [160].

$$HGRE = \frac{1}{S} \sum_i g(i|\theta) i^2 \quad (5.64)$$

The same features are also calculated for the optical density image, also quantized to 16 levels after normalization to the range [0, 255], yielding the features *ODSRE*, *ODLRE*, *ODGLN*, *ODRLN*, *ODRP*, *ODLGRE*, and *ODHGRE* [8, 65, 151, 152, 160]. Each run-length quantity is treated as a vector for  $\theta = 0, \frac{\pi}{4}, \frac{\pi}{2}, \frac{3\pi}{4}$  to avoid the use of orientation-dependent texture features.

- **Wavelet Features:** The energies of the various wavelet decomposition levels are used as a texture descriptor. Several specific wavelets have been tested for this

purpose, including Daubechies, biorthogonal, Gabor, and Mallat's discrete wavelet frames [18, 20, 27, 45]. We choose the CDF 9/7 (Cohen-Daubechies-Feauveau 9/7-tap) wavelet, as used in the JPEG2000 standard, as a good general wavelet transform. This decomposition is implemented in the MatlabCentral function *wavelet* via a lifting scheme. From a 4-level decomposition, the energies of each of the detail images at each level of decomposition as well as the energy of the low-resolution image from the final level are computed. Since this (and many other) wavelet transform is implemented in a separable fashion, it results in three detail images at each level of decomposition, commonly referred to as the horizontal ( $D_h^k$ ), vertical ( $D_v^k$ ), and diagonal ( $D_d^k$ ) detail images at level  $k$ . The average energy is computed for these three detail images for each of the four decomposition levels of the grayscale intensity image:

– **GCDF97Wk:**

$$GCDF97Wk = \frac{1}{3} \left[ \frac{\sum_n \sum_m |D_h^k(n, m)|^2}{N_k M_k} + \frac{\sum_n \sum_m |D_v^k(n, m)|^2}{N_k M_k} + \frac{\sum_n \sum_m |D_d^k(n, m)|^2}{N_k M_k} \right], \quad k = 1, \dots, 4 \quad (5.65)$$

where  $N_k$  and  $M_k$  are the dimensions of the  $D_*^k$  images.

as well as the energy of the final low-resolution image ( $A_{lr}$ ) of the decomposition.

– **GCDF97W5:**

$$GCDF97W5 = \frac{\sum_n \sum_m |A_{lr}(n, m)|^2}{N_4 M_4} \quad (5.66)$$

These measurements are repeated for the normalized optical density image, yielding the features  $ODCDF97W1$ ,  $ODCDF97W2$ ,  $ODCDF97W3$ ,  $ODCDF97W4$ , and  $ODCDF97W5$ .

- **Entropy:** A measure of the information content of the image, calculated as [79]

$$H = - \sum_{i=1}^N P(\alpha_i) \log_2(P(\alpha_i)) \quad (5.67)$$

where  $N$  is the total number of gray levels,  $\alpha_i$  is the  $i$ th gray scale level, and  $P(\alpha_i)$  is the probability of each level.  $P(\alpha_i)$  is generally estimated from the source signal. The entropy is calculated for all image bands, the intensity image, all OD bands, and the OD of the intensity image.

### 5.3.4 Chromatin Density Features

As presented in [152] and [151], these features rely on the segmentation of the optical density image into areas of low, medium, and high chromatin density. This segmentation is controlled by two global thresholds,  $t_1$  and  $t_2$ . Define the low, medium, high, and medium-high chromatin density images as

$$OD_L = (OD < t_1)\Omega \quad (5.68)$$

$$OD_M = (t_1 \leq OD \leq t_2)\Omega \quad (5.69)$$

$$OD_H = (OD > t_2)\Omega \quad (5.70)$$

$$OD_{MH} = OD_M + OD_H \quad (5.71)$$

where  $OD$  is the overall optical density image, normalized to the range  $[0, 255]$ . Note that  $OD_*$  are binary mask images. The thresholds  $t_1$  and  $t_2$  were empirically chosen to be 50 and 100, respectively. These thresholds may warrant further analysis in the future. While these features were specifically designed for feature extraction of cell nuclei, they are included for possible use on other histologic (or non-histologic) entities.

- **Area:** The areas of the low, medium and high images normalized by the total area. [8, 65, 151, 152].

$$LDNAArea = \frac{\sum_n \sum_m OD_L(n, m)}{Area} \quad (5.72)$$

$$MDNAArea = \frac{\sum_n \sum_m OD_M(n, m)}{Area} \quad (5.73)$$

$$HDNAArea = \frac{\sum_n \sum_m OD_H(n, m)}{Area} \quad (5.74)$$

- **IOD:** The integrated optical densities [8, 65, 151, 152].

$$LIOD = \frac{\sum_n \sum_m OD(n, m)OD_L(n, m)}{IOD} \quad (5.75)$$

$$MIOD = \frac{\sum_n \sum_m OD(n, m)OD_M(n, m)}{IOD} \quad (5.76)$$

$$HIOD = \frac{\sum_n \sum_m OD(n, m)OD_H(n, m)}{IOD} \quad (5.77)$$

- **MOD:** The mean optical densities [8, 65, 151, 152].

$$LMMOD = \frac{\sum_n \sum_m OD(n, m) OD_M(n, m)}{\sum_n \sum_m OD(n, m) OD_L(n, m)} \quad (5.78)$$

$$LHMOD = \frac{\sum_n \sum_m OD(n, m) OD_H(n, m)}{\sum_n \sum_m OD(n, m) OD_L(n, m)} \quad (5.79)$$

$$LMHMOD = \frac{\sum_n \sum_m OD(n, m) OD_{MH}(n, m)}{\sum_n \sum_m OD(n, m) OD_L(n, m)} \quad (5.80)$$

- **Number of Regions:** The number of discrete 8-connected regions  $LNum$ ,  $MNum$ , and  $HNum$  [8, 151, 152].
- **Compactness:** The compactness of the various regions [8, 65, 151, 152],

$$LCompactness = \frac{4\pi LDNAArea \cdot Area}{P_L^2} \quad (5.81)$$

$$MCompactness = \frac{4\pi MDNAArea \cdot Area}{P_M^2} \quad (5.82)$$

$$HCompactness = \frac{4\pi HDNAArea \cdot Area}{P_H^2} \quad (5.83)$$

$$MHCompactness = \frac{4\pi MHDNAArea \cdot Area}{P_{MH}^2} \quad (5.84)$$

where  $P_*$  is the perimeter of the given region, computed the same as in Equation (5.17).

- **Distance:** The average distances between all pixels of each chromatin image and the centroid of the object [65, 151, 152].

$$LDist = \frac{\sum_{(n,m) \in OD_L} \sqrt{(n - \bar{x})^2 + (m - \bar{y})^2}}{LDNAArea \cdot Area \cdot \text{mean(Radii)}} \quad (5.85)$$

$$MDist = \frac{\sum_{(n,m) \in OD_M} \sqrt{(n - \bar{x})^2 + (m - \bar{y})^2}}{MDNAArea \cdot Area \cdot \text{mean(Radii)}} \quad (5.86)$$

$$HDist = \frac{\sum_{(n,m) \in OD_H} \sqrt{(n - \bar{x})^2 + (m - \bar{y})^2}}{HDNAArea \cdot Area \cdot \text{mean(Radii)}} \quad (5.87)$$

$$MHDist = \frac{\sum_{(n,m) \in OD_{MH}} \sqrt{(n - \bar{x})^2 + (m - \bar{y})^2}}{MHDNAArea \cdot Area \cdot \text{mean(Radii)}} \quad (5.88)$$

where  $\bar{x}$  and  $\bar{y}$  are the x- and y-coordinates of the object centroid.

- **Center of Mass:** The distance between the center of mass of the optical density image and the object centroid [8, 65, 151, 152]. Calculating the centers of mass for

each of the chromatin images as in Equation (5.32),

$$LCenterMass = \frac{\sqrt{(x_{CML} - \bar{x})^2 + (y_{CML} - \bar{y})^2}}{LDNAArea \cdot Area \cdot \text{mean}(\mathbf{Radii})} \quad (5.89)$$

$$MCenterMass = \frac{\sqrt{(x_{CMM} - \bar{x})^2 + (y_{CMM} - \bar{y})^2}}{MDNAArea \cdot Area \cdot \text{mean}(\mathbf{Radii})} \quad (5.90)$$

$$HCenterMass = \frac{\sqrt{(x_{CMH} - \bar{x})^2 + (y_{CMH} - \bar{y})^2}}{HDNAArea \cdot Area \cdot \text{mean}(\mathbf{Radii})} \quad (5.91)$$

$$MHCenterMass = \frac{\sqrt{(x_{CMMH} - \bar{x})^2 + (y_{CMMH} - \bar{y})^2}}{MHDNAArea \cdot Area \cdot \text{mean}(\mathbf{Radii})} \quad (5.92)$$

### 5.3.5 Fractal Dimension Features

The fractal dimension (FD) of objects differ from the typical integer dimensions of Euclidean geometry (e.g., a 1-D line and 2-D plane). Instead, the FD is a real-valued (not integer) measure of the space-filling properties of the object. These features are based on the work of Mandelbrot [161]. There are many methods for estimation of the FD of objects, many relying on the change in a length or area metric with respect to the size of the ruler grid used for measurement; each suffers from different estimation errors. Based on the recent work published by Behry [162], we choose the power spectrum method of [163] since it produced the most accurate and robust estimation of FD. Previous studies have shied away from the power spectrum method since it requires the computation of an FFT; with the increase in computational power, however, and the fact that the FFT of object boundaries is already computed for other metrics, we do not consider the computational intensity to be too large.

- **PerimeterFD:** This is a measure of the fractal dimension of the object boundary [8, 22, 42, 45, 65, 150]. Defining the power spectrum as

$$\mathbf{P} = |\mathbf{PerimeterFFT}|^2 \quad (5.93)$$

and  $\rho$  as the spatial frequency, we compute the best least-squares linear fit to the log-log plot of  $\mathbf{P}$  versus  $\rho$ . The slope  $b$  of this best-fit line is related to the PerimeterFD via the relationship [162, 163]

$$PerimeterFD = \frac{5 - b}{2} \quad (5.94)$$

Note that for a length  $N$  FFT, the zero frequency term is excluded and only the next  $\frac{N}{2}$  points are used (due to symmetry of the FFT) to compute the best-fit line. It is expected that  $1 \leq PerimeterFD \leq 2$ .

- **AreaFD:** A measure of the FD of the object, treating the grayscale intensity values as a three-dimensional surface. This gives a measure of the space-filling properties of this three-dimensional surface, where the texture of the grayscale image is interpreted as this surface. Let  $A_I$  be the intensity image, computed by averaging the image bands for each pixel. Similar to the *PerimeterFD* case, the 2-D FFT of  $A_I$  is computed over the object mask  $\Omega$ , and the magnitude is squared to yield the 2-D power spectrum  $\mathbf{P}_A$ . Taking the two spatial frequencies  $\mathbf{u}, \mathbf{v} = [1, \dots, N/2]$ , we compute  $\rho = \sqrt{\mathbf{u}(i)^2 + \mathbf{v}(j)^2}$ ,  $i, j = 1, \dots, N/2$ . Calculating the slope  $b$  of the best-fit line for the log-log plot of  $\mathbf{P}_A$  versus  $\rho$ ,

$$AreaFD = \frac{8 - b}{2} \quad (5.95)$$

We expect that  $1 \leq AreaFD \leq 3$ .

- **ODAreaFD:** Similar to *AreaFD*, this is the FD of the optical density image [8, 65, 151, 152].

### 5.3.6 Summary

The object-level features are summarized in Table 5.1. In addition to the features presented above, various statistics measures for any of the vector quantities are also calculated. Thus, the mean, median, minimum, maximum, and standard deviation are calculated for all vector features, with the option to compute the skewness and kurtosis as well. The total number of object-level features extracted is, thus, 1035 for each RGB image object (897 without skewness and kurtosis) and 7171 for a multispectral image object (6201 without skewness and kurtosis). Since the extraction of multispectral features is so much more computationally intensive, the main focus in the feature selection will be on ccd imagery. Multispectral features will be used in limited analyses to demonstrate the differences between RGB and multispectral features.

## 5.4 Spatial-Relation Feature Definitions

In this section we describe the graph and tree structures that we derive from our data, as well as the various features extracted from the structures. A graph is uniquely defined by a set of nodes and edges. Edges connect nodes that are neighbors according to some *a priori* definition of a neighborhood. Edges may be directed, resulting in a directed graph (often called a digraph). Both nodes and edges may be weighted according to some metric (e.g., node degree or edge length). Trees are graphs without cycles, where a cycle is defined as an ordered set of nodes and edges, where the set of edges is unique, and

**Table 5.1.** Summary of object-level features.

Category	Features
<b>Size and Shape</b>	Area
	Elliptical Features: Major and minor axis length, eccentricity, orientation, elliptical deviation
	Convex Hull Features: Convex area, convex deficiency, solidity
	Filled Image Features: Filled area, Euler number
	Bounding Box Features: Extent, aspect ratio
	Boundary Features: Perimeter, radii, perimeter FFT, perimeter curvature, bending energy, perimeter FD
	Other Shape Features: Equivalent diameter, sphericity, compactness, inertia shape
	Center of Mass (gray and OD)
	Reflection Symmetry Features (binary, gray, and OD)
<b>Radiometric and Densitometric</b>	Image Bands
	Intensity
	Optical Density (OD)
	Integrated Optical Density (IOD)
	Mean Optical Density (MOD)
	Hue
<b>Texture</b>	Co-occurrence Matrix Features (gray and OD): Inertia, energy, entropy, homogeneity, maximum probability, cluster shade, cluster prominence, correlation
	Fractal dimension (gray and OD)
	Run-length Features (gray and OD): Short runs emphasis, long runs emphasis, gray-level non-uniformity, run-length non-uniformity, runs percentage, low gray-level runs emphasis, high gray-level runs emphasis
	Wavelet Features (gray and OD): Energies of detail and low resolution images
	Entropy (gray and OD)
<b>Chromatin-Specific</b>	Area, IOD, MOD, number of regions compactness, distance, center of mass

the starting and ending node is the same. We will refer to our graph and tree structures with the generic term “graph” when it will not introduce confusion.

### 5.4.1 Graph and Tree Structures

- **Voronoi Tesselation (VT)<sup>2</sup>:** Given a set of nodes in the plane (e.g., centroids of nuclei), the Voronoi tessellation creates polygonal cells around the nodes such that all pixels within a given cell are closer to the cell node than any other node in the plane. The VT is commonly constructed as the complement (dual graph) of the Delaunay triangulation, and is used for quantification of pathology imagery in [22, 41, 155, 157].
- **Delaunay Triangulation (DT)<sup>3</sup>:** The Delaunay triangulation, in addition to being the dual graph of the Voronoi tessellation, is a commonly used triangulation algorithm. From the set of all possible triangles, a triangle is accepted if its circumcircle contains no other nodes besides the triangle vertices. The DT is used with pathology imagery in [22, 41].
- **Minimum Spanning Tree (MST):** The minimum spanning tree is the tree for which the sum of the edge lengths is minimum. The MST is generally iteratively created. First the two closest points are connected; next the point closest to either of the two included points is added to the tree. This continues until all nodes have been added. The MST is one of the most commonly used tree structures and is used for pathology imagery in [22, 41, 156, 157].
- **O’Callaghan Neighborhood Graph (OCG):** Neighboring nodes are defined based on both a distance and direction constraint. Intuitively, node C is considered a neighbor of node A if the distance between A and C is less than a predefined threshold and if C is not eclipsed by another neighbor of A. More specifically, considering three nodes A, B and C, and assuming that A and B are neighbors, C is a neighbor of A if

$$d(A, C) < t_d \quad (5.96)$$

and

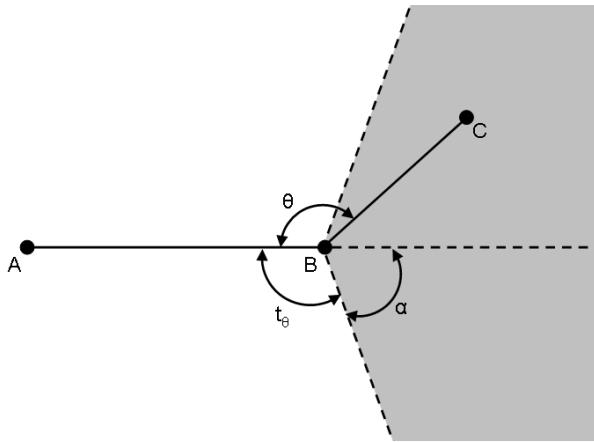
$$\angle ABC < t_\theta \quad (5.97)$$

where  $d(\cdot)$  is the Euclidean distance function,  $t_d$  is the distance threshold, and  $t_\theta$  is the direction constraint. While the direction constraint is generally defined in terms of A, B, and C, a more intuitive explanation of the angle is in terms of an eclipse angle directed from B along the edge connecting A and B. Refer to Figure 5.1 for an illustration of the direction constraint. Experiments have determined that a value

---

<sup>2</sup>Sometimes referred to as the Dirichlet tessellation.

<sup>3</sup>Referred to as the Gabriel Graph in [22] and [41].

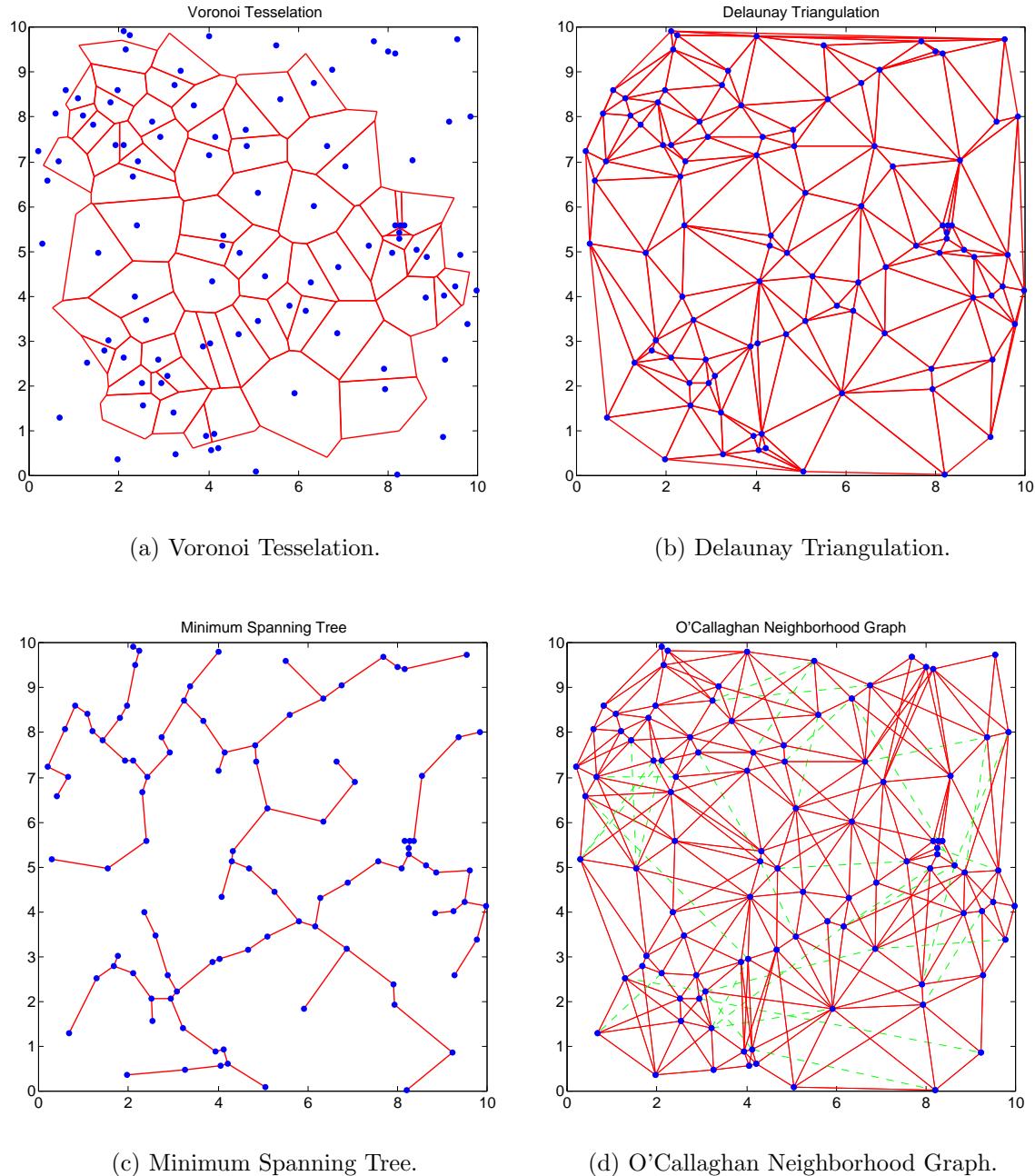


**Figure 5.1.** Illustration of the O'Callaghan direction constraint. For this example, we are assuming that nodes A and B are neighbors, and that nodes A and C satisfy the distance constraint. The direction constraint is specified in terms of threshold  $t_\theta$  for the angle  $\theta = \angle ABC$ . For this example, node C does not satisfy the direction constraint, and is thus not a neighbor of A. The direction constraint can also be thought of as an angle of visual eclipse emanating from node B at a symmetric angle about the line directed from A to B, labeled as  $\alpha$  in this example. Obviously,  $\alpha = 180^\circ - t_\theta$ . A perceptually acceptable value for the direction constraint is  $t_\theta = 120^\circ$  ( $\alpha = 60^\circ$ ) [164].

for  $t_\theta$  of  $120^\circ$  is a perceptually valid threshold [164]. The OCG is constructed in an iterative process similar to that of the MST. It should be noted that the OCG is a digraph since (in fairly rare circumstances) node A may be a neighbor of B but not vice versa. The OCG was used for pathology imagery in [156, 157, 165], although it was incorrectly characterized as a tree in [156].

- **Connected Graph (CG):** In a connected graph, all nodes are connected to all other nodes. In the case of the connected graph used in [36] and [35], the edges are weighted by the Euclidean distance.

Examples of the VT, DT, MST, and OCG for a randomly generated set of nodes are shown in Figure 5.2; the CG is excluded since it is too cluttered with edges for an informative illustration. In addition to the graph and tree structures described above, used for analysis of pathology imagery, we will briefly mention that the k-NN graph was used in [166] and the Relative Neighbor Graph (RNG) in [167]. We forgo the implementation of the k-NN graph since we are more interested examining the features of a graph structure for a neighborhood defined by some other neighborhood constraint, rather than the arbitrary definition of a certain number of neighbors. Additionally, the work presented in [166] was focused on the matching of graph structures extracted from imagery. We are more concerned with the actual extraction process and the features that



**Figure 5.2.** Example graph and tree structures for a randomly generated set of nodes. For the VT, we do not display the edges that end at infinity for the nodes at the border of the graph; we also ignore these edges in computations involving the VT. For the digraph OCG in (d), symmetric edges are displayed as solid red lines, and non-symmetric edges are displayed as green dashed lines.

can be used from these graph structures for further image classification and analysis. We forgo implementation of the RNG in favor of the (simpler to implement) OCG, since both graph structures seek to define a neighborhood based on comparison of metrics with already existing neighbors. In the case of the RNG as presented in [167], the DT is processed to retain an edge only if the nodes are “relatively close.” The use of the k-NN graph and RNG (or any other graph structure) could prove useful in our analysis. The implementation of other graph and tree structures we leave as future work.

### 5.4.2 Spatial-Relation Features

In the following descriptions,  $\mathbf{A}$  is the adjacency matrix of the graph or tree. Additionally, we will use the generic term “graph” to refer to any graph or tree structure. The adjacency matrix is defined such that the  $ij$ -th entry  $a_{ij}$  is the number of edges from node  $i$  to node  $j$ ; there is a one-to-one correspondence between the adjacency matrix and graph structure [168]. For our graph structures, we will have no parallel edges (multiple edges between the same nodes) or loops (edges linking a single node to itself), so that  $a_{ij} \in [0, 1], \forall i, j$ . This means we can define another matrix  $\mathbf{B}$  with edge weights (in our case inter-node distance); this facilitates easy computation of some feature values. Additionally,  $\mathbf{A}$  and  $\mathbf{B}$  will be symmetric for all structures except the OCG; this has some implications for computational simplicity of feature calculations.

The spatial-relation features are summarized in Table 5.2; if the details of feature definitions are not of interest to the reader, the remainder of this section (5.4.2) may be skipped. For each of the features described below, we will make note of the graphs from which we extract the features. We will also note any previous work that has used or recommended the use of these specific graph features.

- **Number of Nodes:** The total number of nodes in the graph [41, 156, 157]. Since  $\mathbf{A} \in \mathbb{Z}^{N_n, N_n}$ , where  $N_n$  is the number of nodes, the number of nodes can be easily ascertained from the size of the adjacency matrix  $\mathbf{A}$ . This feature will be constant for all graph structures.
- **Number of Edges:** The total number of edges in the graph. For any non-digraph (the digraph case is considered later), the sum of the upper (or lower triangle) of  $\mathbf{A}$  is the total number of edges in the graph. For simplicity of computation, it is easier to sum the entire matrix  $\mathbf{A}$  and divide by two (since  $\mathbf{A}$  is symmetric for non-digraphs):

$$N_e = \frac{1}{2} \sum_{i,j} a_{ij} \quad (5.98)$$

For the digraph OCG,  $\mathbf{A}$  may not be symmetric and thus could have an odd value for the sum of the elements of  $\mathbf{A}$ . For now we simply round  $N_e$  to the nearest integer in the case of the OCG to allow for direct comparison of  $N_e$  between the

non-digraphs and the OCG. Use of the digraph characteristics of the OCG may be of use in the future.  $N_e$  is calculated for the VT [157], DT [22], and OCG [156,157].  $N_e$  is not computed for the MST since there will always be  $N_n - 1$  or for the CG since it will always be  $N_n^2$ .

- **Cyclomatic Number:** The cyclomatic number is defined as

$$C = N_e + N_n + 1 \quad (5.99)$$

and is calculated for the VT [157], DT, and OCG [157] (all the graphs for which we calculate  $N_e$ ).  $C$  is normalized by the number of edges,  $N_e$ .

- **Edge Length:** The edge lengths (Euclidean distance) between each pair of neighboring nodes. **EL** is a vector of edge lengths, where an edge length is added according to

$$\max(b_{ij}, b_{ji}); i = 1, \dots, n; j = i, \dots, n; b_{ij} > 0 \text{ or } b_{ji} > 0 \quad (5.100)$$

**EL** is computed for the VT [157], DT [22,41], MST [22,41,156,157], CG [35,36], and OCG [156,157].

- **Degree:** The number of edges incident with a node. This may be calculated from the adjacency matrix **A** by summing over the rows or columns:

$$\deg(n_i) = \sum_{j=1}^{N_n} a_{ij}, i = 1, \dots, n \quad (5.101)$$

The degree of all nodes for the VT, DT [41], MST [41], and OCG are calculated, resulting in a vector,

$$\mathbf{deg} = [\deg(n_1), \dots, \deg(n_{N_n})] \quad (5.102)$$

The CG is excluded since the degree of all nodes will be  $N_n - 1$ . An alternative definition for degree is used in [36] and [35], where degree is the sum of the edge weights for all edges incident with a node:

$$\deg_w(n_i) = \sum_{j=1}^{N_n} b_{ij}, i = 1, \dots, n \quad (5.103)$$

This modified degree is computed for the VT, DT, MST, and OCG, yielding the vector,

$$\mathbf{deg_w} = [\deg_w(n_1), \dots, \deg_w(n_{N_n})] \quad (5.104)$$

In this case edge weights are defined as the Euclidean distance between the two end nodes; this could be extended to include edge weights related to other features, however.

- **Number of Neighbors:** A vector  $\mathbf{NN}$  consisting of the number of nodes with  $k$  neighbors, with  $k = 1, \dots, 5$  for our computations. This feature can be computed from a histogram of the  $\mathbf{deg}$  feature, and is computed for the DT [157] (excluding the VT since it will contain the same information), MST [156, 157], and OCG [156, 157]. The CG is excluded since all nodes will have  $N_n - 1$  neighbors.
- **Randic Index:** The Randic index has been used widely in chemistry applications where it has been shown to be sensitive to shape [157]. Assigning each edge a weight equal to the reciprocal of the square root of the product of the degrees of the two neighboring nodes, the Randic index of a graph is the sum of these edge weights:

$$R = \sum_{i \neq j, a_{ij} \neq 0} \frac{1}{\sqrt{\deg(n_i) * \deg(n_j)}} \quad (5.105)$$

The Randic index is computed for the VT, DT, MST [157], OCG, and CG.  $R$  is normalized by  $N_n^2 / \text{ave}(\mathbf{deg})$ .

- **Balaban Index:** The Balaban index is computed via an iterative pruning process, whereby, at each iteration, points with only one neighbor (i.e., endpoints) are removed [157]. The Balaban index is defined as the sum of the squares of the number of points removed at each pruning step. This feature is a measure of the degree of branching of a tree structure, and is computed for the MST [157], normalized by  $N_n^2$ . This feature is not computed for any graph structures, since there is no guarantee that any endpoints will exist at any given iteration of the pruning process.
- **Wiener Index:** The Wiener index is used in chemistry as a measure of both molecular volume and shape [157] and is defined as the sum of the distances between all pairs of nodes, where distance is defined as the minimum number of edges connecting the two nodes. The Wiener index is computed for the DT, CG, and MST [157]. The OCG is omitted due to computational difficulties with the non-symmetric adjacency matrix; this does not necessarily preclude a different method to compute the Wiener index of the OCG, but is left as future work. Additionally, the VT is omitted since it will yield the same information as the DT. The Wiener index is normalized by  $N_n^2$ .
- **Eccentricity:** The eccentricity of a node is the greatest distance between itself and all other nodes in the graph. Again, distance here is defined as the minimum number of edges connecting two nodes. The eccentricities  $e$  of each node for the DT and MST are computed, yielding the vector  $\mathbf{e}$ . Similar to the two different degree features,  $\mathbf{deg}$  and  $\mathbf{deg}_w$ , we also compute  $\mathbf{e}_w$  using the edge weights rather than number of edges for the DT and MST. The CG is not included here since the eccentricity will always be the largest edge length,  $\max(\mathbf{EL})$ .

- **Number of k-Walks:** A k-walk is an ordered set of nodes and edges

$$W = n_1 e_1 n_2 e_2 \dots n_{k-1} e_k n_k \quad (5.106)$$

such that edge  $e_i$  has nodes  $n_i$  and  $n_{i+1}$  as endpoints [169]. A closed k-walk has  $n_1 = n_k$ . One property of the adjacency matrix is that the  $ij$ -th entry of  $\mathbf{A}^k$  is the number of k-walks from node  $n_i$  to  $n_j$  [168]. Thus the total number of closed k-walks can be computed as

$$N_{\text{kwalks}} = \text{trace}(\mathbf{A}^k) \quad (5.107)$$

The number of k-walks is computed for  $k = 3, 4, 5$  for the VT, DT, OCG, and CG, resulting in the vector  $\mathbf{N}_{\text{kwalks}}$ . The MST is omitted since there will be no closed k-walks in a tree structure. The number of k-cycles is generally of more interest [157], where a k-cycle is a closed k-walk with unique nodes  $n_1, \dots, n_{k-1}$  [169]. The computation of the number of k-cycles from the adjacency matrix, however, is nontrivial and computationally intensive [170]. We substitute the more general number of closed k-walks, although it has not been established in previous work whether this is a useful metric for quantifying spatial arrangement.  $\mathbf{N}_{\text{kwalks}}$  is normalized by  $N_e$ .

- **Number of Triangles:** The number of length-3 cycles in a graph is computed for the VT [157], DT, OCG [157], and CG. Generally, the number of k-cycles of a graph cannot be directly deduced from the adjacency matrix. In the case of 3-cycles, however, the following property holds [170]:

$$N_{\text{tri}} = \frac{\text{trace}(\mathbf{A}^3)}{6} = \frac{N_{\text{3walks}}}{6} \quad (5.108)$$

Again, we normalize  $N_{\text{tri}}$  by  $N_e$ .

- **Spectral Radius:** The spectral radius is defined as the largest absolute value of the eigenvalues of  $\mathbf{A}$ . Along the lines of [36] and [35], both the largest and second largest absolute eigenvalue of  $\mathbf{A}$  are computed for the VT, DT, MST, and CG [35, 36]:

$$\mathbf{SR} = [\lambda_1 \lambda_2] \quad (5.109)$$

where  $\lambda_1$  and  $\lambda_2$  are the first and second largest absolute eigenvalues, respectively.  $\mathbf{SR}$  is not computed for the OCG since the adjacency matrix is not necessarily symmetric, which can result in complex-valued eigenvalues.

- **Eigenexponent:** The eigenexponent is defined as the slope of the sorted eigenvalues in log-log scale; per [36] and [35], the slope from the 3rd largest eigenvalue to the minimum of  $N_n$  or 30 is used. The eigenexponent is computed for the VT, DT, MST, and CG [35, 36].

- **Fractal Dimension:** An image is defined by creating a one-pixel wide line for each of the graph edges, and the same FFT-based method for computation of fractal dimension as presented in Section 5.3 is used. The fractal dimension of the DT [22, 41], MST [22, 41], OCG, and CG is computed. The VT is omitted since the point of the VT is to cover the plane with adjacent Voronoi cells.

- **Area:** The area of the Voronoi cells in the VT [22, 41]:

$$\mathbf{Area} = [Area(X_1), \dots, Area(X_m)] \quad (5.110)$$

where  $X_i$  is the  $i$ -th Voronoi cell, and  $Area$  computes the area in number of pixels. We only consider the Voronoi cells for which there are no vertices at infinity.

- **Average roundness factor:** The average roundness factor of a VT is defined as [155]:

$$RF_{av} = \frac{1}{N} \sum_{i=1}^N \frac{4\pi Area(X_i)}{Perimeter(X_i)^2} \quad (5.111)$$

where  $Perimeter$  is the perimeter of the Voronoi cell.

- **Area Disorder:** The area disorder of a VT is defined in terms of the mean and standard deviation of the Voronoi cell areas,  $m_A$  and  $\sigma_A$ , respectively [155]:

$$AD = 1 - \left(1 + \frac{\sigma_A}{m_A}\right)^{-1} \quad (5.112)$$

- **Roundness Factor Homogeneity:** The roundness factor homogeneity of a VT is defined in terms of the mean and standard deviation of the roundness factor,  $m_{RF}$  and  $\sigma_{RF}$ , respectively [155]:

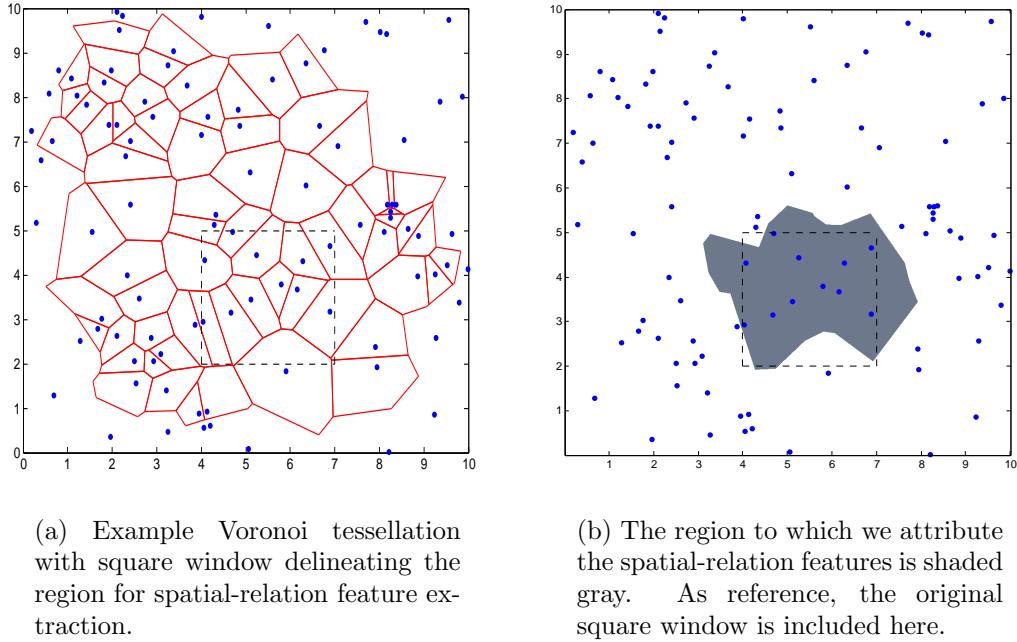
$$RFH = \left(1 + \frac{\sigma_{RF}}{m_{RF}}\right)^{-1} \quad (5.113)$$

### 5.4.3 Summary

The spatial-relation features are summarized in Table 5.2. In addition to the features presented above, various statistics measures for any of the vector quantities are also calculated. Thus, the mean, median, minimum, maximum, and standard deviation are calculated for all vector features, with the option to compute the skewness and kurtosis as well. The total number of spatial-relation features extracted is, thus, 179 (145 without skewness and kurtosis) for each of the graph structures.

**Table 5.2.** Summary of spatial-relation features.

Category	Features
<b>Voronoi Tessellation</b>	Number of nodes, number of edges, cyclomatic number, number of triangles, number of k-walks, spectral radius, eigenexponent, Randic index, area, roundness factor, area disorder, roundness factor homogeneity
<b>Delaunay Triangulation</b>	Number of nodes, edge length, degree, number of edges, cyclomatic number, number of triangles, number of k-walks, spectral radius, eigenexponent, Wiener index, eccentricity, Randic index, fractal dimension
<b>Minimum Spanning Tree</b>	Number of nodes, edge length, degree, number of neighbors, Wiener index, eccentricity, Randic index, Balaban index, fractal dimension
<b>O'Callaghan Neighborhood Graph</b>	Number of nodes, number of edges, cyclomatic number, number of neighbors, number of triangles, number of k-walks, spectral radius, eigenexponent, Randic index, fractal dimension
<b>Connected Graph</b>	Edge length, number of triangles, number of k-walks, spectral radius, eigenexponent, Wiener index, eccentricity, Randic index, fractal dimension



**Figure 5.3.** The sliding window approach to extraction of region-based spatial-relation features. A square window is used to define the current area of interest, but the region to which the features are attributed is defined by the Voronoi cells of all nodes internal to the square window.

## 5.5 Region-Based Spatial-Relation Features

The spatial arrangement of nuclei or other histologic entities will not be consistent across any given image. As such, we would like to quantify the spatial-relation features of an image on a regional basis. The question, however, is how to define a region; ideally there would be no *a priori* constraints on the definition of a region, since that is at the heart of the very issue of extracting spatial-relation features. Realistically, however, there must be some definition of a region for region-based feature extraction. We choose to use a sliding square window and quantify the graph structure that is enclosed within the window. Since this is essentially feature extraction within a truth window, the sliding window starts with a one pixel overlap with the truth window in the upper left and ends with one pixel overlap in the lower right. For each position of the sliding window, the extracted features are associated with each node and the Voronoi cell of each node within the window. Thus, if we were to delineate within an original image the regions with some particular feature value, we would use the Voronoi cell boundaries for this delineation. As an example, refer to Figure 5.3.

In other work, there is reference to the use of spatial-relation features extracted for

a “reference area,” but it is unclear how this reference area is defined. In particular, the number of nodes per reference area is used for the DT [41], MST [41,156], and OCG [156] and the total edge length per reference area is used for the DT [22,41], and MST [22,41].

In order to define a reasonable size for our sliding window, we look at the spatial-relation feature values for differently sized graphs (according to number of nodes). For this we randomly generated two types of graphs designed to be representative of nuclei in our data. For both graph types, assuming that the initial node coordinates  $(x_0, y_0)$  are initialized to some arbitrary value, we:

1. Randomly select from the existing nodes, resulting in the selection of some coordinates  $(x_r, y_r)$ .
2. Calculate a new node at a randomly generated distance from  $(x_r, y_r)$ , i.e.,  $(x_n, y_n) = (x_r + d_x, y_r + d_y)$ .  $d_x$  and  $d_y$  are both independently sampled from a normal distribution of mean 0 and standard deviation 12. We choose 12 to represent a typical minimum distance between cell nuclei, assuming an average nuclear diameter of 25 pixels. These are the same assumptions as used in the implementation of the WSblob nuclear segmentation in Section 4.4.4.

- (a) **Random Graph Type 1:** Accept the new node  $(x_n, y_n)$  only if

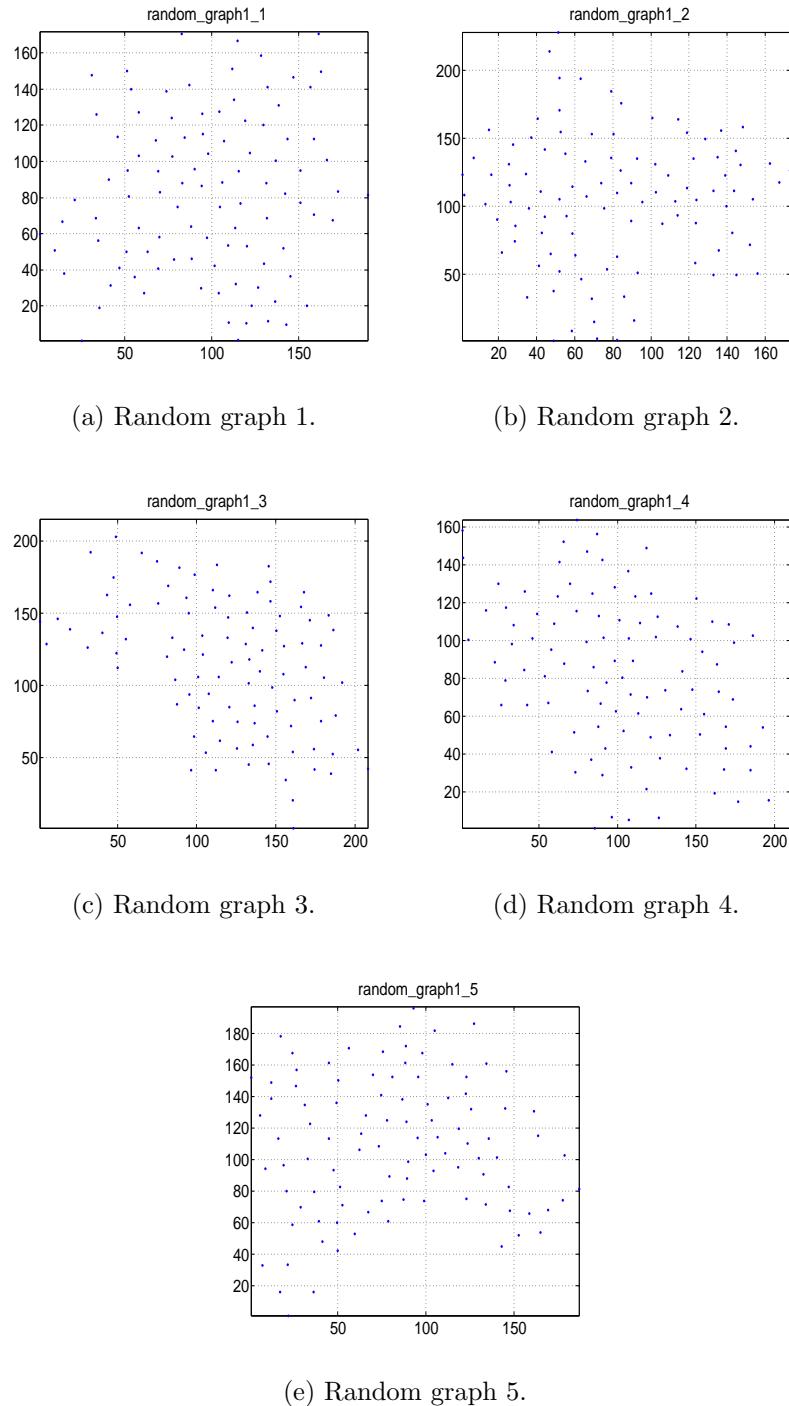
$$\sqrt{(x_n - x_i)^2 + (y_n - y_i)^2} > 10; \quad i = 1, \dots, k \quad (5.114)$$

where  $k$  is the current number of accepted nodes. Thus the new node is rejected if the Euclidean distance between the new node and any other existing node is less than 10 pixels. This imposes a more regular spatial distribution to the random graph, along the lines of what might be expected for normal or benign tissue.

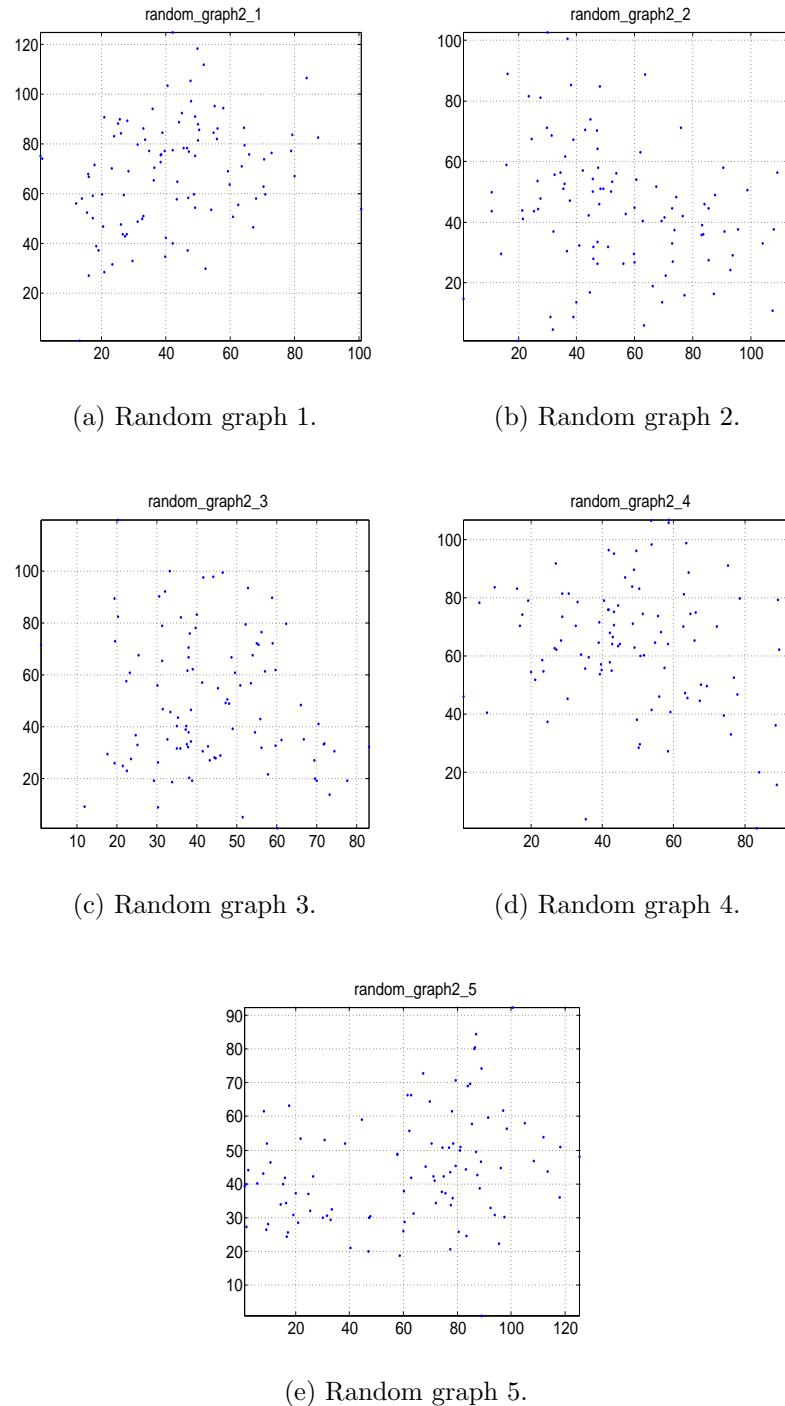
- (b) **Random Graph Type 2:** Accept the new node. This results in the opportunity for nodes to become more crowded, approximating the appearance of some malignant tissue.
3. Repeat step 2  $N - 1$  times, where  $N$  is the total number of nodes desired in the random graph.

Examples of type 1 random graphs are shown in Figure 5.4 and examples of type 2 random graphs in Figure 5.5.

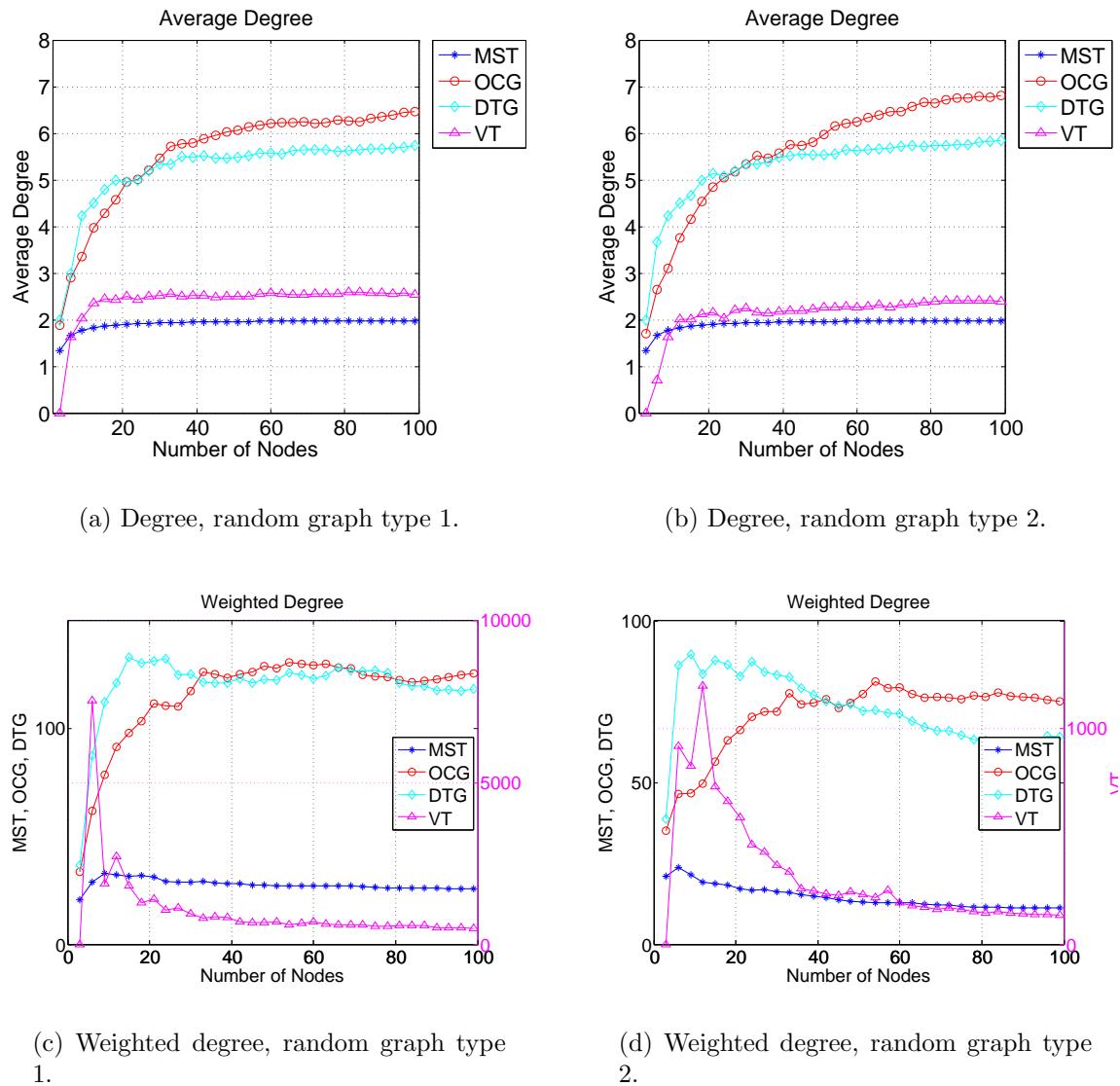
All the described spatial-relation features are extracted from each of the random graphs in Figures 5.4 and 5.5 for number of nodes in the range [3,100], where 100 is the total number of nodes in the graphs. For each graph type, the spatial-relation feature values are averaged and plotted versus the number of nodes. This gives insight into how the spatial-relation features are expected to vary according to the number of nodes of a graph. These plots are presented for some spatial-relation features in Figure 5.6; features

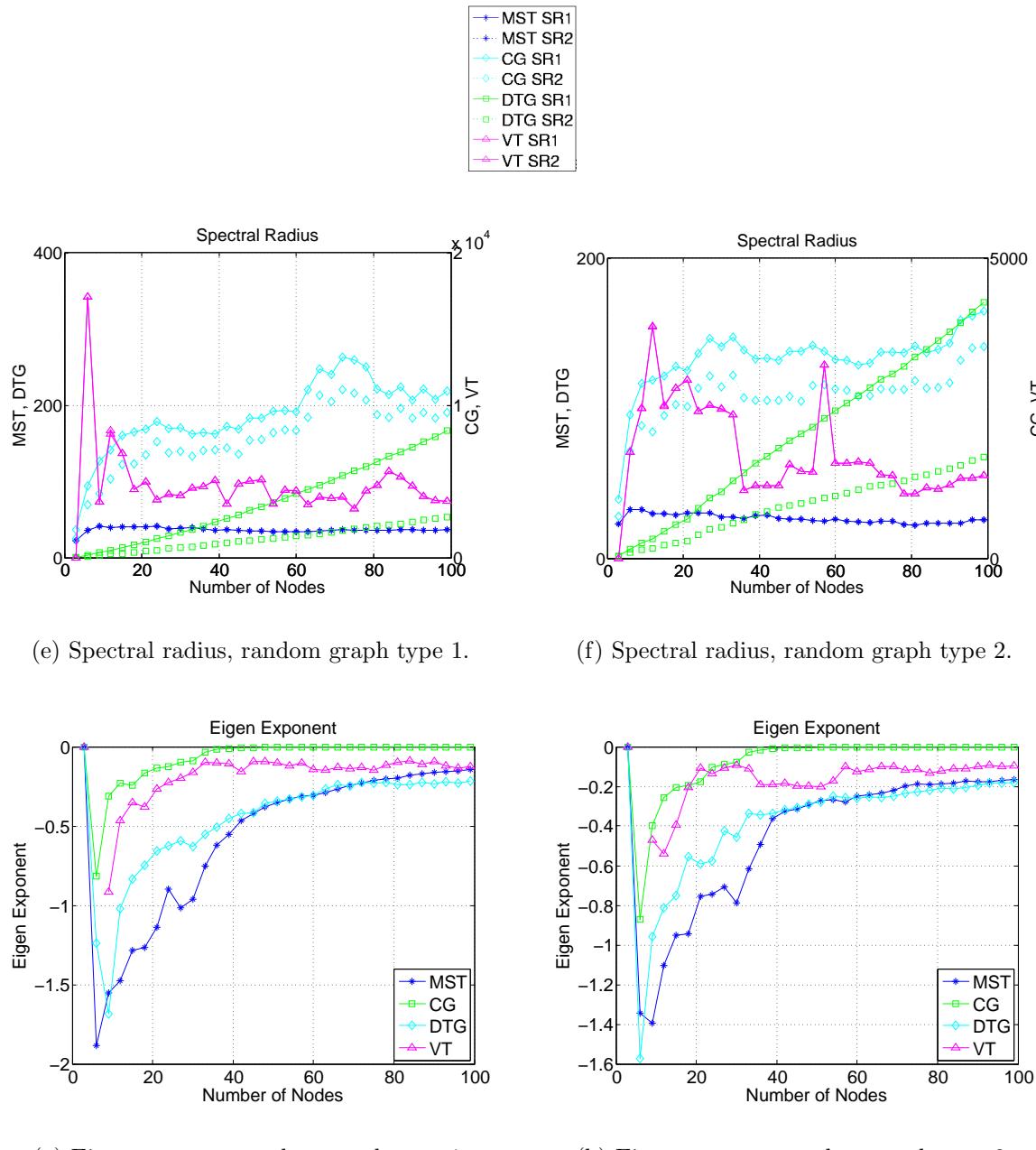


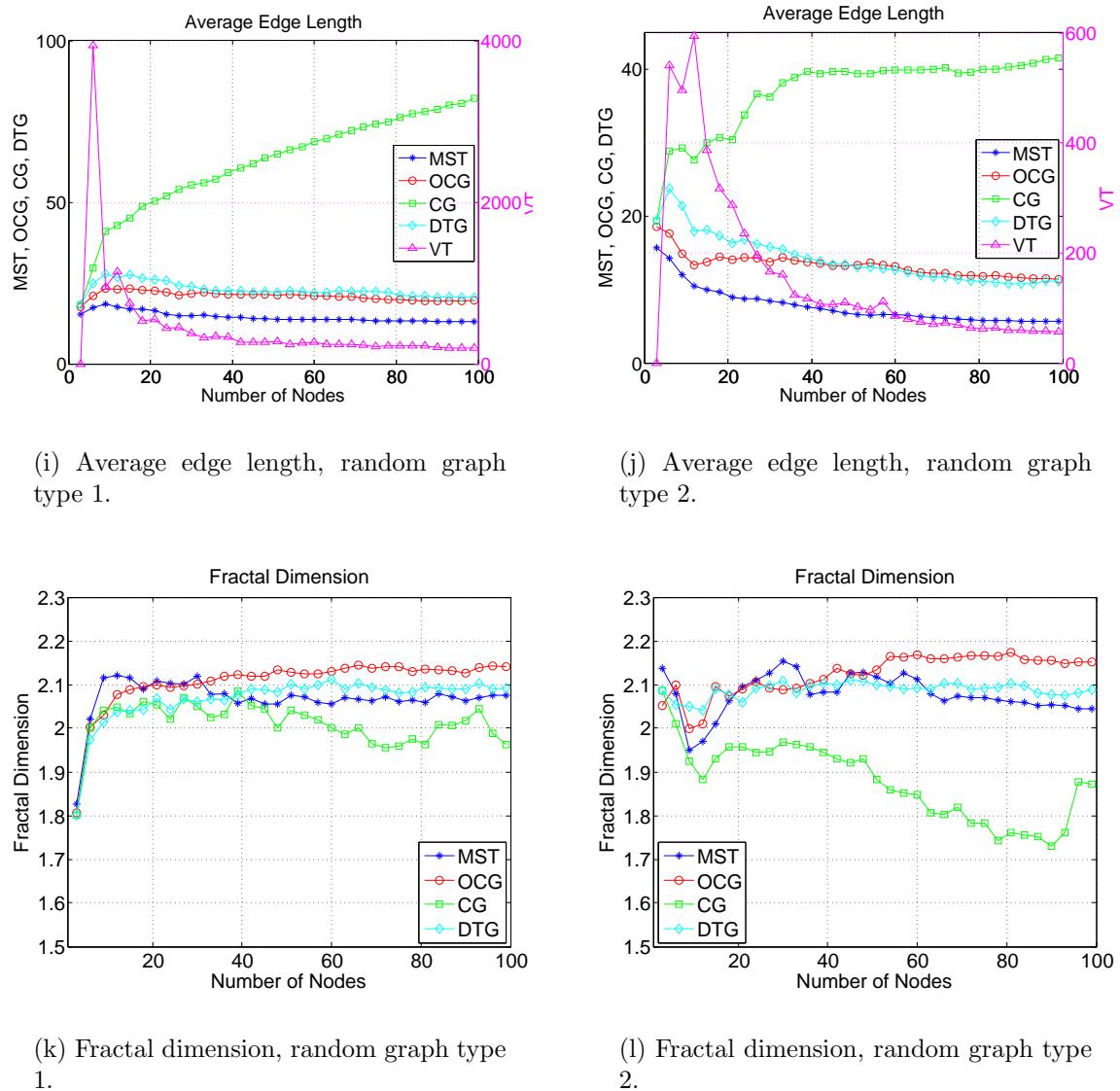
**Figure 5.4.** Random graphs, type 1.



**Figure 5.5.** Random graphs, type 2.

**Figure 5.6.** Random graph features. (continued on next 2 pages)

**Figure 5.6.** (Cont.) Random graph features. (continued on next page)

**Figure 5.6.** (Cont.) Random graph features.

chosen for presentation here are those that are applicable to the widest range of graph structures. Plots for all spatial-relation features are presented in Appendix I.

Since the random graphs are designed to approximate the appearance of histology imagery, we look for the number of nodes for which the majority of the features are close to their “steady-state” value, i.e., the value representative of the entire 100-node graph. From Figure 5.6 note that there are very similar trends in the feature plots of type 1 and type 2 random graphs. Also note that for a number of nodes between 20 and 30, the feature value is generally expected to be close to the steady state value of the total graph. In the choice of the size of our sliding window, we would thus like the window to (most likely) encompass at least 20 to 30 nodes, or approximately  $5^2$  nodes. For each nucleus being approximately 25 pixels, this is a window of about 125x125 pixels. We choose 128x128 pixels as the closest power of 2, since this facilitates efficient computation of some of the features relying on the FFT.

## 5.6 Feature Selection and Classification Performance

In this section we briefly describe the feature selection technique used and then describe results for feature selection and performance.

### 5.6.1 An Overview of Grafting

Grafting (from “gradient feature testing”) [108, 109] is used as the feature selection method for this work. Grafting is chosen based on the elegant formulation of the feature selection problem, whereby the classification of the underlying data and the feature selection process are not separated. Within the grafting framework, a loss function is used that shows preference for classifiers that separate the data with larger margins. Grafting also provides an efficient framework for selection of relevant features.

More specifically, given a set of  $m$  feature vectors  $\mathbf{x}$  with corresponding ground truth labels  $y$ , the grafting framework looks to minimize the criterion

$$C(\theta) = \frac{1}{m} \sum_{i=1}^m L(f_\theta(\mathbf{x}_i), y_i) + \Omega(\theta) \quad (5.115)$$

where  $L(\cdot)$  is a loss function to penalize incorrect classifications,  $f(\cdot)$  is a predictor function (classifier),  $\theta$  is the set of parameters (features) that parametrize the predictor function, and  $\Omega$  is a regularization term to penalize complex (overfit) solutions [108].

The predictor function is assumed to be linear:

$$f(\mathbf{x}) = \sum_{i=1}^n w_i x_i + b \quad (5.116)$$

where  $w_i$  and  $x_i$  are the  $i$ -th entries of weight vector  $\mathbf{w}$  and feature vector  $\mathbf{x}$ , respectively.

The Binomial Negative Log Likelihood (BNLL) loss function is used [108],

$$L_{BNLL} = \frac{1}{m} \sum_{i=1}^m \ln(1 + e^{-\rho_i}) \quad (5.117)$$

where  $\rho$  is the margin of the classifier, defined as

$$\rho_i = y_i f(\mathbf{x}_i) \quad (5.118)$$

since it incorporates the classifier margin, and for other mathematical advantages (refer to [108]). The regularization term is defined as:

$$\Omega_q(\mathbf{w}) = \lambda_q \sum_{i=1}^n \alpha_i |w_i|^q \quad (5.119)$$

From this formulation, commonly  $\alpha_i \in \{0, 1\}$ , such that features are either included or excluded from consideration in the optimization. The choice of  $q$  results in regularizers with different characteristics. As implemented, the  $\Omega_1$  regularizer is used.

In addition to the elegant formulation of feature selection provided by grafting, it also provides for an efficient means to perform the feature selection. Briefly, the features are considered one-by-one and the “best” feature is chosen (i.e., assigned a non-zero weight). A gradient descent optimization is subsequently performed on the current model to re-optimize the weight vector  $\mathbf{w}$ . Refer to [108] for more details on this update process.

### 5.6.2 Object-Level Feature Selection and Performance

In this section we discuss the feature selection and classification results for both object-level and spatial-relation features.

#### Individual Histologic Object Classes

The grafting method of feature selection was applied to the GT, WSHmin ( $h_{min} = 3$ ), and WSGran ( $r = 4$ ) nuclear, combined cytoplasm and stroma, and cytoplasm segmentations. Additionally, feature selection was applied to a simple connected-components analysis (CC) for all previously mentioned classes and also for stroma segmentation.

Note that the GT segmentation is referring to the use of the GT nuclear segmentation for object delineation. Thus the GT segmentation provides a means to analyze the best possible object- and image-level classification performance assuming a perfect underlying nuclear segmentation. The IS and OS performances (percentage of histologic objects correctly classified) are presented in Table 5.3 for a regularization parameter  $\lambda_1$  of 0.05 and in Table 5.4 for  $\lambda_1 = 0.01$ . Within these tables, the nuclei ('N'), combined cytoplasm and stroma ('CS'), cytoplasm ('C'), and stroma ('S') rows correspond to a feature selection with the assumption that all objects in the benign (malignant) images are benign (malignant). For the 'N2' row, further feedback was elicited from the user on individual regions within the malignant images to better assign the class labels within those images.

As a comparison, for nuclear-level classification (benign versus malignant) accuracies, van de Wouwer et al. [45] achieved 67.1% accuracy (leave-one-out, 9-NN classification), and Weyn et al. [18] achieved 76.1% accuracy (leave-one-out, 1-NN classification). It is important to note, however, that the nuclei in [18] were physically extracted prior to imaging. Thus each nucleus was perfectly segmented by design; this result can thus be compared directly to the GT 'N' and 'N2' results in Tables 5.3 and 5.4. The regularization parameter of  $\lambda_1 = 0.01$  appears to provide both better IS and OS performance. It is very interesting to note that in Table 5.4, the IS classification performance on WSGran segmented nuclei approaches that of the GT nuclei. Additionally, the OS WSGran performance is comparable to that of [45]. The use of user feedback for labeling of the malignant image nuclei does not appear to improve performance significantly, if at all.

There exists no work, to our knowledge, that uses features of cytoplasm and stroma. The initial hypothesis was that cytoplasm and stroma features would be poor classifiers in and of themselves, but would slightly increase the classification performance when combined with nuclear features. Instead, as shown in Tables 5.3 and 5.4, the combined cytoplasm and stroma ('CS'), cytoplasm ('C'), and stroma ('S') classes have a performance generally exceeding that of the nuclei ('N' and 'N2'). Indeed, these performances are comparable to or exceed the nuclear classifications reported in [45] and [18]. It should be noted that the sample size used for the CS and C classes are smaller than the sample size used for the N and N2 classes; this is due to the fact that the Voronoi cells with vertices at infinity (i.e., the edge cells) are not considered. Additionally, the sample size for the S class will differ from the N and N2 classes, since the stroma class is segmented in a connected-components analysis, not relying on the nuclear segmentation for node definitions as do the CS and C segmentations.

Interestingly, in all these results, there is no clearly superior segmentation method among CC, WSHmin, and WSGran for OS imagery. This would seem to indicate that a feature selection scheme can make use of the information contained in any (reasonable) given segmentation for classification purposes. Additionally, in several cases, the performance of the CC, WSHmin, and/or WSGran segmentations approaches the best possible classification of the GT segmentation. Thus, it does not appear that too much information is lost or obscured with improperly segmented nuclei. More information about this

**Table 5.3.** Object-level feature subset performance, with regularization parameter  $\lambda_1 = 0.05$ . In the following, ‘N’ designates nuclei, ‘CS’ combined cytoplasm and stroma, ‘C’ cytoplasm, and ‘S’ stroma. ‘N2’ indicates the use of user-defined labels for nuclei in the malignant images, as opposed to the assumption that all benign (malignant) nuclei are (benign) malignant. Bold entries correspond to the best performance across all the segmentations (excluding GT) for each feature subset.

Object	Performance			
	GT	CC	WSHmin	WSGran
N	0.79/0.77	0.62/0.43	<b>0.66/0.65</b>	0.62/0.61
N2	0.78/0.83	0.64/0.55	<b>0.62/0.61</b>	<b>0.65/0.54</b>
CS	0.74/0.72	0.69/0.56	0.67/0.61	<b>0.73/0.63</b>
C	0.73/0.63	0.65/0.50	<b>0.67/0.60</b>	<b>0.74/0.58</b>
S	N/A	<b>0.66/0.50</b>	N/A	N/A
N + CS	0.75/0.79	0.70/0.57	<b>0.68/0.65</b>	<b>0.74/0.63</b>
N + C	0.77/0.76	0.67/0.52	0.69/ <b>0.65</b>	<b>0.73/0.58</b>

**Table 5.4.** Object-level feature subset performance, with regularization parameter  $\lambda_1 = 0.01$ . In the following, ‘N’ designates nuclei, ‘CS’ combined cytoplasm and stroma, ‘C’ cytoplasm, and ‘S’ stroma. ‘N2’ indicates the use of user-defined labels for nuclei in the malignant images, as opposed to the assumption that all benign (malignant) nuclei are (benign) malignant. Bold entries correspond to the best performance across all the segmentations (excluding GT) for each feature subset.

Object	Performance			
	GT	CC	WSHmin	WSGran
N	0.85/0.77	0.74/ <b>0.69</b>	0.78/0.63	<b>0.81/0.67</b>
N2	0.88/0.79	0.73/0.60	<b>0.75/0.66</b>	<b>0.77/0.64</b>
CS	0.90/0.51	0.82/0.70	0.81/0.60	<b>0.88/0.73</b>
C	0.90/0.34	0.79/0.63	0.82/0.63	<b>0.87/0.72</b>
S	N/A	<b>0.81/0.60</b>	N/A	N/A
N + CS	0.94/0.79	0.84/ <b>0.72</b>	0.84/0.62	<b>0.90/0.70</b>
N + C	0.94/0.68	0.83/0.67	0.85/0.65	<b>0.89/0.71</b>

**Table 5.5.** Multispectral versus ccd object-level features,  $\lambda_1 = 0.05$ , for WSGran segmentations (N, CS, C), and CC segmentation (N, CS, C, S). Bold entries correspond to the best performance for each segmentation for each feature subset.

Object	CC Performance		WSGran Performance	
	ccd	multi	ccd	multi
N	0.62/0.43	<b>0.70/0.49</b>	0.62/ <b>0.61</b>	<b>0.77/0.56</b>
CS	0.69/ <b>0.56</b>	<b>0.80/0.55</b>	0.73/ <b>0.63</b>	<b>0.86/0.57</b>
C	0.65/0.50	<b>0.79/0.60</b>	0.74/0.58	<b>0.86/0.61</b>
S	0.66/0.50	<b>0.73/0.56</b>	N/A	N/A

may be gathered from examination of the feature subsets. As such, the feature subsets for nuclei GT, CC, WSHmin, and WSGran are listed in Appendix J, along with a discussion of the differences in selected features. Correspondence between the feature subset and the qualitative observations of general cancer characteristics will be considered for the best feature subset of this chapter.

### Combined Histologic Object Classes

The combination of nuclear and cytoplasm/stroma features is also shown in Tables 5.3 and 5.4. Note the increase in performance when the CS or C class is combined with the N class (for all segmentations excluding GT); this is a performance increase over any of the individual class performances IS, although CS and C perform better OS than the combined classes. For these results, only the nuclei with corresponding CS or C segmentations are included for the feature selection process. Thus the sample size of the N+CS class is the same as the CS class itself, and similarly with the N+C class.

### Use of Multispectral Object-Level Features

Multispectral features were used for WSGran segmentations of nuclei, combined cytoplasm and stroma, and cytoplasm; multispectral features were also used for CC segmentation of nuclei, combined cytoplasm and stroma, cytoplasm, and stroma. Classification results after grafting-based feature selection are shown in Tables 5.5 and 5.6 along with the corresponding results using ccd features for comparison.

While the use of multispectral features improves the IS performance, the OS performance is significantly degraded in many cases (at best it is equivalent to the OS ccd performance). This indicates that the multispectral features provide information that allows the classifier to more readily overfit to the IS data. This degradation in performance can also be understood as the multispectral features being much less generalizable than the corresponding RGB features when applied to unseen imagery. This is an interesting

**Table 5.6.** Multispectral versus ccd object-level features,  $\lambda_1 = 0.01$ , for WSGran segmentations (N, CS, C), and CC segmentation (N, CS, C, S). Bold entries correspond to the best performance for each segmentation for each feature subset.

Object	CC Performance		WSGran Performance	
	ccd	multi	ccd	multi
N	0.74/ <b>0.69</b>	<b>0.91</b> /0.63	0.81/ <b>0.67</b>	<b>0.90</b> /0.58
CS	0.81/ <b>0.75</b>	<b>0.95</b> /0.58	0.86/ <b>0.69</b>	<b>0.96</b> /0.64
C	0.83/ <b>0.73</b>	<b>0.93</b> / <b>0.73</b>	0.85/ <b>0.76</b>	<b>0.95</b> /0.69
S	0.81/ <b>0.60</b>	<b>0.95</b> / <b>0.60</b>	N/A	N/A

**Table 5.7.** Spatial-relation feature subset performance for nuclei. These results are image-level results since there is only one set of features derived per image. Bold entries correspond to the best performance across all the segmentations (excluding GT) for each feature subset.

Object	Performance			
	GT	CC	WSHmin	WSGran
$\lambda_1 = 0.05$	1.00/0.83	<b>1.0</b> /0.55	0.97/ <b>0.69</b>	0.97/0.48
$\lambda_1 = 0.01$	1.00/0.62	<b>1.00</b> /0.55	<b>1.00</b> / <b>0.69</b>	<b>1.00</b> /0.41

result in light of the approximately equivalent performance of multispectral and RGB imagery for the various classification tasks considered in Chapters 2 and 3 and the superiority of RGB imagery for segmentation tasks presented in Chapter 4. Thus, not only does multispectral imagery provide no clear benefit for pixel-level classification, or object-level segmentation, but it also appears to significantly degrade the OS performance of object-level feature selection.

### 5.6.3 Spatial-Relation Feature Selection and Performance

The results from applying grafting to the spatial-relation features is shown in Table 5.7 for feature extraction over the entire truth window (i.e., without implementation of the sliding window approach discussed in Section 5.5). It is important to remember that these results are for classification of the entire image based on spatial-relation features. Results using the region-based approach of Section 5.5 are shown in Table 5.8; thus, these results are object-level classification.

The performance of spatial-relation features on a non-regional basis (Table 5.7) are surprisingly poor. Some of this poor performance may be attributed to the use of a graph that encompasses significantly different spatial attributes, e.g., a graph including portions of normal tissue and portions of a tumor. The use of regional-based spatial-

**Table 5.8.** Region-based spatial-relation feature subset performance for nuclei. These results are object-level results since the region-based spatial approach allows for spatial attributes to be associated with individual image objects. Bold entries correspond to the best performance across all the segmentations (excluding GT) for each feature subset. The CC segmentation is excluded due to computational constraints.

Object	Performance		
	GT	WSHmin	WSGran
$\lambda_1 = 0.05$	0.75/0.58	0.65/ <b>0.60</b>	<b>0.70</b> /0.56
$\lambda_1 = 0.01$	0.84/0.58	0.77/ <b>0.63</b>	<b>0.79</b> /0.50

relation features (Table 5.8), however, which should minimize the inclusion of different spatial attributes, also yields poor performance.

We hypothesize two reasons for the poor performance of spatial-relation features. First, it is possible that the features which have been extracted from the imagery are not appropriate features for elucidation of the spatial arrangement attributes of the tissue. Second, it is possible that the spatial arrangement attributes are not important characteristics in and of themselves to distinguish cancerous from non-cancerous conditions. We will consider each of these possibilities in more detail.

Evidence related to the first possibility, that the spatial-relation features extracted are not appropriate for the classification task, can be found by examining the IS and OS performance of the GT nuclei. Unfortunately, it is difficult to directly compare the results in Tables 5.7 and 5.8 since the former contains image-level classification performances and the latter contains region-level classification performances. In Table 5.7, we see that the OS performance for non-regional spatial-relation classification is actually quite good (0.83) for  $\lambda_1 = 0.05$ . This lends support in favor of the spatial-relation features being appropriate for classifying the cancerous imagery. By noting the poor OS performances for the nuclear segmentations (CC, WSHmin, WSGran), however, it also indicates that these features, unlike the object-level features, are quite reliant on a good underlying nuclear segmentation. Observation of the region-level classification results in Table 5.8, however, result in poor OS performance across all nuclear segmentations, including GT. This indicates that even with a perfect nuclear segmentation, that the spatial-relation features cannot elicit appropriate information about the cancerous nature of the imagery. So, on the one hand there exists evidence to support the utility of the spatial-relation features and also the uselessness of the spatial-relation features. We will return to this issue in our discussion of image-level classification performance in Section 5.6.5.

For the second possibility, that spatial-relation features in and of themselves are not useful in distinguishing cancer from non-cancer, the performance of combined object- and spatial-relation features are analyzed in the following section. First, however, we will address this issue in more general terms. It is well accepted among pathologists that

the spatial arrangement of histologic entities is of utmost importance for many visual tasks, including cancer diagnosis and even classification of tissue origin of cells. The removal of all object-level information, retaining solely spatial information in the form of nodes (i.e., the nuclei centroids) as is done for the extraction of the spatial-relation features removes a lot of the visual cues that the pathologist uses for analysis. As an example of this, consider the example images in Figure 5.7, where nuclei centroids, nuclei ground truth masks, and the underlying RGB images are shown for an example benign and malignant image. The classification of the images is difficult given solely the spatial information, i.e., the nuclei centroids. The addition of object-level information, in the form of binary ground truth object masks, increases the visual cues available, e.g., size of nuclei, irregularity of nuclear boundaries. Finally, the addition of the low-level cues from the RGB image (e.g., color and texture) adds yet another layer to the visual cues that can be used for a classification. This provides an empirical example of the possibility that the spatial-relation features alone may not be particularly indicative of the cancer state of the imagery, and that the inclusion of other features, namely object- and low-level features, are needed for complete description of the imagery. It should be noted that this has not been tested or verified with a pathologist; this could provide for very interesting future work.

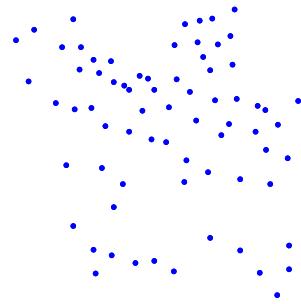
#### 5.6.4 Combining Object-Level and Spatial-Relation Features

As motivated in the previous section, we apply grafting to the combination of spatial-relation and object-level features. Results are shown in Tables 5.9 and 5.10. The inclusion of region-based spatial-relation features provides a slight improvement in some OS GT performance, but does not improve OS performance for the other segmentations as compared to the object-level feature performances in Tables 5.3 and 5.4, especially for  $\lambda_1 = 0.01$ . These results seem to indicate once again the reliance of the spatial-relation features on a very good underlying nuclear segmentation. Assuming a near perfect nuclear segmentation, however, the possible improvement in performance is still very small, and it is unclear that this performance increase is even to be expected.

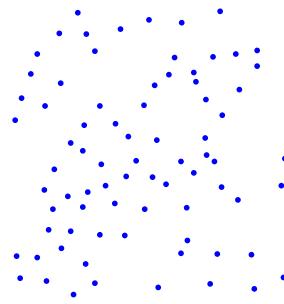
#### 5.6.5 Image-Level Performances

Image-level performances are demonstrated with Receiver Operator Characteristic (ROC) curves, varying the threshold of malignant objects for which an image is considered malignant. ROC curves for OS imagery are shown in Figure 5.8, where the grafting results for object level feature subsets,  $\lambda_1 = 0.01$ , were used. Additionally, the AUC for each ROC curve is shown both in Figure 5.8 and Table 5.11.

The image-level performance using object-level features is very good, resulting in AUC values above 0.9 in many cases (refer to Figure 5.8 and Table 5.11). Again, similar



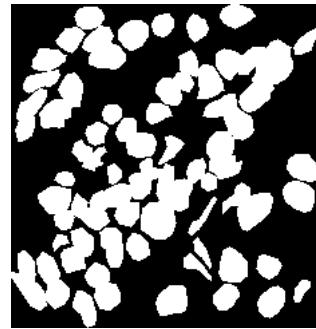
(a) Centroids of benign nuclei.



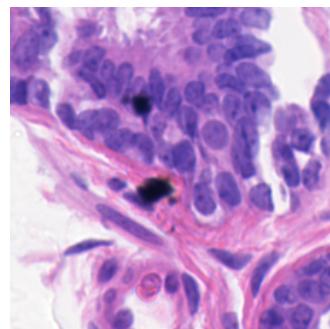
(b) Centroids of malignant nuclei.



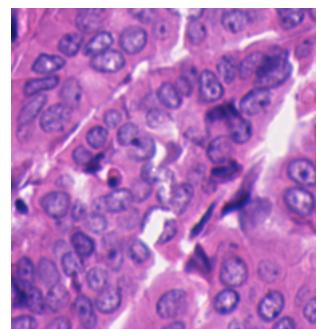
(c) Benign ground truth nuclei.



(d) Malignant ground truth nuclei.



(e) Benign RGB image.



(f) Malignant RGB image.

**Figure 5.7.** Different levels of visual information: nuclei centroids in (a) and (b), binary object masks in (c) and (d), and the RGB images in (e) and (f).

**Table 5.9.** Object-level and spatial-relation feature subset performance,  $\lambda_1 = 0.05$ . In the following, ‘N’ designates nuclei, ‘SPN’ spatial-relation nuclei features, ‘CS’ combined cytoplasm and stroma, ‘C’ cytoplasm, and ‘S’ stroma. Bold entries correspond to the best performance across all the segmentations (excluding GT) for each feature subset. The CC segmentation is excluded due to computational constraints.

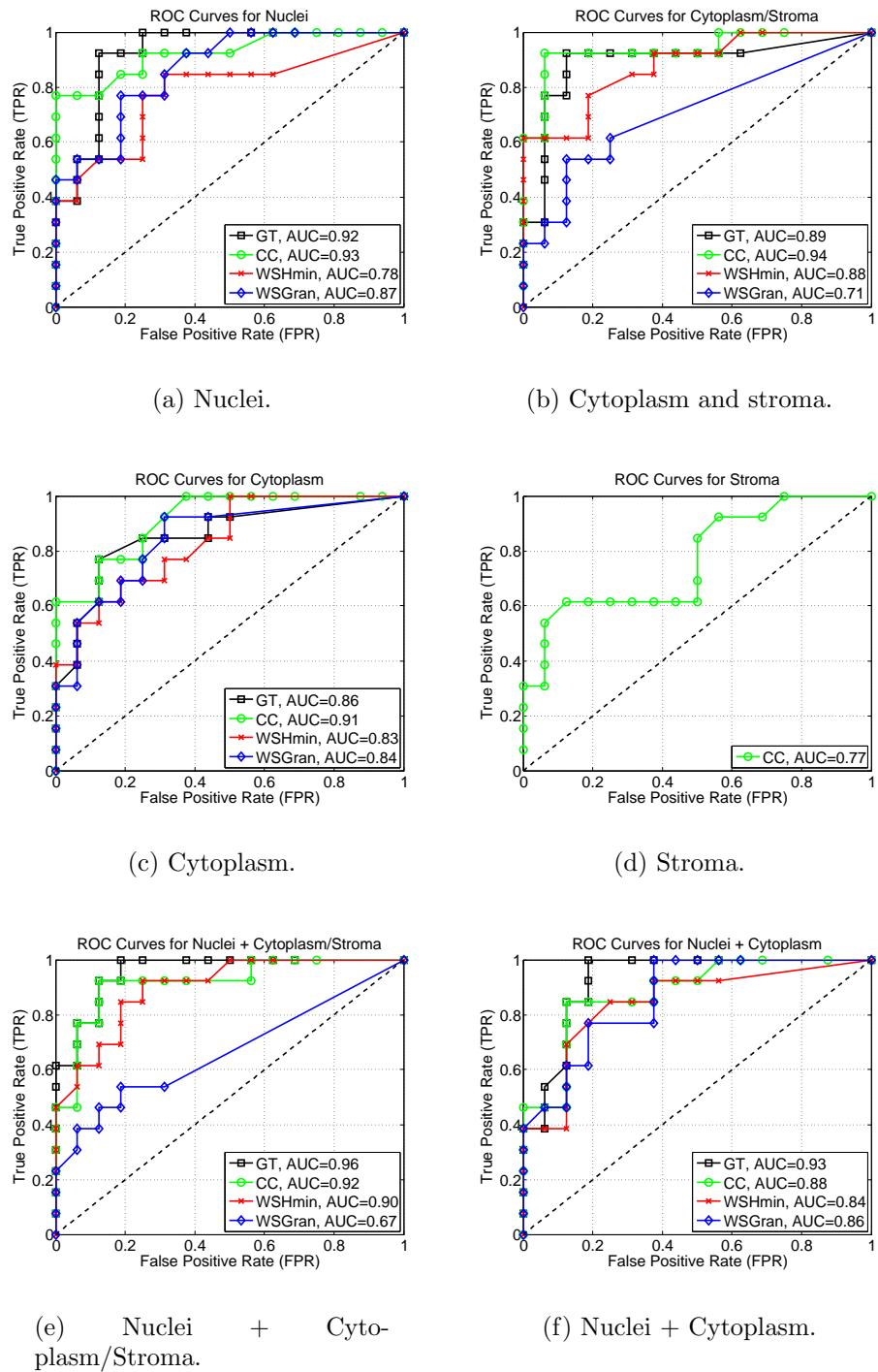
Object	Performance		
	GT	WSHmin	WSGran
N + SPN	0.78/0.71	0.67/ <b>0.67</b>	<b>0.76</b> /0.57
N + CS + SPN	0.82/0.72	0.72/ <b>0.69</b>	<b>0.78</b> /0.61
N + C + SPN	0.80/0.72	0.73/ <b>0.69</b>	<b>0.78</b> /0.59

**Table 5.10.** Object-level and spatial-relation feature subset performance,  $\lambda_1 = 0.01$ . In the following, ‘N’ designates nuclei, ‘SPN’ spatial-relation nuclei features, ‘CS’ combined cytoplasm and stroma, ‘C’ cytoplasm, and ‘S’ stroma. Bold entries correspond to the best performance across all the segmentations (excluding GT) for each feature subset. The CC segmentation is excluded due to computational constraints.

Object	Performance		
	GT	WSHmin	WSGran
N + SPN	0.93/0.78	0.84/ <b>0.63</b>	<b>0.86</b> /0.56
N + CS + SPN	0.96/0.76	0.89/ <b>0.72</b>	<b>0.91</b> /0.64
N + C + SPN	0.96/0.72	0.89/ <b>0.72</b>	<b>0.93</b> /0.65

**Table 5.11.** AUC for OS image-level classification for grafting with  $\lambda_1 = 0.01$ . In the following, ‘N’ designates nuclei, ‘CS’ combined cytoplasm and stroma, ‘C’ cytoplasm, and ‘S’ stroma. Bold entries correspond to the best performance across all the segmentations (excluding GT) for each feature subset.

Object	Performance			
	GT	CC	WSHmin	WSGran
N	0.92	<b>0.93</b>	0.78	0.87
CS	0.89	<b>0.94</b>	0.88	0.71
C	0.86	<b>0.91</b>	0.83	0.84
S	N/A	<b>0.77</b>	N/A	N/A
N + CS	0.96	<b>0.92</b>	0.90	0.67
N + C	0.92	<b>0.88</b>	0.84	0.86



**Figure 5.8.** ROC curves for OS image-level performance (grafting,  $\lambda_1 = 0.01$ ), varying the threshold of malignant objects for which an image is considered malignant.

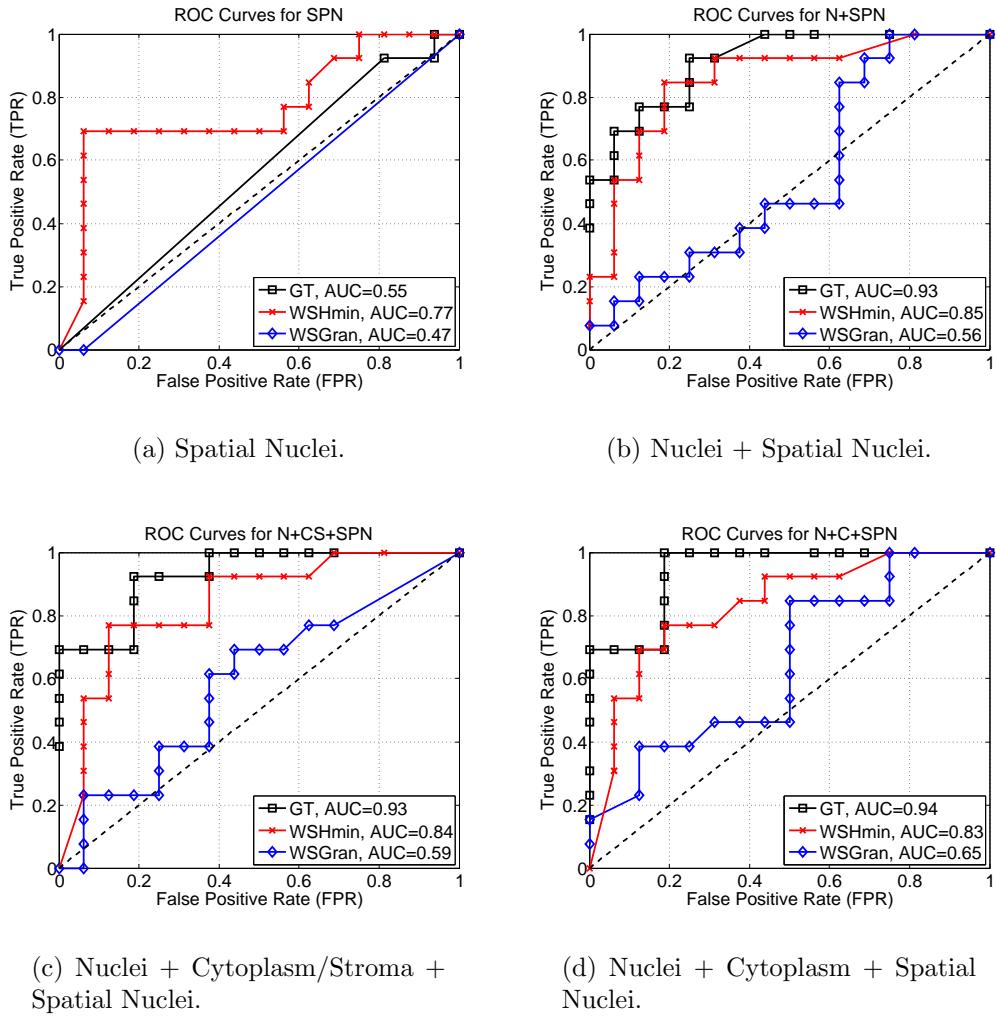
**Table 5.12.** AUC for OS spatial-relation and combination object-level and spatial-relation classification for grafting with  $\lambda_1 = 0.01$ . In the following, ‘N’ denotes object-level nuclei, ‘SPN’ spatial nuclei, ‘CS’ combined cytoplasm and stroma, and ‘C’ cytoplasm. Bold entries correspond to the best performance across all the segmentations (excluding GT) for each feature subset. The CC segmentation is excluded due to computational constraints.

Object	Performance		
	GT	WSHmin	WSGran
SPN	0.55	<b>0.77</b>	0.47
N + SPN	0.93	<b>0.85</b>	0.56
N + CS + SPN	0.93	<b>0.84</b>	0.59
N + C + SPN	0.94	<b>0.83</b>	0.65

to the object-level performance in Tables 5.3 and 5.4, the cytoplasm/stroma and cytoplasm classes provide a better AUC than the nuclei class in most cases. It is interesting that for the image-level classification, the CC segmentation provides consistently better performance than the other segmentations. This is markedly different than the object-level classification performances, where the WSGran segmentation tended to provide the best performance. This is a very interesting result in light of the performance differences for nuclear segmentation of the two classifiers, where WSGran had a best average performance of 0.34 and CC of 0.09.

Additionally, the combined classes here provide worse AUC values than the individual object classes (with the exception of the stroma class). For these results, the combined cytoplasm and stroma class provides the best AUC score of 0.94 for CC segmentation. It is also interesting that the image-level performances for the CC segmentation outperform the GT segmentation.

Looking at the image-level performances of spatial-relation and combined object-level and spatial-relation features, we see very poor AUC values for all segmentations but GT (refer to Figure 5.9 and Table 5.12). This corroborates our hypothesis that the spatial-relation features are highly dependent on the accuracy of the underlying nuclear segmentation, and that the spatial-relation features alone are not useful for classification purposes.



**Figure 5.9.** ROC curves for OS spatial-level and combination object-level and spatial-relation performance (grafting,  $\lambda_1 = 0.01$ ), varying the threshold of malignant objects for which an image is considered malignant.

**Table 5.13.** Statistics of CC object-level feature subsets. Statistics presented include the percentage of OD and non-OD features; R, G, B, I/Hue, and binary features; texture, radiometric and densitometric, and size/shape features; and nuclei and non-nuclei features.

Category	N	CS	C	S	N+CS	N+C
<b>OD</b>	0.84	0.65	0.62	0.64	0.70	0.65
<b>Non-OD</b>	0.16	0.35	0.38	0.36	0.30	0.35
<b>R</b>	0.26	0.25	0.33	0.22	0.24	0.29
<b>G</b>	0.19	0.23	0.16	0.37	0.22	0.17
<b>B</b>	0.26	0.27	0.21	0.13	0.25	0.27
<b>I/Hue</b>	0.28	0.23	0.25	0.25	0.27	0.23
<b>Binary</b>	0.00	0.02	0.05	0.01	0.02	0.04
<b>Texture</b>	0.51	0.59	0.62	0.67	0.58	0.61
<b>Radio-/Densito-metric</b>	0.19	0.18	0.11	0.12	0.14	0.11
<b>Size/Shape</b>	0.30	0.23	0.27	0.21	0.28	0.29
<b>Nuclei</b>	1.00	0.00	0.00	0.00	0.27	0.41
<b>Non-Nuclei</b>	0.00	1.00	1.00	1.00	0.73	0.59

## 5.7 Feature Subsets

### 5.7.1 Object-Level Feature Subsets

The grafting feature subsets for  $\lambda_1 = 0.01$  are presented in Appendix K for each of the objects and combined objects for the CC segmentation. Table 5.13 presents various statistics of the feature subsets, namely the percentage of OD and non-OD features; R, G, B, I/Hue, and binary features; texture, radiometric and densitometric, and size/shape features<sup>4</sup>; and nuclei and non-nuclei features.

All feature subsets display a preference for OD features, particularly the nuclei subset. This is not surprising given the that OD features provide a linear relationship between image intensity and stain density (refer to Appendix H). Additionally, it has been shown that this linear relationship is particularly useful in quantifying characteristics of nuclear chromatin [151, 152].

The preference for different color features (R, G, B, I/Hue) is roughly uniform, with the cytoplasm subset slightly favoring red features, and the stroma subset slightly favoring green features. The preference for red features in the cytoplasm subset carries over into the nuclei and cytoplasm subset, albeit with less preference for red features.

---

<sup>4</sup>For this categorization, we are considering the various chromatin-specific features to fall into either the radiometric/densitometric or size/shape categories.

The texture features are used more often in the CS, C, S subsets (as well as the combined N+CS and N+C subsets), and the size and shape features are used more often in the N and C subsets. This indicates that while nuclear chromatin texture (and cytoplasm texture) is an important characteristic to distinguish cancerous from non-cancerous nuclei (cytoplasm), that size and shape also play an important role. Not surprisingly, given our previous discussion (refer to Section 3.4) regarding the strong texture of stroma, the stroma subset has the largest reliance on texture features of any object-level feature subset.

In the two combined object-level feature subsets, N+CS and N+C, there is a stronger preference for the non-nuclear features than for the nuclear features. Given the good performance of the cytoplasm/stroma and cytoplasm classes, we might expect that there would be a roughly equivalent representation in the feature subsets when combining nuclear and non-nuclear features. There is a slight bias towards cytoplasm features in the nuclear and cytoplasm subset, and a stronger bias towards cytoplasm/stroma features in the nuclear and cytoplasm/stroma feature subset. Thus, in addition to the good performance of the individual cytoplasm/stroma and cytoplasm classes, these features are also very important when combined with nuclear features.

### 5.7.2 Important Object-Level Features

Useful features in all feature subsets include the area-based fractal dimension features (*ODAreaFD*) as well as measures of integrated optical density (*IOD* and *DNAArea* features), number of chromatin regions (*LNum*, *MNum*, and *HNum*), and hue. Interestingly, the inclusion of the hue feature was motivated by the use of hue for characterization of cytoplasm regions in [9], but has been shown to be a useful feature for nuclei classification here. Additionally, some measures of texture are used in many subsets, including preferentially the long-runs emphasis (*LRE*), short-runs emphasis (*SRE*), and low gray-level runs emphasis (*LGRE*). These texture features have a direct relation to the homogeneity of the objects; for nuclei in particular these features are related to the clumping of the chromatin, which is one of the features indicated as indicative of cancer state by pathologists.

### 5.7.3 Spatial-Relation and Combined Object and Spatial Feature Subsets

The grafting feature subsets for  $\lambda_1 = 0.01$  are presented in Appendix K for spatial nuclei and combined object-level and spatial-relation features for the WSGran segmentation. Table 5.14 presents statistics of the feature subsets, namely the percentage of OD and non-OD features; R, G, B, I/Hue, and binary features; texture, radiometric and densitometric, and size/shape features; nuclei and non-nuclei features; VT, DT, MST,

**Table 5.14.** Statistics of WSGran spatial-relation feature subsets. Statistics presented include the percentage of OD and non-OD features; R, G, B, I/Hue, and binary features; texture, radiometric and densitometric, and size/shape features; nuclei and non-nuclei features; VT, DT, MST, OCG, and CG features; and total percentage of object-level and spatial-relation features.

Category	SPN	N+SPN	N+CS+SPN	N+C+SPN
<b>Object-Level</b>				
<b>OD</b>	N/A	0.63	0.58	0.59
<b>Non-OD</b>	N/A	0.37	0.42	0.41
<b>R</b>	N/A	0.31	0.27	0.26
<b>G</b>	N/A	0.19	0.23	0.27
<b>B</b>	N/A	0.23	0.27	0.20
<b>I/Hue</b>	N/A	0.27	0.18	0.19
<b>Binary</b>	N/A	0.00	0.05	0.08
<b>Texture</b>	N/A	0.62	0.61	0.53
<b>Radio-/Densito-metric</b>	N/A	0.19	0.19	0.17
<b>Size/Shape</b>	N/A	0.19	0.20	0.30
<b>Nuclei</b>	N/A	1.00	0.36	0.43
<b>Non-Nuclei</b>	N/A	0.00	0.64	0.57
<b>Spatial-Relation</b>				
<b>VT</b>	0.30	0.32	0.35	0.30
<b>DT</b>	0.09	0.09	0.05	0.05
<b>MST</b>	0.13	0.23	0.20	0.20
<b>OCG</b>	0.35	0.27	0.30	0.35
<b>CG</b>	0.13	0.05	0.10	0.10
<b>Object-Level</b>	0.00	0.70	0.84	0.84
<b>Spatial-Relation</b>	1.00	0.30	0.16	0.16

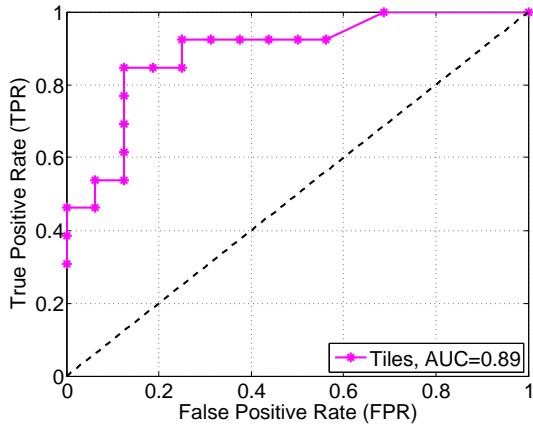
OCG, and CG features; and total percentage of object-level and spatial-relation features.

All object-level statistics remain similar to those presented in Table 5.13. Additionally, choice of graph type for spatial-relation feature extraction is consistent across all the feature subsets. In all the combination object-level and spatial-relation subsets, there is a strong preference for the object-level features over those of the spatial-relation features. Thus, while the spatial-relation features may provide more IS classification accuracy, there is still a preference for the object-level features in the combined object- and spatial-relation feature subsets. Given the IS and OS performance of the spatial-relation features alone, it is likely that the inclusion of the spatial-relation features is what degrades the performance of the combined object-level and spatial-relation classifiers for OS data.

## 5.8 Summary

In this chapter we have described a comprehensive list of object-level features and demonstrated the use of these features on various histologic objects. It has been demonstrated that these object-level features are versatile and general enough to elicit important information from even imperfectly segmented objects. This was demonstrated with the object- and image-level classification performance for the ground truth nuclei versus three different segmentations of nuclei, namely connected-components (CC), WSHmin, and WSGran. The latter two segmentations were described and analyzed in Chapter 4. We have presented the use of non-nuclear features, namely features of cytoplasm and stroma, and their good classification performance, often exceeding that of nuclei. While the use of cytoplasm and/or stroma features has been mentioned in the literature, it has not previously been studied to any extent. Furthermore, we find that the cytoplasm and stroma features have object- and image-level classification performances often exceeding that of nuclei, which are considered the most important histologic objects for cancer characterization. This is a very interesting result that warrants further investigation as well as consultation with pathology experts.

Using object-level features and the grafting method [108] of feature selection, we have shown object-level classification accuracies above 0.70 for OS data. These results are comparable to others presented in the literature, namely [45] and [18], which relied on either the physical extraction of cell nuclei prior to imaging [18] or the interactive selection of well-segmented nuclei [45]. We hypothesize that it is the use of a comprehensive set of features that allows for the use of imperfectly segmented objects, at least for this particular application. As a point of comparison, the feature extraction and grafting-based feature selection for objects defined as non-overlapping tiles of  $50 \times 50$  pixels yields a tile-level performance of 0.92 IS and 0.68 OS for  $\lambda_1 = 0.01$ . This corresponds to an AUC of 0.89; as reference, the ROC curve for tiles is shown in Figure 5.10. Thus, with attention to the pixel-level classification and segmentation of histologic objects, we can expect an



**Figure 5.10.** ROC curve for tiles.

increase in performance (refer to Tables 5.4 and 5.11), but the achievement of near-perfect object-level segmentation may be unnecessary in this application. Other researchers, namely Mar  e et al. [171, 172] have demonstrated the use of “random subwindows” for image classification, similar to this tile-based approach here.

We have also presented a comprehensive list of spatial-relation features for use in quantifying the spatial arrangement aspects of nuclei in our imagery, including a method for assigning these spatial-relation features to a region of an image. It was expected that the spatial-relation features would be particularly useful in the characterization of cancerous versus non-cancerous tissue. We have shown, however, that these features suffer from very poor out-of-sample performance even for the perfect GT segmentation. Additionally, the application of these features to the CC, WSHmin, and WSGran segmentations does not yield good results. It thus appears that the spatial-relation features are either not appropriately defined to quantify the spatial characteristics that human experts are cueing on, or that these features are particularly dependent on a good underlying object segmentation. Evidence, as presented in this chapter, seems to indicate the latter: that these features appear to be dependent on a good object-level segmentation. Even in combination with object-level features, the spatial-relation features tend to degrade the OS performance for the CC, WSHmin, and WSGran segmentations.

Lastly, we have presented and discussed the various feature subsets as determined by the grafting algorithm. We have noted a strong preference for the optical density-related features, and found that texture features are important for classification of all histologic classes considered here. Additionally, for the combined object-level feature subsets, there is a stronger reliance on the non-nuclear features, reflecting the surprisingly better performance of the non-nuclear features when considered alone. Similarly, for the combined object and spatial feature subsets, there is a preference for the object-level features which were shown to perform better in both object- and image-level classification.

## 5.9 Future Directions

The good performance of the cytoplasm and stroma features provides an intriguing possibility for application to other datasets including other tissues (e.g., prostate, colon). The focus on histopathology imagery was originally motivated in part by the desire to study the spatial arrangement aspects of the imagery. Additionally, it will be important to study the use of these object-level features on cytology imagery, especially in light of the poor performance of spatial-relation features.

It would also be very interesting to apply the object-level feature extraction and selection in a completely different application domain, e.g., remote sensing. Of particular interest here, beyond the classification performance of this approach, would be the dependence of classification performance on the accuracy of the underlying object segmentations.

While the spatial-relation features have not been shown to be useful in this study, there is definite cause for further analysis of these features. Other researchers have shown the utility of some spatial-relation features for certain image analysis tasks, e.g., [41] and [22]. Further study of the spatial-relation features and how they may be better tailored to human intuition about spatial arrangement of objects is certainly an open field of research.

# Chapter 6

## Higher-Level Objects

We apply the size and shape features from Chapter 5 to higher-level objects, namely “cells” and “cell clusters.” Thus, we are agglomerating the first-level objects (nuclei, cytoplasm, etc.) into higher-level objects. These higher-level features show promise for increasing the image-level classification performance, but are dependent on a good underlying low-level (i.e., nuclei) segmentation). Additionally, we discuss the plausibility of using a “feature graph” representation for the various features. We conclude that this is not a feasible approach for this particular application, especially given the poor performance of spatial features as demonstrated in Chapter 5. Lastly, we introduce an image representation method in the form of probabilistic graph models and indicate possible further uses for such an image representation.

### 6.1 Motivation

Discussed in the previous chapter were different levels of feature metrics required for object-level analysis, namely object-level (including many that can be considered low-level) and spatial-relation level. In this chapter we extend the use of the object-level features to higher-level image objects. This work is motivated, in part, by the biological evidence for the use of hierarchies (and reverse hierarchies) in human visual perception [173], building from the simple receptive fields of retinal photoreceptors, developing intermediate generalizations of these simple features, and culminating with scene comprehension/perception.

### 6.2 Related Work

Literature searches for similar research yielded only a few results related to this research effort:

- The object-oriented approach to tissue analysis in [65] uses the hierarchical concept of a “Slide-Unit-Cell” structure to facilitate analysis. It does not appear, however, that this technique specifically utilizes a hierarchical feature extraction for analysis tasks. Additionally, this method is not adaptable for different analysis tasks.
- The Cellenger<sup>®</sup> system from Definiens AG (München, Germany), previously described in Section 1.4.1, uses a hierarchical concept of image objects and associates related features with these objects, but is not automated in the general sense [66,69]. Within Cellenger<sup>®</sup>, the initial task of setting up the image analysis requires knowledge and experience in image processing as well as programming skills. Image analysis tasks are specified in a modular fashion using existing analysis tasks common to medical image analysis.

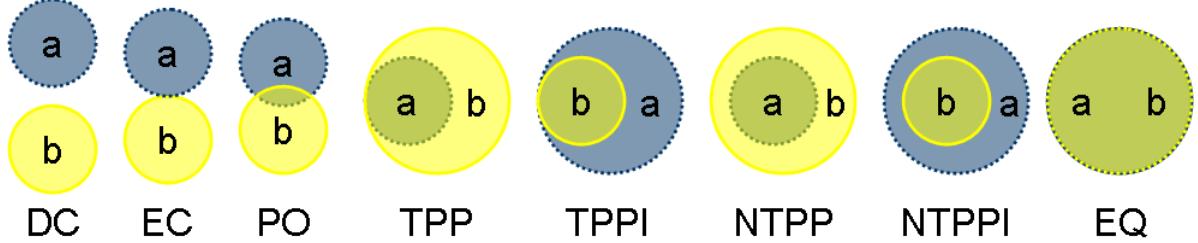
Section 6.3 discusses the segmentation method used to group histologic objects. Section 6.4 discusses the creation and use of higher-level objects in more detail, including results of classification using such objects. A summary is presented in Section 6.5 and future work, including discussion of a probabilistic graph model of the imagery is presented in Section 6.6.

## 6.3 Higher Level Segmentation

For higher-level segmentation, the Region Connection Calculus [174] (RCC) was chosen. RCC<sup>1</sup> is a widely known and studied approach for spatial representation and the general artificial intelligence problem of Qualitative Spatial Reasoning (QSR) [175, 176]. The general term “reasoning” as used in QSR refers to the following problem: given the relationships between regions  $x$  and  $y$  as  $R_1(x, y)$  and between  $y$  and  $z$  as  $R_2(y, z)$ , what can be determined about the relationship  $R_3$  between  $x$  and  $z$ ? This reasoning problem may be NP-complete [175–178] and will not be of initial concern in this work, but instead RCC will be used as a means to represent regions and pairwise relations between regions. The general problem of reasoning over three regions in RCC may be useful in future work, at which point other research on tractable subsets of RCC, e.g., [177–179], may be of use.

---

<sup>1</sup>The acronym RCC was originally derived from the last names of the three authors of the original paper, Randall, Cui, and Cohn [174]. The authors have, however, accepted the more common description of Region Connection Calculus as “a very apt description of [their] spatial formalism.” [175]



**Figure 6.1.** The eight base relations of RCC8.

### 6.3.1 An Introduction to the Region Connection Calculus (RCC)

RCC is based on a single primitive connection operator  $C$ , which may be defined as desired, as long as it is reflexive and symmetric [175], i.e.,

$$\forall x[C(x, x)] \quad (6.1)$$

$$\forall x\forall y[C(x, y) \rightarrow C(y, x)] \quad (6.2)$$

Two common Subsets of RCC, known as RCC8 and RCC5, contain 8 and 5 jointly exhaustive and pairwise disjoint (JEPD) base relations, respectively, developed using  $C$  [175]. The eight base relations of RCC8 are illustrated in Figure 6.1, showing disconnected (DC), externally connected (EC), partial overlap (PO), tangential proper part and its inverse (TPP and TPPI), nontangential proper part and its inverse (NTPP and NTPPI), and equivalent (EQ). For RCC5 relations, the boundary of an object is not considered, such that DC and EC are combined to become DR (discrete), and TPP and NTPP become PP (proper part) [178]. In terms of  $C$ , these relations (plus intermediate relations part (P) and overlap (O)) are defined as [175]:

$$DC(x, y) \equiv \neg C(x, y) \quad (6.3)$$

$$P(x, y) \equiv \forall z[C(z, x) \rightarrow C(z, y)] \quad (6.4)$$

$$O(x, y) \equiv \exists z[P(z, x) \wedge P(z, y)] \quad (6.5)$$

$$EC(x, y) \equiv C(x, y) \wedge \neg O(x, y) \quad (6.6)$$

$$DR(x, y) \equiv \neg O(x, y) \quad (6.7)$$

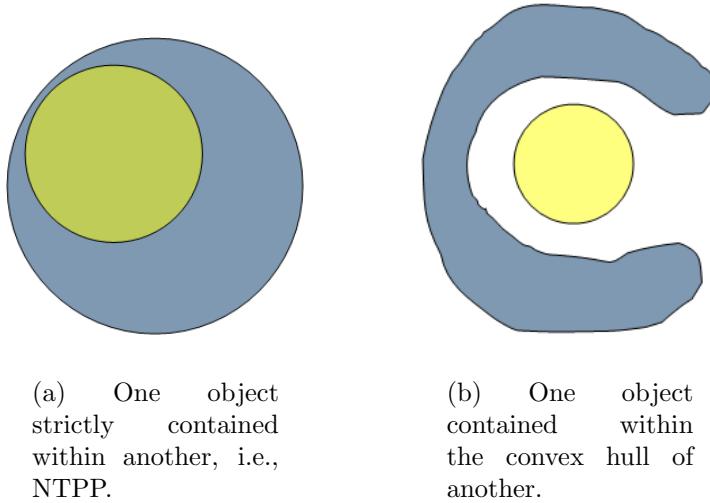
$$PO(x, y) \equiv O(x, y) \wedge \neg P(x, y) \wedge \neg P(y, x) \quad (6.8)$$

$$PP(x, y) \equiv P(x, y) \wedge \neg P(y, x) \quad (6.9)$$

$$TPP(x, y) \equiv PP(x, y) \wedge \exists z[EC(z, x) \wedge EC(z, y)] \quad (6.10)$$

$$NTPP(x, y) \equiv PP(x, y) \wedge \nexists z[EC(z, x) \wedge EC(z, y)] \quad (6.11)$$

$$EQ(x, y) \equiv P(x, y) \wedge P(y, x) \quad (6.12)$$



**Figure 6.2.** Example higher-level spatial reasoning using RCC8 and additional predicates.

The applicability of these operators to biological cells is obvious from the illustration in Figure 6.1. The sets of eight or five relations is commonly expanded with the use of Boolean functions **sum** (+), **compl** (complement), **prod** (\*,product/intersection), and **diff** (-, difference). Indeed, RCC can be reformulated as a Boolean connection algebra [180, 181].

Additional expansions can be achieved by the creation of new predicates by combination of JEPD relations and Boolean functions. Such extensions include methods to describe shape via predicates defined on the dissection-graphs of regions [175]. Another intriguing predicate is the convex hull, which can allow for distinctions between such spatial relations as those shown in Figure 6.2 [175]. The convex hull predicate is defined as [175]:

$$\text{CONV}(x) \equiv \text{EQ}(x, \text{conv}(x)) \quad (6.13)$$

where  $\text{conv}(x)$  denotes the convex hull of region  $x$ .

The use of RCC and extensions has the benefit of a solid mathematical foundation, much of which has been studied extensively [176, 177, 182–185]. Additionally, there is evidence that RCC8 is “conceptual cognitive adequate,” a term defined as “empirical evidence supports the assumption that a system of relations is a model of people’s conceptual knowledge of spatial relationships” [176, 178]. Incidentally, RCC5 was not found to be conceptual cognitive adequate. While there is much literature published on the theoretical aspects of RCC, there are very few publications with demonstrated applications or results [176].

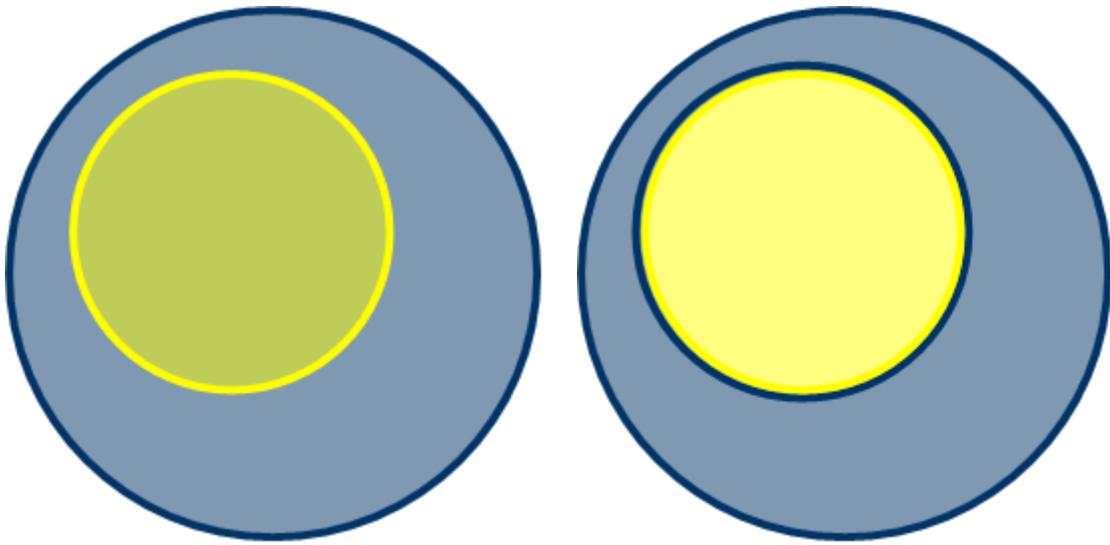
### 6.3.2 Implementation of RCC

While concepts of RCC and other spatial reasoning tend to shy away from pixel-based representations of objects and regions, there are not any proposed methods of avoiding the use of pixels. The spatial relation between two regions,  $x$  and  $y$  can be defined in RCC8 using set operations on the pixels within each region with the following process, where  $p(\cdot)$  is the perimeter of a region, and  $d(\cdot)$  is the Euclidean distance function:

- If  $x \subset y$ 
  - If  $p(x) \cap p(y) = \emptyset \Rightarrow \text{NTPP}(x, y)$
  - Else  $\Rightarrow \text{TPP}(x, y)$
- Else if  $y \subset x$ 
  - If  $p(x) \cap p(y) = \emptyset \Rightarrow \text{NTPPI}(x, y)$
  - Else  $\Rightarrow \text{TPPI}(x, y)$
- Else if  $x \equiv y \Rightarrow \text{EQ}(x, y)$
- Else if  $x \cap y = \emptyset$ 
  - If  $\min(d(x, y)) > 1 \Rightarrow \text{DC}(x, y)$
  - Else  $\Rightarrow \text{EC}(x, y)$
- Else if  $x \cap y \neq \emptyset \Rightarrow \text{PO}(x, y)$

Similarly, the spatial relation between the two convex hulls of regions  $x$  and  $y$ , can be determined by the same process as above, replacing  $x$  and  $y$  by  $\text{conv}(x)$  and  $\text{conv}(y)$ , respectively.

As generally modeled, cytoplasm is assumed to contain the cell nuclei, reflecting an extrapolation of the two objects to three dimensions. Thus, while the segmentation processes described in Chapter 4 represents the cytoplasm as (ideally) an annulus around the nuclei, human intuition about the relationship is not an EC one but rather a NTPP one. Refer to Figure 6.3 for an illustration. Thus, the use of the convex hull operator can allow for more intuitive modeling of relationships between histologic objects, particularly when the objects have been segmented based solely on visual properties and not based on any *a priori* shape information. This may extend into other application domains such as remote sensing; as an analogy, a lake in the middle of a forest would most likely be considered part of the forest rather than externally connected). Note that, in this application, the decision of the RCC relation between two objects is independent of the feature extraction process for those same objects.



(a) NTPP: The blue region contains the same area as the yellow region. This corresponds with common models of cell cytoplasm and nuclei.

(b) EC: The blue region is defined only up to the boundary of the yellow region; thus, they share a boundary and are externally connected. This corresponds with the segmentations of cytoplasm and nuclei as discussed in Chapter 4.

**Figure 6.3.** Proper part versus externally connected, illustrating the need for the convex hull in our implementation of RCC8. As modeled, cytoplasm is assumed to contain the cell nuclei as shown in (a), reflecting an extrapolation of the objects to three dimensions. The segmentation and feature extraction processes of Chapters 4 and 5, however, result in cytoplasm and nuclei as discrete objects as shown in (b). Thus, the use of the convex hull operator can allow for more intuitive modeling of relationships between histologic objects.

### 6.3.3 Results

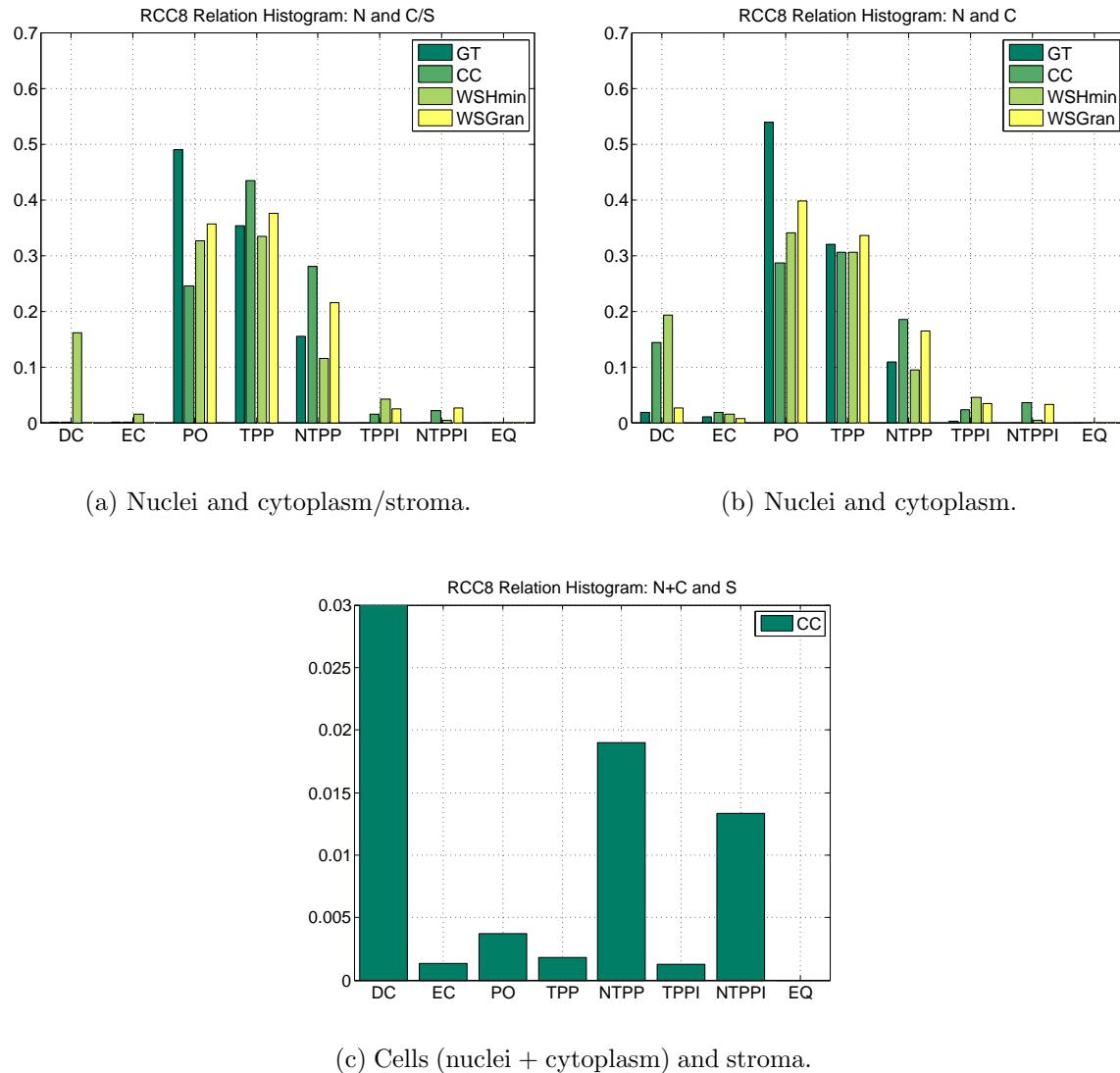
Implementing RCC8 with the convex hull operator, hereafter referred to generically as RCC8, we compute the histograms of RCC8 relations between histologic entities. Results are shown in Figure 6.4 for the relation between nuclei and cytoplasm/stroma, nuclei and cytoplasm, and cells (nuclei and cytoplasm) and stroma. In this implementation, the relationship is determined for the objects based on the order in which they are presented to the algorithm. Thus the relation between nuclei and cytoplasm is interpreted as the relationship of nuclei to cytoplasm, not vice versa. Additionally, it should be noted that there is a threshold for distinguishing between the TPP and TPPI relations versus a PO relation. This was implemented for the empirically determined value of 5 pixels to avoid a bias towards PO for inadequately resolved region boundaries.

Interestingly, nuclei tend to have a PO or TPP relationship with cytoplasm and combined cytoplasm/stroma more often than a NTPP relationship. For the GT segmentation this is not surprising given the prevalence of clustered nuclei, indicating that the nuclei will be tangential to the Voronoi boundary and thus the cytoplasm boundary. It is a bit surprising for the other segmentations. Additionally, eccentrically located nuclei can be a characteristic of breast cancers [2]. However, the difference in RCC8 relations between benign and malignant imagery, as shown in Figure 6.5, is not striking. The DC relation is more common for the segmentation methods that tend to undersegment, namely CC and WSHmin. The few TPPI and NTPPI relations are due to either errors in the nuclear segmentation, since the GT segmentation does not display such relations, or due to configurations of clumped nuclei that contain cytoplasm and/or stroma inside.

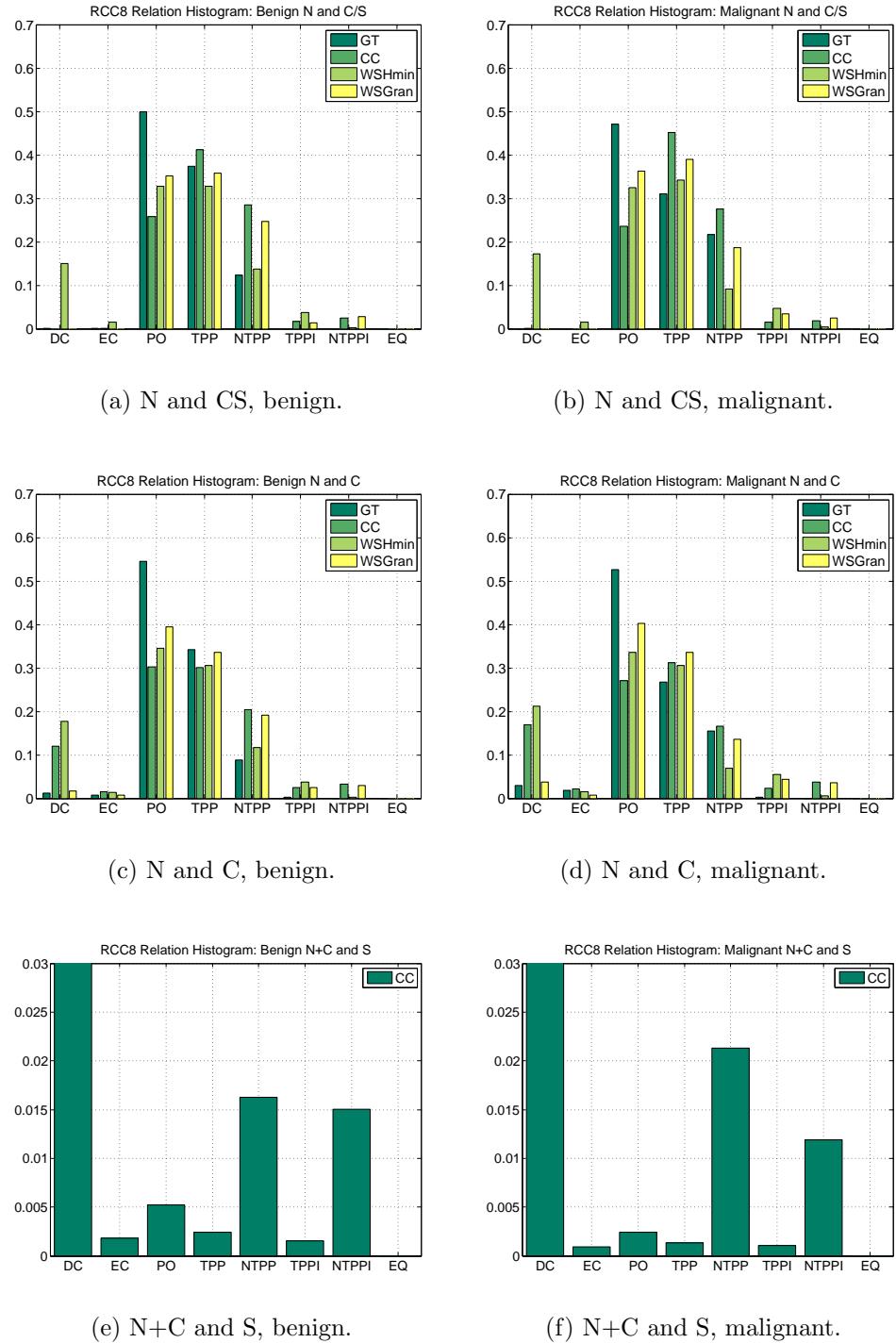
For the RCC8 relationships between cells (nuclei and cytoplasm) and stroma, the relationship is determined for each pair of connected component regions of both cells and stroma. Thus, it is expected that the predominant relationship will be DC; in this case, we are more interested in the possible relationships other than DC. As a result, the y-axis of the cell plots in Figure 6.4 (c) and 6.5 (e) and (f) are plotted in the range [0, 0.03] rather than [0, 0.70] as for the other histograms. Besides the predominant DC relationship, there is a tendency towards NTPP and NTPPI relationships, as well as a few of the remaining relations EC, PO, TPP, and TPPI. There may be more of a tendency towards NTPP rather than NTPPI in malignant imagery (refer to Figure 6.5), but it is difficult to make any strong argument for the reasons behind this.

## 6.4 Use of Higher-Level Objects for Classification

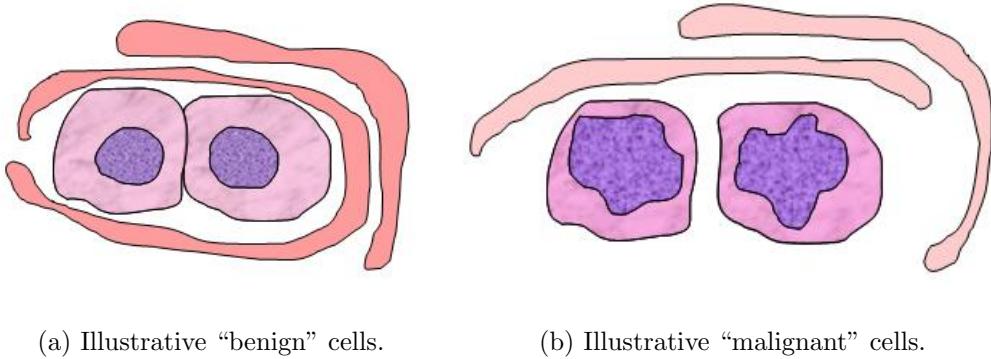
The previous chapter addressed the use of object- and spatial-level features for nuclei, cytoplasm, and stroma for classification of histopathology images as malignant or benign. Here we consider the use of features extracted from higher-level objects for classification.



**Figure 6.4.** Histograms of RCC8 relations between histologic entities. Note that the stroma histogram in (c) is plotted on a different scale for the y-axis.



**Figure 6.5.** Histograms of RCC8 relations between histologic entities, benign and malignant. Note that the stroma histograms in (d) and (e) are plotted on a different scale for the y-axis.



**Figure 6.6.** Illustrative “cytology” images.

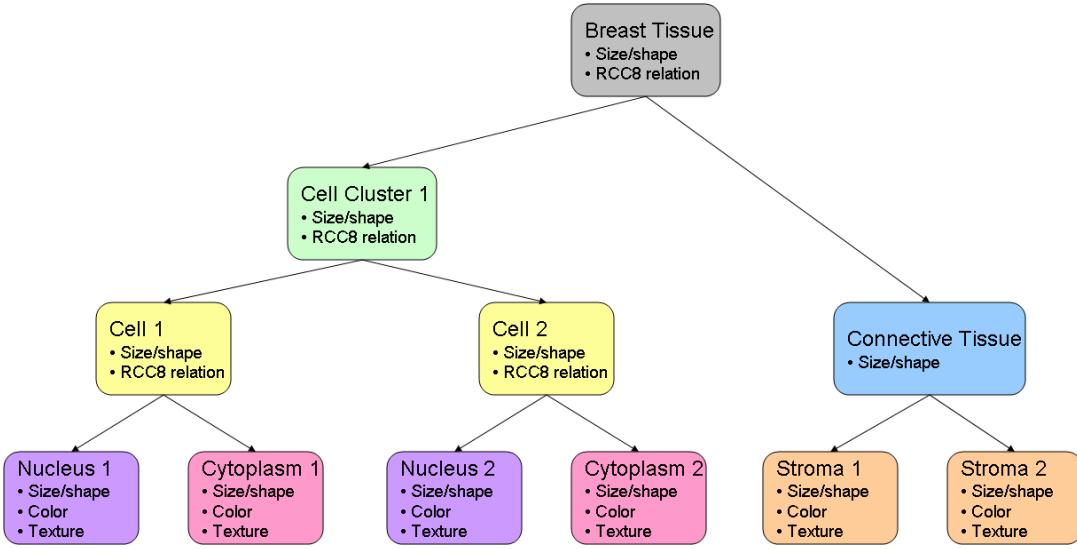
#### 6.4.1 The Plausibility of Using Feature Graphs

It is intuitive to think of histo- and cyto-pathology imagery in a hierarchical fashion. As an example, one could envision a high-level object “tissue region” with object- and architectural-level attributes. This “tissue region” would own several “cells” with their own attributes, which in turn own “nucleus” and “cytoplasm” similarly with attributes. This hierarchy of features embodies the classification and segmentation process and possibly the thought process of the user. Example cartoon cytology images are shown in Figure 6.6, and an example hierarchical representation for these images in Figure 6.7. Note that both Figure 6.6 (a) and (b) share the same basic hierarchical structure since they both consist of the same basic objects in a similar configuration. The differences between these two images would be characterized in the features defined at each node of the hierarchy graph.

The proposal of using feature graphs has recently appeared in the literature, for example [186]. It is tempting, in light of the hierarchical structure of Figure 6.7, to propose the use of the various object-level features in a graphical framework rather than the standard linear feature vector framework. Indeed, it was our assumption in the development of object- and spatial-level features that they would be readily adapted to a graph-like structure. Moreover, we assumed that the use of such “feature graphs” would lend a superior representation to the imagery and yield better classification results.

Upon further and in-depth research using object- and spatial-level features, we conclude that the generalization of the feature vectors of Chapter 5 to feature graphs would not prove useful in our application. We elaborate on specific reasons here.

- The hierarchy of Figure 6.7 is a good tool for visualizing the various object-level features and how they relate to each other in a more abstract fashion. It is unclear, however, what features could be used as edge weights. The edges in Figure 6.7



**Figure 6.7.** Example hierarchy for the images in Figure 6.6. Both images in Figure 6.6 have the same basic hierarchical structure shown here since each image has the same basic objects arranged in a similar fashion. The differences between Figure 6.6 (a) and (b) will be reflected in the feature values defined at each node in the graph. Examples of such features are included here.

signify some abstract relation between the objects, most succinctly summed up as “owns;” thus, a cell “owns” a nucleus and cytoplasm object.

- An original motivation for use of feature graphs was based on the assumption that spatial features would be particularly important for classification. Instead, the spatial features were shown to have very poor performance, and additionally were shown to degrade performance when included with object-level features. Examples of feature graphs in the literature include either randomly generated attributed graphs (edge-attributed, node-attributed, or both) [187–191], or synthetic or real applications in which the node and/or edge information is specifically spatial (usually the coordinate location of points) [188–190, 192–195]. Common applications of feature graphs include symbol and character recognition [192–195] and document clustering and classification [186, 190].
- Bunke [186] motivates feature graphs with the claim that feature vectors cannot model structural relationships. This may be true in the case of features consisting directly of coordinate locations and other very specific (i.e., not relative) spatial information. With the spatial features that we have extracted, and the spatial features that have been reported in the literature for applications in histo- and cyto-pathology, however, they can be easily incorporated in a feature vector, as was presented in Chapter 5.

**Table 6.1.** Summary of high-level object features.

<b>Area:</b> Area
<b>Elliptical Features:</b> Major and minor axis length, eccentricity, orientation, elliptical deviation
<b>Convex Hull Features:</b> Convex area, convex deficiency, solidity
<b>Filled Image Features:</b> Filled area, Euler number
<b>Bounding Box Features:</b> Extent, aspect ratio
<b>Boundary Features:</b> Perimeter, radii, perimeter FFT, perimeter curvature, bending energy, perimeter FD
<b>Other Shape Features:</b> Equivalent diameter, sphericity, compactness, inertia shape
<b>Reflection Symmetry Features:</b> Binary

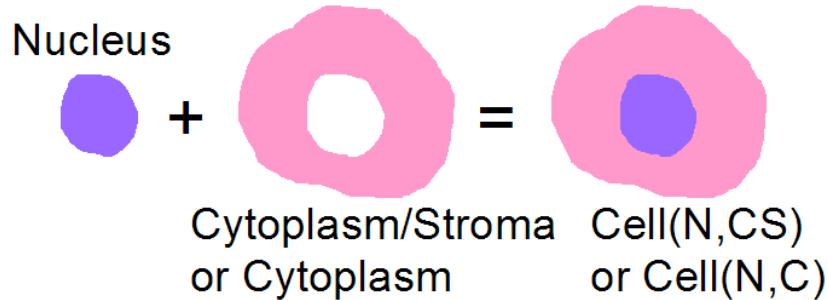
- The graph edit distance, used to compute the distance between two graphs, relies on the assumption that a new, previously unseen graph can be classified based on a cumulative sum of the distance between nodes and edges. This is not a simple sum of distances, since the graph edit distance applies different costs and distance computations for different edit operations, e.g., node substitution and edge deletion [196,197]. The issue of assigning edge labels aside, the use of node features (the object-level features of Chapter 5) in such a distance-based classification is counter to the classification methodology used in the feature selection process. This is not to say that the two methodologies could not be reconciled, but it appears that this would unnecessarily complicate the classification process for this application.

Thus, we forgo the use of feature graphs for representation of higher-level objects.

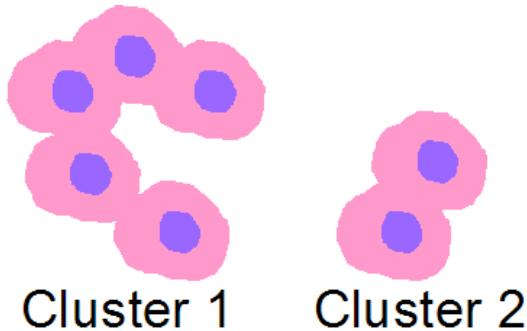
#### 6.4.2 Higher-Level Objects

In this implementation of higher-level objects and the associated features, we create feature vectors consisting of features extracted from the high-level object and all other objects that constitute the high-level object. We extract all the size and shape features as described in Section 5.3.1; these features are also summarized in Table 6.1. We do not compute, however, the center of mass or reflection symmetry features that are computed from anything but the binary mask image. When the various statistics are included, this results in 84 features for each high-level object when skewness and kurtosis are included, and 76 without.

We consider two more levels of objects in this section: cells, and cell clusters, illustrated in Figure 6.8.



(a) Cells.



(b) Cell clusters.

**Figure 6.8.** Illustrative higher-level objects.

**Cells** are defined as either the agglomeration of either nuclei and cytoplasm/stroma ( $\text{Cell}_{(\text{N},\text{CS})}$ ) or nuclei and cytoplasm ( $\text{Cell}_{(\text{N},\text{C})}$ ). The features associated with cell objects includes all the features associated with both constituents (N and CS or N and C), the RCC8 relation between the two, and the size and shape features of Table 6.1 extracted from the union of the two regions.

**Cell clusters** are defined as the connected components of cell objects, much as was done for the computation of RCC8 relations for stroma versus nuclei and cytoplasm in the previous section. Cell clusters consist of either connected components of  $\text{Cell}_{(\text{N},\text{CS})}$  or  $\text{Cell}_{(\text{N},\text{C})}$  objects, resulting in  $\text{Cluster}_{(\text{N},\text{CS})}$  or  $\text{Cluster}_{(\text{N},\text{C})}$  objects. The cluster objects include all features from the constituent objects and the addition of size and shape features of the cell clusters.

**Table 6.2.** High-level object feature subset performance, with regularization parameter  $\lambda_1 = 0.05$ . In the following, ‘Cell<sub>(N,CS)</sub>’ designates cell objects consisting of nuclei and cytoplasm/stroma, ‘Cell<sub>(N,C)</sub>’ cell objects of nuclei and cytoplasm, ‘Cluster<sub>(N,CS)</sub>’ cell cluster objects of nuclei and cytoplasm/stroma, and ‘Cluster<sub>(N,C)</sub>’ cell objects of nuclei and cytoplasm. Bold entries correspond to the best performance across all the feature subsets for each segmentation type. For reference, the results from Table 5.3 from Chapter 5 are included here.

Object	Performance			
	GT	CC	WSHmin	WSGran
Cell <sub>(N,CS)</sub>	0.79/0.78	0.77/0.56	0.66/0.61	<b>0.70</b> /0.58
Cell <sub>(N,C)</sub>	<b>0.81</b> / <b>0.79</b>	<b>0.74</b> / <b>0.69</b>	0.68/0.62	0.76/0.54
Cluster <sub>(N,CS)</sub>	0.81/0.74	<b>0.90</b> /0.53	N/A	N/A
Cluster <sub>(N,C)</sub>	<b>0.82</b> /0.76	0.79/0.55	N/A	N/A
N	0.79/0.77	0.62/0.43	0.66/ <b>0.65</b>	0.62/0.61
CS	0.74/0.72	0.69/0.56	0.67/0.61	0.73/ <b>0.63</b>
C	0.73/0.63	0.65/0.50	0.67/0.60	0.74/0.58
S	N/A	0.66/0.50	N/A	N/A
N + CS	<b>0.75</b> / <b>0.79</b>	0.70/0.57	0.68/ <b>0.65</b>	0.74/ <b>0.63</b>
N + C	0.77/0.76	0.67/0.52	<b>0.69</b> / <b>0.65</b>	0.73/0.58

**Table 6.3.** High-level object feature subset performance, with regularization parameter  $\lambda_1 = 0.01$ . In the following, ‘Cell<sub>(N,CS)</sub>’ designates cell objects consisting of nuclei and cytoplasm/stroma, ‘Cell<sub>(N,C)</sub>’ cell objects of nuclei and cytoplasm, ‘Cluster<sub>(N,CS)</sub>’ cell cluster objects of nuclei and cytoplasm/stroma, and ‘Cluster<sub>(N,C)</sub>’ cell objects of nuclei and cytoplasm. Bold entries correspond to the best performance across all the feature subsets for each segmentation type. For reference, the results from Table 5.4 from Chapter 5 are included here.

Object	Performance			
	GT	CC	WSHmin	WSGran
Cell <sub>(N,CS)</sub>	0.97/0.78	0.92/0.53	0.87/0.65	<b>0.94</b> /0.63
Cell <sub>(N,C)</sub>	<b>0.97</b> / <b>0.82</b>	0.95/0.69	<b>0.88</b> / <b>0.69</b>	<b>0.94</b> /0.66
Cluster <sub>(N,CS)</sub>	<b>1.00</b> /0.67	<b>1.00</b> /0.57	N/A	N/A
Cluster <sub>(N,C)</sub>	0.99/0.65	0.95/0.57	N/A	N/A
N	0.85/0.77	0.74/0.69	0.78/0.63	0.81/0.67
CS	0.90/0.51	0.82/0.70	0.81/0.60	0.88/ <b>0.73</b>
C	0.90/0.34	0.79/0.63	0.82/0.63	0.87/0.72
S	N/A	0.81/0.60	N/A	N/A
N + CS	0.94/0.79	0.84/ <b>0.72</b>	0.84/0.62	0.90/0.70
N + C	0.94/0.68	0.83/0.67	0.85/0.65	0.89/0.71

**Table 6.4.** AUC for OS image-level classification for high-level objects for grafting with  $\lambda_1 = 0.01$ . In the following, ‘Cell<sub>(N,CS)</sub>’ designates cell objects consisting of nuclei and cytoplasm/stroma, ‘Cell<sub>(N,C)</sub>’ cell objects of nuclei and cytoplasm, ‘Cluster<sub>(N,CS)</sub>’ cell cluster objects of nuclei and cytoplasm/stroma, and ‘Cluster<sub>(N,C)</sub>’ cell objects of nuclei and cytoplasm. Bold entries correspond to the best performance across all the feature subsets for each segmentation type. For reference, the results from Table 5.11 from Chapter 5 are included here.

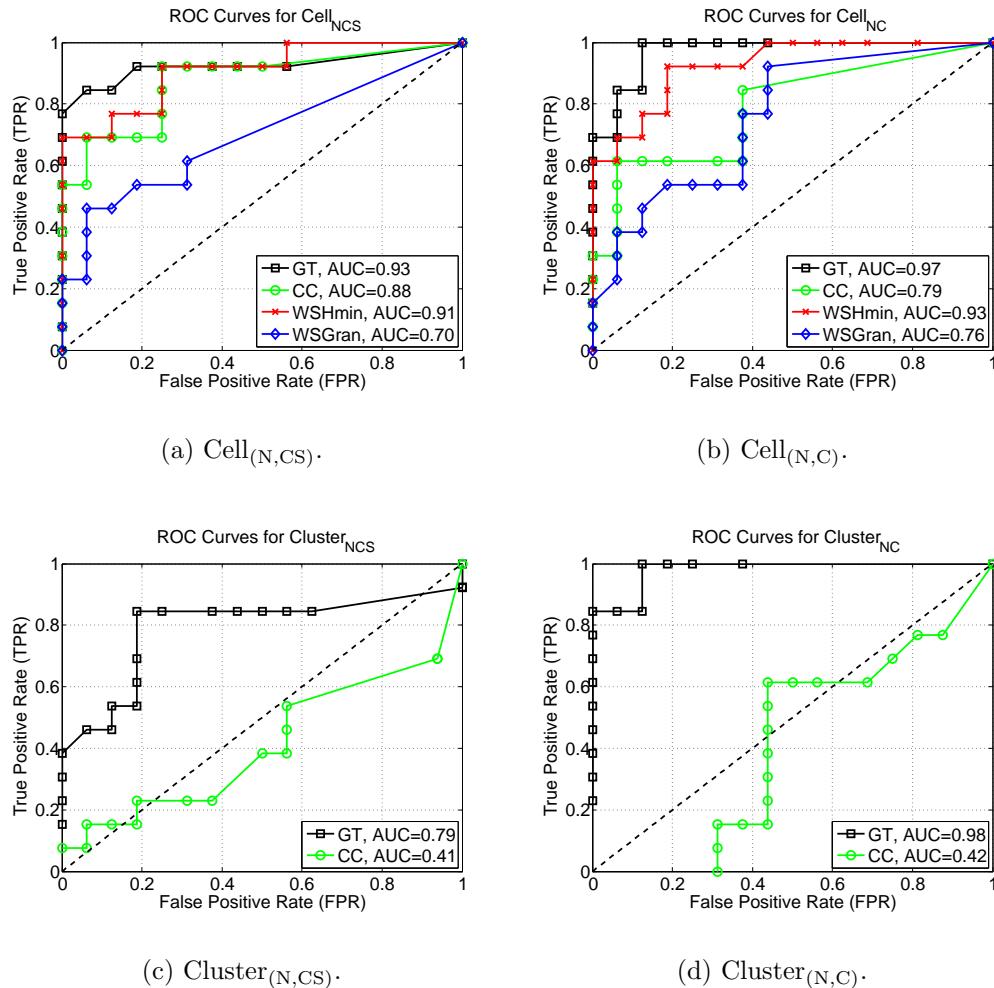
Object	Performance			
	GT	CC	WSHmin	WSGran
Cell <sub>(N,CS)</sub>	0.93	0.88	0.91	0.70
Cell <sub>(N,C)</sub>	0.97	0.79	<b>0.93</b>	0.76
Cluster <sub>(N,CS)</sub>	0.79	0.41	N/A	N/A
Cluster <sub>(N,C)</sub>	<b>0.98</b>	0.42	N/A	N/A
N	0.92	0.93	0.78	<b>0.87</b>
CS	0.89	<b>0.94</b>	0.88	0.71
C	0.86	0.91	0.83	0.84
S	N/A	0.77	N/A	N/A
N + CS	0.96	0.92	0.90	0.67
N + C	0.92	0.88	0.84	0.86

### 6.4.3 Results

Object-level classification performances using the grafting method of feature selection are shown in Tables 6.2 and 6.3 for  $\lambda_1 = 0.05$  and  $\lambda = 0.01$ . The addition of higher-level object features almost tends to improve the IS performance for all the segmentation types considered here; the OS performance, however, tends to suffer. The highest performance when all segmentations are considered are still for the object-level feature subsets of Chapter 5.

Image-level OS classification performances for the higher-level object subsets are shown via ROC curves and AUC measures in Figure 6.9; AUC measures are also summarized in Table 6.4. From these results it is apparent that many of these higher-level object subsets are improving the image-level classification for the GT segmentation, but cannot exceed the image-level performance for any of the realistic segmentations. Again, similar to the object-level performances, while higher-level object features improve the image-level classification performance of certain segmentations, they do not exceed lower-level object performance for image classification.

From these results it appears that the use of higher-level image objects can improve classification performance, but that these higher-level features are dependent on an accurate underlying lower-level object segmentation.



**Figure 6.9.** ROC curves for OS image-level performance for high-level objects (grafting,  $\lambda_1 = 0.01$ ), varying the threshold of malignant objects for which an image is considered malignant.

**Table 6.5.** Statistics of GT and WSHmin Cell<sub>NC</sub> feature subsets. Statistics presented include the percentage of OD and non-OD features; R, G, B, I/Hue, and binary features; texture, radiometric and densitometric, and size/shape features; and nuclei, non-nuclei, and cell features.

Category	GT	WSHmin
<b>OD</b>	0.56	0.58
<b>Non-OD</b>	0.44	0.42
<b>R</b>	0.24	0.22
<b>G</b>	0.23	0.22
<b>B</b>	0.25	0.23
<b>I/Hue</b>	0.17	0.19
<b>Binary</b>	0.11	0.14
<b>Texture</b>	0.60	0.46
<b>Radio-/Densito-metric</b>	0.15	0.15
<b>Size/Shape</b>	0.25	0.40
<b>Nuclei</b>	0.53	0.43
<b>Non-Nuclei</b>	0.43	0.50
<b>Cell</b>	0.04	0.07

#### 6.4.4 Higher-Level Object Feature Subsets

The grafting feature subsets for  $\lambda_1 = 0.01$  are presented in Appendix L for the Cell<sub>NC</sub> GT and WSHmin objects. Table 6.5 presents various statistics of the Cell<sub>NC</sub> feature subset for GT and WSHmin, namely the percentage of OD and non-OD features; R, G, B, I/Hue, and binary features; texture, radiometric and densitometric, and size/shape features<sup>2</sup>; and nuclei, non-nuclei, and cell features. The presentation of GT and WSHmin is motivated by the increased OS performance with the higher-level object features for these segmentations (refer to Table 6.3).

These feature subsets display less preference for OD features than the subsets described in Chapter 5, most likely due to the more prominent role of the size/shape features of the higher-level Cell<sub>NC</sub> object. This also explains the higher usage of binary features with respect to the other image bands. Interestingly, the preference for cytoplasm over nuclei features is diminished when the Cell<sub>NC</sub> features are included, even while the percentage of Cell<sub>NC</sub> features is rather small.

---

<sup>2</sup>For this categorization, we are considering the various chromatin-specific features to fall into either the radiometric/densitometric or size/shape categories.

## 6.5 Summary

In this chapter we have applied the size and shape features from Chapter 5 to higher-level objects, namely “cells” and “cell clusters.” These higher-level features display a potential for increasing both object- and image-level classification performance, but seem to be dependent on an accurate underlying lower-level object segmentation.

Additionally, we have discussed the plausibility of using a “feature graph” representation for the various features. We have concluded that this is not a feasible approach for this particular application, especially given the poor performance of spatial features as demonstrated in Chapter 5.

## 6.6 Future Directions

### 6.6.1 Higher-Level Objects

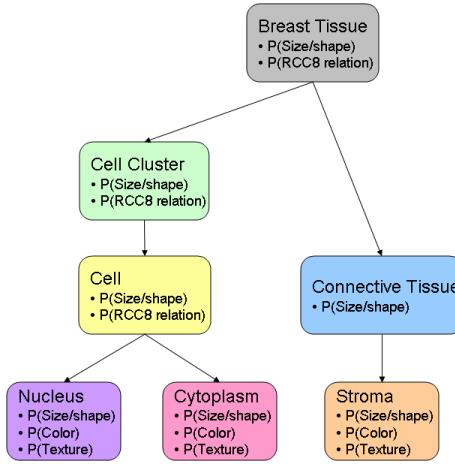
If better lower-level object segmentations could be developed, the use of higher-level objects could be important for other image analysis tasks. As such, it will be important to explore the characteristics of lower-level agglomeration to form higher-level objects. Voting over a region by different level objects could provide another means of image classification. Preliminary work in this area using lower-level objects did not indicate that this was a promising approach.

### 6.6.2 Introduction to Probabilistic Graph Models (PGMs)

Using the features that have been extracted for all the various objects illustrated in Figure 6.7, a probabilistic model of the histopathology imagery can be developed; we call this representation a Probabilistic Graph Model (PGM).

#### Creation of a PGM

A PGM can be derived in a fairly straight-forward fashion from a feature hierarchy such as the one shown in Figure 6.7. In such a derivation, multiple instantiations of objects within the feature hierarchy define a probability distribution of features. An example PGM derived from the feature hierarchy shown in Figure 6.7 is shown in Figure 6.10. Note that this derived PGM has a similar structure to the feature hierarchy, but the feature values have instead been replaced with a probability distribution of feature values. Another way to view the difference is to think of the feature hierarchy as a representation of a specific image, whereas the derived PGM is a general description of imagery belonging to a certain category (e.g., benign or malignant). This sort of analysis,



**Figure 6.10.** Example probabilistic graph model for the hierarchy in Figure 6.7.

taking multiple instantiations of objects in a model and formulating a class-based model, does not appear to have been researched yet.

While we have avoided the use of feature graphs as such, we return to a graph-like representation here for visualization purposes. Thus, while the features of the various image objects are actually implemented and classified in a feature vector form, we are visualizing the features and their abstract relation to each other in a graphical model.

### Further Uses of the PGMs

The representation of histopathology imagery in a probabilistic graph model could lend itself to the generation of synthetic imagery with different feature distributions. These synthesized images could be further analyzed by the expert to gain insight into the continuum of conditions between truly benign and frankly malignant (e.g., generate the image in Figure 6.6 (b) by modifying feature distributions of Figure 6.6 (a)).

The synthesis of an image, provided a probabilistic graph model, is a very intriguing area of future research. Particularly for this application in histo- and cyto-pathology imagery, this could allow for very important insights into the characteristics of cancerous and non-cancerous tissue. More importantly, this could allow for the exploration of the continuum between benign and malignant conditions (the atypias) and the characteristics that may indicate a malignancy potential.



# Chapter 7

## Summary and Conclusions

This chapter summarizes our research contributions for this dissertation and concludes with a broader view of the potential of this research in future work. The summary is organized in terms of Chapters 2-6.

### 7.1 Summary of Research Contributions

#### 7.1.1 Chapter 2: Multispectral Analysis of Pixel-Level Nuclear Classification

##### Summary

Our datasets are relatively unique in histo- and cyto-pathology in that they are imaged multispectrally (usually with 29 bands covering the visible spectrum from 420 nm to 700 nm, though capability exists for other bands). While some researchers have found advantages to spectral imaging for some applications in medical image analysis [39, 70–73], the value of multispectral analysis for routine histo/cytopathology preparations has not been previously studied. Throughout our proposed research we have striven to analyze the utility of the multispectral nature of our imagery. Chapter 2 presented analyses of our multispectral data for the pixel-level classification of nuclei.

##### Research Contributions

- Quantitative and in-depth analysis of multispectral versus RGB imagery for pixel-level nuclear classification of H&E stained imagery, a common histopathology image analysis task.

## Results

Results indicate only slight performance differences using multispectral imagery as opposed to derived RGB imagery. These performance differences are not, however, statistically significant in many cases. These conclusions hold for both classification using all available image bands as well as using single image bands, indicating that the multispectral bands do not contain any more useful spectral information than do the RGB bands for this nuclear classification task.

## Future Work

- Further research could molecularly specific pixel-level classification [76] to identify subtle differences in the spectra associated with different dye-protein interactions.
- While we have presented analysis for the RGB color space, there may be other color/information spaces more suitable for this and other classification tasks such as YUV, HSV, HSI, optical density (OD), Principal Components Analysis (PCA) and Independent Components Analysis (ICA).
- A very important research direction is the incorporation of feedback to the system. In this case, one could “fine-tune” the results of the pixel-level classification based on the results of a further higher-level analysis. In particular, the results of higher-level analysis can be incorporated into the performance metric for the pixel-level classification.

### 7.1.2 Chapter 3: Pixel-Level Classification of Cytoplasm and Stroma

#### Summary

While the characteristics of cell nuclei are well established as useful for diagnostic purposes, it is expected that the characteristics of cytoplasm and stroma will be similarly useful. The discrimination between cytoplasm and stroma, however, is not commonly addressed in quantitative pathology studies, especially for standard H&E stained imagery. In Chapter 3 we presented the development and analysis of a classifier for cytoplasm and stroma.

#### Research Contributions

- Development of a pixel-level classifier for discrimination of cytoplasm and stroma.
- Analysis of features useful for the cytoplasm/stroma discrimination problem.

## Results

In this chapter we presented the development and analysis of a classifier for cytoplasm and stroma. We find our best overall average classification accuracy to be approximately 0.86. It appears that the use of nuclei is particularly useful for characterizing cytoplasm versus stroma.

## Future Work

- It would be of great use to utilize a dataset for which a more comprehensive ground truth could be developed for cytoplasm and stroma, e.g., serial sections classified with vibrational spectroscopy [29] or stained with other more specific stains. This would allow for a more comprehensive analysis of the performance of various cytoplasm and stroma classification methods.
- There may well be other features better suited for this classification problem. Some of these well-suited features may stem from a higher-level conceptual analysis of the imagery, leading back to the need for feedback from higher levels of analysis. In particular, it would be interesting to further investigate the potential use of other nuclear proximity features, e.g., a simple dilation of the binary nuclei.
- Further investigation of multiple independent Genie Pro algorithms for cytoplasm and stroma discrimination could be of use either directly for the discrimination process, or for further insight into relevant features.

### 7.1.3 Chapter 4: Nuclei Segmentation: Methods and Evaluation Metrics

#### Summary

The segmentation of cell nuclei at an object level is a very difficult problem. While there is much literature addressing this problem, we sought a solution that is feasible for our H&E stained imagery and that avoids too many underlying assumptions about the nuclear characteristics (e.g., parameterization as an ellipse). We also presented the development of an object-level segmentation metric applicable to our nuclear segmentations, as well as other application areas.

#### Research Contributions

- A new method to quantify segmentation performance.
- Comparison of several common segmentation methods using the new metric.

## Results

We developed a new object-level metric for segmentation evaluation and have shown its correspondence to qualitative observations of general segmentation characteristics. This metric was used to compare several methods for delineation of cell nuclei, and to illustrate the dependence of this higher-level segmentation on the accuracy of the underlying pixel-level classification. We have shown this segmentation metric and its constituent terms to correspond well with the qualitative observations of segmentation accuracy, including the general tendency of an algorithm to over- or under-segment an image. This metric also allows for a direct quantitative comparison between the outputs of different segmentation algorithms. While the metric defines a single performance, we have shown the usefulness of observing the performance of the individual metric terms. Analysis of multispectral versus RGB imagery was continued, with the result that for nuclear segmentation, RGB imagery performs better than the multispectral.

## Future Work

- While we have avoided any segmentations that rely on assumptions of the underlying object shape, e.g., ellipse fitting, such algorithms could be used to compute markers for subsequent segmentation.
- It would be interesting to quantify any differences in the characteristics of “true” concavities and the artifactual ones and similarly for holes. A better discrimination of “true” concavities may allow for the concavity-based segmentation performance to greatly improve.
- Most of the segmentation methods presented have a tendency to undersegment, even for ideal pixel-level input. There may be ways to combine the results of an undersegmented and oversegmented output that could provide a nice separation of individual nuclei. This would be a region merging algorithm, taking into account *a priori* knowledge in the form of the undersegmented output as well as other information (e.g., concavities, shape).

### 7.1.4 Chapter 5: Feature Extraction and Selection

#### Summary

In Chapter 5, we described the various object-level and spatial-relation features that we extract from our various image objects. We provided a comprehensive list of features compiled from a variety of published research on histo- and cytological image analysis. We also presented our results on the feature selection and classification performance using various categories of features.

## Research Contributions

- Comprehensive list of features for object- and spatial-level feature extraction.
- Use of spatial arrangement features.
- Use of non-nuclear features, namely cytoplasm and stroma features.
- Feature subsets for several classification tasks.
- Use of and analysis of performance for imperfectly segmented objects in a feature extraction, selection, and classification framework.

## Results

We have described a comprehensive list of object-level features and demonstrated the use of these features on various histologic objects. It has been demonstrated that these object-level features are versatile and general enough to elicit important information from even imperfectly segmented objects. We have presented the use of non-nuclear features, namely features of cytoplasm and stroma, and their good classification performance, often exceeding that of nuclei. Using object-level features and the grafting method [108] of feature selection, we have shown object-level classification accuracies above 0.70 for OS object-level classification, and an AUC above 0.90 for OS image-level classification. We have also presented a comprehensive list of spatial-level features for use in quantifying the spatial arrangement aspects of nuclei in our imagery, including a method for assigning these spatial features to a region of an image. These features, however, were shown to have poor classification performance, especially for less than perfect nuclear segmentation.

## Future Work

- The good performance of the cytoplasm and stroma features provides an intriguing possibility for application to other datasets including other tissues (e.g., prostate, colon).
- It will be important to study the use of these object-level features to cytology imagery.
- It would also be very interesting to apply the object-level feature extraction and selection in a completely different application domain, e.g., remote sensing. Of particular interest here, beyond the classification performance of this approach, would be the dependence of classification performance on the accuracy of the underlying object segmentations.

- While the spatial features have not been shown to be useful in this study, there is definite cause for further analysis of these features. Further study of the spatial features and how they may be better tailored to human intuition about spatial arrangement of objects is certainly an open field of research.

### 7.1.5 Chapter 6: Higher-Level Objects

#### Summary

We used the Qualitative Spatial Reasoning (QSR) formulation Region Connection Calculus (RCC) as a means to segment, i.e., agglomerate higher-level image objects. In Chapter 6 we used these higher-level image objects for classification as well as the construction of a probabilistic image representation.

#### Research Contributions

- Characterization of higher-level segmentation via RCC.
- Object- and image-level classification using higher-level objects.
- Introduction to probabilistic graph models.

#### Results

We applied the size and shape features from Chapter 5 to higher-level objects, namely “cells” and “cell clusters.” These higher-level features display a potential to increase both object- and image-level classification performance, but seem to be dependent on an accurate underlying lower-level object segmentation. Additionally, we discussed the plausibility of using a “feature graph” representation for the various features. We concluded that this is not a feasible approach for this particular application, especially given the poor performance of spatial features as demonstrated in Chapter 5. Lastly, we introduced an image representation method in the form of probabilistic graph models and indicated possible further uses for such an image representation.

#### Future Work

- If better lower-level object segmentations could be developed, the use of higher-level objects could be important for other image analysis tasks, such as image representation by a probabilistic graph model. As such, it will be important to explore the characteristics of lower-level agglomeration to form higher-level objects.

- The synthesis of an image, provided a probabilistic graph model, is a very intriguing area of future research. Particularly for this application in histo- and cyto-pathology imagery, this could allow for very important insights into the characteristics of cancerous and non-cancerous tissue. More importantly, this could allow for the exploration of the continuum between benign and malignant conditions (the atypias) and the characteristics that may indicate a malignancy potential.

## 7.2 Conclusion

We have presented here an approach for quantitative analysis of imagery, Quantitative Object- and spatial Arrangement-Level Analysis (QOALA), using expert (pathologist) input to guide the classification process. QOALA has yielded some very good object- and image-level classification performances. Future work could continue with QOALA both for application to different datasets (including, possibly, different application areas) as well as for extension of some of the concepts therein. Specifically, now that the framework is in place, analysis of atypical histo- and cyto-pathology specimens could be considered. For this future work, the use of a probabilistic image representation is an intriguing area of research.



# Bibliography

- [1] “Probability of breast cancer in American women.” <http://www.cancer.gov/cancertopics/breast>.
- [2] R. M. DeMay, *The Art and Science of Cytopathology: Exfoliative Cytology/Aspiration Cytology*. American Society for Clinical Pathology, 1996.
- [3] “Breast cancer facts & figures 2005-2006.” American Cancer Society, Inc.
- [4] A. M. Hanby, “The pathology of breast cancer and the role of the histopathology laboratory,” *Clinical Oncology*, vol. 17, pp. 234–239, 2005.
- [5] G. D. Leonard and S. M. Swain, “Ductal carcinoma in situ, complexities and challenges,” *Journal of the National Cancer Institute*, vol. 96, pp. 906–920, June 2004.
- [6] G. Viale, “Histopathology of primary breast cancer 2005,” *Breast*, vol. 14, pp. 487–492, 2005.
- [7] U. Veronesi, P. Boyle, A. Goldhirsch, R. Orecchia, and G. Viale, “Breast cancer,” *Lancet*, vol. 365, pp. 1727–1741, May 2005.
- [8] A. Neher, G. Öfner, E. Appenroth, and A. Gschwendtner, “High-resolution image cytometry on smears of normal oral mucosa: A possible approach for the early detection of laryngopharyngeal cancers,” *Head & Neck*, vol. 26, pp. 694–700, 2004.
- [9] D. J. Zahniser, L. M. Isenstein, and M. L. Hutchinson, “Computerized contextual analysis in diagnosis,” *Cancer Letters*, vol. 77, pp. 139–144, 1994.
- [10] T. Mairinger, G. Mikuz, and A. Gschwendter, “Nuclear chromatin texture analysis of nonmalignant tissue can detect adjacent prostatic adenocarcinoma,” *Prostate*, vol. 41, pp. 12–19, 1999.
- [11] “On-line medical dictionary.” <http://cancerweb.ncl.ac.uk/omd>. Published by the Centre for Cancer Education at the Department of Medical Oncology, University of Newcastle upon Tyne, UK.

## Bibliography

---

- [12] Personal communication with Dr. David Rimm, Department of Pathology, Yale University, Jan. 11, 2006.
- [13] “Varispec brochure.” <http://www.cri-inc.com/files/VariSpec-Brochure.pdf>.
- [14] M. H. Ross, G. I. Kaye, and W. Pawlina, *Histology: A Text and Atlas With Cell and Molecular Biology*. Lippincott Williams & Wilkins, 4th ed., 2003.
- [15] F. Schnorrenberg, C. S. Pattichis, K. C. Kyriacou, and C. N. Schizas, “Computer-aided detection of breast cancer nuclei,” *IEEE Transactions on Information Technology B*, vol. 1, pp. 128–140, June 1997.
- [16] A. Hoque, S. M. Lippman, I. V. Boiko, E. N. Atkinson, N. Sneige, M. D. Lagios, R. Schwarting, W. J. Colburn, K. Dhingra, M. Follen, G. J. Kelloff, C. W. Boone, and W. N. Hittelman, “Quantitative nuclear morphometry by image analysis for prediction of recurrence of ductal carcinoma in situ of the breast,” *Cancer Epidemiology, Biomarkers, & Prevention*, vol. 10, pp. 249–259, Mar. 2001.
- [17] L. Latson, B. Sebek, and K. A. Powell, “Automated cell nuclear segmentation in color images of hematoxylin and eosin-stained breast biopsy,” *Analytical and Quantitative Cytology and Histology*, vol. 25, pp. 321–331, Dec. 2003.
- [18] B. Weyn, G. van de Wouwer, A. van Daele, P. Scheunders, D. van Dyck, E. van Marck, and W. Jacob, “Automated breast tumor diagnosis and grading based on wavelet chromatin texture description,” *Cytometry*, vol. 33, pp. 32–40, 1998.
- [19] N. H. Anderson, P. W. Hamilton, P. H. Bartels, D. Thompson, R. Montironi, and J. M. Sloan, “Computerized scene segmentation for the discrimination of architectural features in ductal proliferative lesions of the breast,” *Journal of Pathology*, vol. 181, pp. 374–380, 1997.
- [20] B. Weyn, G. van de Wouwer, M. Koprowski, A. van Daele, K. Dhaene, P. Scheunders, W. Jacob, and E. van Marck, “Value of morphometry, texture analysis, densitometry, and histometry in the differential diagnosis and prognosis of malignant mesothelioma,” *Journal of Pathology*, vol. 189, pp. 581–589, 1999.
- [21] W. B. Spillman, J. L. Robertson, W. R. Huckle, B. S. Govindan, and K. E. Meissner, “Complexity, fractals, disease time, and cancer,” *Physics Review E*, vol. 70, pp. 1–12, 2004.
- [22] B. Weyn, W. A. A. Tjalma, P. Vermeylen, A. van Daele, E. van Marck, and W. Jacob, “Determination of tumour prognosis based on angiogenesis-related vascular patterns measured by fractal and syntactic structure analysis,” *Clinical Oncology*, vol. 16, pp. 307–316, 2004.

## Bibliography

---

- [23] R. Montironi, W. F. Whimster, Y. Collan, P. W. Hamilton, D. Thompson, and P. H. Bartels, “How to develop and use a Bayesian Belief Network,” *Journal of Clinical Pathology*, vol. 49, pp. 194–201, 1996.
- [24] M. L. Morrison, W. G. McCluggage, G. J. Price, J. Diamond, M. R. M. Sheeran, K. M. Mulholland, M. Y. Walsh, R. Montironi, P. H. Bartels, D. Thompson, and P. W. Hamilton, “Expert system support using a Bayesian belief network for the classification of endometrial hyperplasia,” *Journal of Pathology*, vol. 197, pp. 403–414, 2002.
- [25] J. W. Hoppin, M. A. Kupinski, G. A. Kastis, E. Clarkson, and H. H. Barrett, “Objective comparison of quantitative imaging modalities without the use of a gold standard,” *IEEE Transactions on Medical Imaging*, vol. 21, pp. 441–449, May 2002.
- [26] V. Sharifi-Salmation, B. Pesquet-Popescu, J. Simony-Lafontaine, and J. P. Rigaut, “Index for spatial heterogeneity in breast cancer,” *Journal of Microscopy-Oxford*, vol. 216, pp. 110–122, Nov. 2004.
- [27] T. Zhao, E. S. Wachman, and D. L. Farkas, “A novel scheme for abnormal cell detection in Pap smear images,” in *Proc. SPIE*, vol. 5318, pp. 151–162, 2004.
- [28] “Infrared spectroscopy from Wikipedia, the free encyclopedia.” [http://en.wikipedia.org/wiki/Infrared\\_spectroscopy](http://en.wikipedia.org/wiki/Infrared_spectroscopy).
- [29] D. C. Fernandez, R. Bhargava, S. M. Hewitt, and I. W. Levin, “Infrared spectroscopic imaging for histopathologic recognition,” *Nature Biotechnology*, vol. 23, pp. 469–474, Apr. 2005.
- [30] R. A. Shaw, J. R. Mansfield, S. P. Rempel, S. Low-Ying, V. V. Kupriyanov, and H. H. Mantsch, “Analysis of biomedical spectra and images: From data to diagnosis,” *Journal of Molecular Structure*, vol. 500, pp. 129–138, 2000.
- [31] G. Lin, M. K. Chawla, K. Olson, J. F. Guzowski, C. A. Barnes, and B. Roysam, “Hierarchical, model-based merging of multiple fragments for improved three-dimensional segmentation of nuclei,” *Cytometry Part A*, vol. 63A, pp. 20–33, 2005.
- [32] C. Wählby, I.-M. Sintorn, F. Erlandsson, G. Borgefors, and E. Bengtsson, “Combining intensity, edge and shape information for 2D and 3D segmentation of cell nuclei in tissue sections,” *Journal of Microscopy*, vol. 215, pp. 67–76, July 2004.
- [33] B. L. Luck, K. D. Carlson, A. C. Bovik, and R. R. Richards-Kortum, “An image model and segmentation algorithm for reflectance confocal images of in vivo cervical tissue,” *IEEE Transactions on Image Processing*, vol. 14, pp. 1265–1276, Sept. 2005.

## Bibliography

---

- [34] F. Ciocchetta, R. Dell'Anna, F. Demichelis, A. Sboner, A. P. Dhillon, A. Dhillon, A. Godfrey, and A. Quaglia, "Knowledge discovery in support of early diagnosis of hepatocellular carcinoma," in *Proc. Int Joint Conf on Neural Networks*, vol. 1, pp. 177–181, July 2003.
- [35] C. Demir, S. H. Gultekin, and B. Yener, "Augmented cell-graphs for automated cancer diagnosis," *Bioinformatics*, vol. 21, no. Suppl. 2, pp. ii7–ii12, 2005.
- [36] C. Gunduz, B. Yener, and S. H. Gultekin, "The cell graphs of cancer," *Bioinformatics*, vol. 20, no. Suppl. 1, pp. i145–i151, 2004.
- [37] M. Sammouda, R. Sammouda, N. Niki, N. Yamaguchi, and N. Moriyama, "Cancerous nuclei detection on digitized pathological lung color images," *Journal of Biomedical Informatics*, vol. 35, pp. 92–98, 2005.
- [38] M. A. Roula, A. Bouridane, F. Kurugollu, and A. Amira, "A quadratic classifier based on multispectral texture features for prostate cancer diagnosis," in *Proc. Int Symp on Signal Processing and Its Applications*, vol. 2, pp. 37–40, July 2003.
- [39] C. Angeletti, N. R. Harvey, V. Khomitch, A. H. Fischer, R. M. Levenson, and D. L. Rimm, "Detection of malignancy in cytology specimens using spectral-spatial analysis," *Laboratory Investigation*, vol. 85, pp. 1555–1564, 2005.
- [40] M. A. Brewer, J. Ranger-Moore, M. H. Greene, D. S. Alberts, Y. Liu, H. G. Bartels, A. C. Baruch, and P. H. Bartels, "Preneoplastic changes in ovarian tissues," *Analytical and Quantitative Cytology and Histology*, vol. 26, pp. 207–216, Aug. 2004.
- [41] B. Weyn, G. van de Wouwer, S. Kumar-Singh, A. van Daele, P. Scheunders, E. van Marck, and W. Jacob, "Computer-assisted differential diagnosis of malignant mesothelioma based on syntactic structure analysis," *Cytometry*, vol. 35, pp. 23–29, 1999.
- [42] L. Ballerini and L. Franzén, "Classification of microscopic images of breast tissue," in *Proc. SPIE*, vol. 5370, pp. 960–971, 2004.
- [43] N. R. Harvey, R. M. Levenson, and D. L. Rimm, "Investigation of automated feature extraction techniques for applications in cancer detection from multispectral histopathology images," in *Proc. SPIE*, vol. 5032, pp. 557–566, 2003.
- [44] K.-M. Lee and W. N. Street, "An adaptive resource-allocating network for automated detection, segmentation, and classification of breast cancer nuclei topic area: Image processing and recognition," *IEEE Transactions on Neural Networks*, vol. 14, pp. 680–687, May 2003.

## Bibliography

---

- [45] G. van de Wouwer, B. Weyn, P. Scheunders, W. Jacob, E. van Marck, and D. van Dyck, “Wavelets as chromatin texture descriptors for the automated identification of neoplastic nuclei,” *Journal of Microscopy*, vol. 197, pp. 25–35, Jan. 2000.
- [46] C. Herrera-Espiñeira, C. Marcos-Muñoz, and J. Esquivias, “Automated segmentation of cell nuclei in fine needle aspirates of the breast,” *Analytical and Quantitative Cytology and Histology*, vol. 20, pp. 29–35, Feb. 1998.
- [47] D. Thompson, P. H. Bartels, H. G. Bartels, and R. Montironi, “Image segmentation of cribriform gland tissue,” *Analytical and Quantitative Cytology and Histology*, vol. 17, pp. 314–322, Oct. 1995.
- [48] L. Jia and L. Kitchen, “Object-based image similarity computation using inductive learning of contour-segment relations,” *IEEE Transactions on Image Processing*, vol. 9, pp. 80–87, Jan. 2000.
- [49] L. Jia and L. Kitchen, “Object-based image content characterisation for semantic-level image similarity calculation,” *Pattern Analysis Applications*, vol. 4, pp. 215–226, 2001.
- [50] Y. Xu, P. Duygulu, E. Saber, A. M. Tekalp, and F. T. Yarman-Vural, “Object based image retrieval based on multi-level segmentation,” in *Proc. IEEE Int Conf on Acoustics, Speech, and Signal Processing*, vol. 6, pp. 2019–2022, 2000.
- [51] S. Newsam, S. Bhagavathy, C. Kenney, and B. S. Manjunath, “Object-based representations of spatial images,” *Acta Astronautica*, vol. 48, no. 5-12, pp. 567–577, 2001.
- [52] G. G. Hazel, “Object-level processing of spectral imagery for detection of targets and changes using spatial-spectral-temporal techniques,” in *Proc. SPIE*, vol. 4381, pp. 380–390, 2001.
- [53] B. Ko and H. Byun, “Integrated region-based image retrieval using region’s spatial relationships,” in *Proc. ICPR*, vol. 1, pp. 196–199, 2002.
- [54] F. L. Ber and A. Napoli, “Object-based representation and classification of spatial structures and relations,” in *Proc. IEEE Int Conf on Tools with Artificial Intelligence*, pp. 268–275, 2002.
- [55] U. Braga-Neto and J. Goutsias, “Object-based image analysis using multiscale connectivity,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 27, pp. 892–907, June 2005.
- [56] G. Cong and B. Parvin, “Model-based segmentation of nuclei,” *Pattern Recognition*, vol. 33, pp. 1383–1393, 2000.

## Bibliography

---

- [57] T. McInerney and D. Terzopoulos, “Deformable models in medical image analysis: A survey,” *Medical Image Analysis*, vol. 1, no. 2, pp. 91–108, 1996.
- [58] S. Schüpp, A. Elmoataz, J. Fadili, P. Herlin, and D. Bloyet, “Image segmentation via multiple active contour models and fuzzy clustering with biomedical applications,” in *Proc. ICPR*, vol. 1, pp. 622–625, 2000.
- [59] H.-S. Wu, J. Barba, and J. Gil, “A parametric fitting algorithm for segmentation of cell images,” *IEEE Transactions on Biomedical Engineering*, vol. 45, pp. 400–407, Mar. 1998.
- [60] N. Malpica, C. Ortiz de Solàrzano, J. J. Vaquero, A. Santos, I. Vallcorba, J. M. García-Sagredo, and F. del. Pozo, “Applying watershed algorithms to the segmentation of clustered nuclei,” *Cytometry*, vol. 28, pp. 289–297, 1997.
- [61] D. L. Pham, C. Xu, and J. L. Prince, “Current methods in medical image segmentation,” *Annual Review of Biomedical Engineering*, vol. 2, pp. 315–337, 2000.
- [62] D. Hoiem, R. Sukthankar, H. Schneiderman, and L. Huston, “Object-based image retrieval using the statistical structure of images,” in *Proc. CVPR*, vol. 2, pp. 490–497, 2004.
- [63] N. Oswald and P. Levi, “Cooperative object recognition,” *Pattern Recogn Letters*, vol. 22, pp. 1273–1282, 2001.
- [64] H. Wersing and E. Körner, “Learning optimized features for hierarchical models of invariant object recognition,” *Neural Computation*, vol. 15, pp. 1559–1588, 2003.
- [65] R. Kamalov, M. Guillaud, D. Haskins, A. Harrison, R. Kemp, D. Chiu, M. Follen, and C. MacAulay, “A Java application for tissue section image analysis,” *Computer Methods and Programs in Biomedicine*, vol. 77, pp. 99–113, 2005.
- [66] “Definiens Cellenger—automated multivariate image analysis for cell-based assays.” <http://www.definiens.com/products/cellenger.php>.
- [67] “Enterprise products—Definiens eCognition server.” [http://www.definiens.com/products/ecognition\\_server.php](http://www.definiens.com/products/ecognition_server.php).
- [68] P. Biberthaler, M. Athelogou, S. Langer, R. Leiderer, and K. Messmer, “Evaluation of murine liver transmission electron micrographs by an innovative object-based quantitative image analysis system (Cellenger®),” *European Journal of Medical Research*, vol. 8, pp. 275–282, July 2003.
- [69] “Comparison of automated and manual methods for counting cells in cell proliferation studies.” [http://www.definiens.com/pdf/documents/PublicationsLIFE/Persohn\\_Baatz%202004.pdf](http://www.definiens.com/pdf/documents/PublicationsLIFE/Persohn_Baatz%202004.pdf).

## Bibliography

---

- [70] R. Levenson, P. J. Cronin, and K. K. Pankratov, "Spectral imaging for brightfield microscopy," in *Proc. SPIE*, vol. 4959, pp. 27–33, 2003.
- [71] R. Jaganath, C. Angeletti, R. Levenson, and D. L. Rimm, "Diagnostic classification of urothelial cells in urine cytology specimens using exclusively spectral information," *Cancer Cytopathology*, vol. 102, pp. 186–191, June 2004.
- [72] S. M. Gentry and R. Levenson, "Biomedical applications of the information-efficient spectral imaging sensor (ISIS)," in *Proc. SPIE*, vol. 3603, pp. 129–142, 1999.
- [73] R. M. Levenson, E. S. Wachman, W. Niu, and D. L. Farkas, "Spectral imaging in biomedicine: A selective overview," in *Proc. SPIE*, vol. 3438, pp. 300–312, 1998.
- [74] L. E. Boucheron, N. R. Harvey, and B. S. Manjunath, "Utility of multispectral imaging for analysis of routine clinical histopathology imagery," in *Presented at the 2006 Workshop on Multiscale Biological Imaging, Data Mining, & Informatics*, (Santa Barbara, CA), Sept. 2006.
- [75] L. E. Boucheron, Z. Bi, N. R. Harvey, and B. S. Manjunath, "Utility of multispectral imaging for nuclear classification of routine clinical histopathology imagery," *BMC Cell Biology*, vol. 8(Suppl 1):S8, 10 July 2007. <http://www.biomedcentral.com/1471-2121/8/S1/S8>.
- [76] R. M. Levenson, "Spectral imaging and pathology: seeing more," *Laboratory Medicine*, vol. 35, pp. 244–251, Apr. 2004.
- [77] I. Barshak, J. Kopolovic, and C. Rothmann, "Spectral morphometric characterization of breast carcinoma cells," *British Journal of Cancer*, vol. 79, no. 9/10, pp. 1613–1619, 1999.
- [78] L. Zeng and Q. Wu, "Fast segmentation of cervical cells by using spectral imaging analysis techniques," in *Lecture Notes in Computer Science*, vol. 4222, pp. 734–741, 2006.
- [79] R. C. Gonzalez and R. E. Woods, *Digital Image Processing*. Prentice-Hall, 2nd ed., 2002.
- [80] A. C. Ruifrok, R. L. Katz, and D. A. Johnston, "Comparison of quantification of histochemical staining by Hue-Saturation-Intensity (HSI) transformation and color-deconvolution," *Applied Immunohistochemistry and Molecular Morphology*, vol. 11, no. 1, pp. 85–91, 2003.
- [81] H. Stark and J. W. Woods, *Probability and Random Processes with Applications to Signal Processing*. Prentice Hall, 3rd ed., 2002.
- [82] C. M. Bishop, *Neural Networks for Pattern Recognition*. Oxford University Press, 1995.

## Bibliography

---

- [83] N. R. Harvey, J. Theiler, S. P. Brumby, S. Perkins, J. J. Szymanski, J. J. Bloch, R. B. Porter, M. Galassi, and A. C. Young, “Comparison of GENIE and conventional supervised classifiers for multispectral image feature extraction,” *IEEE Transactions on Geoscience and Remote Sensing*, vol. 40, pp. 393–404, 2002.
- [84] S. Theodoridis and K. Koutroumbas, *Pattern Recognition*. Elsevier Academic Press, 2nd ed., 2003.
- [85] “Svm-light support vector machine.” <http://svmlight.joachims.org/>.
- [86] “Wilcoxon signed-rank test, from Wikipedia, the free encyclopedia.” [http://en.wikipedia.org/wiki/Wilcoxon\\_signed\\_rank\\_test](http://en.wikipedia.org/wiki/Wilcoxon_signed_rank_test).
- [87] “Bonferroni correction, from Wikipedia, the free encyclopedia.” <http://en.wikipedia.org/wiki/Bonferroni>.
- [88] “Principal components analysis, from Wikipedia, the free encyclopedia.” [http://en.wikipedia.org/wiki/Principal\\_components\\_analysis](http://en.wikipedia.org/wiki/Principal_components_analysis).
- [89] “Independent component analysis, from Wikipedia, the free encyclopedia.” [http://en.wikipedia.org/wiki/Independent\\_components\\_analysis](http://en.wikipedia.org/wiki/Independent_components_analysis).
- [90] A. J. Sims, M. K. Bennett, and A. Murray, “Image analysis can be used to detect spatial changes in the histopathology of pancreatic tumours,” *Physics in Medicine and Biology*, vol. 48, pp. N183–N191, 2003.
- [91] A. J. Sims, M. K. Bennett, and A. Murray, “Comparison on semi-automated image analysis and manual methods for tissue quantification in pancreatic carcinoma,” *Physics in Medicine and Biology*, vol. 47, pp. 1255–1266, 2002.
- [92] R. M. Levenson and D. L. Farkas, “Digital spectral imaging for histopathology and cytopathology,” in *Proc. SPIE*, vol. 2983, pp. 123–135, 1997.
- [93] Personal communication with Dr. David Rimm, Department of Pathology, Yale University, Apr. 19, 2007.
- [94] K. Masood, N. Rajpoot, H. Qureshi, and K. Rajpoot, “Co-occurrence and morphological analysis for colon tissue biopsy classification,” in *Proc. 4th Int Workshop on Frontiers of Information Technology (FIT’06)*, 2006.
- [95] M. Teverovskiy, V. Kumar, J. Ma, A. Kotsianti, D. Verbel, A. Tabesh, H.-Y. Pang, Y. Vengrenyuk, S. Fogarasi, and O. Saidi, “Improved prediction of prostate cancer recurrence based on an automated tissue image analysis system,” in *Proc. ISBI*, pp. 257–260, 2004.

## Bibliography

---

- [96] A. Tabesh, M. Teverovskiy, H.-Y. Pang, V. P. Kumar, D. Verbel, A. Kotsianti, and O. Saidi, “Multi-feature prostate cancer diagnosis and gleason grading of histological images,” *To appear: IEEE Transactions on Medical Imaging*, 2007.
- [97] “K-means algorithm, from Wikipedia, the free encyclopedia.” <http://en.wikipedia.org/wiki/Kmeans>.
- [98] J. G. Daugman, “Complete discrete 2-D Gabor transforms by neural networks for image analysis and compression,” *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 36, pp. 1169–1179, July 1988.
- [99] J. Daugman, “Two-dimensional spectral analysis of cortical receptive field profiles,” *Vision Research*, vol. 20, pp. 847–856, 1980.
- [100] B. S. Manjunath and W. Y. Ma, “Texture features for browsing and retrieval of image data,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 18, pp. 837–842, Aug. 1996.
- [101] P. Wu, B. S. Manjunath, S. Newsam, and H. D. Shin, “A texture descriptor for browsing and similarity retrieval,” *Signal Processing-Image Communication*, vol. 16, pp. 33–43, 2000.
- [102] A. W. Whitney, “A direct method of nonparametric measurement selection,” *IEEE Transactions on Computers*, vol. 20, pp. 1100–1103, 1971.
- [103] T. Marill and D. M. Green, “On the effectiveness of receptors in recognition systems,” *IEEE Transactions on Information Theory*, vol. 9, pp. 11–17, 1963.
- [104] P. Pudil, J. Novovičová, and J. Kittler, “Floating search methods in feature selection,” *Pattern Recognition Letters*, vol. 15, pp. 1119–1125, Nov. 1994.
- [105] R. E. Shapire, *Nonlinear Estimation and Classification*, ch. The boosting approach to machine learning: An overview. Springer, 2003.
- [106] Y. Freund and R. E. Shapire, “A decision-theoretic generalization of on-line learning and an application to boosting,” *Journal of Computer and System Sciences*, vol. 55, no. 1, pp. 119–139, 1997.
- [107] R. E. Shapire and Y. Singer, “Improved boosting algorithms using confidence-rated predictions,” *Machine Learning*, vol. 37, no. 3, pp. 297–336, 1999.
- [108] S. Perkins, K. Lacker, and J. Theiler, “Grafting: Fast, incremental feature selection by gradient descent in function space,” *Journal of Machine Learning Research*, vol. 3, pp. 1333–1356, 2003.
- [109] S. Perkins and J. Theiler, “Online feature selection using grafting,” in *Proc. ICML*, pp. 592–599, 2003.

## Bibliography

---

- [110] J. Theiler and K. Glocer, “Sparse linear filters for detection and classification in hyperspectral imagery,” in *Proc. SPIE*, vol. 6233, 2006.
- [111] A. Jain and D. Zongker, “Feature selection: Evaluation, application, and small sample performance,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 19, pp. 153–158, Feb. 1997.
- [112] R. P. W. Duin, P. Juszczak, P. Paclik, E. Pekalska, D. de Ridder, and D. M. J. Tax, “PRTools4: A Matlab toolbox for pattern recognition,” tech. rep., Delft University of Technology, 2004. <http://www.prtools.org>.
- [113] J. Gil, H. Wu, and B. Y. Wang, “Image analysis and morphometry in the diagnosis of breast cancer,” *Microscopy Research and Technique*, vol. 59, pp. 109–118, 2002.
- [114] T. W. Nattkemper, “Automatic segmentation of digital micrographs: A survey,” in *Proc. MEDINFO*, 2004.
- [115] V. Chalana and Y. Kim, “A methodology for evaluation of boundary detection algorithms on medical images,” *IEEE Transactions on Medical Imaging*, vol. 16, pp. 642–652, Oct. 1997.
- [116] R. Unnikrishnan, C. Pantofaru, and M. Hebert, “A measure for objective evaluation of image segmentation algorithms,” in *Proc. CVPR*, p. CHANGE THIS, 2005.
- [117] J. K. Udupa, V. R. LeBlanc, Y. Zhuge, C. Imielinska, H. Schmidt, L. M. Currie, B. E. Hirsch, and J. Woodburn, “A framework for evaluating image segmentation algorithms,” *Computerized Medical Imaging and Graphics*, vol. 30, pp. 75–87, 2006.
- [118] A. Hoover, G. Jean-Baptiste, X. Jiang, P. J. Flynn, H. Bunke, D. B. Goldgof, K. Bowyer, D. W. Eggart, A. Fitzgibbon, and R. B. Fisher, “An experimental comparison of range image segmentation algorithms,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 18, pp. 673–689, July 1996.
- [119] J. S. Cardoso and L. Corte-Real, “Toward a generic evaluation of image segmentation,” *IEEE Transactions on Image Processing*, vol. 14, pp. 1773–1782, Nov. 2005.
- [120] W. A. Yasnoff, J. K. Mui, and J. W. Bacus, “Error measures for scene segmentation,” *Pattern Recognition*, vol. 9, pp. 217–231, 1977.
- [121] Y. J. Zhang, “A survey on evaluation methods for image segmentation,” *Pattern Recognition*, vol. 29, no. 8, pp. 1335–1346, 1996.
- [122] Y. J. Zhang and J. J. Gerbrands, “Segmentation evaluation using ultimate measurement accuracy,” in *Proc. SPIE*, vol. 1657, pp. 449–460, 1992.
- [123] Q. Huang and B. Dom, “Quantitative methods of evaluating image segmentation,” in *Proc. ICIP*, pp. 53–56, 1995.

## Bibliography

---

- [124] O. Lezoray and H. Cardot, “Cooperation of color pixel classification schemes and color watershed: A study for microscopic images,” *IEEE Transactions on Image Processing*, vol. 11, pp. 783–789, July 2002.
- [125] L. Cohen, P. Vinet, P. T. Sander, and A. Gagalowicz, “Hierarchical region based stereo matching,” in *Proc. CVPR*, pp. 416–421, 1989.
- [126] A. Fenster and B. Chiu, “Evaluation of segmentation algorithms for medical imaging,” in *Proc. IEEE Engineering in Medicine and Biology*, pp. 7186–7189, Sept. 2005.
- [127] M. Everingham, H. Muller, and B. Thomas, “Evaluating image segmentation algorithms using the pareto front,” in *Lecture Notes in Computer Science*, vol. 2353, pp. 34–48, 2002.
- [128] D. Glotsos, P. Spyridonos, D. Cavouras, P. Ravazoula, P.-A. Dadioti, and G. Nikiforidis, “Automated segmentation of routinely hematoxylin-eosin-stained microscopic images by combining support vector machine clustering and active contour models,” *Analytical and Quantitative Cytology and Histology*, vol. 26, pp. 331–340, Dec. 2006.
- [129] G. J. Price, W. G. McCluggage, M. L. Morrison, G. McClean, L. Venkatraman, J. Diamond, H. Bharucha, R. Montirnoni, P. H. Bartels, D. Thompson, and P. W. Hamilton, “Computerized diagnostic decision support system for the classification of preinvasive cervical squamous lesions,” *Human Pathology*, vol. 34, pp. 1193–1203, Nov. 2003.
- [130] V. Canzonieri, S. Monfardini, and A. Carbone, “Defining prognostic factors in malignancies through image analysis,” *European Journal of Cancer*, vol. 34, no. 4, pp. 451–458, 1998.
- [131] H. Honda, “Geometrical models for cells in tissues,” *International Review of Cytology*, vol. 81, pp. 191–248, 1983.
- [132] T. R. Jones, A. Carpenter, and P. Golland, “Voronoi-based segmentation of cells on image manifolds,” in *Lecture Notes in Computer Science*, vol. 3765, pp. 535–543, 2005.
- [133] L. E. Boucheron, N. R. Harvey, and B. S. Manjunath, “A quantitative object-level metric for segmentation performance and its application to cell nuclei,” in *Proc: International Symposium on Visual Computing, Lecture Notes in Computer Science*, vol. 4841, pp. 208–219, Nov. 2007.
- [134] D. P. Huttenlocher, G. A. Klanderman, and W. J. Rucklidge, “Comparing images using the Hausdorff distance,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 15, pp. 850–863, Sept. 1993.

## Bibliography

---

- [135] E. P. Vivek and N. Sudha, “Robust Hausdorff distance measure for face recognition,” *Pattern Recognition*, vol. 40, pp. 431–442, 2007.
- [136] E. Yörük, E. Konukoglu, B. Sankur, and J. Darbon, “Shape-based hand recognition,” *IEEE Transactions on Image Processing*, vol. 15, pp. 1803–1815, July 2006.
- [137] C. R. Giardina and E. R. Dougherty, *Morphological Methods in Image Processing*. Prentice Hall, 1988.
- [138] J. T. Newell and E. R. Dougherty, “Maximum-likelihood morphological granulometric classifiers,” in *Proc. SPIE*, vol. 1657, pp. 386–395, 1992.
- [139] J. Byun, M. R. Verardo, B. Sumengen, G. P. Lewis, B. S. Manjunath, and S. K. Fisher, “Automated tool for nuclei detection in digital microscopic images: Application to retinal images,” *Molecular Vision*, vol. 12, pp. 949–960, Aug. 2006.
- [140] S. Kumar, S. H. Ong, S. Ranganath, T. C. Ong, and F. T. Chew, “A rule-based approach for robust clump splitting,” *Pattern Recognition*, vol. 39, pp. 1088–1098, 2006.
- [141] T. T. E. Yeo, X. C. Jin, S. H. Ong, Jayasoорiah, and R. Sinniah, “Clump splitting through concavity analysis,” *Pattern Recognition Letters*, vol. 15, pp. 1013–1018, 1993.
- [142] G. Fernandez, M. Kunt, and J. P. Zyrd, “A new plant image segmentation algorithm,” in *Lecture Notes in Computer Science*, vol. 974, pp. 229–234, 1995.
- [143] J. Liang, “Intelligent splitting in the chromosome domain,” *Pattern Recognition*, vol. 22, no. 5, pp. 519–522, 1989.
- [144] W. X. Wang, “Binary image segmentation of aggregates based on polygonal approximation and classification of concavities,” *Pattern Recognition*, vol. 31, pp. 1503–1524, 1998.
- [145] A. Rosenfeld, “Measuring the sizes of concavities,” *Pattern Recognition Letters*, vol. 3, pp. 71–75, 1985.
- [146] J. Diamond, N. H. Anderson, D. Thompson, P. H. Bartels, and P. W. Hamilton, “A computer-based training system for breast fine needle aspiration cytology,” *Journal of Pathology*, vol. 196, pp. 113–121, 2002.
- [147] P. W. Hamilton, N. H. Anderson, J. Diamond, P. H. Bartels, J. B. Gregg, D. Thompson, and R. J. Millar, “An interactive decision support system for breast fine needle aspiration cytology,” *Analytical and Quantitative Cytology and Histology*, vol. 18, pp. 185–190, June 1996.

## Bibliography

---

- [148] P. W. Hamilton, N. Anderson, P. H. Bartels, and D. Thompson, “Expert system support using Bayesian belief networks in the diagnosis of fine needle aspiration biopsy specimens of the breast,” *Journal of Clinical Pathology*, vol. 47, pp. 329–336, 1994.
- [149] R. Malka and B. Lerner, “Classification of fluorescence in situ hybridization images using belief networks,” *Pattern Recognition Letters*, vol. 25, pp. 1777–1785, 2004.
- [150] R. Nafe, B. Yan, W. Schlote, and B. Schneider, “Application of different methods for nuclear shape analysis with special reference to the differentiation of brain tumors,” *Analytical and Quantitative Cytology and Histology*, vol. 28, pp. 69–77, Apr. 2006.
- [151] A. Doudkine, C. MacAulay, N. Pouplin, and B. Palcic, “Nuclear texture measurements in image cytometry,” *Pathologica*, vol. 87, pp. 286–299, 1995.
- [152] B. Palcic, C. E. MacAulay, S. A. Harrison, S. Lam, P. W. Payne, D. M. Garner, and A. Doudkine, “System and method for automatically detecting malignant cells and cells having malignancy associated changes,” Tech. Rep. 6026174, United States Patent, Feb. 2000.
- [153] E. C. M. Mommers, N. Poulin, J. Sangulin, C. J. L. M. Meijer, J. P. A. Baak, and P. J. van Diest, “Nuclear cytometric changes in breast carcinogenesis,” *Journal of Pathology*, vol. 193, pp. 33–39, 2001.
- [154] K. Rodenacker and E. Bengtsson, “A feature set for cytometry on digitized microscopic images,” *Analytical Cellular Pathology*, vol. 25, pp. 1–36, 2003.
- [155] R. Marcelpoil and Y. Usson, “Methods for the study of cellular sociology: Voronoi diagrams and parametrization of the spatial relationships,” *Journal of Theoretical Biology*, vol. 154, pp. 359–369, 1992.
- [156] R. Albert, T. Schindewolf, I. Baumann, and H. Harms, “Three-dimensional image processing for morphometric analysis of epithelium sections,” *Cytometry*, vol. 13, pp. 759–765, 1992.
- [157] P. J. van Diest, “Syntactic structure analysis,” *Pathologica*, vol. 87, pp. 255–262, June 1993.
- [158] W. H. Wolberg, W. N. Street, and O. L. Mangasarian, “Machine learning techniques to diagnose breast cancer from image-processed nuclear features of fine needle aspirates,” *Cancer Letters*, vol. 77, pp. 163–171, 1994.
- [159] M. Bevk and I. Kononenko, “A statistical approach to texture description of medical images: A preliminary study,” in *Proc. IEEE Symposium on Computer-Based Medical Systems*, pp. 239–244, 2002.

## Bibliography

---

- [160] F. Albrigtsen, “Statistical texture measures computed from gray level run length matrices,” tech. rep., Image Processing Laboratory, Department of Informatics, University of Oslo, Nov. 1995.
- [161] B. B. Mandelbrot, *The Fractal Geometry of Nature*. Freeman and Company, 1982.
- [162] G. M. Behry, “Simulation of fractal dimension evaluations,” *International Journal of Modelling and Simulation*, vol. 26, no. 2, pp. 91–97, 2006.
- [163] P. P. Ohanian and R. C. Dubes, “Performance evaluation for four classes of textural features,” *Pattern Recognition*, vol. 25, no. 8, pp. 819–833, 1992.
- [164] J. F. O’Callaghan, “Computing the perceptual boundaries of dot patterns,” *Computer Graphics and Image Processing*, vol. 3, pp. 141–162, 1974.
- [165] J. F. O’Callaghan, “An alternative definition for “neighborhood of a point”,” *IEEE Transactions on Computers*, pp. 1121–1125, Nov. 1975.
- [166] J.-M. Geusebroek, A. W. M. Smeulders, F. Cornelissen, and H. Geerts, “Segmentation of tissue architecture by distance graph matching,” *Cytometry*, vol. 35, pp. 11–22, 1999.
- [167] R. Fernandez-Gonzalez, M. H. Barcellos-Hoff, and C. O. de Solórzano, “A tool for the quantitative spatial analysis of complex cellular systems,” *IEEE Transactions on Image Processing*, vol. 14, pp. 1300–1313, Sept. 2005.
- [168] B. Liu and H.-J. Lai, *Matrices in Combinatorics and Graph Theory*. Kluwer Academic Publishers, 2000.
- [169] J. Clark and D. A. Holton, *A First Look at Graph Theory*. World Scientific Publishing Co., 1991.
- [170] L. W. Beineke and R. J. Wilson, eds., *Graph Connections: Relationships between Graph Theory and other Areas of Mathematics*. Clarendon Press, Oxford, 1997.
- [171] R. Marèe, P. Geurts, J. Piater, and L. Wehenkel, “Random subwindows for robust image classification,” in *Proc. CVPR*, 2005.
- [172] R. Marèe, P. Geurts, J. Piater, and L. Wehenkel, “Biomedical image classification with random subwindows and decision trees,” in *Lecture Notes in Computer Science*, vol. 3765, pp. 220–229, 2005.
- [173] S. Hochstein and M. Ahissar, “View from the top: Hierarchies and reverse hierarchies in the visual system,” *Neuron*, vol. 36, pp. 791–804, Dec. 2002.
- [174] D. A. Randell, Z. Cui, and A. G. Cohn, “A spatial logic based on regions and connection,” in *Proc. 3rd Int Conf on Knowledge Representation and Reasoning*, pp. 165–176, 1992.

## Bibliography

---

- [175] A. G. Cohn, B. Bennett, J. Gooday, and N. M. Gotts, “Qualitative spatial representation and reasoning with the region connection calculus,” *GeoInformatica*, vol. 1, pp. 275–316, 1997.
- [176] J. Renz, ed., *Qualitative Spatial Reasoning with Topological Information*, vol. 2293 of *Lecture Notes in Computer Science*. Springer Berlin/Heidelberg, 2002.
- [177] S. Li and M. Ying, “Region Connection Calculus: Its models and composition table,” *Artificial Intelligence*, vol. 145, pp. 121–146, 2003.
- [178] J. Renz and B. Nebel, “Spatial reasoning with topological information,” in *Proc. Spatial Cognition, An Interdisciplinary Approach to Representing and Processing Spatial Knowledge*, pp. 351–371, 1997.
- [179] J. Renz and B. Nebel, “On the complexity of qualitative spatial reasoning: A maximum tractable fragment of the Region Connection Calculus,” *Artificial Intelligence*, vol. 108, pp. 69–123, 1999.
- [180] J. G. Stell, “Boolean connection algebras: A new approach to the Region-Connection Calculus,” *Artificial Intelligence*, vol. 122, pp. 111–136, 2000.
- [181] I. Düntsch, H. Wang, and S. McCloskey, “A relation–algebraic approach to the region connection calculus,” *Theoretical Computer Science*, vol. 255, pp. 63–83, 2001.
- [182] S. Li and Y. Li, “On the complemented disk algebra,” *Journal of Logic and Algebraic Programming*, vol. 66, pp. 195–211, 2006.
- [183] I. Düntsch and M. Winter, “A representation theorem for Boolean contact algebras,” *Theoretical Computer Science*, vol. 347, pp. 498–512, 2005.
- [184] J. G. Stell and M. F. Worboys, “The algebraic structure of sets of regions,” *Lecture Notes in Computer Science*, vol. 1329, pp. 163–174, Aug. 2005.
- [185] J. G. Stell, “Part and complement: Fundamental concepts in spatial relations,” *Annals of Mathematics and Artificial Intelligence*, vol. 41, pp. 1–17, 2004.
- [186] H. Bunke, “Graph-based tools for data mining and machine learning,” in *Lecture Notes in Computer Science*, vol. 2734, pp. 7–19, 2003.
- [187] A. Hlaoui and S. Wang, “A new median graph algorithm,” in *Lecture Notes in Computer Science*, vol. 2726, pp. 225–234, 2003.
- [188] A. Hlaoui and S. Wang, “Median graph computation for graph clustering,” *Soft Computing*, vol. 10, pp. 47–53, 2006.

## Bibliography

---

- [189] A. Hlaoui and S. Wang, “A new algorithm for graph matching with application to content-based image retrieval,” in *Lecture Notes in Computer Science*, vol. 2396, pp. 291–300, 2002.
- [190] X. Jiang, A. Münger, and H. Bunke, “On median graphs: Properties, algorithms, and applications,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 23, pp. 1144–1151, Oct. 2001.
- [191] M. Ferrer, F. Serratosa, and A. Sanfeliu, “Synthesis of median spectral graph,” in *Lecture Notes in Computer Science*, vol. 3523, pp. 139–146, 2005.
- [192] M. Ferrer, F. Serratosa, and E. Valveny, “Evaluation of spectral-based methods for median graph computation,” in *Lecture Notes in Computer Science*, vol. 4478, pp. 580–587, 2007.
- [193] M. Ferrer, E. Valveny, and F. Serratosa, “Spectral median graphs applied to graphical symbol recognition,” in *Lecture Notes in Computer Science*, vol. 4225, pp. 774–783, 2006.
- [194] S. Günter and H. Bunke, “Validation indices for graph clustering,” *Pattern Recognition Letters*, vol. 24, pp. 1107–1113, 2003.
- [195] S. Günter and H. Bunke, “Self-organizing map for clustering in the graph domain,” *Pattern Recognition Letters*, vol. 23, pp. 405–417, 2002.
- [196] D. Conte, P. Foggia, C. Sansone, and M. Vento, “Thirty years of graph matching in pattern recognition,” *International Journal of Pattern Recognition*, vol. 18, no. 3, pp. 265–298, 2004.
- [197] A. Shokoufandeh and S. Dickinson, “Graph-theoretical methods in computer vision,” in *Lecture Notes in Computer Science*, vol. 2292, pp. 148–174, 2002.
- [198] “Cribriiform, from Wikipedia, the free encyclopedia.” <http://en.wikipedia.org/wiki/Cribriiform>.
- [199] “Papanicolaou stain, from Wikipedia, the free encyclopedia.” [http://en.wikipedia.org/wiki/Papanicolaou\\_stain](http://en.wikipedia.org/wiki/Papanicolaou_stain).
- [200] C. D. Meyer, *Matrix Analysis and Applied Linear Algebra*. SIAM, 2000.
- [201] A. C. Ruifrok and D. A. Johnston, “Quantification of histochemical staining by color deconvolution,” *Analytical and Quantitative Cytology and Histology*, vol. 23, pp. 291–299, Aug. 2001.
- [202] “Beer-Lambert law from Wikipedia, the free encyclopedia.” [http://en.wikipedia.org/wiki/Beer-Lambert\\_law](http://en.wikipedia.org/wiki/Beer-Lambert_law).

## Bibliography

---

- [203] T. F. Chan and L. A. Vese, “Active contours without edges,” *IEEE Transactions on Image Processing*, vol. 10, pp. 266–277, Feb. 2001.
- [204] B. Sumengen and B. S. Manjunath, “Graph partitioning active contours (GPAC) for image segmentation,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 28, pp. 509–521, Apr. 2006.
- [205] L. Bertelli, J. Byun, and B. S. Manjunath, “A variational approach to exploit prior information in object-background segregation: Application to retinal images,” in *To appear: Proc. ICIP*, Sept. 2007.
- [206] Personal communication with NGA, Apr. 11, 2005.
- [207] M. A. O’Brien, “Performance evaluation of the GENIE system,” in *Proc. American Society for Photogrammetry and Remote Sensing*, May 2003.

---

*Appendix*

# Appendix A

## Glossary of Pathology Terminology

**actin:** A cytoskeletal protein.

**adenocarcinoma:** A carcinoma originating in glandular tissue.

**aspirative cytology:** Cytology specimens obtained via syringe.

**atypical:** Cells or tissue displaying some characteristics of a malignancy, but not considered either malignant or benign. The diagnosis of atypia generally requires a more comprehensive (and possibly invasive) follow-up to determine the true diagnosis.

**benign:** A condition which will not metastasize and is not harmful in and of itself. Treatment/removal can alleviate symptoms (e.g., pressure on surrounding organs), and treatment/removal is considered sufficient for complete recovery.

**brightfield microscopy:** Microscopy techniques using a broad spectrum light source to visualize the specimen.

**carcinoma:** A cancer of the epithelium.

**CD31:** An immunostain which targets epithelial cells.

**cellularity:** A qualitative description of the abundance of cells present in a specimen.

**chromatin:** Nuclear material that is readily stained, consisting of the nucleic acids and associated proteins.

**confocal:** Confocal microscopy images different focal planes through the specimen.

**counterstain:** A stain used as contrast to another, generally more specific, stain.

**cribriform:** Latin for perforation [198].

**cytology:** The study of cells at a microscopic level, generally via a light microscopy technique.

**cytopathology:** The study of diseased cells at the microscopic level.

**densitometry:** Measurements related to the optical density of a sample.

**drosophila:** A genus of fly, commonly known as the fruit fly.

**ductal carcinoma:** Carcinoma originating in ductal structures.

**dysplasia:** Abnormalities in the characteristics of cells and tissues; often used interchangeably with hyperplasia.

**endothelium:** Specific type of epithelial cells which line the vessels and organs of the cardiovascular system and serous cavities.

**eosin:** A pink-staining acidic dye that stains membranes and fibers.

**epithelium:** The internal and external lining of cavities within the body; also the external covering (skin).

**exfoliative cytology:** Cytology specimens obtained via scraping (e.g., cervical cytology) or normal exfoliation of cells (e.g., urine cytology).

**Feulgen:** A stain specific to DNA which lends a purple color.

**fibroadenosis:** A benign cause of many breast lumps.

**fine needle aspiration:** An procedure using a small needle inserted into the lesion and drawing a small amount of cellular material into a syringe; a form of aspirative cytology.

**fluorescence imagery:** Fluorescent dyes are attached to antibodies specific to some feature of interest (e.g., certain proteins) and imaged by exciting the fluorescence of the dyes with appropriate incident light. This method can very specifically target certain molecular attributes of a biological specimen.

**Fourier transform spectroscopy:** Similar to infrared spectroscopy, where instead of an array of energies, an interferogram is produced; the Fourier transform of this interferogram produces the energies [28].

**glycogen:** Polymer of glucose used for short-term storage of energy [11].

**Gleason grading:** A grading for prostate cancer, characterizing the tumor into one of 5 categories based on tumor differentiation.

**gross:** Macroscopic.

**hematoxylin:** A blue-staining basic dye that stains genetic material; this is mainly seen in nuclear material, although some components of cytoplasmic and extracellular material is also stained.

**histology:** The study of tissue at a microscopic level, generally via a light microscopy technique.

**histopathology:** The study of diseased tissue at the microscopic level.

**hyperchromasia:** An overall increase in staining intensity.

**hyperplasia:** Abnormalities in the characteristics of cells and tissues, generally including an increase in cellularity and/or mitosis; often used interchangeably with dysplasia.

**immunohistochemistry:** See immunostain.

**immunostain:** Immunostains use antibodies to specifically target molecules of interest, similar to fluorescence imaging, but use standard dyes for viewing with light microscopy.

**infrared spectroscopy:** Uses infrared light to excite vibrations in the molecular structure of a specimen. These are reflected as energies which give insight into the underlying molecular constituents [28].

**in situ:** Within normal boundaries, not invading surrounding tissues.

**in vivo:** Living tissue in its natural environment.

**karyometry:** Nuclear characteristics, generally texture.

**lobular carcinoma:** A type of adenocarcinoma.

**malignant:** A condition which will eventually lead to death if untreated. Malignant conditions tend to metastasize, grow uncontrollably, and lack proper tissue differentiation.

**mesothelium:** Squamous (flattened cells) epithelium originating in the mesoderm (specific layer of the epithelium).

**metastasis:** The spread of cancer from the originating tissue to other parts of the body.

**microarray:** Tissue microarrays align many (hundreds or thousands) of tissue core samples on a single slide; this allows for simultaneous analysis of all samples and is commonly used in high-throughput operations.

**nucleolus:** A small, round sub-organelle within the cell nucleus.

**optical density:** Provides a linear relationship between image intensity and stain density, based on Lambert-Beer's law describing the intensity of light transmitted through a specimen.

**Orange G, Orange II:** Stains keratin orange.

**Papanicolaou (Pap):** A common cytology stain, consisting of several individual stains, not consistently standardized. Most Pap-stains contain Hematoxylin, Orange G, Eosin, Light Green (stains cytoplasm of metabolically active cells blue), and Bismark Brown (stains mucins and cartilage and other things) [199].

**pathology:** The study of disease, with emphasis on disease structure and the effects on the body as a whole.

**pleomorphic:** Containing more than one stage of the life cycle.

**premalignancy:** A diseased state that, while not considered cancerous, will progress to cancer if left untreated.

**reflectance microscopy:** Captures the reflected light from a specimen (rather than the transmitted light as through a thin slice specimen).

**stroma:** Connective tissue.

**transmission microscopy:** A broad spectrum light source is directed underneath a specimen, and the transmitted light is viewed and/or captured with a camera setup.

**urothelium:** Epithelial lining of the bladder.

## Appendix B

# Glossary of Machine Learning and Computer Vision Terminology

**accuracy:** The probability of a correct decision,  $N_C/N$  where  $N_C$  is the number of correctly determined cases and  $N$  is the total number of cases.

**classification:** The process of assigning a label to an image object.

**connected component:** A set of pixels that are connected, generally 4- or 8-connected.

**cycle:** An ordered set of nodes and edges, where the set of edges is unique, and the starting and ending node is the same.

**densitometry:** Metrics related to the optical density of an image or image object.

**detection rate:** The fraction of true (foreground) pixels classified correctly.

**eigendecomposition:** The eigenvectors  $\mathbf{x}$  and eigenvalues (scalars)  $\lambda$  of a square matrix  $\mathbf{A}$  satisfy the relationship  $\mathbf{Ax} = \lambda\mathbf{x}$ . The name comes from the German word “eigen” which can be translated as “owned by” or “peculiar to” [200].

**entropy:** A measure of the information content of an image or signal.

**error:** The probability of an incorrect decision,  $N_I/N$  where  $N_I$  is the number of incorrectly determined cases and  $N$  is the total number of cases. Note that  $Accuracy = 1 - Error$ .

**false alarm rate:** The fraction of false (background) pixels classified incorrectly.

**feature selection:** A means to select the relevant and important features from a large set of features, many of which may be redundant, irrelevant, or not particularly useful (and possibly detrimental) to the classification performance.

**fractal dimension:** The fractal dimension of objects differ from the typical integer dimensions of Euclidean geometry (e.g., a 1-D line and 2-D plane). Instead, the FD is a real-valued (not integer) measure of the space-filling properties of the object, based on the work of Mandelbrot [161].

**Gabor wavelet:** Gabor functions are Gaussians modulated by complex sinusoids, and have been proposed as an apt representation of human visual receptor fields [98,99].

**grafting:** (from “gradient feature testing”) [108,109] is a feature selection method where the classification of the underlying data and the feature selection process are not separated. Within the grafting framework, a loss function is used that shows preference for classifiers that separate the data with larger margins. Grafting also provides an efficient framework for selection of relevant features.

**granulometry:** A common technique to elicit information about the size distribution of objects in an image. Granulometries are calculated by applying successively larger structuring elements in a morphological image operation and analyzing the residue image.

**graph:** A graph structure is uniquely defined by a set of nodes and edges. Edges connect nodes that are neighbors according to some *a priori* definition of a neighborhood. Edges may be directed, resulting in a directed graph or digraph. Both nodes and edges may be weighted according to some metric (e.g., node degree or edge length).

**ground truth:** The correct/desired output (i.e., truth) for an image analysis algorithm. Originally a term from the remote sensing community, reflecting the truth obtained from a ground-based survey.

**Hausdorff distance:** The deviation between two sets of points is defined as the largest distance between any two corresponding points [115,134].

**in-sample:** Data for which an algorithm was trained on; often called training data.

**k-means:** A widely known and used unsupervised classification algorithm which clusters data into  $k$  clusters while minimizing the intracluster variance [97].

**Laplacian of Gaussian:** A function which is composed of a Laplacian (edge detector) and Gaussian (smoothing) filter to reduce the effects of noise in the edge detection process [79].

**margin:** In classifiers, the minimum separation (i.e., distance) between samples from different classes.

**markup:** The specification of ground truth, often obtained from an expert by the physical marking of an image for regions of interest, etc.

**multiplespectral imagery:** Imagery with more than 3 bands and generally less than 100.

**optical density:** For transmission microscopy of stained tissue, optical density provides a linear relationship between image intensity and staining density. This is based on Lambert-Beer's law describing the intensity of light transmitted through a specimen [201, 202].

**out-of-sample:** Data which was not used in the training of an algorithm, often called unseen data or test data.

**oversegmentation:** A segmentation that assigns too many regions to an image or image object.

**pattern spectrum:** In granulometric analysis, defining  $\Psi_k$  as the area (number of pixels) in the  $k$ -th residue image, the function  $\phi(k) = 1 - \Psi_k/\Psi_0$  is used to calculate the size distribution of image  $I$  [137, 138]. The derivative  $\phi'(k)$  is often called the pattern spectrum of image  $I$ .

**Principal Components Analysis (PCA):** An orthogonal linear transformation such that the largest variance in the data lies along the first coordinate direction, the second largest variance on the second coordinate direction, etc. [88].

**radiometry:** Metrics related to the spectral information of an image or image object.

**Receiver Operating Characteristic (ROC) curve:** ROC curves provide a graphical representation of the tradeoff between sensitivity and specificity (sensitivity on the y-axis and 1-specificity on the x-axis). Equivalently, this may also be plotted as true positive rate versus false positive rate. The Area Under the Curve (AUC) is often used as a single number to quantify a classifier; a perfect classifier has an AUC of 1.0 and the random guess an AUC of 0.5.

**segmentation:** The process of delineating an image object.

**sensitivity:** The probability that a positive case will be correctly decided,

$$N_{TP}/(N_{TP} + N_{FN})$$

where  $N_{TP}$  is the number of true positives and  $N_{FN}$  is the number of false negatives.

**specificity:** The probability that a negative case will be correctly decided,

$$N_{TN}/(N_{TN} + N_{FP})$$

where  $N_{TN}$  is the number of true negatives and  $N_{FP}$  is the number of false positives.

**tree:** A graph without cycles.

**undersegmentation:** A segmentation that assigns too few regions to an image or image object.

**Voronoi tessellation:** Given a set of nodes in the plane, the Voronoi tessellation creates polygonal cells around the nodes such that all pixels within a given cell are closer to the cell node than any other node in the plane.

**watershed transform:** A method of segmentation whereby the image is viewed as a 3-dimensional landscape; assuming a hole is created in each regional minimum, the landscape is immersed and as the basins fill with water, watershed lines are created where the basins eventually merge to become one.

**Wilcoxon test:** An alternative to the common Student's t-test for situations in which the underlying data distribution cannot be assumed to be normal and/or the classes cannot be assumed to have the same variance. The paired Wilcoxon signed-rank test is a two-sided test used to compare the distributions of two related measurements, and specifically tests for a zero median value of the pairwise sample differences. The p-value is the probability that the samples are symmetrically distributed about a single central (i.e., median) point [86].

# Appendix C

## Glossary of Acronyms and Symbols

### C.1 Acronyms

<b>AFE</b>	Automated Feature Extraction
<b>AUC</b>	Area Under the Curve
<b>BNLL</b>	Binomial Negative Log Likelihood
<b>bpp</b>	bits per pixel
<b>C</b>	Cytoplasm class
<b>CA</b>	Concavity Alignment
<b>CC</b>	Concavity-Concavity alignment
<b>CCD</b>	Charge Coupled Device
<b>CD</b>	Concavity Depth
<b>CG</b>	Connected Graph
<b>CL</b>	Concavity-Line alignment
<b>CR</b>	Concavity Ratio
<b>CS</b>	combined Cytoplasm and Stroma class
<b>DR</b>	Detection Rate
<b>DT</b>	Delaunay Triangulation
<b>EP</b>	Extra Pixels
<b>FAR</b>	False Alarm Rate
<b>FD</b>	Fractal Dimension
<b>FIM</b>	Feature Intensity Map
<b>FLDA</b>	Fisher Linear Discriminant Analysis
<b>FNA</b>	Fine Needle Aspiration
<b>GA</b>	Genetic Algorithm
<b>GENIE</b>	GENetic Imagery Exploitation
<b>GG</b>	Gabriel's Graph
<b>GPAC</b>	Graph Partitioning Active Contours

---

<b>GT</b>	Ground Truth
<b>H&amp;E</b>	Hematoxylin & Eosin
<b>HSI</b>	Hue, Saturation, Intensity
<b>HSV</b>	Hue, Saturation, Value
<b>ICA</b>	Independent Components Analysis
<b>IOD</b>	Integrated Optical Density
<b>IS</b>	In Sample
<b>JEPD</b>	Jointly Exhaustive and Pairwise Disjoint
<b>k-NN</b>	k-Nearest Neighbor
<b>KU</b>	University of Kansas
<b>LCTF</b>	Liquid Crystal Tunable Filter
<b>LoG</b>	Laplacian of Gaussian
<b>MED</b>	Minimum Euclidean Distance
<b>ML</b>	Maximum Likelihood
<b>MOD</b>	Mean Optical Density
<b>MST</b>	Minimum Spanning Tree
<b>N</b>	Nuclei class
<b>NGA</b>	National Geospatial-Intelligence Agency
<b>OCG</b>	O'Callaghan's Neighborhood Graph
<b>OD</b>	Optical Density
<b>OS</b>	Out-of-Sample
<b>P</b>	Performance
<b>PCA</b>	Principal Components Analysis
<b>PGM</b>	Probabilistic Graph Model
<b>PM</b>	Pixels Missed
<b>QOALA</b>	Quantitative Object- and spatial Arrangement-Level Analysis
<b>QS</b>	Quadrant Sum
<b>RCC</b>	Region Connection Calculus
<b>RCC8</b>	Region Connection Calculus, 8 JEPD relations
<b>RGB</b>	Red, Green, Blue
<b>ROC</b>	Receiver Operating Characteristics
<b>S</b>	Stroma class
<b>SA</b>	Saliency
<b>SAM</b>	Spectral Angle Mapper
<b>SE</b>	Structuring Element
<b>SPN</b>	SPatial Nuclei class
<b>SR</b>	Segmentation Regions
<b>SSA</b>	Syntactic Structure Analysis
<b>SVM</b>	Support Vector Machine
<b>LSVM</b>	Linear Support Vector Machine
<b>NLSVM</b>	Nonlinear Support Vector Machine
<b>SBS</b>	Sequential Backward Selection

---

<b>SFBS</b>	Sequential Floating Backward Selection
<b>SFFS</b>	Sequential Floating Forward Selection
<b>SFS</b>	Sequential Forward Selection
<b>UMA</b>	Ultimate Measurement Accuracy [120]
<b>VT</b>	Voronoi Tessellation
<b>WSBlob</b>	Watershed using Blobdetector-based markers
<b>WSCDT</b>	Watershed on the Complemented Distance Transform
<b>WSCM</b>	Watershed using Concavity-based Markers
<b>WSGran</b>	Watershed using Granulometry-based markers
<b>WSHmin</b>	Watershed on the H-minima transform
<b>YUV</b>	luminance (Y) and chrominance (UV) color space

## C.2 Symbols

$\forall$	for all
$\exists$	there exists
$\nexists$	there does not exist
$\neg$	complement
$\rightarrow$	implies
$\wedge$	and
$\cap$	intersect
$\cup$	union
$^\top$	transpose
$ $	given
$\angle$	angle
$\in$	‘element of’ or ‘in’
$\notin$	‘not an element of’ or ‘not in’
$\propto$	proportional to
$\circ$	morphological opening
$\ominus$	morphological erosion
$\sim$	binary complement
$\equiv$	‘defined as’ or ‘equivalent to’
$\mathbb{Z}$	the set of integers
$\emptyset$	empty set

## Appendix D

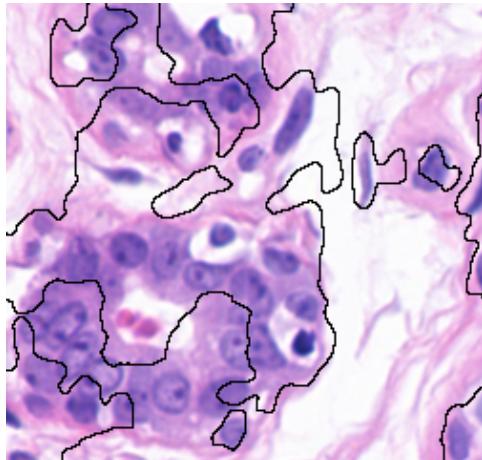
# Color- and Texture-Based Classification of Nuclei

The curve evolution and active contour methods are popular image segmentation methods, which evolve an image segmentation curve based on various energies included in the cost function. As a simple example, one cost function term may penalize sharp variations in curvature throughout the boundary. A commonly used technique is that of Chan and Vese, presented in [203], where the curve evolution method is simplified by the use of a piecewise constant approximation of the cost function. The Graph Partitioning Active Countours (GPAC) approach in [204] uses pairwise pixel similarities for construction of the cost function, as well as efficient implementation techniques for the minimization of the cost function. The modification of GPAC presented in [205] uses prior information (in the form of reference/training images) within the same variational framework. Moreover, the modified GPAC method introduces and uses a dissimilarity measure based on color and texture information. The color dissimilarity is the Euclidean distance in the three-channel (RGB) color space, and the texture dissimilarity is based on a distance in the Gabor filter space using 5 scales and 6 orientations [205].

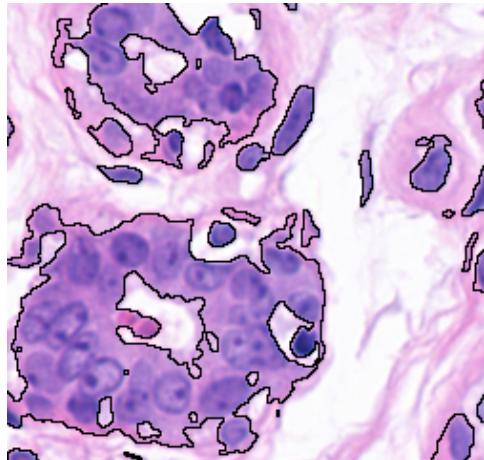
We have applied the modified GPAC method to our imagery to gauge the effectiveness of the inclusion of texture information for our nuclear classification task.<sup>1</sup> Results are shown for an example image in Figure D.1, where we note that the segmentation using solely texture information (Figure D.1 (a)) is a much poorer segmentation than that using solely color information (Figure D.1 (b)). The segmentation using equally weighted color and texture information (Figure D.1 (c)) indicates that the texture term does not add any useful information for the curve evolution, and may actually degrade performance slightly. Even using solely color information, however, this method is inaccurate for the classification of nuclear pixels, as it classifies much cytoplasm as nuclei.

---

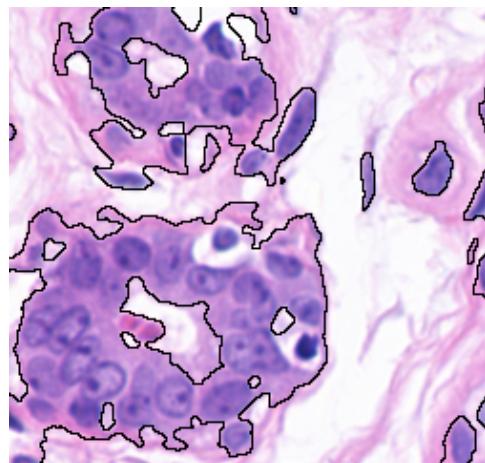
<sup>1</sup>Many thanks to Luca Bertelli, first author of [205], for his work applying the modified GPAC method to our imagery, and for technical consultation regarding his method.



(a) Classification using only texture information.



(b) Classification using only color information.



(c) Classification using color and texture information.

**Figure D.1.** Example outputs of the modified GPAC method, showing the differences in nuclear classification utilizing solely texture information (a), solely color information (b), and both texture and color information (equally weighted) (c). Note that the texture information does not appear to improve the classification performance, and may actually be detrimental.

## Appendix E

# An Introduction to GENIE and Genie Pro

The Yale pathologists have been using the GENIE and Genie Pro systems for quantitative analysis of their multispectral imagery and results have been promising (see [39, 43]). Additionally, Genie Pro was included in a recent National Geospatial-Intelligence Agency (NGA) “bake-off” comparing automated feature extraction tools. The NGA surveyed all existing feature extraction tools and chose four for the final analysis: Feature Analyst (Visual Learning Systems, Inc., Missoula, MT), eCognition (Definiens AG, München, Germany), Genie Pro (Los Alamos National Laboratory, Los Alamos, NM), and Neural Fusion (BAE Systems, PLC, London, UK) [206, 207]. The NGA decided that “The results showed that GENIE Pro had the best overall performance in terms of total extraction time. User feedback for GENIE Pro was also very favorable, indicating that GENIE performed well and was easy to use” [206].

GENetic Imagery Exploitation (GENIE) software (Los Alamos National Laboratory, Los Alamos, NM) is an evolutionary computation-based software system designed to evolve feature extraction algorithms for multispectral imagery, specifically for remote sensing imagery. Algorithms consist of spatial and spectral operators designed to extract features of interest from the data. GENIE operates in a manner analogous to the reproduction of biological systems, namely that “chromosomes” (sequences of image processing operators, “genes”) with high fitness (good performance in classifying training data) have a higher probability of “reproducing” to form the next generation of “chromosomes”. While this biological analogy is a good tool to understand the mechanisms behind genetic algorithms (GAs), the chromosome/gene terminology has fallen out of favor in some applications, particularly in biomedical analysis due to the confusion with actual chromosomes and genes of cells. We will, therefore, refer to the GENIE “chromosomes” as image processing algorithms (or algorithms) and the GENIE “genes” as image processing operators (or operators).

GENIE is designed to explore the entire feature space of multispectral data, and evolve a solution best fit for the classification task. More practically speaking, GENIE selects a subset from the available data planes and develops an algorithm, consisting of spectral and spatial operators, to extract various feature planes which are subsequently fed to a standard classifier backend. GENIE selects an initial set of algorithms consisting of randomly selected operators and randomly selected data planes as input. Throughout the evolution process, only appropriate algorithms with appropriate data input will survive, as quantified by the classification performance of the algorithms. Classification performance is defined as an equal tradeoff between detection rate and false alarm rate:

$$P = 0.5(R_d + (1 - R_f)), \quad (\text{E.1})$$

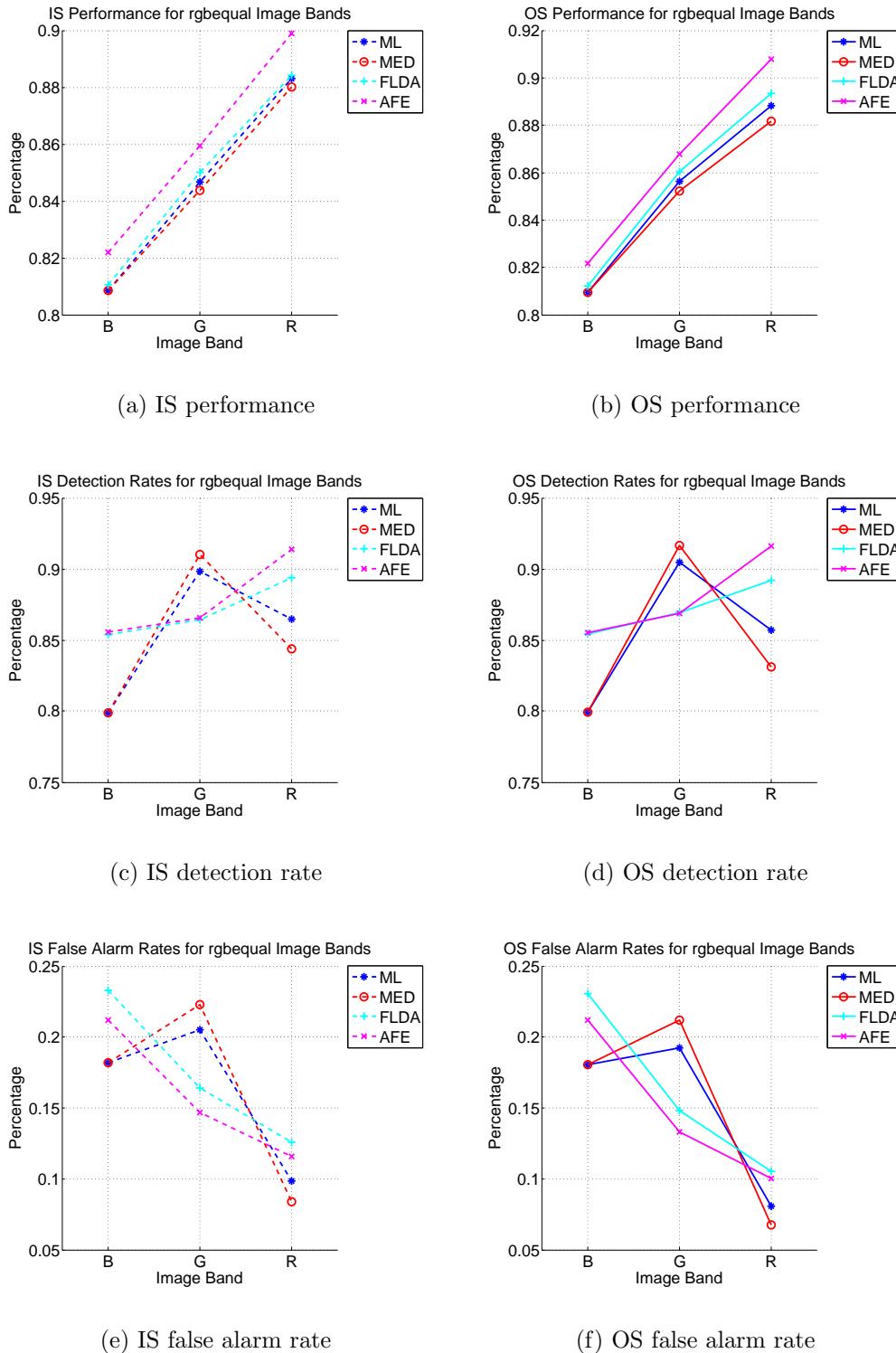
where  $R_d$  is the fraction of true pixels classified correctly (detection rate),  $R_f$  is the fraction of false pixels classified incorrectly (false alarm rate), and the factor of 0.5 scales the metric to the range [0, 1]. Note that a perfect segmentation will yield a performance score of 1 (100%), while a score of 0.5 (50%) can be obtained by a trivial solution of all pixels labeled true (or false).

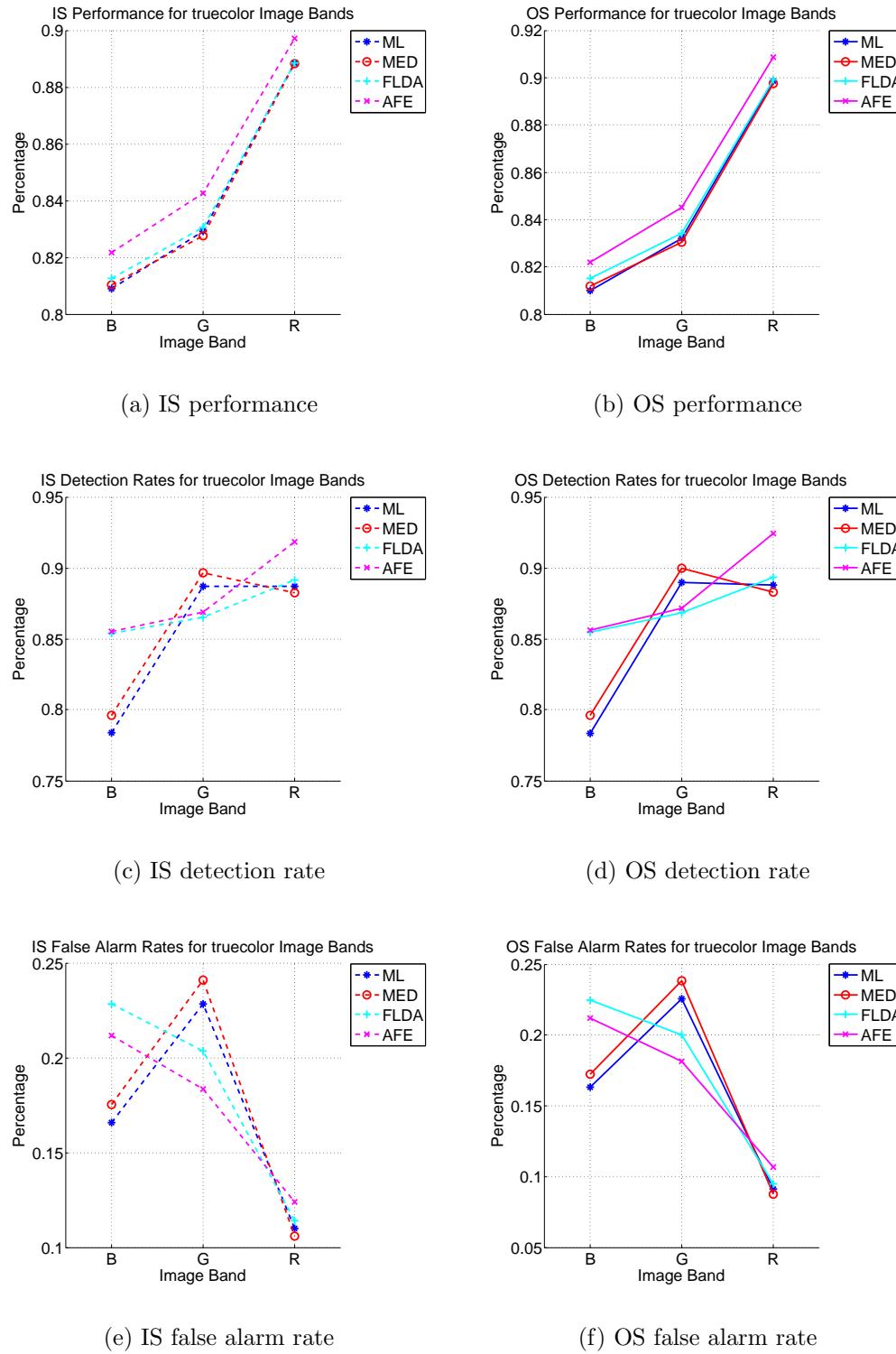
Genie Pro is a newer, commercial-quality version of GENIE, where the genetic algorithm has been replaced with a greedy hill climbing algorithm to help speed the training process. Additionally, Genie Pro searches for the least complex algorithm among any with equal performance. The pool of operators in Genie Pro has been pared down with respect to GENIE, although the possibility exists to implement new operators.

Operators implemented in GENIE and GeniePro include spectral, textural, and mathematical morphology-derived operators. Additionally, the user interface for GENIE/GeniePro has been developed to efficiently elicit information from a human expert. This is facilitated through a set of tools that allows the user to “paint” the image regions corresponding to various object classes. The number of object classes is variably selected by the user for the task at hand. For more information, see Reference [83].

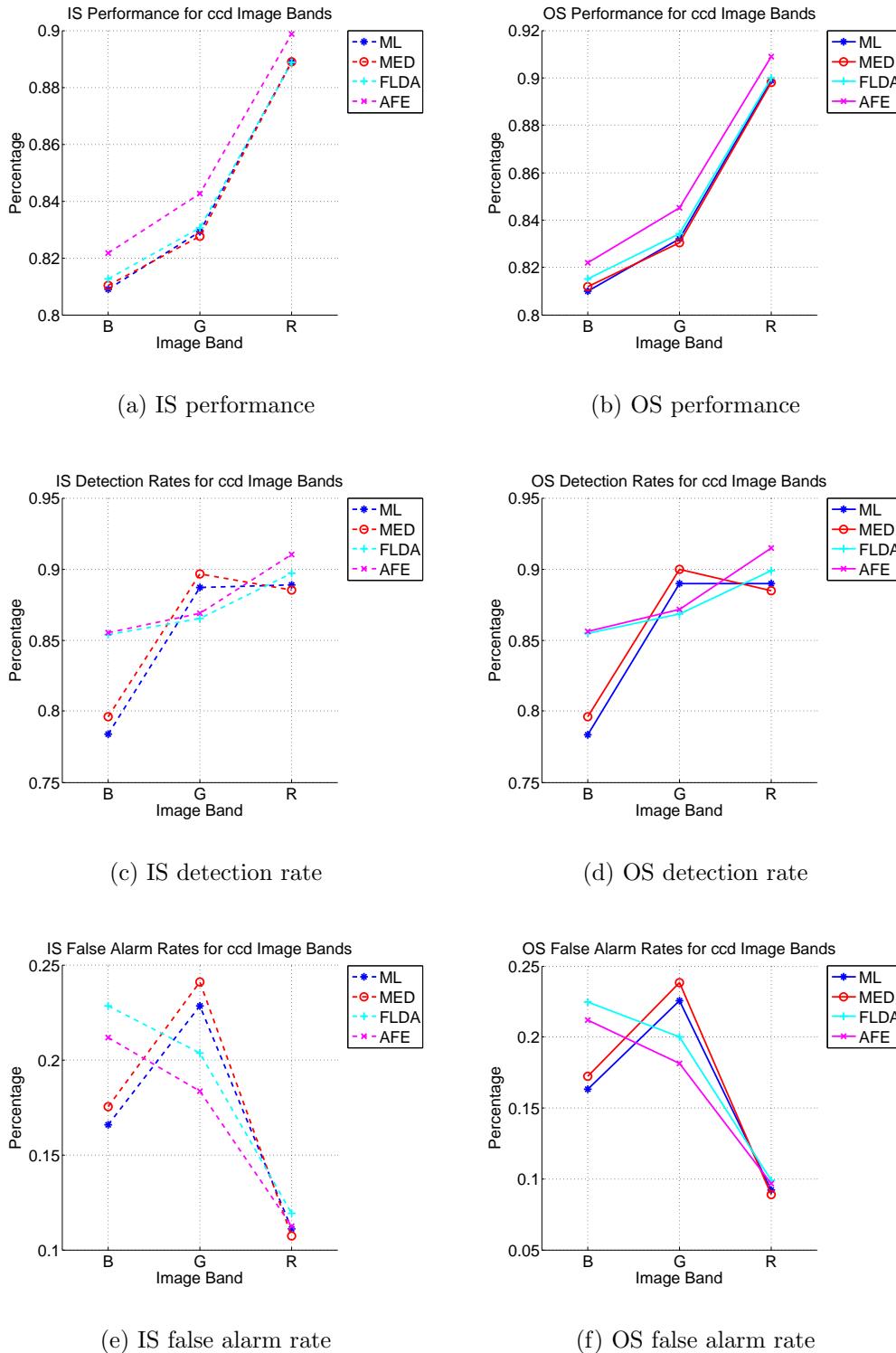
## Appendix F

### Extra Plots: Classification of Individual RGB Image Bands

**Figure F.1.** Classification P, DR, and FAR for individual rgbequal bands.

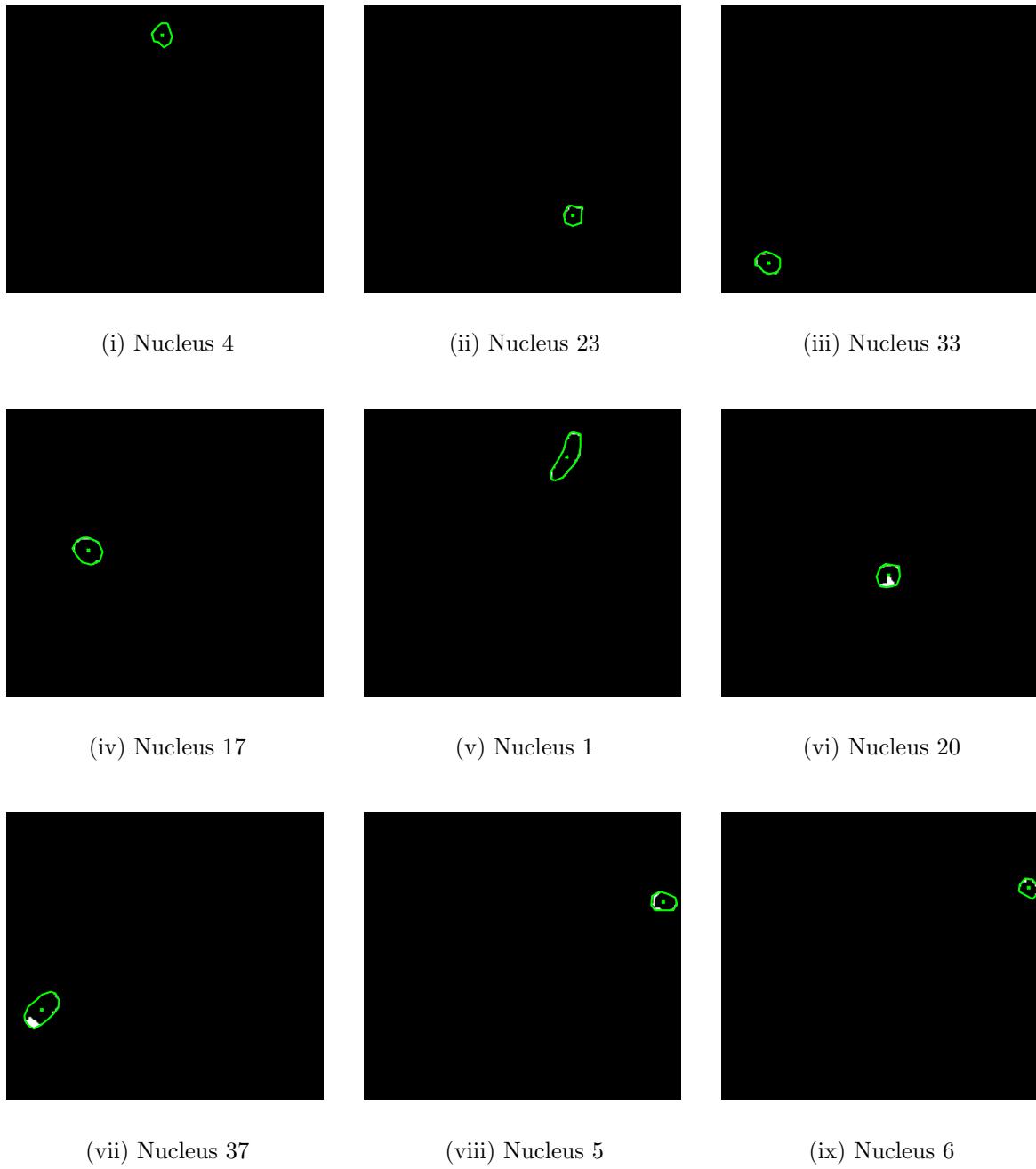


**Figure F.2.** Classification P, DR, and FAR for individual truecolor bands.

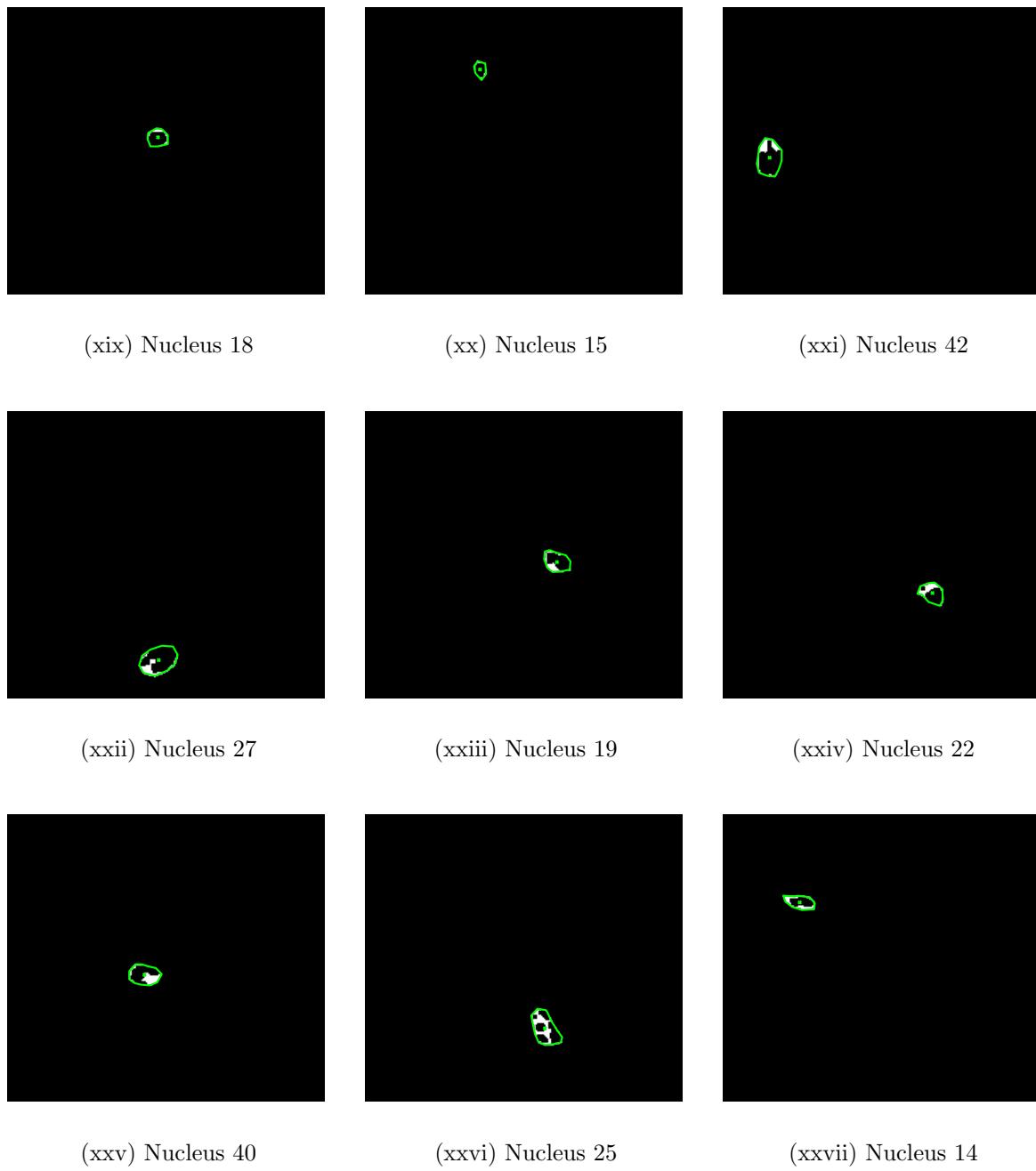
**Figure F.3.** Classification P, DR, and FAR for individual ccd bands.

## Appendix G

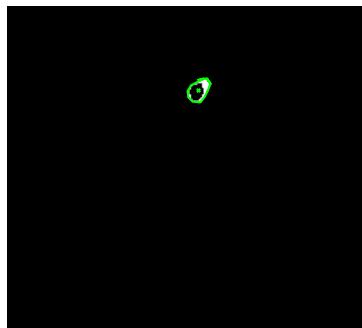
### Ranking of Extra Pixels and Pixels Missed by Nucleus



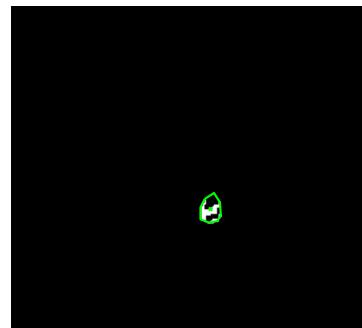
**Figure G.1.** Pixels missed (PM) ranked by nucleus according to size and quadrant sum, displayed in ascending order (best to worst).



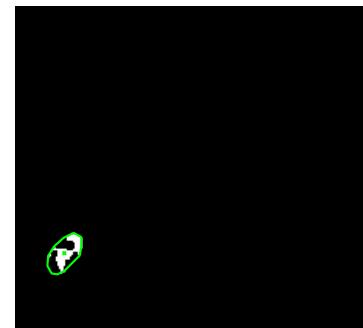
**Figure G.1.** (Cont.) Pixels missed (PM) ranked by nucleus according to size and quadrant sum, displayed in ascending order (best to worst).



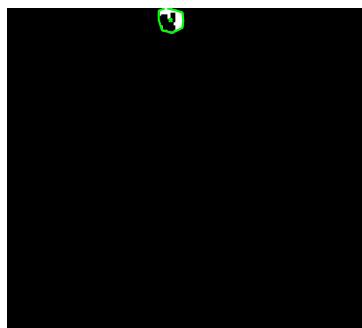
(xlvi) Nucleus 2



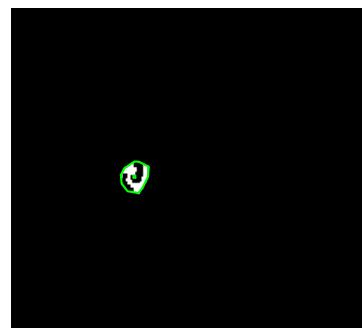
(xlvii) Nucleus 21



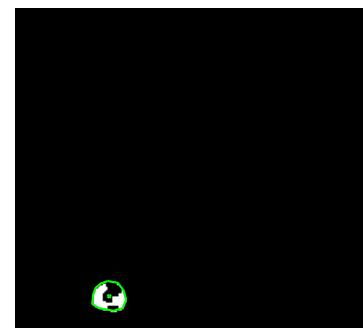
(xlviii) Nucleus 34



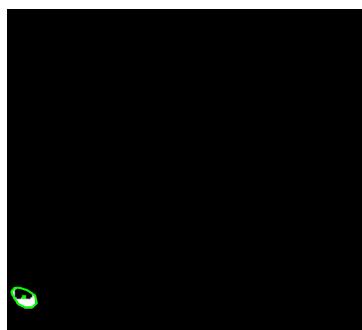
(xlix) Nucleus 8



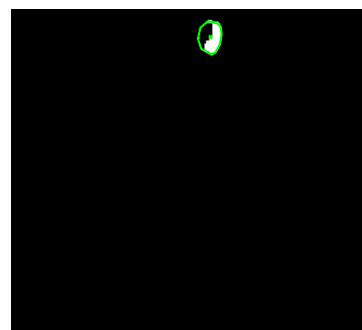
(l) Nucleus 41



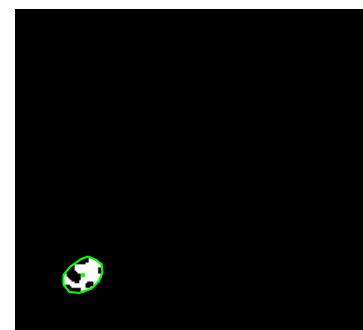
(li) Nucleus 31



(lii) Nucleus 36

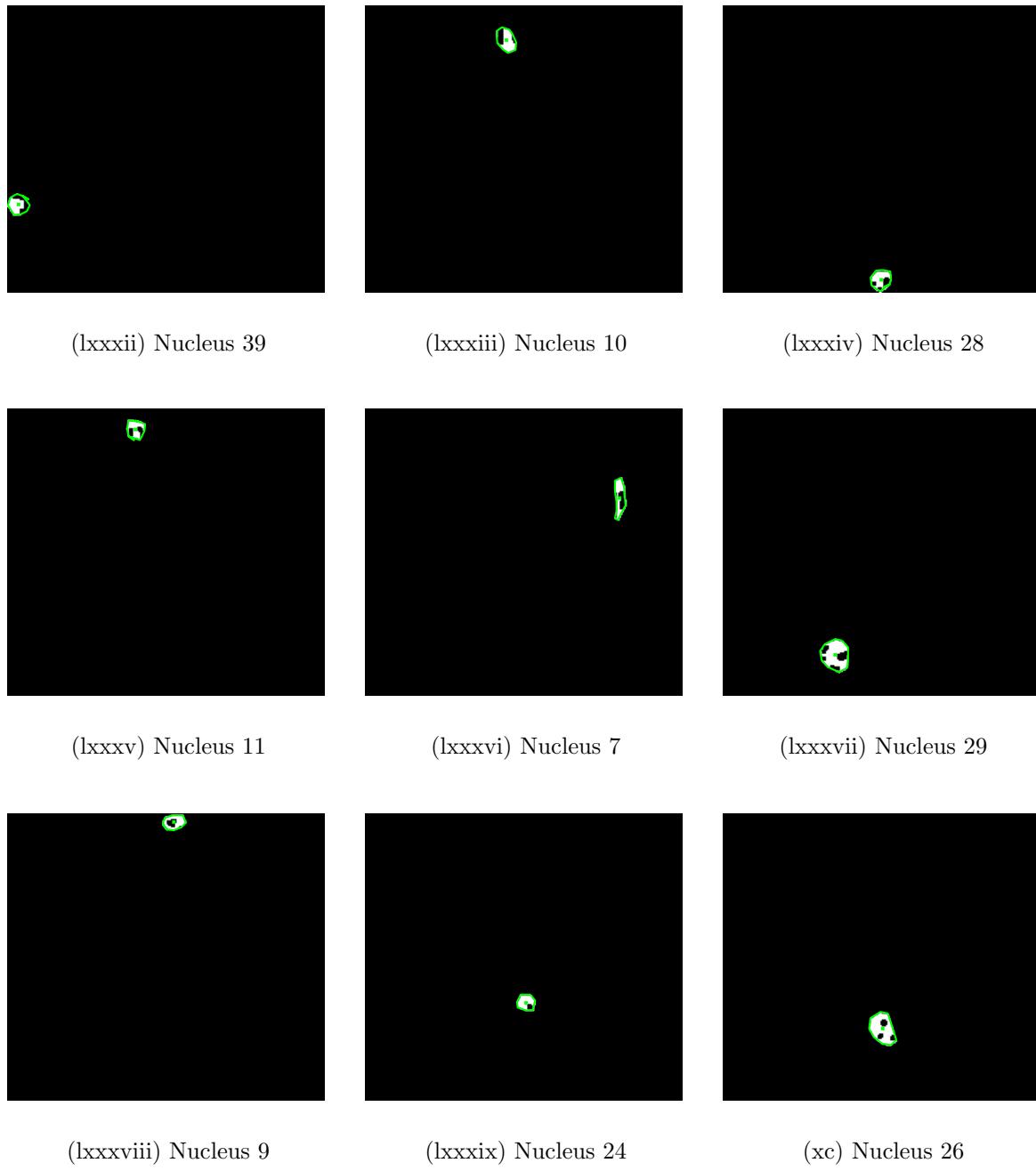


(liii) Nucleus 3

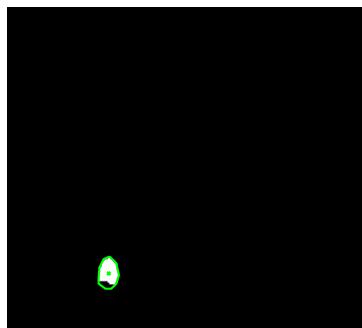


(liv) Nucleus 32

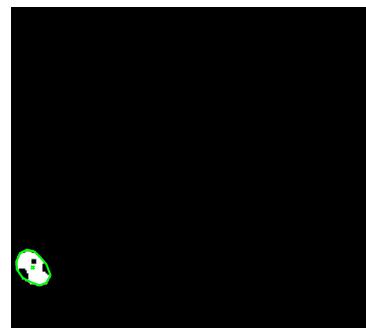
**Figure G.1.** (Cont.) Pixels missed (PM) ranked by nucleus according to size and quadrant sum, displayed in ascending order (best to worst).



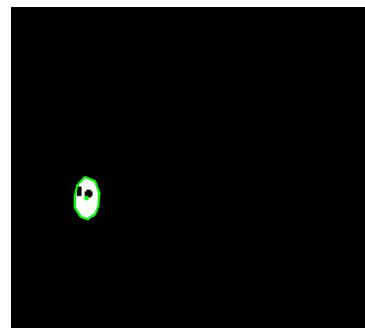
**Figure G.1.** (Cont.) Pixels missed (PM) ranked by nucleus according to size and quadrant sum, displayed in ascending order (best to worst).



(cxxvii) Nucleus 30



(cxxviii) Nucleus 35

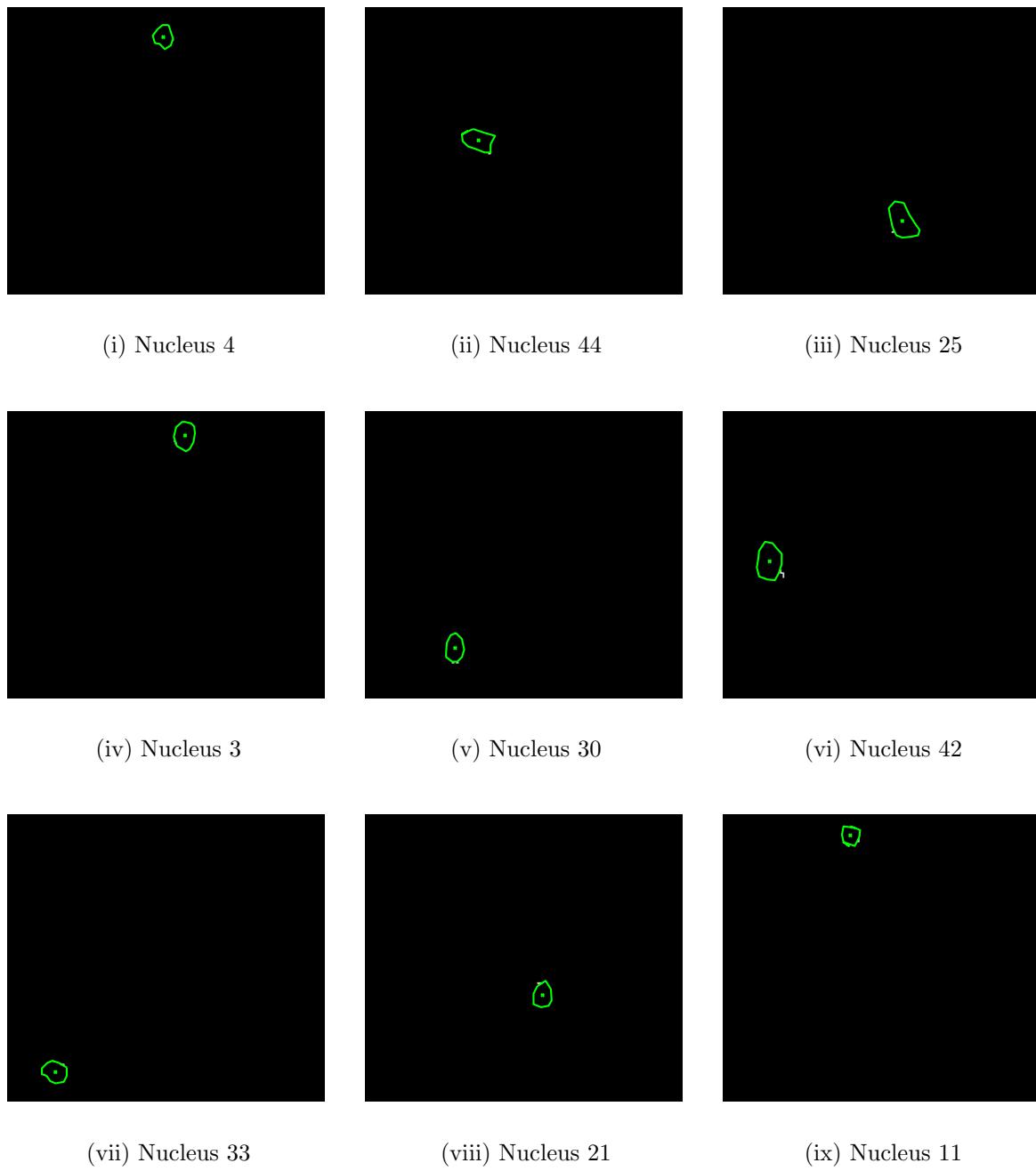


(cxxix) Nucleus 43

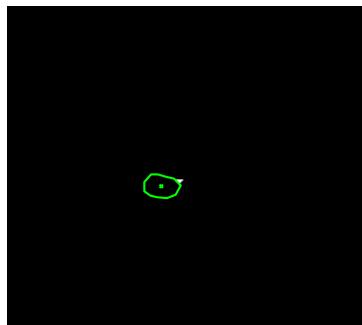


(cxxxi) Nucleus 44

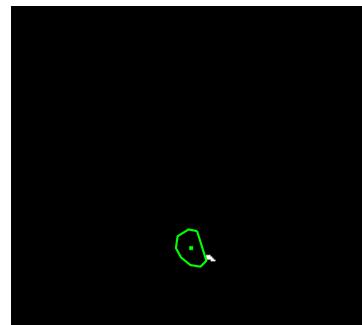
**Figure G.1.** (Cont.) Pixels missed (PM) ranked by nucleus according to size and quadrant sum, displayed in ascending order (best to worst).



**Figure G.2.** Extra pixels (EP) ranked by nucleus according to size and quadrant sum, displayed in ascending order (best to worst).



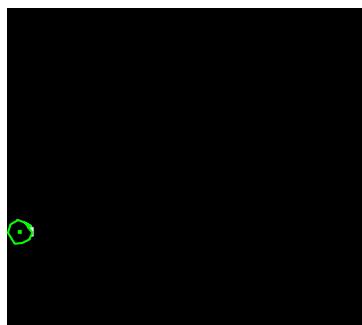
(xix) Nucleus 40



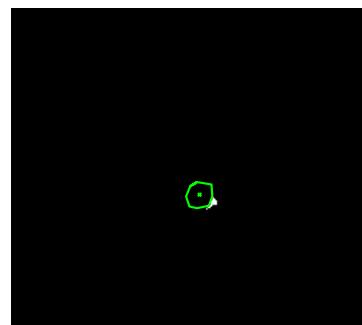
(xx) Nucleus 26



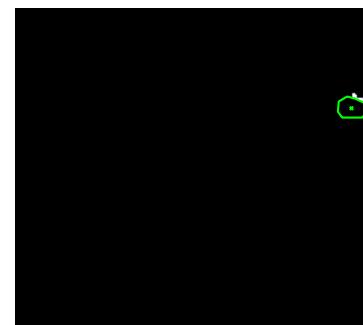
(xxi) Nucleus 35



(xxii) Nucleus 39



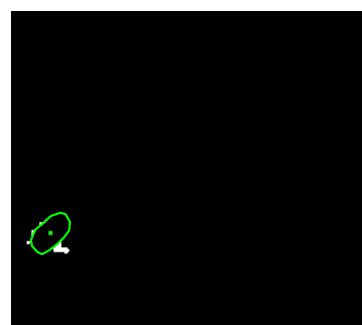
(xxiii) Nucleus 20



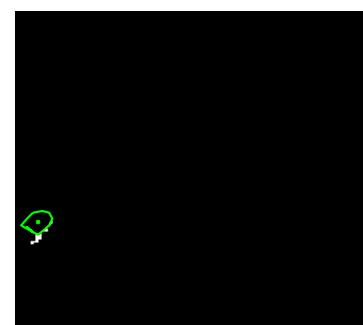
(xxiv) Nucleus 5



(xxv) Nucleus 34

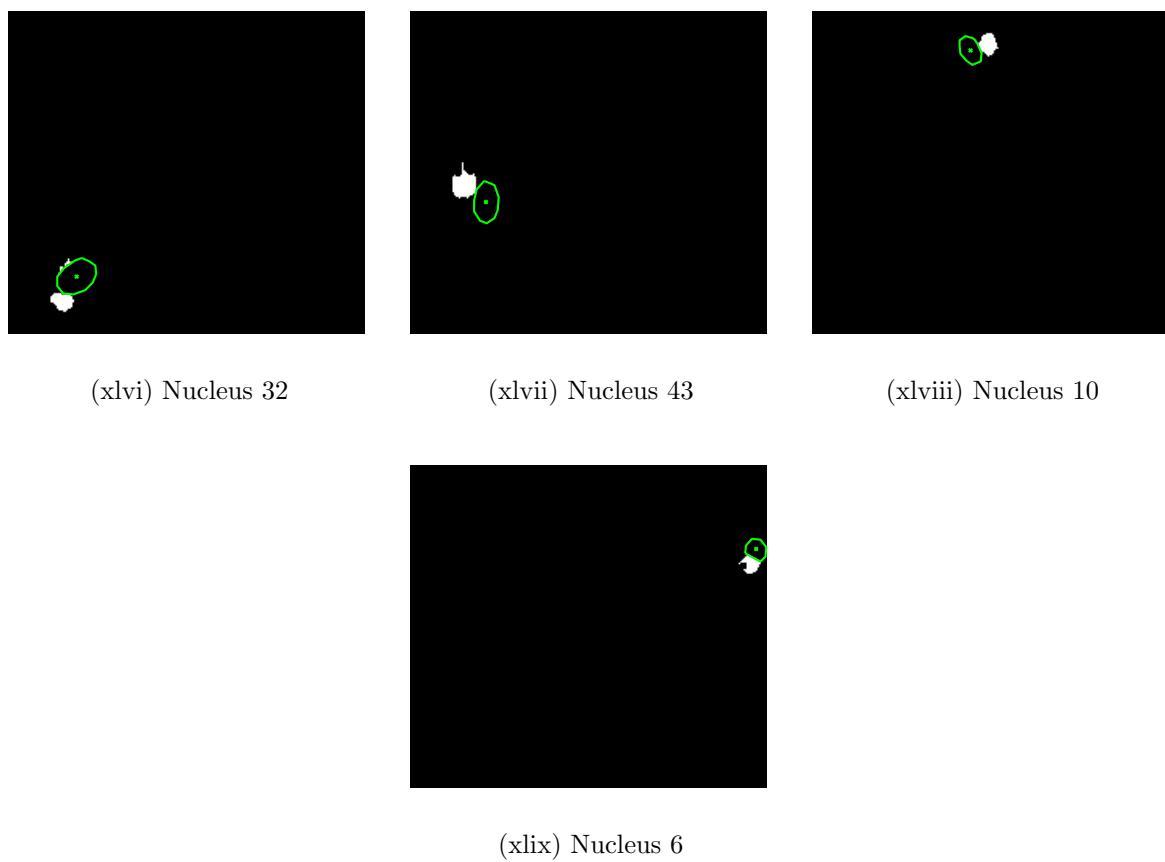


(xxvi) Nucleus 37



(xxvii) Nucleus 38

**Figure G.2.** (Cont.) Extra pixels (EP) ranked by nucleus according to size and quadrant sum, displayed in ascending order (best to worst).



**Figure G.2.** (Cont.) Extra pixels (EP) ranked by nucleus according to size and quadrant sum, displayed in ascending order (best to worst).



## Appendix H

# Optical Density of Transmission Microscopy Images

For transmission microscopy of stained tissue, optical density (OD) provides a linear relationship between image intensity and staining density. This is based on Lambert-Beer's law describing the intensity of light transmitted through a specimen:

$$I = I_0 * 10^{-A*c*l} \quad (\text{H.1})$$

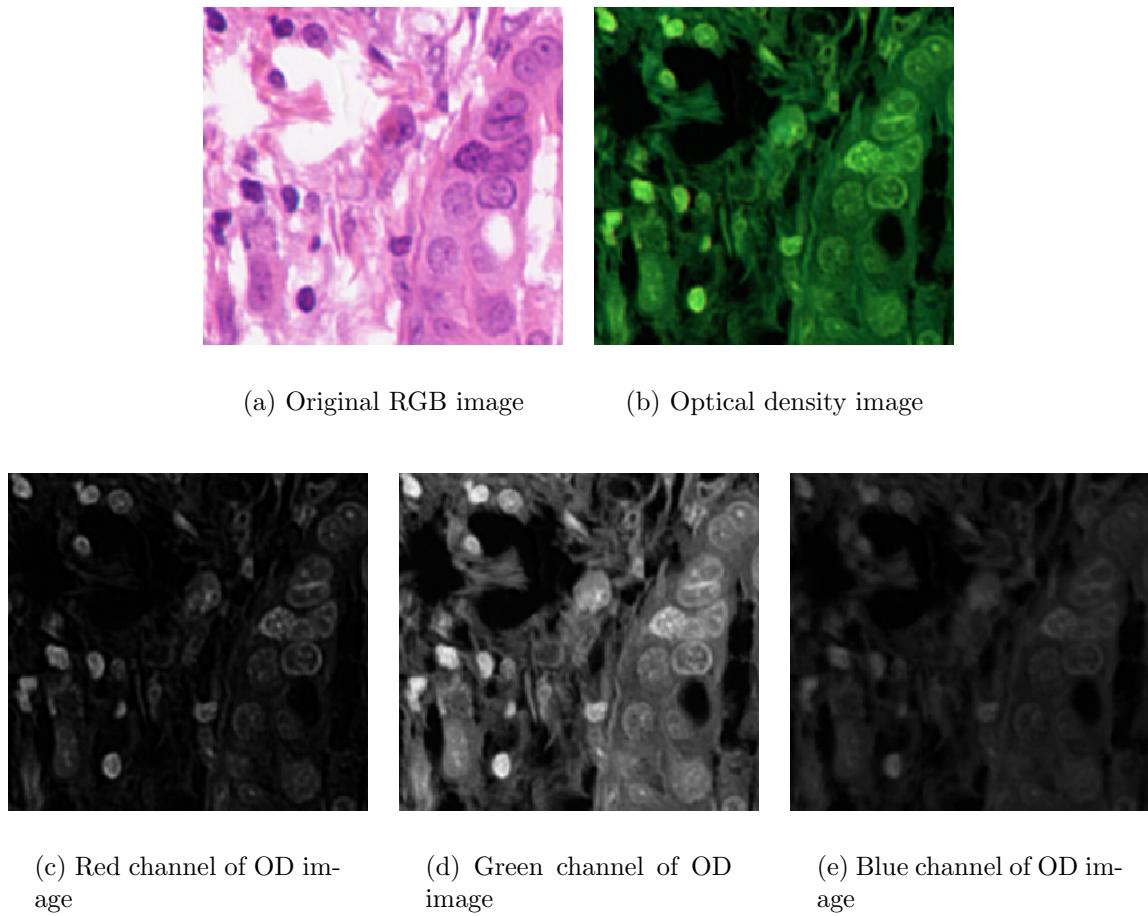
where  $I$  is the intensity of the observed light,  $I_0$  is the intensity of incident light,  $A$  is the amount of stain,  $c$  is the absorption factor of the stain, and  $l$  is the distance traveled through the sample [201, 202]. OD reduces this relationship to a linear one [201, 202]:

$$OD = -\frac{1}{l} \log_{10}\left(\frac{I}{I_0}\right) = A * c \quad (\text{H.2})$$

OD is calculated on a channel-by-channel basis and can be used in applications other than stained medical imagery, anytime the absorption of light by a material is of importance; the equations are generally specified in a similar fashion.

An example RGB image and the corresponding OD image can be seen in Figure H.1, where we estimate the incident light intensity  $I_0$  for each channel by the brightest pixel. Note that, in the OD image, brighter pixels correspond (linearly) to larger amounts of stain. For the image of Figure H.1, there are two stains present, hematoxylin and eosin, each with a different absorption factor  $c$ , left over from the original formulation of Beer-Lambert's law. Without images of singly-stained tissue, we cannot directly separate the individual contributions of each stain, but will instead get an estimate of the total stain present.

The analysis in [201], however, presents relative contributions of hematoxylin and eosin to the R, G, and B channels as [0.1 0.20 0.08] and [0.01 0.13 0.01], respectively.



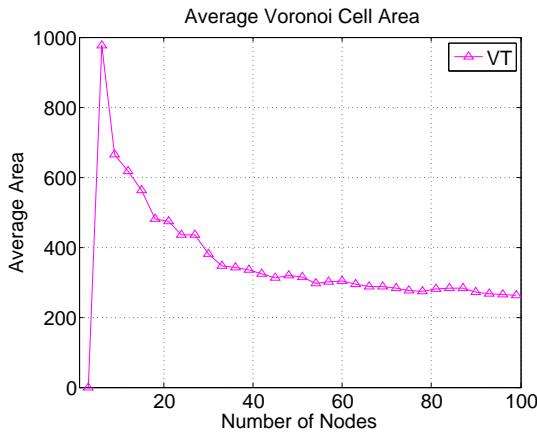
**Figure H.1.** Example optical density image.

Note that these relative contributions are qualitatively expressed in the optical density image of Figure H.1. The individual channels of Figure H.1 (b) are shown in (c)-(e) as reference. The hematoxylin (which stains nuclei) is represented in both the red and green channels, eosin (which stains cytoplasm and fibers) is represented mostly in the green channel, and the blue channel has very little response for either stain; this corroborates with the quantitative responses given in [201].

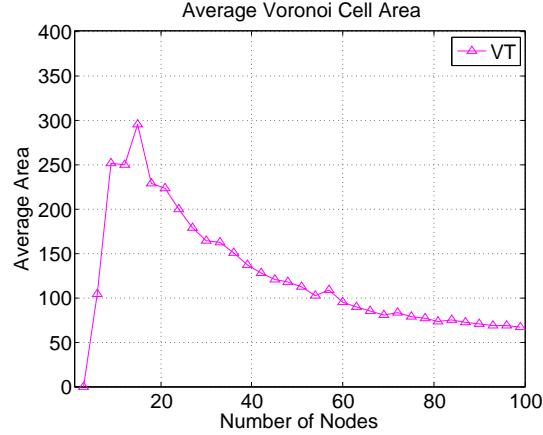
While we have presented here the optical density corresponding to a 3-channel RGB image, the same computation can be performed for each channel of a multispectral image stack. This operation is widely used in cyto/histopathology analysis.

# Appendix I

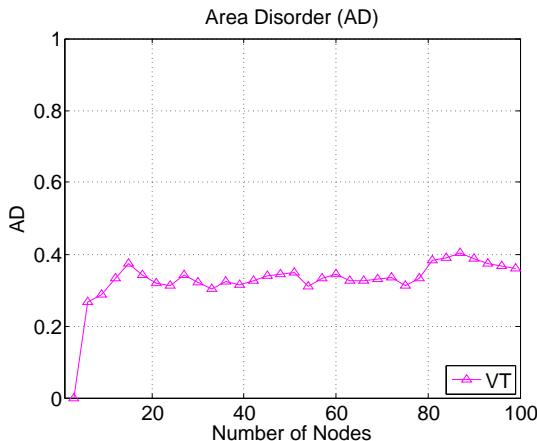
## Extra Plots: Random Graph Features



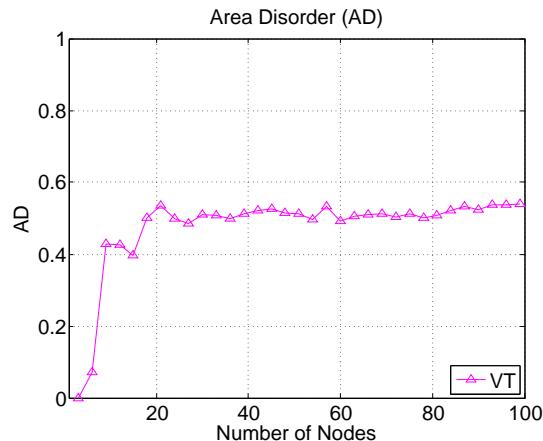
(i) Average Voronoi cell area, random graph type 1.



(ii) Average Voronoi cell area, random graph type 2.

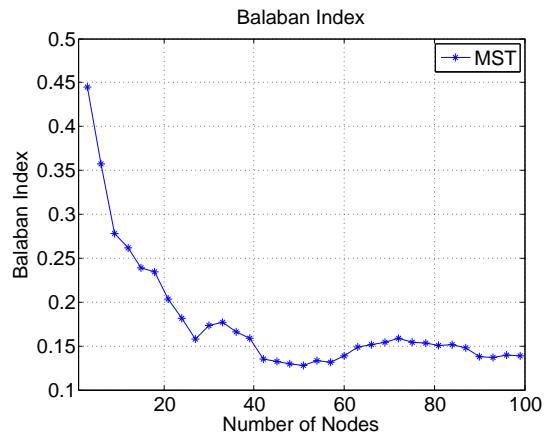


(iii) Area disorder, random graph type 1.

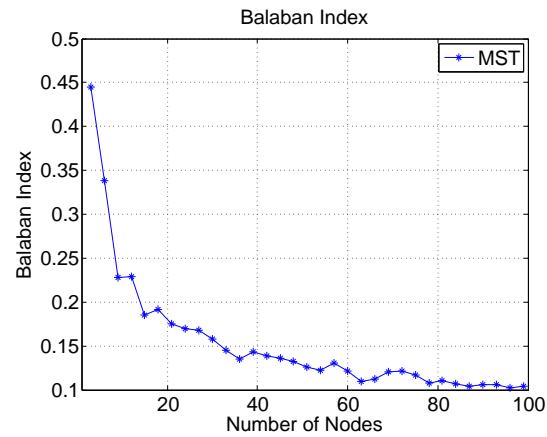


(iv) Area disorder, random graph type 2.

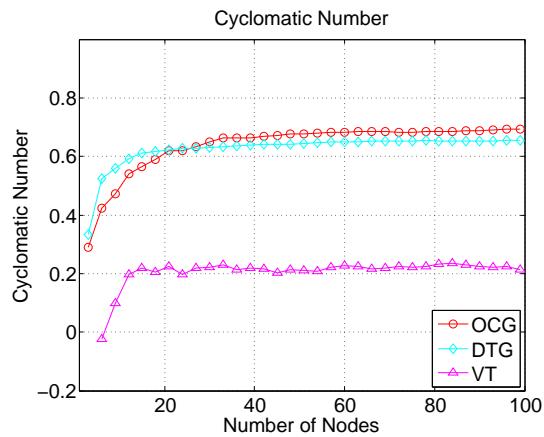
**Figure I.1.** Random graph features. (continued on next page)



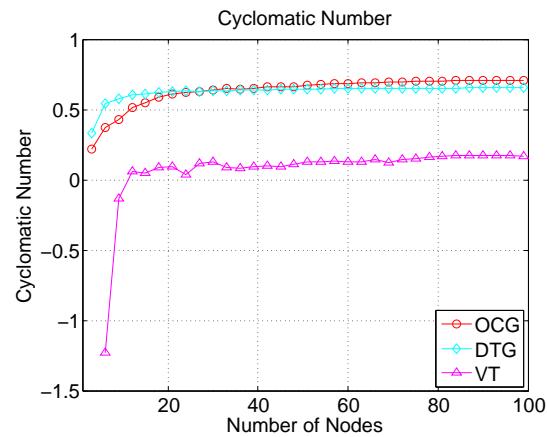
(v) Balaban index, random graph type 1.



(vi) Balaban index, random graph type 2.

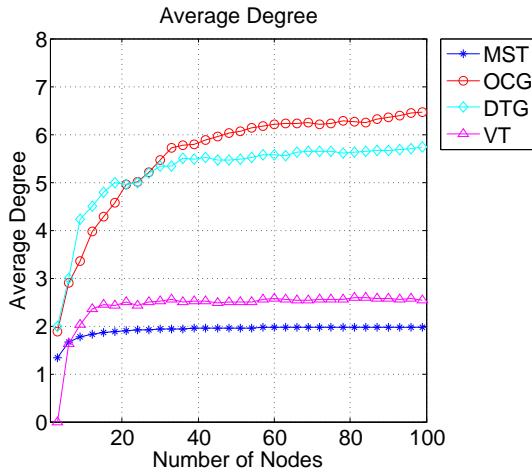


(vii) Cyclomatic number, random graph type 1.

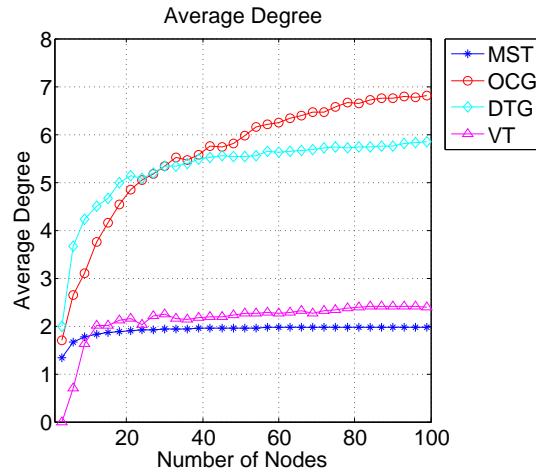


(viii) Cyclomatic Number, random graph type 2.

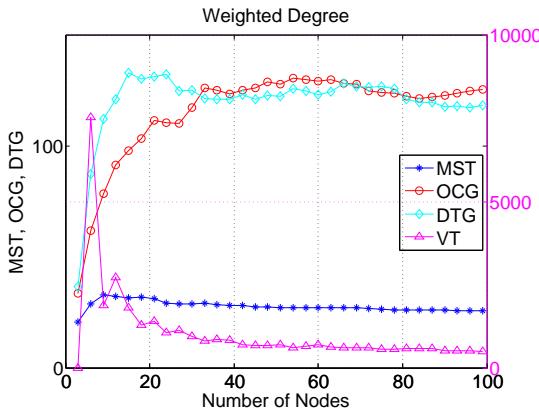
**Figure I.1.** (Cont.) Random graph features. (continued on next page)



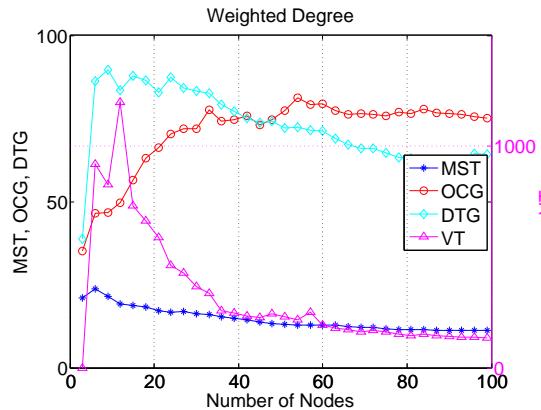
(ix) Average degree, random graph type 1.



(x) Average degree, random graph type 2.

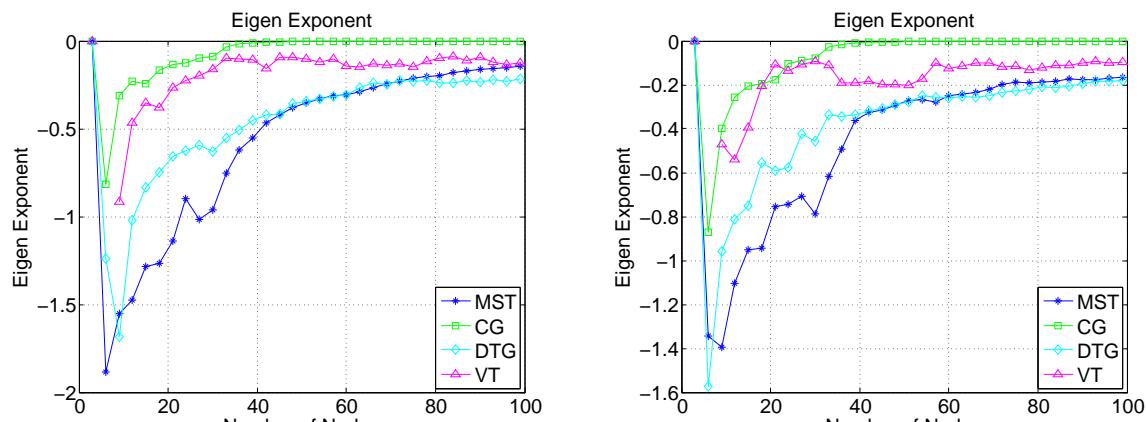
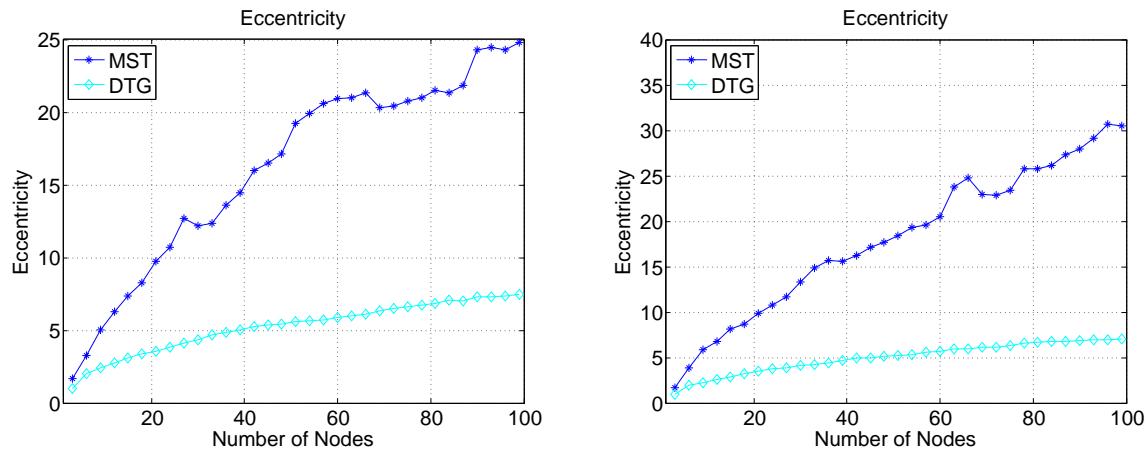


(xi) Average weighted degree, random graph type 1.

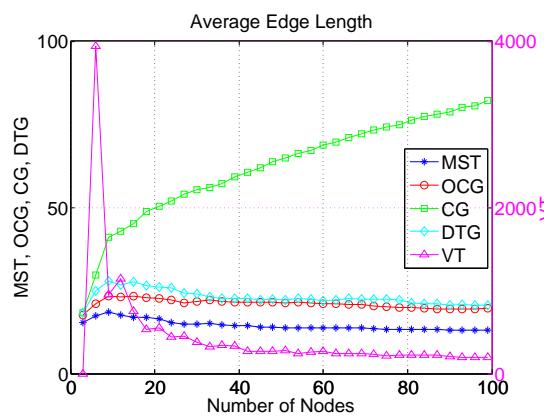


(xii) Average weighted degree, random graph type 2.

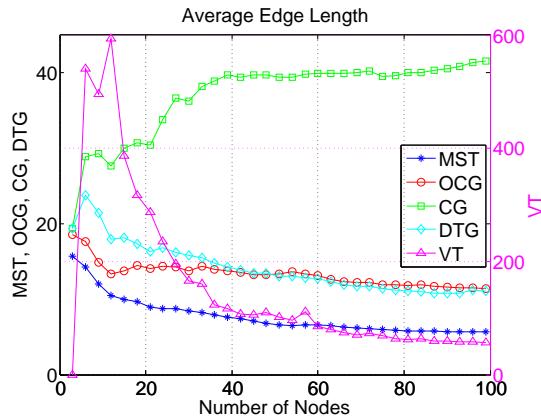
**Figure I.1.** (Cont.) Random graph features. (continued on next page)



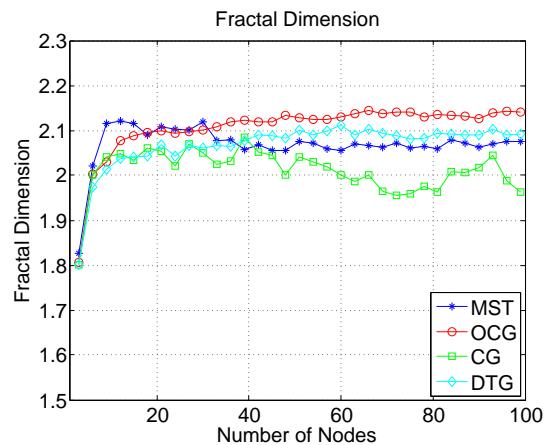
**Figure I.1.** (Cont.) Random graph features. (continued on next page)



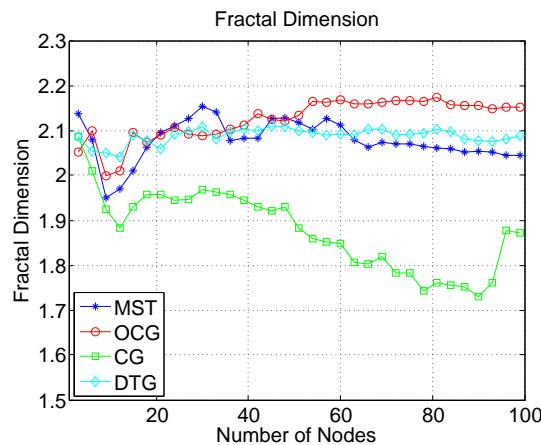
(xvii) Average edge length, random graph type 1.



(xviii) Average edge length, random graph type 2.

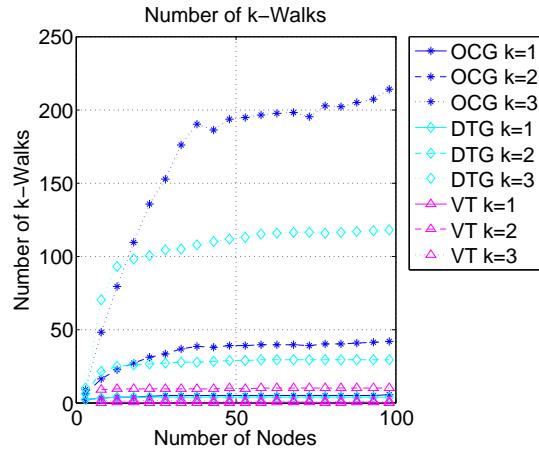


(xix) Fractal dimension, random graph type 1.

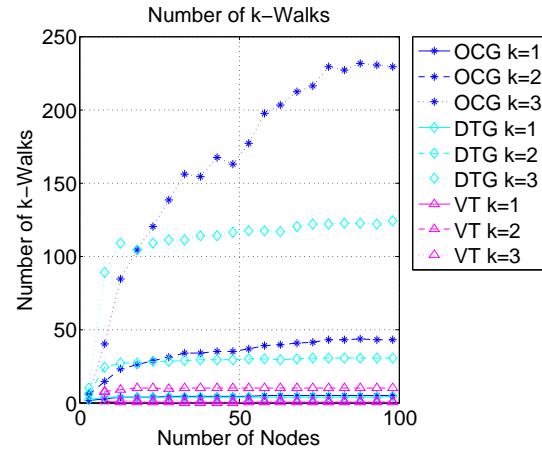


(xx) Fractal dimension, random graph type 2.

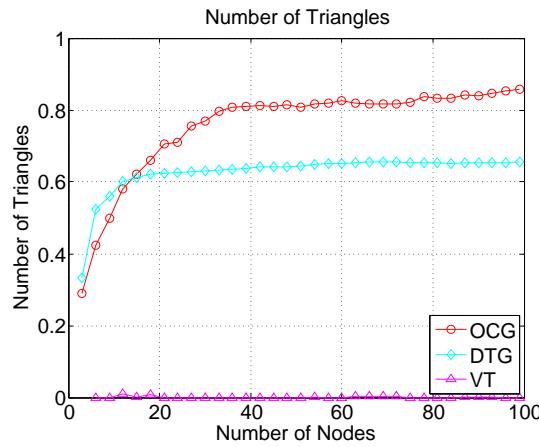
**Figure I.1.** (Cont.) Random graph features. (continued on next page)



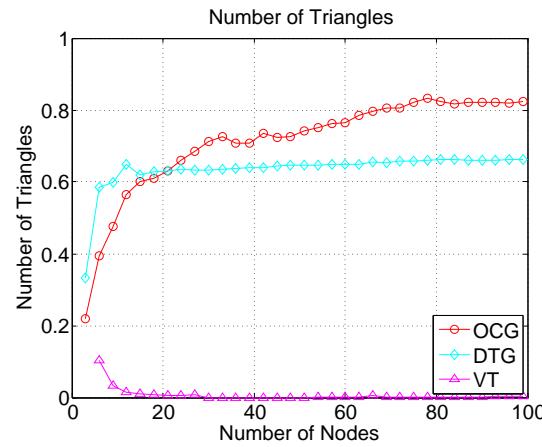
(xxi) Average number of k-walks, random graph type 1.



(xxii) Average number of k-walks, random graph type 2.

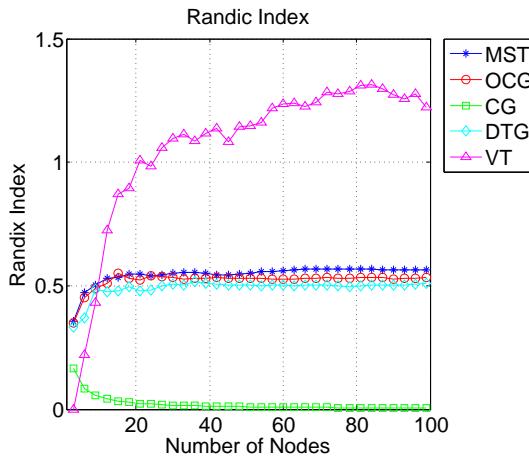


(xxiii) Average number of triangles, random graph type 1.

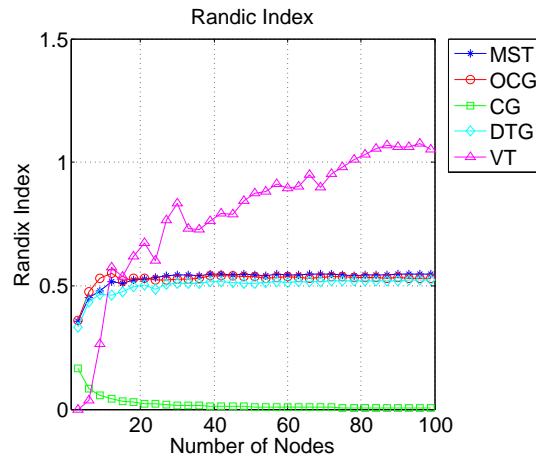


(xxiv) Average number of triangles, random graph type 2.

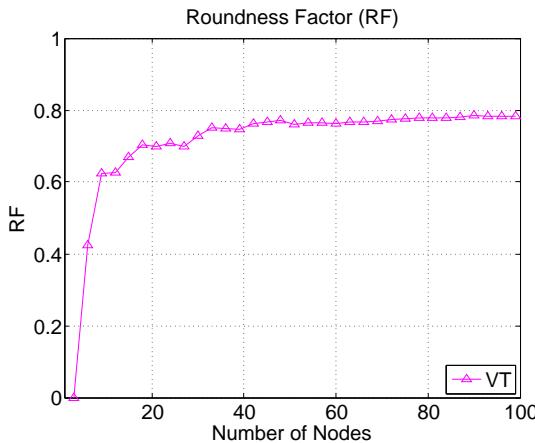
**Figure I.1.** (Cont.) Random graph features. (continued on next page)



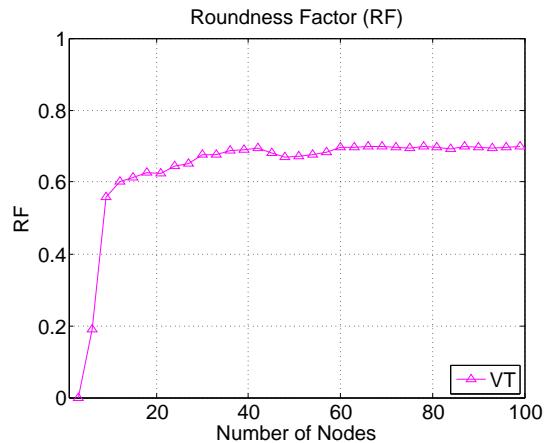
(xxv) Randic index, random graph type 1.



(xxvi) Randic index, random graph type 2.

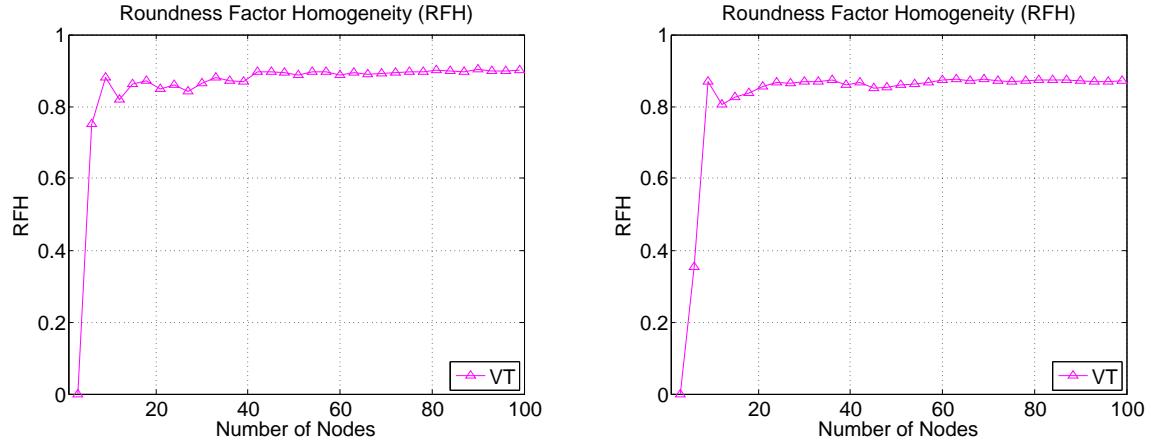


(xxvii) Roundness factor, random graph type 1.



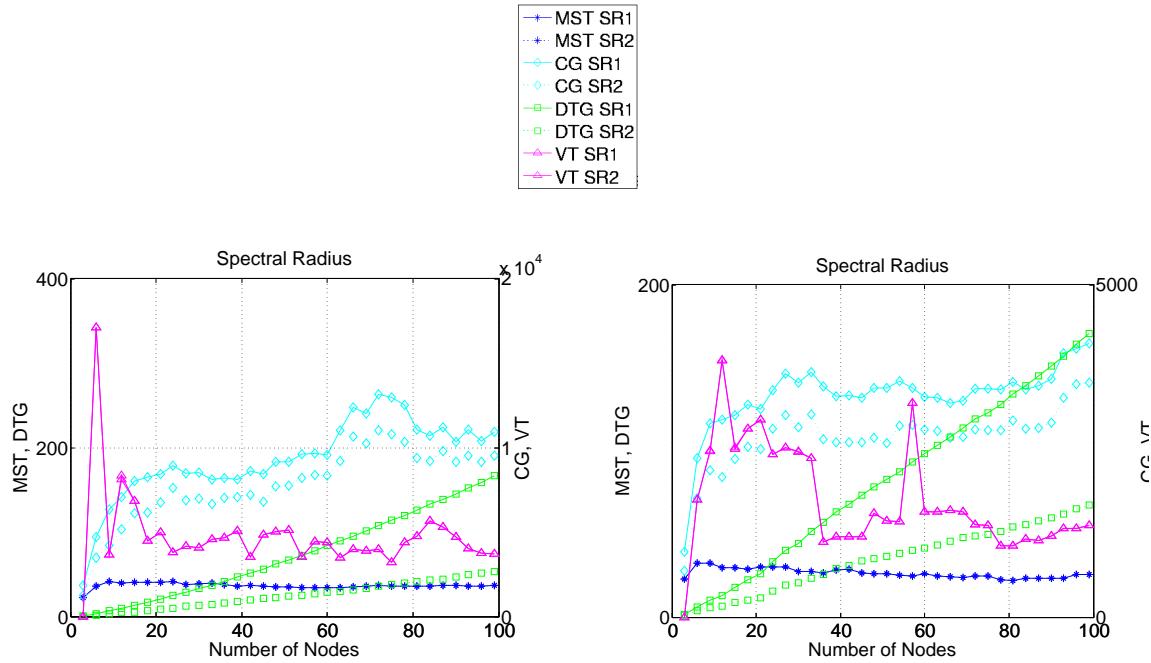
(xxviii) Roundness factor, random graph type 2.

**Figure I.1.** (Cont.) Random graph features. (continued on next page)



(xxix) Roundness factor homogeneity, random graph type 1.

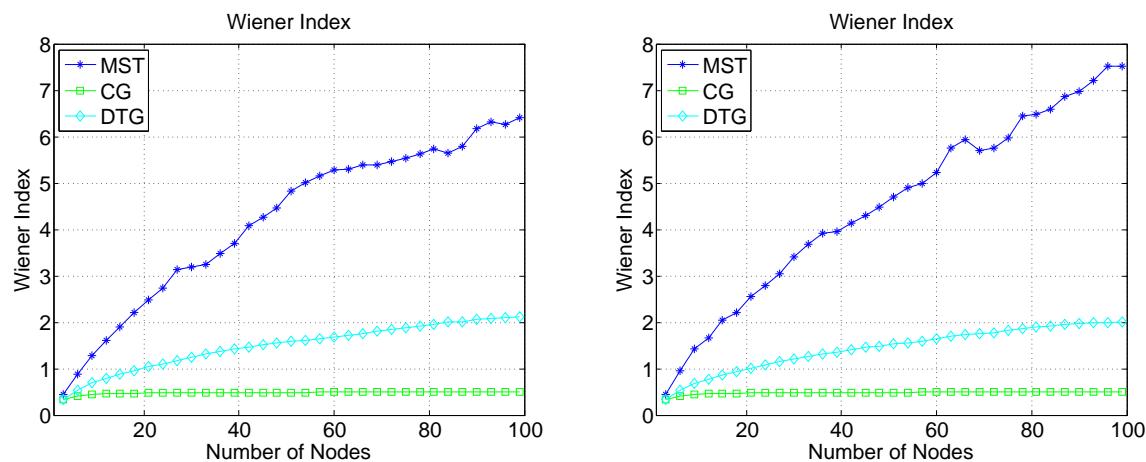
(xxx) Roundness factor homogeneity, random graph type 2.



(xxxi) Spectral radius, random graph type 1.

(xxxii) Spectral radius, random graph type 2.

**Figure I.1.** (Cont.) Random graph features. (continued on next page)



(xxxiii) Wiener index, random graph type 1.

(xxxiv) Wiener index, random graph type 2.

**Figure I.1.** (Cont.) Random graph features.

# Appendix J

## Grafting Feature Subsets: Nuclei

The grafting feature subsets are presented here for each of the four segmentations, GT, CC, WSHmin, and WSGran for nuclei. Statistics about the subsets are presented in Table J.1. Subsets obtained with a regularization parameter of  $\lambda_1 = 0.05$  are presented in Tables J.2-J.5, and for  $\lambda_1 = 0.01$  in Tables J.6- J.9. Features are presented in descending order according to the absolute value of the feature weight as returned from the grafting algorithm. The feature names are presented in the general form `name_band_stat`, where `name` is the name of the variable as defined in Section 5.3; `band` is the band from which the feature is extracted, if applicable, and `b1` corresponds to the red channel; and `stat` is the statistic (mean, median, etc.), if applicable. The co-occurrence matrix features have the form `name_band_distance`.

Surprisingly, as presented in Table J.1, all subsets display similar trends in the distribution of feature categories (texture, radiometric, size and shape), band choice (R, G, B, intensity or hue, binary), and a strong preference towards the optical density related features ( $\sim 80\%$  OD features in each subset). It was expected that the different segmentations would yield preference for a different types of features, due to the differences in segmentation output (e.g, WSHmin tends to have many smaller regions than CC or WSGran). It appears from these results, however, that the feature selection process is well-equipped to make use of the full variety of features and tailor the linear classifier to the given problem. This is an encouraging result in light of the less than perfect nuclear segmentation results in Chapter 4.

**Table J.1.** Statistics of object-level nuclei subsets. Statistics presented include the percentage of OD and non-OD features; R, G, B, I/Hue, and binary features; and texture, radiometric, and size/shape features.

Category	GT	CC	WSHmin	WSGran
<b>OD</b>	0.66	0.84	0.82	0.67
<b>Non-OD</b>	0.34	0.16	0.18	0.33
<b>R</b>	0.20	0.25	0.35	0.24
<b>G</b>	0.29	0.19	0.22	0.16
<b>B</b>	0.33	0.26	0.27	0.34
<b>I/Hue</b>	0.15	0.28	0.15	0.26
<b>Binary</b>	0.03	0.00	0.02	0.00
<b>Texture</b>	0.61	0.51	0.58	0.56
<b>Radiometric</b>	0.13	0.19	0.05	0.19
<b>Size/Shape</b>	0.27	0.30	0.37	0.26

**Table J.2.** Feature subset for GT nuclei,  $\lambda_1 = 0.05$ .

Rank	Weight	Feature Name
1	2.162633e-01	ODRLN_b1_mean
2	-1.558952e-01	HIOD_b1
3	1.273255e-01	RLN_b3_std
4	-1.253635e-01	MDist_b1
5	-1.183385e-01	ODCMClusterShade_b1_d1
6	1.114229e-01	LNum_b3
7	8.388252e-02	ODAreaFD_b2
8	5.950405e-02	ODV_b3_skewness
9	5.412770e-02	Hue_median
10	-4.475546e-02	ODCMClusterShade_b1_d2
11	1.972148e-02	HNum_b1
12	-9.974922e-03	ODLRE_b2_std
13	2.648506e-03	HNum_b3
14	1.949134e-03	MNum_b2

**Table J.3.** Feature subset for CC nuclei,  $\lambda_1 = 0.05$ .

Rank	Weight	Feature Name
1	2.377164e-01	ODAreaFD_b2
2	-4.034047e-02	ODCenterMass_b3
3	-3.549975e-02	ImageBands_b3_std
4	1.186530e-02	MDNAArea_b3

**Table J.4.** Feature subset for WSHmin nuclei,  $\lambda_1 = 0.05$ .

Rank	Weight	Feature Name
1	-1.484512e-01	ODCMClusterProminence_b1_d5
2	-1.470192e-01	ODLGRE_b3_std
3	-1.458370e-01	ODAreaFD_b3
4	-4.984972e-02	ODLGRE_b1_std
5	-3.318718e-02	ODCMClusterProminence_b1_d3
6	-1.624376e-02	HIOD_b1
7	-8.471932e-03	ODLRE_b3_std

**Table J.5.** Feature subset for WSGran nuclei,  $\lambda_1 = 0.05$ .

Rank	Weight	Feature Name
1	-2.198194e-01	ODAreaFD_b3
2	-9.472627e-02	LHMOD_b1
3	7.018033e-02	ODAreaFD_b2
4	5.307831e-02	LRE_b1_std
5	-4.900884e-02	ODRP_L_std
6	-4.731632e-02	SRE_b2_median
7	2.407060e-02	LRE_b2_std
8	1.695792e-02	MDNAArea_b3

**Table J.6.** Feature subset for GT nuclei,  $\lambda_1 = 0.01$ . (continued on next 2 pages)

Rank	Weight	Feature Name
1	-1.841640e+00	HIOD_b1
2	-1.390113e+00	MIOD_b1
3	7.324112e-01	IOD_b2
4	5.955182e-01	ODAreaFD_b2
5	5.771943e-01	IOD_b3
6	5.153647e-01	ODRLN_b1_median
7	4.235984e-01	GLN_b3_min
8	3.486472e-01	ODLGRE_b3_kurtosis
9	-3.022236e-01	ODLRE_b2_std
10	-2.982332e-01	MDist_b1
11	2.823205e-01	MCompactness_b1
12	2.519286e-01	HDNAArea_b3
13	2.379657e-01	ODLGRE_b1_skewness
14	2.047873e-01	HNum_b1
15	-1.940368e-01	SRE_b2_median
16	-1.714055e-01	HDist_b3
17	1.501765e-01	HDNAArea_I
18	-1.460353e-01	ImageBands_b3_skewness
19	-1.444637e-01	LIOD_b2
20	1.337584e-01	HNum_b3
21	1.332487e-01	ODRLN_b3_std
22	1.327974e-01	GLN_b1_skewness
23	-1.281736e-01	ODCMEntropy_b3_d1
24	1.251639e-01	SRE_b3_std
25	1.225147e-01	GCDF97W3_b2
26	1.216660e-01	LNum_b3
27	1.139369e-01	LRE_b2_skewness
28	1.072908e-01	MHCompactness_b2
29	1.010967e-01	ODGLN_b2_std
30	-1.003844e-01	GLCMHomogeneity_b1_d5
31	-9.286826e-02	ODCMCorrelation_b3_d1
32	-9.194899e-02	ODRLN_b2_skewness
33	-8.908218e-02	GLCMIInertia_b1_d1
34	8.878753e-02	MNum_b2
35	8.726396e-02	LRE_b1_skewness
36	8.421284e-02	MHCenterMass_I
37	8.413377e-02	HDNAArea_b2
38	8.269569e-02	GLCMCorrelation_b2_d5
39	8.018098e-02	GLCMClusterProminence_b2_d5

**Table J.6.** (Cont.) Feature subset for GT nuclei,  $\lambda_1 = 0.01$ . (continued on next page)

Rank	Weight	Feature Name
40	-7.963609e-02	GCDF97W1_b2
41	-7.665365e-02	Hue_max
42	7.419035e-02	ODCDF97W1_b3
43	7.364729e-02	ODReflSymmetry_b2
44	-7.213576e-02	Extent
45	7.148506e-02	ODRLN_b3_median
46	7.132117e-02	RP_b2_kurtosis
47	7.018066e-02	ODRLN_b1_std
48	-6.414370e-02	ODRP_b2_skewness
49	6.293047e-02	ODRLN_b2_median
50	-5.574011e-02	ODSRE_b3_std
51	-5.537077e-02	LDNAArea_b3
52	5.257629e-02	ODLGRE_I_kurtosis
53	4.710128e-02	MNum_I
54	4.440954e-02	GCDF97W4_b2
55	-3.899401e-02	ImageBands_I_kurtosis
56	-3.838133e-02	ODCenterMass_b3
57	3.824348e-02	ODRLN_b2_std
58	-3.700358e-02	ODCMMaxProb_I_d2
59	3.595686e-02	FFTk_9
60	3.199575e-02	ODCDF97W2_b3
61	3.182542e-02	SRE_I_kurtosis
62	3.136833e-02	ODCDF97W3_I
63	-3.061666e-02	LCenterMass_b2
64	3.033058e-02	PerimeterCurvature_min
65	2.970763e-02	RLN_b2_kurtosis
66	-2.938836e-02	GLCMHomogeneity_b1_d2
67	2.830165e-02	LCompactness_b1
68	2.787912e-02	ODRP_b3_kurtosis
69	-2.690397e-02	ODSRE_b2_std
70	2.574180e-02	MDNAArea_b3
71	2.405891e-02	ODCMEnergy_b3_d1
72	2.281913e-02	SRE_I_skewness
73	-2.069192e-02	RLN_b3_kurtosis
74	2.030969e-02	GLCMMaxProb_b3_d1
75	1.890607e-02	GLCMCorrelation_b1_d1
76	-1.786484e-02	ODLRE_b2_kurtosis
77	-1.562675e-02	ODSRE_b3_kurtosis
78	1.347224e-02	ODV_I_skewness

**Table J.6.** (Cont.) Feature subset for GT nuclei,  $\lambda_1 = 0.01$ .

<b>Rank</b>	<b>Weight</b>	<b>Feature Name</b>
79	-1.123048e-02	ODSRE_I_std
80	1.108909e-02	HNum_I
81	-1.066431e-02	LIOD_b1
82	8.525423e-03	LRE_b3_skewness
83	8.231778e-03	SRE_b2_skewness
84	-7.066230e-03	ODSRE_I_kurtosis
85	-5.728343e-03	ImageBands_b3_max
86	4.933196e-03	ODCDF97W1_b1
87	9.633620e-04	ODLRE_b3_skewness

**Table J.7.** Feature subset for CC nuclei,  $\lambda_1 = 0.01$ .

<b>Rank</b>	<b>Weight</b>	<b>Feature Name</b>
1	1.113797e-01	ODLGRE_b2_median
2	1.055120e-01	ODCMHomogeneity_I_d3
3	-1.007202e-01	ODLRE_b2_std
4	-8.702572e-02	ODV_b1_median
5	-8.565925e-02	HDist_b3
6	7.686655e-02	SRE_b3_std
7	-7.656663e-02	ODCMClusterProminence_b3_d3
8	-7.215335e-02	MIOD_b1
9	-6.841609e-02	ODSRE_b3_std
10	-6.467407e-02	ODCenterMass_b3
11	-6.447093e-02	SRE_b2_min
12	-6.412156e-02	Hue_max
13	-6.260592e-02	HDNAArea_b1
14	5.241304e-02	ODCMHomogeneity_b3_d3
15	-4.572463e-02	ODCMIInertia_b1_d4
16	4.447202e-02	LMMOD_b3
17	3.975185e-02	GLCMClusterShade_b2_d2
18	3.502014e-02	LRE_I_std
19	3.467747e-02	MIOD_I
20	3.218665e-02	MCompactness_b3
21	-3.095161e-02	LCompactness_I
22	3.009057e-02	MDNAArea_I
23	-2.537402e-02	ODSRE_I_std
24	2.418323e-02	LMHMOD_b1
25	-2.389915e-02	LIOD_b2
26	-2.321081e-02	ODSRE_b2_std
27	-1.966638e-02	ODCDF97W2_b1
28	-1.876088e-02	ODCMClusterProminence_b1_d3
29	1.875712e-02	MHCompactness_b3
30	1.510172e-02	RP_b1_std
31	-1.295929e-02	LRE_b3_min
32	1.106105e-02	ODCDF97W1_I
33	9.327655e-03	HNum_b1
34	8.169018e-03	ODCMHomogeneity_b2_d1
35	-5.146398e-03	ODCMClusterProminence_b1_d4
36	-5.116055e-03	ODCMClusterProminence_b3_d1
37	4.951072e-03	LRE_b2_std
38	-2.246809e-03	LMMOD_b2

**Table J.8.** Feature subset for WSHmin nuclei,  $\lambda_1 = 0.01$ . (continued on next page)

Rank	Weight	Feature Name
1	1.773061e+00	ODAreaFD_b2
2	-1.649383e+00	ODAreaFD_b1
3	-6.464460e-01	HIOD_b1
4	-4.833969e-01	ODAreaFD_b3
5	-4.268108e-01	MIOD_b1
6	-2.929890e-01	LDNAArea_I
7	-2.484377e-01	HDNAArea_b1
8	-2.221832e-01	MHDist_b1
9	1.989248e-01	LGRE_b2_min
10	-1.957615e-01	LDNAArea_b3
11	1.924735e-01	LGRE_b1_min
12	1.893189e-01	HDNAArea_b3
13	1.615248e-01	ODLGRE_b2_min
14	1.424650e-01	IOD_b2
15	1.298895e-01	LRE_b1_std
16	-1.231708e-01	ODLRE_b2_std
17	-1.182566e-01	ODLGRE_b3_std
18	1.150640e-01	ODRLN_b1_median
19	-1.143140e-01	ODCMClusterProminence_b1_d3
20	1.125800e-01	HNum_b1
21	-1.116663e-01	ODCMClusterProminence_b1_d5
22	-1.107394e-01	ODCDF97W1_b1
23	1.098446e-01	HNum_b3
24	1.084308e-01	GLN_b3_min
25	-1.054795e-01	MDist_b2
26	-1.045476e-01	LDNAArea_b2
27	9.655921e-02	GCDF97W4_b2
28	-8.095066e-02	ODCMClusterProminence_b1_d4
29	7.697195e-02	ODRP_I_median
30	7.605640e-02	ODRLN_I_std
31	-7.342969e-02	ODCDF97W5_b1
32	7.032747e-02	ODReflSymmetry_b3
33	-6.628056e-02	GLCMHomogeneity_b1_d3
34	-6.322837e-02	ODLGRE_b1_kurtosis
35	6.008535e-02	MCompactness_b3
36	5.749801e-02	ODGLN_b2_std
37	5.749231e-02	FFTk_9
38	5.376181e-02	MHCompactness_b1
39	5.356879e-02	ODRLN_b2_median

**Table J.8.** (Cont.) Feature subset for WSHmin nuclei,  $\lambda_1 = 0.01$ .

<b>Rank</b>	<b>Weight</b>	<b>Feature Name</b>
40	4.968789e-02	LNum_b3
41	4.427622e-02	ODCMHomogeneity_b3_d2
42	-4.054931e-02	MHDist_b2
43	3.798543e-02	ODRP_b1_median
44	3.662451e-02	MNum_b3
45	3.589571e-02	RLN_b1_std
46	-2.917107e-02	MDist_I
47	2.571653e-02	ODHGRE_b2_min
48	-2.570221e-02	LGRE_b3_std
49	2.565934e-02	MHCompactness_I
50	2.050302e-02	HCompactness_b3
51	-1.877079e-02	ODCDF97W2_b1
52	1.658247e-02	MDNAArea_b1
53	1.633439e-02	RP_I_std
54	-1.471283e-02	ODLRE_I_std
55	1.280882e-02	HNum_I
56	-1.167106e-02	ODLRE_b3_std
57	-6.777826e-03	GLCMClusterProminence_b3_d3
58	1.665726e-03	ODLRE_b2_min
59	-1.120004e-03	ODHGRE_b3_std
60	9.446366e-04	MNum_I

**Table J.9.** Feature subset for WSGran nuclei,  $\lambda_1 = 0.01$ . (continued on next page)

Rank	Weight	Feature Name
1	1.504659e+00	ODAreaFD_b2
2	-1.140390e+00	ODAreaFD_b1
3	-6.802808e-01	ODAreaFD_b3
4	2.764476e-01	LGRE_b1_skewness
5	-2.740522e-01	Hue_max
6	-2.737837e-01	Hue_min
7	2.683098e-01	ODLGRE_b2_median
8	-2.681761e-01	ImageBands_b3_max
9	-2.548641e-01	HDNAArea_b1
10	2.496469e-01	ODRP_I_min
11	-2.400566e-01	ODLRE_b2_std
12	-2.267429e-01	LDNAArea_b3
13	-2.164828e-01	LHMOD_b1
14	2.121254e-01	LGRE_b3_min
15	2.056475e-01	MNum_I
16	-1.779260e-01	GLCMClusterProminence_b3_d1
17	1.757503e-01	LRE_b2_std
18	1.746772e-01	MDNAArea_b2
19	-1.734724e-01	LIOD_b1
20	-1.574551e-01	HIOD_b1
21	-1.569690e-01	ODCDF97W5_b1
22	-1.336989e-01	LGRE_I_std
23	1.298546e-01	LIOD_I
24	1.269518e-01	ODEntropy_b3
25	-1.232697e-01	HDNAArea_b2
26	-1.231057e-01	ODCMClusterProminence_b1_d3
27	1.198082e-01	ODCDF97W2_b3
28	1.104252e-01	ODLGRE_I_median
29	1.102292e-01	LNum_b1
30	-1.089608e-01	LGRE_I_kurtosis
31	1.065812e-01	LRE_b2_median
32	9.947861e-02	HDNAArea_b3
33	-9.909156e-02	HDist_b3
34	9.882499e-02	ODLGRE_b2_min
35	9.491284e-02	MDNAArea_b3
36	-9.418599e-02	SRE_I_max
37	9.370859e-02	LCompactness_b3
38	-9.054158e-02	MIOD_b1
39	8.230639e-02	LRE_I_std

**Table J.9.** (Cont.) Feature subset for WSGran nuclei,  $\lambda_1 = 0.01$ .

<b>Rank</b>	<b>Weight</b>	<b>Feature Name</b>
40	-7.300993e-02	ODCDF97W3_b1
41	-6.075880e-02	SRE_I_mean
42	5.983667e-02	LRE_b1_std
43	5.862455e-02	LMMOD_b1
44	-5.717765e-02	ODCenterMass_b3
45	5.508303e-02	LRE_I_max
46	5.476791e-02	ODCMHomogeneity_b3_d4
47	-4.886133e-02	GLCMIertia_b3_d5
48	-4.737774e-02	LDist_b2
49	-4.288871e-02	ODCMClusterProminence_b1_d5
50	4.071061e-02	LNum_b3
51	-3.734846e-02	Hue_mean
52	-3.604142e-02	ODSRE_I_std
53	3.488105e-02	ODCDF97W1_b3
54	-2.694101e-02	ODCDF97W4_b1
55	2.412775e-02	HNum_b3
56	-2.388443e-02	ODLRE_I_std
57	-2.152077e-02	LMHMOD_I
58	2.121263e-02	HCenterMass_I
59	1.902337e-02	ODV_b1_min
60	-1.676677e-02	MDist_b1
61	9.355875e-03	ODReflSymmetry_b3
62	9.193190e-03	RP_b3_std
63	8.443953e-03	LMMOD_b3
64	-7.247392e-03	LGRE_b3_std
65	-6.605253e-03	SRE_I_min
66	-6.380032e-03	HGRE_b3_std
67	-4.729233e-03	MCenterMass_b2
68	-1.823611e-03	ODHGRE_b3_std
69	-1.585781e-03	GLCMEnergy_b2_d2
70	1.143871e-03	GLCMHomogeneity_b3_d3



# Appendix K

## Grafting Feature Subsets: Object- and Spatial-Level

Subsets for CC segmentations for nuclei (N) are presented in Table K.1, for combined cytoplasm and stroma (CS) in Table K.2, for cytoplasm (C) in Table K.3, for nuclei and cytoplasm/stroma (N+CS) in Table K.5, and for nuclei and cytoplasm (N+C) in Table K.6. Features are presented in descending order according to the absolute value of the feature weight as returned from the grafting algorithm. The feature names are presented in the general form `name_band_stat`, where `name` is the name of the variable as defined in Section 5.3; `band` is the band from which the feature is extracted, if applicable, and `b1` corresponds to the red channel; and `stat` is the statistic (mean, median, etc.), if applicable. The co-occurrence matrix features have the form `name_band_distance`. Additionally, for the combined feature sets, the feature names are prepended by the class abbreviation for which the feature is extracted (N, CS, C, or S).

Subsets for WSGran segmentations for spatial nuclei (SPN) are presented in Table K.7; for combined nuclei and spatial nuclei (N+SPN) in Table K.8; for nuclei, cytoplasm/stroma, and spatial nuclei (N+CS+SPN) in Table K.9; and for nuclei, cytoplasm, and spatial nuclei (N+C+SPN) in Table K.10. Features are presented in descending order according to the absolute value of the feature weight as returned from the grafting algorithm. The object feature names are presented in the same form as Tables K.1-K.6. When disambiguation is required, the feature names are prepended by the class abbreviation for which the feature is extracted (N, CS, C, or SPN). The spatial features have the general form `graph_name_stat` or `graph_name_i` where `i` is the index into the feature vector.

**Table K.1.** Feature subset for CC nuclei,  $\lambda_1 = 0.01$ .

Rank	Weight	Feature Name
1	1.113797e-01	ODLGRE_b2_median
2	1.055120e-01	ODCMHomogeneity_I_d3
3	-1.007202e-01	ODLRE_b2_std
4	-8.702572e-02	ODV_b1_median
5	-8.565925e-02	HDist_b3
6	7.686655e-02	SRE_b3_std
7	-7.656663e-02	ODCMClusterProminence_b3_d3
8	-7.215335e-02	MIOD_b1
9	-6.841609e-02	ODSRE_b3_std
10	-6.467407e-02	ODCenterMass_b3
11	-6.447093e-02	SRE_b2_min
12	-6.412156e-02	Hue_max
13	-6.260592e-02	HDNAArea_b1
14	5.241304e-02	ODCMHomogeneity_b3_d3
15	-4.572463e-02	ODCMIertia_b1_d4
16	4.447202e-02	LMMOD_b3
17	3.975185e-02	GLCMClusterShade_b2_d2
18	3.502014e-02	LRE_I_std
19	3.467747e-02	MIOD_I
20	3.218665e-02	MCompactness_b3
21	-3.095161e-02	LCompactness_I
22	3.009057e-02	MDNAArea_I
23	-2.537402e-02	ODSRE_I_std
24	2.418323e-02	LMHMOD_b1
25	-2.389915e-02	LIOD_b2
26	-2.321081e-02	ODSRE_b2_std
27	-1.966638e-02	ODCDF97W2_b1
28	-1.876088e-02	ODCMClusterProminence_b1_d3
29	1.875712e-02	MHCompactness_b3
30	1.510172e-02	RP_b1_std
31	-1.295929e-02	LRE_b3_min
32	1.106105e-02	ODCDF97W1_I
33	9.327655e-03	HNum_b1
34	8.169018e-03	ODCMHomogeneity_b2_d1
35	-5.146398e-03	ODCMClusterProminence_b1_d4
36	-5.116055e-03	ODCMClusterProminence_b3_d1
37	4.951072e-03	LRE_b2_std
38	-2.246809e-03	LMMOD_b2

**Table K.2.** Feature subset for CC cytoplasm/stroma,  $\lambda_1 = 0.01$ . (continued on next 2 pages)

Rank	Weight	Feature Name
1	1.372266e+00	ODAreaFD_b2
2	8.530343e-01	ODEntropy_b3
3	-8.369316e-01	GEntropy_b3
4	-8.326799e-01	ODAreaFD_b1
5	-5.588271e-01	LIOD_b1
6	-5.237117e-01	HIOD_b1
7	-3.692752e-01	MIOD_b1
8	2.876006e-01	ODRLN_b1_median
9	2.728857e-01	IOD_b1
10	-2.466509e-01	MIOD_b3
11	-2.254520e-01	ImageBands_b3_max
12	-2.049502e-01	ODLGRE_b2_kurtosis
13	-2.044104e-01	SRE_b2_min
14	-2.002985e-01	ODLGRE_b3_skewness
15	1.905710e-01	ODV_b2_std
16	1.894088e-01	ODRLN_I_min
17	-1.799665e-01	ODCMClusterProminence_b2_d3
18	1.580825e-01	MNum_I
19	-1.450038e-01	Hue_max
20	-1.375057e-01	ODSRE_I_std
21	1.352631e-01	ODV_b2_kurtosis
22	-1.350663e-01	Hue_min
23	-1.337829e-01	RP_b2_min
24	-1.325952e-01	SRE_b3_mean
25	-1.293787e-01	ODLGRE_b1_skewness
26	-1.257452e-01	LGRE_b3_kurtosis
27	1.153152e-01	LNum_b1
28	-1.134925e-01	LDist_b1
29	-1.133535e-01	ODLRE_b2_std
30	1.132006e-01	HNum_I
31	-1.009022e-01	MDist_b3
32	-1.005183e-01	GLCMClusterProminence_b1_d1
33	9.563276e-02	ODCMHomogeneity_b2_d5
34	-8.897396e-02	ODV_b3_kurtosis
35	-8.886528e-02	ODGLN_b3_kurtosis
36	8.247380e-02	MHCompactness_b2
37	-8.123379e-02	ODCDF97W2_b1
38	-7.630384e-02	LGRE_I_kurtosis

**Table K.2.** (Cont.) Feature subset for CC cytoplasm/stroma,  $\lambda_1 = 0.01$ . (continued on next page)

Rank	Weight	Feature Name
39	-7.039326e-02	LDist_I
40	-7.033877e-02	ImageBands_b1_skewness
41	6.685520e-02	MNum_b1
42	6.673634e-02	ODV_b1_max
43	-6.566642e-02	Hue_mean
44	6.507854e-02	LNum_b3
45	6.387991e-02	ODLGRE_b2_min
46	-6.235837e-02	SRE_I_min
47	6.092144e-02	ODCenterMass_b1
48	-5.930142e-02	LCompactness_I
49	5.851990e-02	ODReflSymmetry_b1
50	-5.144015e-02	SRE_I_mean
51	-4.934010e-02	ODCDF97W3_b1
52	4.890886e-02	GLCMClusterShade_b2_d5
53	-4.650724e-02	RP_I_kurtosis
54	-4.578416e-02	GLCMMMaxProb_b3_d1
55	4.003334e-02	GLN_b3_skewness
56	3.658819e-02	ODRP_b3_kurtosis
57	3.566955e-02	SRE_I_skewness
58	3.358232e-02	ODLRE_b2_skewness
59	-3.208658e-02	ODGLN_b3_skewness
60	3.048181e-02	GLN_b1_kurtosis
61	-2.936986e-02	ODRLN_b2_skewness
62	2.895763e-02	ODRP_I_kurtosis
63	2.686536e-02	ODCMCorrelation_b3_d5
64	-2.685315e-02	SRE_b3_max
65	2.576725e-02	LIOD_I
66	-2.309164e-02	ODRP_b3_std
67	2.136719e-02	SRE_b2_kurtosis
68	2.041652e-02	MDNAArea_b2
69	-1.922366e-02	GLN_I_kurtosis
70	-1.911864e-02	GCDF97W5_b3
71	1.875467e-02	ODV_b1_min
72	-1.861724e-02	LCompactness_b1
73	-1.709158e-02	ODSRE_b3_kurtosis
74	-1.588635e-02	SRE_b2_mean
75	1.502142e-02	LRE_b3_std
76	1.327288e-02	Compactness

**Table K.2.** (Cont.) Feature subset for CC cytoplasm/stroma,  $\lambda_1 = 0.01$ .

Rank	Weight	Feature Name
77	1.240371e-02	GLN_b2_skewness
78	1.203883e-02	MHCompactness_b3
79	1.135211e-02	EulerNumber
80	-1.064627e-02	LMMOD_I
81	-1.003690e-02	ODGLN_b2_skewness
82	-8.745233e-03	ODAreaFD_b3
83	-7.759486e-03	ODLRE_b1_skewness
84	-6.698439e-03	LGRE_b2_std
85	6.511758e-03	HCenterMass_b1
86	-6.379481e-03	RLN_b1_kurtosis
87	-6.293002e-03	ODSRE_b3_std
88	-5.855633e-03	ODLRE_I_std
89	5.309267e-03	LNum_I
90	3.058990e-03	MHCompactness_I
91	-2.362016e-03	HGRE_b3_kurtosis
92	2.123988e-03	LNum_b2

**Table K.3.** Feature subset for CC cytoplasm,  $\lambda_1 = 0.01$ . (continued on next page)

Rank	Weight	Feature Name
1	1.407838e+00	ODAreaFD_b2
2	-9.286198e-01	LIOD_b1
3	-8.918492e-01	ODAreaFD_b1
4	-6.438109e-01	HIOD_b1
5	5.136531e-01	ODLGRE_b2_min
6	-4.649106e-01	MIOD_b1
7	3.174187e-01	RLN_b1_std
8	2.786994e-01	ODRP_I_min
9	2.710272e-01	GLN_b1_min
10	2.692515e-01	GLN_b3_min
11	-2.445367e-01	SRE_b3_median
12	2.431562e-01	MNum_I
13	-2.010231e-01	SRE_b2_max
14	-1.931867e-01	SRE_I_max
15	-1.592960e-01	SRE_I_median
16	-1.555246e-01	ODCMEntropy_b3_d1
17	1.486767e-01	ODReflSymmetry_b3
18	1.485899e-01	IOD_b1
19	-1.343150e-01	ODLRE_b2_std
20	-1.320431e-01	ODLGRE_b1_std
21	1.196498e-01	ODCMIInertia_b1_d2
22	-1.172548e-01	ODAreaFD_b3
23	1.122880e-01	ODCMCorrelation_b3_d5
24	9.591729e-02	MCompactness_b1
25	-9.032558e-02	ODLRE_I_std
26	8.613028e-02	LRE_b1_std
27	-8.314118e-02	Hue_min
28	8.269799e-02	GLN_b1_max
29	8.183227e-02	LRE_b1_median
30	-7.764044e-02	ODLGRE_b3_std
31	-7.735280e-02	ODCDF97W2_b1
32	7.581957e-02	LNum_I
33	7.438173e-02	ODV_b3_min
34	7.418293e-02	ODCenterMass_b2
35	6.114643e-02	RLN_b1_max
36	-5.677301e-02	GLCMHomogeneity_b1_d1
37	5.640536e-02	ODGLN_b2_std
38	-4.551607e-02	ODSRE_I_std
39	4.375312e-02	AspectRatio

**Table K.3.** (Cont.) Feature subset for CC cytoplasm,  $\lambda_1 = 0.01$ .

Rank	Weight	Feature Name
40	-4.106057e-02	LHMOD_b1
41	-4.103328e-02	MHCenterMass_b3
42	3.697805e-02	LNum_b1
43	-3.696764e-02	GLCMHomogeneity_b1_d2
44	-3.535553e-02	LGRE_b3_kurtosis
45	-2.969576e-02	ODCDF97W3_b1
46	2.523243e-02	FFTk_18
47	-2.417334e-02	MHDist_b3
48	-2.299077e-02	GLCMMMaxProb_b3_d3
49	-2.007447e-02	HGRE_b3_kurtosis
50	-1.813756e-02	ODCMClusterProminence_b1_d3
51	1.810955e-02	GLN_b2_max
52	-1.739925e-02	ODHGRE_b3_std
53	-1.669021e-02	GLCMHomogeneity_I_d2
54	-1.564372e-02	SRE_b2_min
55	1.476383e-02	EulerNumber
56	-1.242704e-02	GLCMClusterProminence_b1_d5
57	-1.102079e-02	ODCMClusterProminence_b1_d4
58	9.459049e-03	ODCMEnergy_b3_d5
59	8.426404e-03	MDNAArea_b2
60	7.671378e-03	ODCenterMass_I
61	-7.650493e-03	GEntropy_b3
62	7.537261e-03	ODRLN_I_min
63	4.807110e-03	MCenterMass_I
64	4.768580e-03	MHCenterMass_b2
65	-4.693899e-03	LGRE_I_kurtosis
66	4.331801e-03	MHCompactness_I
67	-4.122372e-03	RP_I_std
68	-3.223252e-03	ODCMMMaxProb_b1_d1
69	-3.062314e-03	PerimeterCurvature_max
70	1.543347e-03	MNum_b3
71	-1.468795e-03	LHMOD_b2
72	-1.239501e-03	HDist_b2
73	9.456942e-04	MHCenterMass_I

**Table K.4.** Feature subset for CC stroma,  $\lambda_1 = 0.01$ . (continued on next page)

Rank	Weight	Feature Name
1	2.537943e+00	ODAreaFD_b2
2	-1.976310e+00	ODAreaFD_b1
3	-7.103831e-01	ODAreaFD_b3
4	5.242084e-01	ODLGRE_b2_median
5	-2.903803e-01	LDist_b2
6	-2.558657e-01	Hue_max
7	-2.408627e-01	HDNAArea_b1
8	2.390086e-01	LGRE_b2_min
9	-2.160964e-01	Hue_mean
10	-2.140332e-01	GLCMCorrelation_b1_d5
11	-2.068375e-01	GLCMClusterProminence_b2_d5
12	1.798453e-01	ODGLN_b2_std
13	1.710979e-01	MDNAArea_b3
14	1.679345e-01	LRE_b1_std
15	-1.572163e-01	MDist_b1
16	1.446849e-01	RP_b2_std
17	1.349345e-01	ODLGRE_b2_min
18	-1.154299e-01	ODCMHomogeneity_I_d5
19	-1.094936e-01	ODLRE_b2_std
20	1.078371e-01	LGRE_b2_median
21	-1.069329e-01	ODLGRE_b1_std
22	-1.036038e-01	ODCDF97W3_b2
23	-1.004553e-01	ODCDF97W3_b1
24	-1.003676e-01	ODRP_I_std
25	9.572589e-02	ODCMHomogeneity_b1_d2
26	-9.098253e-02	GLCMHomogeneity_b2_d1
27	8.528006e-02	LCompactness_b2
28	-8.152533e-02	GLCMHomogeneity_b2_d4
29	8.095450e-02	GLCMHomogeneity_b3_d4
30	7.657782e-02	GCDF97W5_b2
31	-7.165878e-02	HIOD_b2
32	6.673824e-02	LIOD_I
33	-6.429912e-02	MHCenterMass_b2
34	6.000862e-02	RP_b3_std
35	-5.694571e-02	ODCMHomogeneity_I_d4
36	-5.683829e-02	LDist_I
37	5.408498e-02	GCenterMass_I
38	5.259442e-02	SRE_b1_std
39	4.930673e-02	LMHMOD_I

**Table K.4.** (Cont.) Feature subset for CC stroma,  $\lambda_1 = 0.01$ .

Rank	Weight	Feature Name
40	4.913513e-02	MIOD_I
41	-4.715422e-02	MDist_b2
42	-4.670712e-02	ODCMClusterProminence_b1_d1
43	-4.529034e-02	ODCMClusterProminence_b1_d2
44	-4.463394e-02	ODCMClusterProminence_b2_d1
45	4.267370e-02	ODRP_L_median
46	-4.200086e-02	ODCMHomogeneity_b2_d2
47	-3.892544e-02	SRE_I_max
48	-3.836966e-02	GLCMIertia_b2_d3
49	3.690871e-02	GCDF97W4_b2
50	3.383571e-02	ODRP_b1_median
51	3.211979e-02	GLCMHomogeneity_b3_d5
52	-3.066840e-02	EllipticalDeviation
53	-2.776812e-02	HCompactness_b2
54	-2.530716e-02	LDNAArea_b3
55	-2.508065e-02	HDNAArea_b2
56	2.113497e-02	GLCMIertia_b3_d5
57	-2.090449e-02	ODLGRE_L_std
58	-1.896869e-02	GLCMClusterProminence_L_d5
59	-1.728119e-02	LHMOD_b3
60	1.485969e-02	ODCMIertia_b3_d2
61	-1.401400e-02	ODSRE_L_std
62	1.113028e-02	ODSRE_b1_min
63	-9.880508e-03	ODCDF97W4_b1
64	-7.668128e-03	SRE_L_std
65	4.739166e-03	LHMOD_I
66	3.915037e-03	GCenterMass_b2
67	5.261036e-04	ODRP_b1_min

**Table K.5.** Feature subset for CC nuclei and cytoplasm/stroma,  $\lambda_1 = 0.01$ . (Continued on next 3 pages)

Rank	Weight	Feature Name
1	1.053949e+00	N_ODAreaFD_b2
2	-6.576984e-01	CS_LIOD_b1
3	-5.739820e-01	CS_HIOD_b1
4	-4.616726e-01	N_ODAreaFD_b1
5	-4.049259e-01	CS_MIOD_b1
6	3.590346e-01	CS_ODRLN_b1_median
7	-3.078129e-01	CS_RP_b2_min
8	2.592028e-01	CS_IOD_b1
9	-2.571441e-01	CS_SRE_I_mean
10	-2.564260e-01	CS_ODLGRE_b3_skewness
11	-2.545059e-01	CS_Hue_mean
12	-2.524557e-01	CS_MIOD_b3
13	-2.447282e-01	N_MDist_b1
14	2.381545e-01	CS_MNum_I
15	2.031970e-01	CS_ODRLN_I_min
16	1.920703e-01	N_HDNAArea_I
17	-1.822790e-01	CS_ImageBands_b3_max
18	-1.763452e-01	CS_ODLGRE_b2_kurtosis
19	1.638891e-01	N_MHCompactness_I
20	-1.471625e-01	CS_SRE_b3_mean
21	-1.384717e-01	CS_Hue_min
22	-1.348936e-01	CS_ODSRE_I_std
23	-1.345640e-01	CS_ODLRE_b2_std
24	1.334027e-01	CS_ODV_b2_std
25	1.249558e-01	CS_ODV_b2_kurtosis
26	-1.229008e-01	CS_GLCMMaxProb_b3_d1
27	-1.196820e-01	CS_LGRE_b3_kurtosis
28	-1.088703e-01	CS_SRE_b2_min
29	1.082188e-01	CS_MNum_b1
30	-1.057158e-01	N_ImageBands_b3_max
31	-1.050976e-01	CS_LGRE_I_kurtosis
32	-1.024885e-01	CS_GEntropy_b3
33	1.013400e-01	N_MDNAArea_b3
34	-9.978822e-02	N_HDist_b3
35	9.949468e-02	CS_HNum_I
36	9.492799e-02	CS_MHCompactness_b2
37	-9.367480e-02	N_ODCenterMass_b3
38	9.342009e-02	CS_ODLGRE_b2_min

**Table K.5.** (Cont.) Feature subset for CC nuclei and cytoplasm/stroma,  $\lambda_1 = 0.01$ .  
 (continued on next 2 pages)

Rank	Weight	Feature Name
39	-9.098498e-02	CS_ODV_b3_kurtosis
40	-9.094574e-02	CS_ImageBands_b1_skewness
41	-8.773484e-02	CS_ODGLN_b3_kurtosis
42	8.303784e-02	CS_ODEntropy_b3
43	8.151405e-02	CS_ODRP_I_min
44	7.899184e-02	CS_LNum_b3
45	-7.761245e-02	CS_ODCDF97W2_b1
46	7.603328e-02	N_MDNAArea_b2
47	-7.533644e-02	CS_ODCMClusterProminence_b2_d3
48	7.512152e-02	CS_LNum_b1
49	-7.460612e-02	CS_ODCDF97W3_b1
50	7.292792e-02	N_ODCMHomogeneity_I_d3
51	7.178129e-02	CS_SRE_I_skewness
52	6.911142e-02	CS_ODReflSymmetry_b1
53	6.667859e-02	N_LRE_I_std
54	-6.455022e-02	N_ODCDF97W5_I
55	6.385717e-02	CS_ODCMHomogeneity_b2_d5
56	-5.935099e-02	CS_ODCMClusterProminence_I_d2
57	5.859325e-02	CS_ODLRE_b2_skewness
58	5.384344e-02	CS_ODSRE_b1_median
59	-5.330937e-02	N_ODCMClusterProminence_b3_d3
60	-5.251014e-02	N_ODCDF97W2_b1
61	5.198734e-02	CS_ODCMHomogeneity_b3_d5
62	-5.153757e-02	N_ODLGRE_b2_std
63	-5.124655e-02	CS_Hue_max
64	-4.715189e-02	CS_RP_I_kurtosis
65	4.486134e-02	N_ODLRE_I_min
66	-4.477632e-02	CS_ODRP_b1_std
67	-4.422391e-02	CS_ODLRE_I_std
68	4.336292e-02	CS_ODRP_I_kurtosis
69	-4.332445e-02	CS_LCompactness_I
70	-4.298579e-02	N_ODLRE_b2_std
71	-4.260715e-02	CS_LDist_I
72	4.233224e-02	N_LIOD_I
73	4.088949e-02	CS_GLCMEnergy_b3_d5
74	3.949617e-02	CS_MIOD_I
75	-3.818348e-02	N_ODSRE_b3_std
76	3.525204e-02	CS_LNum_b2

**Table K.5.** (Cont.) Feature subset for CC nuclei and cytoplasm/stroma,  $\lambda_1 = 0.01$ . (continued on next page)

Rank	Weight	Feature Name
77	3.440645e-02	CS__GLN_b2_kurtosis
78	3.220187e-02	CS__ODV_b2_max
79	3.095541e-02	N_LDNAArea_b1
80	-3.036425e-02	CS__SRE_I_min
81	2.835895e-02	N_RP_I_std
82	2.832560e-02	CS__ODRP_b3_kurtosis
83	-2.736809e-02	N_LDist_I
84	2.671440e-02	N_MCompactness_b3
85	-2.641126e-02	CS__GLN_I_kurtosis
86	2.584416e-02	CS__LNum_I
87	-2.564011e-02	CS__RLN_b1_kurtosis
88	-2.520161e-02	CS__ODSRE_b3_std
89	-2.491105e-02	CS__LDist_b1
90	-2.355646e-02	CS__ODRLN_b2_skewness
91	-2.247180e-02	CS__LCompactness_b1
92	2.185892e-02	CS__GLN_b1_min
93	2.119723e-02	CS__GLN_b1_kurtosis
94	2.040979e-02	N_MNum_I
95	1.975475e-02	N_GCenterMass_b2
96	1.969992e-02	CS__Compactness
97	-1.918475e-02	N_LCompactness_I
98	1.730535e-02	CS__MDNAArea_I
99	1.688664e-02	N_SRE_b2_std
100	-1.688464e-02	N_LCenterMass_b1
101	1.661050e-02	N_ODLGRE_b1_min
102	-1.349395e-02	CS__ODRP_b3_skewness
103	-1.340011e-02	CS__ODLRE_b1_skewness
104	-1.282328e-02	CS__LIOD_b2
105	-1.212517e-02	CS__LGRE_b2_std
106	-1.155534e-02	CS__ODSRE_b1_skewness
107	-1.084457e-02	N_PerimeterCurvature_mean
108	-1.079300e-02	CS__HGRE_I_kurtosis
109	-9.920936e-03	CS__ODHGRE_b3_skewness
110	-9.823512e-03	CS__MDist_b3
111	-9.551226e-03	CS__SRE_b3_min
112	-9.339508e-03	CS__SRE_b2_mean
113	7.451152e-03	CS__GLN_b2_std
114	-6.961526e-03	CS__ODRP_b3_std

**Table K.5.** (Cont.) Feature subset for CC nuclei and cytoplasm/stroma,  $\lambda_1 = 0.01$ .

Rank	Weight	Feature Name
115	6.414969e-03	N_MHCompactness_b1
116	-5.705879e-03	CS_GLCMClusterProminence_b1_d1
117	5.438041e-03	CS_ODGLN_b2_std
118	5.123104e-03	CS_MCompactness_b1
119	-4.073111e-03	N_GLCMInertia_b1_d4
120	3.743285e-03	N_ODCDF97W1_b3
121	3.224854e-03	CS_LRE_b3_std
122	-2.638115e-03	CS_MHDist_b3
123	-1.681523e-03	CS_HCenterMass_b2
124	1.045633e-03	CS_ODAreaFD_b2
125	-9.600046e-04	CS_HGRE_b3_kurtosis

**Table K.6.** Feature subset for CC nuclei and cytoplasm,  $\lambda_1 = 0.01$ . (continued on next 2 pages)

Rank	Weight	Feature Name
1	1.251702e+00	N_ODAreaFD_b2
2	-9.858650e-01	C_LIOD_b1
3	-6.820145e-01	C_HIOD_b1
4	-6.565903e-01	N_ODAreaFD_b1
5	-5.463131e-01	C_MIOD_b1
6	3.794900e-01	C_GLN_b1_min
7	3.792767e-01	C_ODLGRE_b2_min
8	2.994245e-01	C_GLN_b3_min
9	2.264070e-01	C_RLN_b1_std
10	2.201208e-01	C_ODReflSymmetry_b3
11	2.117552e-01	C_ODRP_I_min
12	2.095972e-01	C_MNum_I
13	-2.060013e-01	C_SRE_b2_max
14	-2.017680e-01	C_SRE_b3_median
15	-1.952699e-01	C_SRE_I_max
16	1.918907e-01	N_MHCompactness_I
17	1.676926e-01	C_RLN_b1_max
18	-1.648723e-01	N_ODCenterMass_b3
19	1.598578e-01	N_MDNAArea_b2
20	1.560903e-01	C_IOD_b1
21	-1.455506e-01	N_ImageBands_b3_max
22	-1.440702e-01	C_Hue_min
23	-1.370185e-01	N_ODAreaFD_b3
24	-1.360255e-01	C_ODCMEntropy_b3_d1
25	-1.347419e-01	C_ODLGRE_b3_std
26	1.337330e-01	N_GLCMClusterShade_b1_d2
27	1.290199e-01	N_LIOD_I
28	-1.281096e-01	N_HDist_b3
29	-1.197638e-01	C_SRE_I_median
30	1.110771e-01	N_HDNAArea_I
31	1.087624e-01	N_GCenterMass_b2
32	1.054937e-01	N_LRE_I_std
33	-1.048900e-01	C_ODLRE_b2_std
34	-1.044880e-01	N_Hue_max
35	-1.022770e-01	C_GLCMHomogeneity_b1_d1
36	-1.017113e-01	N_MDist_b1
37	-9.956403e-02	N_ODLRE_b2_std
38	9.921076e-02	C_MCompactness_b1

**Table K.6.** (Cont.) Feature subset for CC nuclei and cytoplasm,  $\lambda_1 = 0.01$ . (continued on next page)

Rank	Weight	Feature Name
39	9.541216e-02	C__ODCMInertia_b1_d2
40	-9.282623e-02	N__ODCMClusterProminence_b3_d3
41	8.895169e-02	N__MDNAArea_b3
42	8.519903e-02	C__ConvexArea
43	7.971208e-02	C__ODCMCorrelation_b3_d5
44	7.490504e-02	C__ODGLN_b2_std
45	-7.394597e-02	C__LHMOD_b1
46	-7.083600e-02	C__ODSRE_I_std
47	-6.901476e-02	N__SRE_b2_min
48	6.815954e-02	N__ODCDF97W1_b3
49	6.570788e-02	C__LRE_b1_std
50	6.370750e-02	N__ODLGRE_b1_min
51	-6.318366e-02	N__ODRP_b3_std
52	-5.964482e-02	C__ODLRE_I_std
53	-5.930790e-02	N__MDNAArea_b1
54	-5.830436e-02	N__LCompactness_b2
55	4.849972e-02	C__ODRLN_I_min
56	-4.751412e-02	C__ODCDF97W2_b1
57	-4.378894e-02	N__ODRP_b2_std
58	4.354749e-02	N__ODCMHomogeneity_b3_d3
59	-4.198131e-02	N__LCompactness_I
60	4.165918e-02	C__ODCMInertia_I_d2
61	4.134608e-02	N__ODCMHomogeneity_I_d3
62	4.025929e-02	N__HDNAArea_b3
63	-3.929634e-02	N__ODCDF97W2_b1
64	3.871546e-02	N__LRE_b2_std
65	-3.193664e-02	N__LHMOD_b1
66	3.164583e-02	N__ODLRE_I_min
67	-3.051417e-02	C__GLCMIInertia_b3_d5
68	-3.034242e-02	N__ODCDF97W3_b1
69	-2.967081e-02	C__GLCMMaxProb_b3_d3
70	-2.929939e-02	N__ODCDF97W1_b1
71	2.843784e-02	C__ODCenterMass_b2
72	2.797285e-02	C__GLN_I_std
73	2.686572e-02	C__HDNAArea_b3
74	-2.683164e-02	N__SRE_I_min
75	-2.517503e-02	C__ODCDF97W3_b1
76	-2.337644e-02	C__MHDist_b3

**Table K.6.** (Cont.) Feature subset for CC nuclei and cytoplasm,  $\lambda_1 = 0.01$ .

Rank	Weight	Feature Name
77	-2.219181e-02	C_GLCMHomogeneity_I_d2
78	-2.142377e-02	N_HDist_b1
79	-2.068372e-02	C_GLCMClusterProminence_b1_d2
80	1.998749e-02	C_LNum_I
81	-1.955534e-02	C_MHCenterMass_b3
82	1.700223e-02	C_ODCenterMass_I
83	1.641796e-02	N_LMHMOD_I
84	-1.604921e-02	C_RP_I_std
85	1.523584e-02	N_LRE_I_max
86	-1.522668e-02	C_PerimeterCurvature_max
87	1.431586e-02	N_ODCMHomogeneity_b1_d2
88	-1.236985e-02	C_GLCMClusterProminence_b1_d1
89	-1.203575e-02	C_ODRP_b3_std
90	-1.186359e-02	C_HCenterMass_I
91	-1.085220e-02	C_GCDF97W4_b3
92	1.056413e-02	C_Compactness
93	1.028205e-02	C_LNum_b2
94	-9.386847e-03	C_HGRE_b3_kurtosis
95	9.025984e-03	C_MHCenterMass_b2
96	-8.793596e-03	C_LGRE_b3_kurtosis
97	-8.366224e-03	C_ODCMClusterProminence_b1_d3
98	-8.183801e-03	C_LGRE_b1_std
99	7.652556e-03	C_ODCMEnergy_b2_d2
100	6.143225e-03	C_GLCMEnergy_b3_d5
101	-6.055648e-03	N_LMHMOD_b1
102	-5.852721e-03	N_LDNAArea_I
103	5.551662e-03	N_PerimeterCurvature_max
104	-5.162097e-03	C_GEntropy_b3
105	-5.117293e-03	N_ODLGRE_b2_std
106	4.494290e-03	C_PerimeterFFT_min
107	2.987789e-03	N_LDNAArea_b1
108	-2.665136e-03	C_MCompactness_b3
109	2.471786e-03	C_ODCDF97W4_b2
110	1.798200e-03	N_ODCMHomogeneity_b3_d2
111	-1.618008e-03	C_ODAreaFD_b3
112	8.564049e-04	C_ODCMEnergy_b3_d4
113	-5.546928e-04	C_HGRE_b1_std
114	-3.190818e-04	N_ODCMIertia_b1_d4

**Table K.7.** Feature subset for WSGran spatial nuclei,  $\lambda_1 = 0.01$ .

Rank	Weight	Feature Name
1	1.433220e+00	CG_NumkWalks_3
2	-9.820955e-01	OCG_EdgeLength_std
3	-5.130411e-01	OCG_NumNeighbors_5
4	4.971090e-01	CG_EdgeLength_skewness
5	4.744610e-01	VT_NumTriangles
6	4.594567e-01	DTG_EdgeLength_kurtosis
7	4.404952e-01	VT_Degree_std
8	-3.571838e-01	MST_Degree_skewness
9	-3.153648e-01	OCG_WeightedDegree_std
10	3.073803e-01	OCG_NumNeighbors_3
11	-2.562353e-01	OCG_NumTriangles
12	2.337469e-01	VT_CyclomaticNumber
13	-2.234441e-01	OCG_WeightedDegree_skewness
14	-2.216045e-01	DTG_WeightedDegree_skewness
15	1.845952e-01	CG_SpectralRadius_2
16	1.731443e-01	VT_RF_min
17	-1.729189e-01	VT_Area_std
18	1.658975e-01	VT_Area_min
19	-1.501533e-01	MST_WeightedDegree_skewness
20	1.495022e-01	VT_NumkWalks_3
21	1.292297e-01	OCG_NumNeighbors_1
22	7.622500e-02	MST_EdgeLength_median
23	2.991934e-02	OCG_NumNeighbors_2

**Table K.8.** Feature subset for WSGran nuclei and spatial nuclei,  $\lambda_1 = 0.01$ . (continued on next page)

Rank	Weight	Feature Name
1	1.353502e+00	N__ODAreaFD_b2
2	-1.191605e+00	N__ODAreaFD_b1
3	8.932419e-01	SPN__CGNumWalks_3
4	-7.854484e-01	SPN__OCGEdgeLength_std
5	-7.111480e-01	N__ODAreaFD_b3
6	4.308323e-01	SPN__VTNumTriangles
7	-3.931756e-01	SPN__OCGNumNeighbors_5
8	3.596382e-01	SPN__CGEdgeLength_skewness
9	-3.441300e-01	SPN__MSTDegree_skewness
10	-3.216486e-01	SPN__OCGWeightedDegree_std
11	3.175511e-01	SPN__VTDegree_std
12	-2.672809e-01	N__ODLRE_b2_std
13	2.574076e-01	SPN__VTArea_min
14	-2.510505e-01	N__LDNAArea_b3
15	2.344822e-01	N__LRE_b2_std
16	2.190354e-01	SPN__DTGEdgeLength_kurtosis
17	-2.084176e-01	N__Hue_max
18	-2.069853e-01	SPN__DTGWeightedDegree_skewness
19	2.052665e-01	N__ODLGRE_b2_min
20	-2.019364e-01	N__LHMOD_b1
21	-1.976259e-01	N__Hue_min
22	1.970064e-01	SPN__OCGNumNeighbors_3
23	1.948461e-01	N__ODRP_I_min
24	1.930981e-01	N__MDNAArea_b2
25	1.808876e-01	SPN__OCGNumNeighbors_1
26	-1.715475e-01	SPN__MSTWeightedDegree_skewness
27	-1.697914e-01	N__ImageBands_b3_max
28	-1.516665e-01	N__LGRE_I_std
29	-1.369464e-01	SPN__OCGWeightedDegree_skewness
30	-1.346968e-01	N__HDNAArea_b1
31	-1.345674e-01	N__GLCMClusterProminence_b3_d1
32	1.301399e-01	SPN__MSTEEdgeLength_median
33	1.300312e-01	N__LRE_b1_std
34	1.298048e-01	SPN__VTRF_min
35	-1.289063e-01	N__ODLGRE_b2_std
36	1.228508e-01	N__LNum_b1
37	-1.114896e-01	N__MIOD_b1
38	1.089374e-01	N__LIOD_I

**Table K.8.** (Cont.) Feature subset for WSGran spatial nuclei and nuclei,  $\lambda_1 = 0.01$ .

Rank	Weight	Feature Name
39	1.076537e-01	SPN_VTCyclomaticNumber
40	9.325616e-02	N_MDNAArea_b3
41	-9.200293e-02	N_HIOD_b1
42	8.194485e-02	N_LGRE_b1_skewness
43	6.784439e-02	N_ODCDF97W2_I
44	-6.662233e-02	N_LIOD_b2
45	-6.465396e-02	SPN_MSTEdgeLength_skewness
46	-5.998979e-02	N_ODCMClusterProminence_b1_d5
47	-5.931180e-02	N_ODCDF97W3_b1
48	-5.842754e-02	N_SRE_I_mean
49	5.783340e-02	N_ODCDF97W1_I
50	5.657927e-02	N_LNum_b2
51	5.555807e-02	N_LRE_I_std
52	5.349659e-02	N_GLCMHomogeneity_b3_d4
53	-5.128658e-02	N_ODCDF97W5_b1
54	-4.995358e-02	N_ODCMClusterProminence_b1_d4
55	-4.657032e-02	N_ODCMClusterProminence_b1_d3
56	-4.557845e-02	N_SRE_I_max
57	-4.241251e-02	N_LGRE_I_kurtosis
58	-3.854045e-02	SPN_VTWeightedDegree_std
59	3.648365e-02	N_ODV_b1_min
60	3.487694e-02	N_LMMOD_b1
61	3.443637e-02	N_MNum_I
62	3.301699e-02	SPN_VTArea_median
63	3.028585e-02	N_LNum_b3
64	2.926121e-02	N_GLCMHomogeneity_b3_d3
65	-2.861893e-02	SPN_MSTDegree_kurtosis
66	2.525174e-02	N_LRE_b2_median
67	-2.510757e-02	N_SRE_b3_median
68	-2.164717e-02	N_SRE_b1_max
69	-8.903142e-03	N_ODSRE_I_std
70	8.404052e-03	N_HDNAArea_b3
71	7.628763e-03	N_ODReflSymmetry_b3
72	-4.781991e-03	N_GLCMMaxProb_b2_d2
73	-4.373068e-03	N_GLCMInertia_b3_d5
74	7.803042e-04	N_ODSRE_I_min

**Table K.9.** Feature subset for WSGran nuclei, cytoplasm/stroma, and spatial nuclei,  $\lambda_1 = 0.01$ . (continued on next 3 pages)

Rank	Weight	Feature Name
1	9.219192e-01	SPN__CGNumkWalks_3
2	8.591580e-01	N__ODAreaFD_b2
3	-8.221530e-01	N__ODAreaFD_b1
4	-7.205722e-01	N__ODAreaFD_b3
5	-4.780049e-01	SPN__OCGNumNeighbors_5
6	-4.211430e-01	CS__HIOD_b1
7	3.649713e-01	CS__ODEntropy_b3
8	-3.640612e-01	SPN__OCGWeightedDegree_max
9	-3.320639e-01	CS__Hue_mean
10	-2.931073e-01	CS__Hue_min
11	2.928478e-01	SPN__VTNumTriangles
12	-2.653949e-01	CS__SRE_b2_min
13	-2.408082e-01	CS__MIOD_b1
14	2.341711e-01	SPN__VTCyclomaticNumber
15	2.314865e-01	CS__ImageBands_b1_std
16	-2.148643e-01	CS__GEntropy_b3
17	2.133934e-01	CS__LNum_b1
18	2.122797e-01	SPN__CGEdgeLength_skewness
19	-2.064428e-01	SPN__MSTDegree_skewness
20	-2.045280e-01	SPN__OCGEdgeLength_std
21	-2.025859e-01	SPN__OCGWeightedDegree_skewness
22	-1.945072e-01	N__ImageBands_b1_max
23	-1.914469e-01	CS__GLCMEntropy_b3_d1
24	-1.911118e-01	N__ODLRE_b2_std
25	-1.729921e-01	CS__SRE_b3_median
26	-1.720368e-01	SPN__MSTWeightedDegree_skewness
27	-1.657401e-01	SPN__MSTEEdgeLength_skewness
28	-1.595818e-01	CS__ODSRE_b3_std
29	-1.573675e-01	N__HIOD_b1
30	1.571009e-01	CS__MinorAxisLength
31	-1.561577e-01	CS__ODCMClusterProminence_b1_d3
32	-1.504490e-01	CS__MHCompactness_b1
33	1.451456e-01	N__LRE_b2_std
34	1.443607e-01	SPN__VTArea_min
35	-1.439916e-01	CS__GCDF97W5_b3
36	-1.336737e-01	CS__ODCMCorrelation_b1_d2
37	1.318643e-01	CS__ODRP_I_skewness
38	-1.307727e-01	CS__ODLGRE_b2_kurtosis

**Table K.9.** (Cont.) Feature subset for WSGran nuclei, cytoplasm/stroma, and spatial nuclei,  $\lambda_1 = 0.01$ . (continued on next 2 pages)

Rank	Weight	Feature Name
39	-1.306235e-01	CS__ODCMCorrelation_b1_d1
40	-1.280196e-01	CS__ODRP_I_std
41	-1.258643e-01	N__GLCMClusterProminence_b3_d1
42	1.213039e-01	N__ODV_b1_min
43	-1.199673e-01	N__ImageBands_b3_max
44	1.192056e-01	SPN__VTRF_min
45	1.187919e-01	CS__ODLGRE_b2_min
46	1.183410e-01	N__LNum_b1
47	-1.159897e-01	CS__ODRP_b1_std
48	1.128121e-01	N__ODCDF97W1_b3
49	-1.111728e-01	N__ODSRE_I_std
50	9.871352e-02	SPN__VTDegree_std
51	-9.723357e-02	N__Hue_skewness
52	9.672993e-02	N__MDNAArea_b2
53	-9.446738e-02	SPN__VTWeightedDegree_std
54	9.147790e-02	CS__LIOD_I
55	-9.106999e-02	CS__ODLRE_b2_std
56	-9.048746e-02	N__ODCMClusterProminence_b1_d4
57	-8.805524e-02	N__ODCMClusterProminence_b1_d5
58	8.638512e-02	CS__RP_b2_std
59	-8.601399e-02	N__LGRE_I_std
60	-8.555913e-02	CS__ODRP_b3_std
61	-8.174037e-02	N__LHMOD_b1
62	8.144471e-02	SPN__OCGNumNeighbors_1
63	-8.141367e-02	CS__ODCDF97W3_b1
64	8.017953e-02	CS__HNum_b1
65	7.752299e-02	N__LRE_b3_max
66	7.531869e-02	CS__LNum_I
67	7.259549e-02	N__LNum_I
68	-7.254003e-02	CS__LGRE_I_std
69	7.046510e-02	CS__RLN_b2_skewness
70	6.953772e-02	CS__GCDF97W2_b1
71	-6.744181e-02	CS__GCDF97W4_b1
72	6.448770e-02	SPN__MSTWeightedDegree_min
73	6.287227e-02	N__LRE_b3_std
74	-6.191219e-02	CS__GLN_b3_kurtosis
75	-5.550484e-02	N__ODCMClusterProminence_b1_d3
76	-5.458177e-02	CS__ODSRE_I_std

**Table K.9.** (Cont.) Feature subset for WSGran nuclei, cytoplasm/stroma, and spatial nuclei,  $\lambda_1 = 0.01$ . (continued on next page)

Rank	Weight	Feature Name
77	-5.005125e-02	CS_HCompactness_I
78	4.994990e-02	CS_LRE_b3_std
79	-4.945253e-02	N_Hue_min
80	-4.875975e-02	CS_GLN_b3_skewness
81	-4.592865e-02	SPN_DTGWeightedDegree_skewness
82	-4.562034e-02	CS_LIOD_b2
83	4.495147e-02	CS_ODV_b3_min
84	4.480360e-02	CS_ODLRE_b1_kurtosis
85	-4.232635e-02	CS_GLN_b2_kurtosis
86	4.160046e-02	CS_ImageBands_b2_max
87	-3.848219e-02	CS_PerimeterCurvature_median
88	3.726935e-02	N_GLCMCorrelation_b2_d2
89	3.625060e-02	CS_ODSRE_I_skewness
90	-3.382495e-02	CS_ODV_b3_skewness
91	-3.254520e-02	CS_MCompactness_b3
92	-3.209882e-02	CS_ODCMClusterProminence_b2_d2
93	-3.097719e-02	CS_GLCMMaxProb_b3_d1
94	-3.074413e-02	CS_LGRE_b3_kurtosis
95	-2.987131e-02	N_LHMOD_b2
96	2.729591e-02	CS_SRE_b2_kurtosis
97	2.610011e-02	CS_HNum_b3
98	-2.454967e-02	CS_GLCMHomogeneity_b1_d2
99	-2.325623e-02	CS_RP_b3_min
100	2.247649e-02	CS_RP_b3_std
101	-1.977827e-02	CS_ODLRE_b2_skewness
102	-1.937860e-02	CS_SRE_b3_min
103	-1.926013e-02	CS_ODCDF97W4_b1
104	1.878564e-02	CS_PerimeterFFT_min
105	-1.850700e-02	N_LGRE_b3_std
106	1.795542e-02	N_MDNAArea_I
107	1.783353e-02	CS_HIOD_b3
108	1.775005e-02	CS_GLCMEnergy_b3_d5
109	-1.731813e-02	N_LGRE_I_kurtosis
110	-1.486366e-02	CS_ODSRE_b1_skewness
111	1.408216e-02	SPN_VTWeightedDegree_median
112	-1.280598e-02	CS_HDNAArea_b1
113	1.191954e-02	N_MCompactness_b2
114	1.177361e-02	CS_MNum_b2

**Table K.9.** (Cont.) Feature subset for WSGran nuclei, cytoplasm/stroma, and spatial nuclei,  $\lambda_1 = 0.01$ .

Rank	Weight	Feature Name
115	-1.105157e-02	N_LIOD_b2
116	-9.826770e-03	N_ODCDF97W3_b1
117	-9.564738e-03	N_ODCMMaxProb_I_d4
118	-9.328367e-03	N_ODCMIInertia_b2_d2
119	9.024592e-03	SPN_OCGNumNeighbors_3
120	-8.501128e-03	CS_HCenterMass_b2
121	8.083548e-03	N_PerimeterCurvature_median
122	5.062619e-03	N_LRE_b1_median
123	4.800467e-03	CS_MNum_I
124	-3.002135e-03	N_Eccentricity
125	-2.197020e-03	N_ODCMClusterProminence_b2_d4

**Table K.10.** Feature subset for WSGran nuclei, cytoplasm, and spatial nuclei,  $\lambda_1 = 0.01$ . (continued on next 3 pages)

Rank	Weight	Feature Name
1	9.781533e-01	N_ODAreaFD_b2
2	-9.483967e-01	N_ODAreaFD_b1
3	9.251866e-01	SPN_CGNumkWalks_3
4	-7.374574e-01	N_ODAreaFD_b3
5	-4.709234e-01	SPN_OCGNumNeighbors_5
6	-4.232100e-01	C_MIOD_b1
7	-3.506210e-01	C_HIOD_b1
8	3.499069e-01	SPN_VTNumTriangles
9	-3.081870e-01	SPN_OCGWeightedDegree_max
10	-2.872939e-01	N_ODLRE_b2_std
11	-2.867729e-01	C_Hue_min
12	2.520167e-01	N_ODV_b1_min
13	-2.489204e-01	SPN_OCGEdgeLength_std
14	-2.450167e-01	SPN_MSTDegree_skewness
15	-2.387001e-01	C_GLCMHomogeneity_b1_d2
16	-2.345011e-01	C_ODLGRE_b2_kurtosis
17	-2.297051e-01	N_HIOD_b1
18	-2.273935e-01	C_ODRP_b3_std
19	2.243810e-01	SPN_VTDegree_std
20	2.170180e-01	C_LNum_b1
21	2.162101e-01	C_LRE_b3_std
22	2.142909e-01	SPN_CGEdgeLength_skewness
23	-2.137077e-01	SPN_OCGWeightedDegree_skewness
24	-2.000755e-01	C_GEntropy_b3
25	1.941439e-01	N_LRE_b2_std
26	1.866234e-01	C_ImageBands_b2_std
27	1.801789e-01	C_ODGLN_b2_std
28	1.643820e-01	SPN_VTCyclomaticNumber
29	-1.537093e-01	SPN_MSTWeightedDegree_skewness
30	1.505681e-01	C_ODCenterMass_b2
31	-1.451887e-01	C_MIOD_b2
32	1.448871e-01	C_ODRP_I_min
33	-1.431871e-01	SPN_MSTEEdgeLength_skewness
34	1.431669e-01	N_LRE_b3_std
35	1.374897e-01	SPN_VTArea_min
36	-1.348755e-01	N_GLCMClusterProminence_b3_d1
37	1.340154e-01	C_FilledArea
38	1.296744e-01	C_RLN_b2_std

**Table K.10.** (Cont.) Feature subset for WSGran nuclei, cytoplasm, and spatial nuclei,  $\lambda_1 = 0.01$ . (continued on next 2 pages)

Rank	Weight	Feature Name
39	-1.296363e-01	C_ODLRE_b2_std
40	1.295547e-01	C_ODLGRE_b2_min
41	1.285267e-01	C_GLCMEnergy_b3_d5
42	-1.273095e-01	C_SRE_b3_median
43	1.248769e-01	C_LRE_b1_std
44	1.210796e-01	SPN_VTRF_min
45	-1.207868e-01	N_LGRE_I_std
46	-1.122382e-01	C_ODCMClusterProminence_I_d2
47	-1.107021e-01	N_ODSRE_I_std
48	-1.082092e-01	N_ODCMClusterProminence_b1_d3
49	1.049334e-01	C_ODReflSymmetry_b3
50	-1.008694e-01	C_GLCMMMaxProb_b3_d2
51	-9.722073e-02	N_ImageBands_b3_max
52	-9.567047e-02	C_LIOD_b1
53	9.451971e-02	N_LNum_b1
54	9.277175e-02	C_PerimeterFFT_max
55	-9.147188e-02	N_Hue_skewness
56	9.146152e-02	C_ODCMIInertia_b2_d3
57	-9.023131e-02	N_Hue_min
58	8.737524e-02	C_MDNAArea_b2
59	8.385797e-02	N_MDNAArea_I
60	8.323950e-02	C_GCDF97W1_b2
61	7.645512e-02	N_MDNAArea_b2
62	-7.560977e-02	C_GLCMInertia_b3_d5
63	7.472523e-02	C_MNum_b3
64	-7.208453e-02	C_SRE_b2_min
65	-6.892390e-02	SPN_OCGNumTriangles
66	-6.710253e-02	SPN_VTWeightedDegree_std
67	-6.690431e-02	C_ODCMClusterProminence_b1_d5
68	6.667242e-02	C_FFTk_min
69	6.488578e-02	C_HDNAArea_b3
70	6.456159e-02	SPN_OCGNumNeighbors_1
71	-6.435266e-02	C_SRE_I_max
72	-6.379043e-02	C_ODCMClusterProminence_I_d5
73	-6.297665e-02	C_SRE_b1_median
74	6.244464e-02	N_LNum_I
75	-6.191660e-02	SPN_DTGWeightedDegree_skewness
76	-5.935430e-02	N_ODCMClusterProminence_b1_d4

**Table K.10.** (Cont.) Feature subset for WSGran nuclei, cytoplasm, and spatial nuclei,  $\lambda_1 = 0.01$ . (continued on next page)

Rank	Weight	Feature Name
77	-5.828272e-02	C_SRE_b2_median
78	5.746360e-02	N_LRE_b1_mean
79	-5.006769e-02	N_ODCMClusterProminence_b2_d4
80	4.587393e-02	C_ODSRE_I_min
81	-4.515761e-02	N_Hue_max
82	-4.182965e-02	N_ODCMClusterProminence_b1_d5
83	-4.094695e-02	C_GCDF97W5_b3
84	-3.656617e-02	N_MHCompactness_b3
85	3.599177e-02	N_ODRP_b3_std
86	-3.444574e-02	C_MHCompactness_b1
87	3.293710e-02	C_MinorAxisLength
88	3.070967e-02	N_ODReflSymmetry_b3
89	2.942729e-02	N_MCompactness_b2
90	-2.577301e-02	C_ODSRE_b1_std
91	-2.406649e-02	C_ODRP_b1_std
92	-2.397892e-02	C_ODCDF97W4_b1
93	-2.396995e-02	N_HGRE_I_std
94	2.340520e-02	N_LNum_b3
95	-2.296886e-02	N_LGRE_I_kurtosis
96	2.295203e-02	N_ODCDF97W1_b3
97	-2.194126e-02	C_ODSRE_I_std
98	-2.021824e-02	N_ODCMMaxProb_I_d4
99	1.827500e-02	C_LNum_I
100	1.652990e-02	N_ODRP_I_min
101	1.607653e-02	C_ImageBands_b2_max
102	-1.463570e-02	N_ODCMIInertia_b2_d2
103	1.252659e-02	SPN_OCGNumNeighbors_3
104	-1.123760e-02	N_LHMOD_b2
105	-1.079279e-02	C_LGRE_b1_std
106	1.010016e-02	C_GCDF97W2_b2
107	9.178845e-03	C_GReflSymmetry_b2
108	8.779657e-03	N_Orientation
109	-8.148793e-03	C_ODCMClusterProminence_I_d4
110	7.935389e-03	N_PerimeterCurvature_median
111	-7.723514e-03	C_HCompactness_b2
112	7.378422e-03	C_HNum_b1
113	-7.219415e-03	N_HDNAArea_b1
114	5.183049e-03	C_FFTk_2

**Table K.10.** (Cont.) Feature subset for WSGran nuclei, cytoplasm, and spatial nuclei,  $\lambda_1 = 0.01$ .

Rank	Weight	Feature Name
115	4.754695e-03	C_ImageBands_b2_median
116	2.582191e-03	N_MDNAArea_b3
117	1.962732e-03	SPN_MSTWeightedDegree_min
118	-1.879942e-03	C_GCDF97W4_b1
119	1.621643e-03	N_MDNAArea_b1
120	-1.550982e-03	C_HDist_b2
121	-1.219770e-03	N_LHMOD_b1
122	7.890785e-04	N_GReflSymmetry_b2
123	7.540850e-04	N_LMMOD_b1



## Appendix L

### Grafting Feature Subsets: Higher-Level Objects

**Table L.1.** Feature subset for GT Cell<sub>NC</sub>,  $\lambda_1 = 0.01$ . (continued on next 2 pages)

Rank	Weight	Feature Name
1	-8.160287e-01	C_HIOD_b1
2	-7.844025e-01	N_MIOD_b1
3	-7.678485e-01	N_HIOD_b1
4	-7.189662e-01	C_LGRE_b1_std
5	5.673338e-01	N_ODAreaFD_b2
6	4.379211e-01	C_ODRLN_b1_median
7	3.907667e-01	N_SRE_b2_skewness
8	-3.826657e-01	C_SRE_b3_median
9	3.708245e-01	N_IOD_b2
10	-3.496470e-01	C_MIOD_b1
11	-3.307434e-01	C_ODLRE_b2_std
12	-3.278040e-01	C_HDist_I
13	3.243743e-01	C_ODV_b1_min
14	2.920634e-01	N_LRE_b1_std
15	2.798733e-01	N_ODEntropy_b3
16	2.690174e-01	N_Radii_min
17	2.678069e-01	CELLNC_FilledArea
18	2.648872e-01	C_LNum_b3
19	2.495305e-01	N_ODCDF97W1_b3
20	2.441716e-01	N_ODCDF97W3_I
21	2.377962e-01	N_HNum_b1
22	-2.347988e-01	N_SRE_b2_max
23	-2.288592e-01	C_ImageBands_b2_skewness
24	-2.180971e-01	C_BWReflSymmetry
25	2.152027e-01	N_FFTk_32
26	2.062688e-01	N_MNum_I
27	1.985098e-01	N_ODLGRE_b3_kurtosis
28	1.886843e-01	C_ODEntropy_b3
29	-1.852344e-01	N_ODCMClusterProminence_b3_d1
30	1.824316e-01	C_ODCDF97W1_b3
31	-1.799781e-01	C_Hue_std
32	-1.777254e-01	N_ODLRE_b2_std
33	1.726001e-01	N_ODRLN_b1_median
34	1.632274e-01	N_SRE_I_kurtosis
35	1.631924e-01	N_RP_b2_kurtosis
36	1.620142e-01	N_ODRLN_I_median
37	-1.559363e-01	N_GLCMMaxProb_I_d3
38	1.275820e-01	N_MIOD_I
39	1.237306e-01	N_ODGLN_b2_std

**Table L.1.** (Cont.) Feature subset for GT Cell<sub>NC</sub>,  $\lambda_1 = 0.01$ . (continued on next page)

Rank	Weight	Feature Name
40	1.217239e-01	C_ODRLN_b1_std
41	1.195572e-01	N_ODV_b3_skewness
42	-1.185514e-01	N_ODSRE_b1_skewness
43	1.112794e-01	N_GLN_b1_skewness
44	-1.102793e-01	CELLNC_FFTk_kurtosis
45	1.089147e-01	N_GLCMCorrelation_b1_d4
46	-1.085744e-01	N_ODV_b1_kurtosis
47	1.007086e-01	C_SRE_b3_std
48	9.718332e-02	C_ODAreaFD_I
49	-9.495781e-02	N_MHDist_b2
50	9.034832e-02	N_ODAreaFD_I
51	-8.768732e-02	N_HDist_I
52	8.518046e-02	C_HNum_I
53	-8.504872e-02	N_RP_b2_skewness
54	8.129641e-02	C_ImageBands_b2_std
55	-8.090097e-02	N_RP_b1_kurtosis
56	7.884266e-02	N_GLN_b3_skewness
57	-7.872932e-02	N_ODCenterMass_b3
58	-7.601192e-02	N_GLN_b2_skewness
59	-7.460881e-02	N_GCDF97W1_b2
60	-7.363267e-02	C_GLCMMaxProb_b1_d1
61	7.322559e-02	C_ODSRE_I_skewness
62	6.610530e-02	C_FFTk_20
63	6.604539e-02	C_RP_b2_kurtosis
64	6.369649e-02	C_Radii_std
65	6.074374e-02	C_HNum_b3
66	5.420442e-02	C_MHCompactness_I
67	-5.353743e-02	C_ODSRE_I_kurtosis
68	5.161576e-02	C_FFTk_14
69	-4.966046e-02	N_ODRP_b2_kurtosis
70	4.651878e-02	C_PPerimeter
71	4.555359e-02	C_MCompactness_b1
72	4.360405e-02	N_ODCDF97W4_b2
73	4.272141e-02	N_RLN_b1_min
74	3.841127e-02	C_GLCMCorrelation_b3_d5
75	3.617106e-02	C_GLCMClusterShade_b2_d4
76	3.544090e-02	C_SRE_b3_skewness
77	3.435932e-02	N_ODCMHomogeneity_b3_d2
78	3.271589e-02	N_ODCMHomogeneity_b3_d5

**Table L.1.** (Cont.) Feature subset for GT Cell<sub>NC</sub>,  $\lambda_1 = 0.01$ .

<b>Rank</b>	<b>Weight</b>	<b>Feature Name</b>
79	2.977610e-02	C_SRE_b2_skewness
80	2.708510e-02	N_LNum_b3
81	-2.696053e-02	N_SRE_b2_median
82	2.616285e-02	N_GCDF97W4_I
83	2.448410e-02	N_MNum_b1
84	-2.412499e-02	N_RP_b3_skewness
85	-2.215215e-02	N_RP_b1_skewness
86	2.195098e-02	C_RLN_b3_kurtosis
87	-2.166344e-02	N_MCompactness_b3
88	2.127358e-02	N_MNum_b3
89	2.080034e-02	C_ODCMHomogeneity_b3_d1
90	-1.966765e-02	C_ODCMMaxProb_b1_d2
91	-1.870099e-02	N_ODRLN_b2_skewness
92	1.753060e-02	N_GCDF97W4_b2
93	-1.539217e-02	C_ODLRE_b1_std
94	1.275383e-02	CELLNC_PerimeterCurvature_min
95	-8.176203e-03	C_ODV_I_kurtosis
96	7.487404e-03	N_ODCMEnergy_b3_d1
97	5.711174e-03	CELLNC_Orientation
98	-5.563365e-03	C_ODSRE_b1_skewness
99	-5.494433e-03	C_ODLRE_b1_kurtosis
100	-2.292509e-03	N_LMMOD_b3
101	-1.677802e-03	C_ODV_b3_kurtosis
102	1.072681e-03	C_LGRE_b2_kurtosis
103	9.946978e-04	C_GCDF97W4_b2

**Table L.2.** Feature subset for WSHmin Cell<sub>NC</sub>,  $\lambda_1 = 0.01$ . (continued on next 3 pages)

Rank	Weight	Feature Name
1	1.220544e+00	N_ODAreaFD_b2
2	-1.053911e+00	N_ODAreaFD_b1
3	-6.873962e-01	C_LIOD_b1
4	5.850526e-01	C_GLN_b3_min
5	-5.499661e-01	C_HIOD_b1
6	-3.855832e-01	C_MIOD_b1
7	3.668061e-01	N_LDNAArea_b1
8	3.048425e-01	CELLNC_AspectRatio
9	-2.575350e-01	C_ODLRE_b2_std
10	-2.522256e-01	N_ODAreaFD_b3
11	2.420228e-01	C_RLN_b3_std
12	-2.369523e-01	N_LDNAArea_I
13	2.364603e-01	CELLNC_ConvexArea
14	-2.189740e-01	C_SRE_b2_mean
15	-2.160113e-01	N_LGRE_b2_std
16	2.114274e-01	C_ODGLN_b2_std
17	-1.883402e-01	N_LDNAArea_b2
18	-1.866988e-01	N_HIOD_b1
19	-1.695946e-01	C_ODLGRE_I_kurtosis
20	-1.673250e-01	N_ODCMClusterProminence_b3_d1
21	-1.438847e-01	N_ODCDF97W2_I
22	-1.409956e-01	N_GLCMHomogeneity_b1_d5
23	1.382479e-01	N_HNum_b1
24	1.379554e-01	C_ODRLN_I_min
25	-1.345586e-01	CELLNC_PerimeterFFT_kurtosis
26	1.328763e-01	C_LGRE_b3_kurtosis
27	-1.315591e-01	C_LGRE_b2_std
28	-1.312598e-01	C_LGRE_I_std
29	1.308670e-01	C_GLCMCorrelation_b2_d5
30	1.250873e-01	C_FFTk_min
31	1.230001e-01	C_LNum_b3
32	-1.225976e-01	N_GLCMClusterProminence_b3_d3
33	1.147955e-01	N_ODRLN_b2_min
34	1.127848e-01	C_ODCenterMass_b2
35	1.076979e-01	CELLNC_PerimeterCurvature_median
36	-1.071099e-01	N_ODCDF97W1_b1
37	1.060829e-01	C_MCompactness_b3
38	1.028368e-01	C_Solidity
39	9.933178e-02	N_MNum_I

**Table L.2.** (Cont.) Feature subset for WSHmin Cell<sub>NC</sub>,  $\lambda_1 = 0.01$ . (continued on next 2 pages)

Rank	Weight	Feature Name
40	-9.759831e-02	N_MHDist_I
41	9.522927e-02	N_ODV_b2_std
42	9.509541e-02	N_LNum_b3
43	-9.298273e-02	N_HGRE_b2_std
44	-8.923023e-02	N_ODCMClusterProminence_b1_d1
45	8.848998e-02	C_LNum_I
46	8.667730e-02	CELLNC_Solidity
47	-8.653904e-02	N_ODLRE_b2_std
48	8.461227e-02	C_LRE_b1_std
49	8.261215e-02	N_HCenterMass_b3
50	8.097480e-02	C_LCenterMass_I
51	-7.974716e-02	N_MIOD_b1
52	7.868244e-02	C_HCompactness_b3
53	-7.854501e-02	C_ODLGRE_b1_kurtosis
54	7.783348e-02	C_ODV_b2_kurtosis
55	-7.757616e-02	C_ODCMClusterProminence_b1_d5
56	7.700310e-02	C_AspectRatio
57	-7.571468e-02	N_Hue_kurtosis
58	-7.557698e-02	C_Hue_min
59	7.237996e-02	C_LCenterMass_b2
60	-6.994746e-02	N_MDist_I
61	6.726145e-02	C_LGRE_b1_skewness
62	6.342034e-02	C_ODLGRE_b2_min
63	-6.261742e-02	C_LMMOD_b1
64	-6.121030e-02	C_ImageBands_b1_skewness
65	6.099212e-02	C_PerimeterCurvature_min
66	-6.072021e-02	C_HCompactness_I
67	-5.699138e-02	N_ImageBands_I_kurtosis
68	-5.666445e-02	C_ODCMMaxProb_b3_d4
69	5.655431e-02	C_GReflSymmetry_b2
70	5.634539e-02	C_ODGLN_I_std
71	-5.510737e-02	C_SRE_b1_median
72	5.483319e-02	C_GLCMClusterProminence_b2_d1
73	-5.465824e-02	CELLNC_BWReflSymmetry
74	5.412360e-02	N_RP_b1_std
75	5.281421e-02	CELLNC_FFTk_skewness
76	5.248490e-02	C_ODHGRE_b2_min
77	-5.146589e-02	N_EulerNumber

**Table L.2.** (Cont.) Feature subset for WSHmin Cell<sub>NC</sub>,  $\lambda_1 = 0.01$ . (continued on next page)

Rank	Weight	Feature Name
78	-5.070011e-02	CELLNC_PerimeterCurvature_max
79	-5.021789e-02	N_HIOD_I
80	-4.878699e-02	N_HCenterMass_I
81	-4.877315e-02	C_LDist_b2
82	-4.861992e-02	N_LGRE_b1_std
83	-4.857970e-02	N_GLCMHomogeneity_b3_d3
84	-4.639771e-02	N_LDist_b3
85	-4.606996e-02	N_GLCMInertia_I_d3
86	4.372939e-02	CELLNC_FFTk_32
87	4.351796e-02	N_MDNAArea_b2
88	-4.120825e-02	C_MHDist_b1
89	4.099682e-02	C_HCenterMass_b3
90	4.094361e-02	N_LGRE_b3_skewness
91	-3.888219e-02	C_ODCMClusterProminence_b2_d3
92	3.859910e-02	N_LRE_I_std
93	-3.726338e-02	C_ODCMMaxProb_b3_d5
94	3.699295e-02	N_LCompactness_I
95	-3.619673e-02	N_LHMOD_b3
96	3.600630e-02	C_ODV_b3_mean
97	3.577699e-02	N_ODRP_b1_median
98	3.424721e-02	CELLNC_MajorAxisLength
99	3.353080e-02	C_ODV_b1_min
100	3.296647e-02	N_Hue_std
101	3.182520e-02	N_PerimeterFFT_min
102	-3.116044e-02	N_GLCMMaxProb_b2_d5
103	2.922453e-02	N_ODCDF97W4_b2
104	-2.742087e-02	C_PerimeterFFT_min
105	2.642781e-02	C_FFTk_15
106	-2.624575e-02	N_SRE_b3_median
107	2.613195e-02	N_MNum_b3
108	-2.583653e-02	N_GLCMHomogeneity_b1_d3
109	2.551547e-02	C_MHCompactness_b2
110	2.516091e-02	N_LCompactness_b1
111	2.502463e-02	C_ODV_b2_skewness
112	-2.477562e-02	C_ODAreaFD_b3
113	2.476889e-02	N_RP_I_std
114	2.440756e-02	N_LCenterMass_b1
115	-2.398193e-02	C_LGRE_I_max

**Table L.2.** (Cont.) Feature subset for WSHmin Cell<sub>NC</sub>,  $\lambda_1 = 0.01$ .

Rank	Weight	Feature Name
116	2.332806e-02	C_HGRE_b3_kurtosis
117	2.329523e-02	C_LGRE_b2_kurtosis
118	-2.315149e-02	N_MDist_b3
119	-2.216584e-02	N_ODCDF97W2_b1
120	-2.184168e-02	C_HGRE_I_max
121	2.132375e-02	CELLNC_FFTk_21
122	2.063411e-02	N_ODRP_b1_max
123	2.046244e-02	N_ODCMHomogeneity_b3_d2
124	-2.034621e-02	C_ImageBands_b3_max
125	-2.000478e-02	C_LHMOD_b2
126	-1.919148e-02	C_ODReflSymmetry_b2
127	1.912222e-02	C_ODCDF97W2_b3
128	1.853116e-02	C_GCDF97W3_b3
129	1.844933e-02	N_ODCMHomogeneity_b3_d5
130	-1.629662e-02	C_ODLRE_I_std
131	-1.562251e-02	C_HNum_b2
132	1.435046e-02	N_Orientation
133	-1.368310e-02	C_ODCDF97W3_b1
134	1.334274e-02	C_ODGLN_b3_std
135	1.318131e-02	C_PerimeterFFT_std
136	-1.304601e-02	C_SRE_b2_min
137	1.206957e-02	C_MCompactness_b2
138	1.176983e-02	N_HGRE_b3_skewness
139	1.158396e-02	C_MHCenterMass_I
140	1.031649e-02	N_LRE_b1_std
141	-9.167682e-03	N_LCenterMass_b3
142	8.085717e-03	C_HNum_b1
143	-6.290405e-03	N_LIOD_b1
144	6.234072e-03	N_ODV_I_skewness
145	2.373537e-03	C_SRE_b3_std
146	2.070886e-03	C_ODCMIInertia_b3_d3