

Medical Image Analysis of Gastric Cancer in Digital Histopathology: Methods, Applications and Challenges

vorgelegt von

Master of Technology

Harshita SHARMA

geb. in Neu Delhi, Indien

von der Fakultät IV-Elektrotechnik und Informatik
der Technischen Universität Berlin
zur Erlangung des akademischen Grades

Doktorin der Ingenieurwissenschaften

- Dr.-Ing -

genehmigte Dissertation

Promotionsausschuss :

Vorsitzender: Prof. Dr.-Ing. Thomas SIKORA

Gutachter: Prof. Dr.-Ing. Olaf HELLWICH

Prof. Dr.-Ing. Niels GRABE

Prof. Dr. rer. nat. Peter HUFNAGL

Tag der wissenschaftlichen Aussprache: 25. April 2017

Berlin 2017

Dedicated to my family and loved ones.

“I have become my own version of an optimist. If I can’t make it through one door, I’ll go through another door – or I’ll make a door. Something terrific will come no matter how dark the present.”

— **Rabindranath Tagore**

Acknowledgements

“At times our own light goes out and is rekindled by a spark from another person. Each of us has cause to think with deep gratitude of those who have lighted the flame within us.”

— **Albert Schweitzer**

The undertaking and completion of this dissertation has proved to be an eventful journey, filled with tribulations and challenges, both professional and personal. With proper coordination and guidance I was able to overcome the hurdles and realize my goals. Hence, it is with immense gratitude that I acknowledge those, who have provided me with inspiration, financial and moral support, and technical direction.

There are no words which can adequately express my gratitude for the invaluable guidance and encouragement, which I have been extremely fortunate to receive from my supervisor **Prof. Dr. -Ing. Olaf Hellwich**, Professor and Head, Computer Vision and Remote Sensing Group, Department of Computer Engineering and Microelectronics, Technical University Berlin, Berlin, Germany. He kindly welcomed me in his group, and guided me throughout the course of my investigation. His generous support and stimulating suggestions have been the driving force for me throughout my work. Besides being an excellent guide, he has been conferring valuable hints to improve various aspects of my research. It was the inspiration provided by him that gave me the confidence to pursue this innovative topic and work it through to a successful completion.

I am greatly indebted and express my sincere gratitude to **Prof. Dr. rer. nat. Peter Hufnagl**, Professor and Head, Department of Digital Pathology and IT, Institute of Pathology “Rudolf-Virchow-House”, Charité University Hospital, Berlin, Germany. He granted the approval to conduct my research in collaboration with the group at Charité University Hospital, and provided me necessary resources for the same. He offered me additional office space to perform my work at Charité University Hospital, and has been really welcoming, encouraging and supportive towards my ideas. His in-depth domain knowledge and consistent motivation have been most essential for a deeper understanding of the problem and deciding the next course of action. Furthermore, he advised me to utilize parts of his doctorate research comprising of the object-level shape-based feature extraction for my studies, which proved an important ingredient for the success of my work.

I sincerely thank **Dipl.-Inf. Norman Zerbe**, Project Manager, Department of Digital Pathology and IT, Institute of Pathology “Rudolf-Virchow-House”, Charité University Hospital, Berlin, Germany for his extremely useful suggestions during my research. His wide knowledge and logical way of thinking have been truly inspiring. He has always been forthcoming to provide his assistance during various steps of my research. I also thank him for his semi-automatic whole slide image registration implementation, and for providing a subset of cell nuclei shape features.

I am immensely thankful to **Prof. Dr. med. Christoph Röcken**, Director, Department of Pathology, University Medical Center Schleswig Holstein (UKSH), Campus Kiel, Kiel, Germany for collaboration with Charité, and allowing me to acquire the necessary gastric cancer whole slide images used in my research. A former PhD candidate under his supervision, **Dr. rer. medic. Hans-Michael Behrens**, has been in direct contact with the research group at Charité, assisted me during the data acquisition stage and provided me with the associated knowledge, to whom I am extremely thankful.

I am grateful to **Dr. med. Christine Böger**, Pathologist, University Medical Center Schleswig Holstein (UKSH), Campus Kiel, Kiel, Germany, for giving me her precious time by performing the review of ground truth data for the cell nuclei classification step. I sincerely thank **Dr. med. Iris Klempert**, Pathologist, Institute of Pathology, Charité University Hospital, Berlin, Germany for creating the ground truth for the necrosis detection step. I am also thankful to **PD Dr. med. Fredrick Klauschen**, Pathologist, Institute of Pathology, Charité University Hospital, Berlin, Germany for introducing me to the histological architecture and various cell nuclei types in our datasets, and also for visually identifying necrosis in the tissue.

I truly admire the support and assistance from all the members of Department of Digital Pathology and IT, Institute of Pathology, Charité University Hospital, Berlin, Germany. Especially, I would like to thank **Dr. rer. medic. Stephan Wienert**, Postdoctoral Scientist, for providing his implementation of cell nuclei segmentation as required for our H&E stained gastric cancer images, and to introduce me to the functionality of *CognitionMaster* software. I thank **M.Sc. Björn Lindequist**, Scientist, for his contribution in the preparatory stage and his implementation of four object-level gray level co-occurrence matrix texture-based features. I also thank **M.Sc. Daniel Heim**, Scientist, for his optimized version of AdaBoost algorithm as required for my work. I am extremely thankful to **M.Sc. Sebastian Lohmann**, Scientist, to introduce me to certain texture-based approaches such as Gabor filter-banks, and also to a few graph-based state-of-the-art approaches. I am grateful to **VMscope GmbH**, Berlin, Germany, for providing me individual licenses of two of their software tools, namely *Virtual Slide Access SDK 4.0* and *VM Slide Explorer 2010*, in order to access whole slide image data, as required for this work.

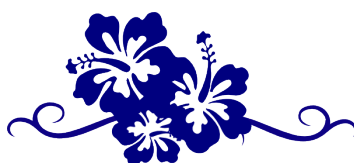
My sincere thanks goes to all the members of Remote Sensing and Computer Vision Group, Department of Computer Engineering and Microelectronics, Technical University, Berlin, Berlin Germany. I would specially like to thank **Dr.-Ing. Ronny Hänsch**, Postdoctoral Scientist, for constructive feedback and comments about my work, especially for the presentations and dissertation. Furthermore, my sincere thanks goes to **Dipl.-Ing. Wolfgang Stinner**, our technical

colleague who provided hardware or software assistance at the office, that helped me to perform the requisite computer-based tasks uninterruptedly in the duration of my studies.

I am indebted and filled with gratitude towards **DAAD (Deutscher Akademischer Austausch Dienst: German Academic Exchange Service)** for giving me this wonderful opportunity to conduct my PhD studies in Berlin, Germany, and providing the requisite financial support for the same.

I wish to thank my family for the inspirational impetus and moral support during the course of work. A big thanks to my parents and sister, for the unconditional affection and care I received from them in my good and bad times. My special thanks to Rahul who encouraged and supported me throughout this journey. I also want to express gratitude towards my friends and relatives who helped directly or indirectly for the completion of this work. Last but not the least, I would like to express my deepest gratitude to the Almighty for showering blessings on me.

Thank you. Danke. Gracias. Grazie. Merci. Shukriya.



Medical Image Analysis of Gastric Cancer in Digital Histopathology: Methods, Applications and Challenges

Abstract: Medical image analysis in digital histopathology is a currently expanding and exciting field of scientific research. In this work, histopathological image analysis is extensively studied and a systematic framework for computer-based analysis in H&E stained whole slide images of *gastric carcinoma* is proposed. The exhaustive experimental study comprises of three fundamental stages, namely, *preparation of materials*, *image pre-analysis* and *analysis of cancer regions*. These stages collectively incorporate the understanding, formulation, implementation and evaluation of suitable image analysis tasks required to achieve the defined research objectives in each stage, for example, registration and annotation transformation between whole slide images, cell nuclei segmentation and classification, multiresolution combination of visual information for segmentation enhancement, appearance-based necrosis detection, and cancer classification based on immunohistochemistry. Computerized applications are also demonstrated as an outcome of the conducted research, including computer-aided diagnosis, content-based image retrieval and automatic determination of tissue composition.

The research focuses on the development of methods for effective representation and subsequent classification of regions of interest in histopathological image datasets. For this purpose, two image analysis routes, namely, traditional route and deep learning route are investigated. The traditional route consists of handcrafted feature extraction to describe textural, color, intensity, morphological and architectural properties of tissues followed by traditional machine learning methods including support vector machines, AdaBoost ensemble learning and random forests. In this domain, graph-based methods are extensively explored due to their ability to suitably represent the spatial arrangements and neighborhood relationships in tissue regions. A novel graph-based image description method called *cell nuclei attributed relational graph* is proposed along with multiple variants, for knowledge description of individual component characteristics, spatial interactions and underlying tissue architecture in histopathological images. In the deep learning route, convolutional neural networks are thoroughly investigated. A self-designed *convolutional neural network architecture* is introduced and analyzed for cancer classification and necrosis detection, also compared with a widely-known framework and ensemble of deep networks.

During the detailed investigation, system performance is rigorously analyzed using different experimental aspects, for instance, algorithm parameters, feature configurations and classification strategies. The entire proposed framework is quantitatively evaluated at each stage using a set of performance metrics. A reasonable performance is achieved using the described methods, comparing favorably, and even outperforming the state-of-the-art techniques in certain occasions. The empirical observations, corresponding conclusions, scientific implications and future directions of the research are thoroughly discussed.

Collectively, the discussed histopathological image analysis methods in H&E stained gastric cancer whole slide images aim to reduce manual preparation efforts, inspection times, and inter-and intra-observer variability. Moreover, the developed methods can potentially improve the current state of technology through applications such as automatic classification, content-based image retrieval, archiving, bio-banking, marker quantification, detecting malignant changes over time, providing second opinions to pathologists, thereby contributing towards diagnosis, prognosis, education and research in biology and medicine.

Keywords: Digital histopathology, gastric carcinoma, whole slide images, cancer classification, necrosis detection, handcrafted feature extraction, attributed relational graphs, traditional machine learning, deep learning, convolutional neural networks, computer-aided diagnosis.

Medizinische Bildanalyse von Magenkarzinom in der Digitalen Histopathologie: Methoden, Anwendungen und Herausforderungen

Zusammenfassung: Medizinische Bildanalyse in der digitalen Histopathologie ist ein derzeit expandierendes und spannendes Forschungsfeld. In dieser Arbeit wird die histopathologische Bildanalyse umfassend untersucht und ein systematischer Ansatz für computerbasierte Analysen H&E-gefärbter virtueller Schnitte von *Magenkarzinom* vorgeschlagen. Die erschöpfende experimentelle Studie besteht aus drei grundlegenden Stadien, nämlich *Vorbereitung von Materialien*, *Bildvoranalyse* und *Analyse von Krebsregionen*. Diese Phasen umfassen das Verständnis, die Formulierung, die Implementierung und die Bewertung von geeigneten Bildanalyseaufgaben, die erforderlich sind, um die definierten Forschungsziele in jedem Stadium zu erreichen, beispielsweise die Registrierung und die Annotationstransformation zwischen virtuellen Schnitten, Zellkernsegmentierung und Klassifikation, Kombination visueller Informationen unterschiedlicher Auflösung zur Segmentierungsverbesserung, aussehensbasierte Nekrose-Erkennung, und Krebs-Klassifizierung auf der Grundlage der Immunhistochemie. Als Ergebnis werden computergestützte Anwendungen der durchgeführten Forschung gezeigt, einschließlich computergestützter Diagnose, inhaltsbasierter Bildwiedergewinnung und automatischer Bestimmung der Gewebezusammensetzung.

Die durchgeführte Forschung betrifft die Entwicklung von Methoden für eine effektive Darstellung und anschließende Klassifizierung in histopathologischen Bilddatensätzen. Dazu werden zwei Bildanalysewege untersucht, nämlich ein traditioneller und ein Deep Learning Ansatz. Der traditionelle Ansatz besteht aus einer Merkmalsextraktion, um Textur, Farbe, Intensität, morphologische und architektonische Eigenschaften von Geweben zu beschreiben, gefolgt von traditionellen maschinellen Lernmethoden, einschließlich der Support Vector Maschinen, AdaBoost Ensemble Learning und Random Forests. Auf diesem Gebiet werden graphenbasierte Verfahren aufgrund ihrer Fähigkeit, die räumlichen Anordnungen und Nachbarschaftsbeziehungen in Gewebebereichen geeignet darzustellen, umfassend erforscht. Ein neues graphisches Bildbeschreibungsverfahren namens *Cell-nuclei Attributed Relational Graph* wird zusammen mit mehreren Varianten zur Wissensbeschreibung einzelner Komponentenmerkmale, räumlicher Interaktionen und zugrunde liegenden Gewebearchitekturen in histopathologischen Bildern vorgeschlagen. Im Deep Learning Bereich werden Convolutional Neural Networks gründlich untersucht. Eine *Convolutional Neural Network Architektur* wird eingeführt und für die Krebsanalyse und Nekrose-Detektion verwendet, und mit einer weithin bekannten Netzwerk Architektur und Ensemble von Deep Netzwerken verglichen.

Während der Detailuntersuchungen wurde die Systemleistung unter unterschiedlichen Aspekten wie Algorithmenparametern, Merkmalskonfigurationen und Klassifizierungsstrategien analysiert. Der vorgeschlagene Ansatz wird in jedem Stadium quantitativ unter Verwendung eines Satzes von Leistungsmetriken bewertet. Mit den entwickelten Methoden wird eine Leistung erzielt, die den Stand der Technik erreicht oder besser ist. Die empirischen Beobachtungen, die entsprechenden Schlussfolgerungen, die wissenschaftlichen Implikationen und die zukünftigen Forschungsrichtungen werden eingehend erörtert.

Zusammenfassend kann festgestellt werden, dass die diskutierten histopathologischen Bildanalyseverfahren in H&E-gefärbten virtuellen Schnitten von Magenkrebs manuelle Vorbereitung, Prüfzeiten und Ergebnisvariabilität reduzieren. Darüber hinaus können die entwickelten Methoden den aktuellen Stand der Technik durch Anwendungen, wie beispielsweise automatische Klassifizierung, inhaltsbasierte Bildsuche, Archivierung, Bio-Banking und Marker-Quantifizierung verbessern. Bei der Analyse, bösartiger Veränderungen im Zeitverlauf bieten sie eine zweite Meinung zu der der Pathologen, und dadurch einen Beitrag zur Diagnose und Prognose in Biologie und Medizin.

Schlüsselwörter: Digitale Histopathologie, Magenkarzinom, Virtuelle Schnitte, Krebs-Klassifikation, Nekrose-Erkennung, Merkmalsextraktion, Attributierte Relationale Graphen, Maschinelles Lernen, Deep Learning, Convolutional Neural Networks, computergestützte Diagnose.

Contents

List of Figures	xvii
List of Tables	xxvii
List of Abbreviations	xxxi
1 Introduction	1
1.1 Background	1
1.2 Study Objectives	2
1.3 Motivations	3
1.4 Contributions	5
1.4.1 Scientific Contributions	5
1.4.2 Application Areas	6
1.5 Challenges in Study	7
1.6 Organization of Thesis	8
2 Theoretical Background	9
2.1 Introduction	9
2.2 Domain Description	10
2.2.1 Overview of Digital Histopathology	10
2.2.2 Introduction to Gastric Histopathology	12
2.3 Feature Extraction in Digital Histopathology	14
2.3.1 Low-level (Pixel-based) Methods	15
2.3.2 Object-level Methods	21
2.3.3 High-level (Architectural) Methods	22
2.4 Machine Learning in Digital Histopathology	33
2.4.1 Support Vector Machines	34
2.4.2 AdaBoost Ensemble Learning	34
2.4.3 Random Forests	34
2.4.4 Deep Convolutional Neural Networks	35
2.5 Summary	36

3	Related Work in Digital Histopathology	37
3.1	Introduction	37
3.2	Feature Extraction in Digital Histopathology	38
3.2.1	Low-level (Pixel-based) Methods	38
3.2.2	Object-level Methods	39
3.2.3	High-level (Architectural) Methods	40
3.3	Machine Learning in Digital Histopathology	43
3.3.1	Support Vector Machines	43
3.3.2	AdaBoost Ensemble Learning	44
3.3.3	Random Forests	44
3.3.4	Deep Convolutional Neural Networks	44
3.4	Image Analysis in Gastric Cancer	45
3.5	Summary	46
4	Overview of Research Methodology	47
4.1	Introduction	47
4.2	Stage 1: Preparation of Materials	47
4.3	Stage 2: Image Pre-analysis	49
4.4	Stage 3: Analysis of Cancer Regions	49
4.5	Summary	50
5	Stage 1: Preparation of Materials	51
5.1	Introduction	51
5.2	Whole Slide Image Acquisition	52
5.2.1	Specimen Properties	52
5.2.2	Scanner Details	52
5.3	Annotations: HER2 whole slide images	54
5.3.1	Labeling Process	54
5.3.2	Semi-automatic WSI Registration	55
5.3.3	Annotation Transformation	57
5.4	Initial Working Datasets	59
5.4.1	Annotations for Cell Nuclei Segmentation Evaluation	60
5.4.2	Annotations for Cell Nuclei Classification	60
5.5	Annotations for Necrosis Detection	64
5.5.1	Datasets for SVM-based Method	66
5.5.2	Datasets for Deep Learning Methods	66
5.6	Summary	66
6	Stage 2: Image Pre-analysis	69
6.1	Introduction	69
6.2	Necrosis Detection	70

6.2.1	Background	70
6.2.2	Motivation	71
6.2.3	Feature Extraction	72
6.2.4	Machine Learning	73
6.3	Cell Nuclei Segmentation and Evaluation	79
6.3.1	Segmentation Algorithm	79
6.3.2	Segmentation Evaluation	82
6.4	Multiresolution Segmentation Enhancement	84
6.4.1	Background	84
6.4.2	Cell Nuclei Classification	85
6.4.3	Multiresolution Combination	93
6.5	Applications	95
6.5.1	Automatic Necrosis Detection	95
6.5.2	Automatic Determination of Histological Tissue Composition	96
6.6	Summary	98
7	Stage 3: Analysis of Cancer Regions	99
7.1	Introduction	100
7.2	Ground Truth Dataset Generation	101
7.2.1	Datasets for Traditional Machine Learning Methods	101
7.2.2	Datasets for Deep Learning Methods	102
7.3	Image Description Methods	105
7.3.1	Low-level State-of-the-art Features	105
7.3.2	High-level State-of-the-art Features	107
7.3.3	High-level Handcrafted Features	108
7.3.4	Combination of Features	121
7.4	Machine Learning Methods	123
7.4.1	Traditional Methods	123
7.4.2	Deep Convolutional Neural Networks	125
7.4.3	Classification Strategies	135
7.5	Applications	136
7.5.1	Computer-aided Diagnosis: Cancer Classification based on Immunohistochemistry	136
7.5.2	Content-based Image Retrieval	137
7.6	Summary	141
8	Performance Evaluation	143
8.1	Introduction	143
8.1.1	Evaluation Methods	144
8.1.2	Performance Metrics	146
8.2	Comparative Evaluation: Necrosis Detection	149

8.3	Comparative Evaluation: Cell Nuclei Classification	151
8.4	Comparative Evaluation: Analysis of Cancer Regions	153
8.4.1	Concept of Comparative Evaluation	153
8.4.2	Low-level State-of-the-art Features	157
8.4.3	High-level State-of-the-art Features	163
8.4.4	High-level Handcrafted Features	164
8.4.5	Combination of Features	173
8.4.6	Deep Learning Methods	175
8.4.7	Overall Observations	177
8.5	Summary	182
9	Statistical Analysis of Stains: Observations and Challenges	183
9.1	Introduction	183
9.2	Visualizing Image Statistics: HER2 Stain	184
9.2.1	Visual Inspection	184
9.2.2	Feature Heatmaps	185
9.3	Visualizing Image Statistics: H&E Stain	186
9.3.1	Visual Inspection	186
9.3.2	Feature Heatmaps	187
9.4	Summary	189
10	Summary	191
10.1	Conclusions	191
10.2	Future Work and Recommendations	195
10.2.1	Future Work	195
10.2.2	Recommendations	198
Appendix A	Implementation Details	201
A.1	Hardware Specifications	201
A.2	Operating Systems	202
A.3	Programming Languages	202
A.4	Supporting Software	203
A.5	Computational Requirements	205
A.6	Examples of Metadata Screenshots	212
Appendix B	Detailed Experimental Results	215
B.1	Confusion Matrices	215
B.1.1	Comparative Evaluation: Necrosis Detection	215
B.1.2	Comparative Evaluation: Cell Nuclei Classification	216
B.1.3	Comparative Evaluation: Analysis of Cancer Regions	218
B.2	Box and Whisker Diagrams	224

B.2.1	Comparative Evaluation: Necrosis Detection	224
B.2.2	Comparative Evaluation: Cell Nuclei Classification	225
B.2.3	Comparative Evaluation: Analysis of Cancer Regions	226
Bibliography		238

List of Figures

2.1	A schematic overview of tissue specimen preparation	10
2.2	Pipeline of virtual microscopy. Adapted from [Saeger 2009].	11
2.3	Pictorial representation of the stomach. Source: [Wikipedia 2017]	13
2.4	Example of Gabor filter-bank with varying frequencies and orientations	17
2.5	The LBP code generated from an example image	18
2.6	The RFS filter-bank to obtain MR8 filter responses in the Varma-Zisserman approach. Constructed using the open-source MATLAB filter implementation in [Varma 2002]	19
2.7	Example showing (a) an image (b) gray model and image gray histogram (c) RGB color model and image component histograms (d) HSV color model and image component histograms (generated in MATLAB)	20
2.8	Voronoi Diagram of a set of random points	23
2.9	Delaunay Triangulation corresponding to the Voronoi diagram in Figure 2.8 showing empty circle and duality properties	24
2.10	Gabriel Graph corresponding to Delaunay graph in Figure 2.9 showing empty circle property	24
2.11	Relative Neighborhood Graph corresponding to Delaunay graph in Figure 2.9 showing empty lune property	25
2.12	Euclidean minimum spanning tree corresponding to Delaunay graph in Figure 2.9	25
2.13	Nearest Neighbor Graph corresponding to Delaunay graph in Figure 2.9	26
2.14	A Voronoi diagram and its corresponding Ulam Tree	27
2.15	β neighborhoods (shaded regions) with different values of β .(a) Lune-based and circle-based $\beta = 0.5$ (b) Lune-based and circle-based $\beta = 1$ (c) Lune-based $\beta = 2$ (d) Circle-based $\beta = 2$	28
2.16	Johnson-Mehl tessellation for a set of random points (a) growth of particles at $t = t_1$ (b) growth of particles at $t = t_1 + 50$	29
2.17	O’Callaghan neighborhood graph	29

2.18	Representative cell graphs for small regions of H&E stained images (a) Simple cell graph with $D = 200$ for gastric tissue region (b) Probabilistic cell graph with $\alpha = 2$ for gastric tissue region (c) Hierarchical cell graph for breast tissue where each vertex represents a cell cluster with a corresponding simple cell graph built on the constituent cells (denoted by rectangle)	30
2.19	Attributed Relational Graph	31
4.1	Schematic overview of the experimental pipeline	48
5.1	Examples of corresponding sections in (a) HER2 and (b) H&E stains.	53
5.2	Screenshot of virtual microscopy program used by pathologists to create annotations in HER2 WSI based on immunohistochemical response	54
5.3	Screenshot of ViSPEe program for semi-automatic registration in one HER2 and H&E WSI pair	57
5.4	Mapping by linear triangular interpolation of (a) triangle $P_1P_2P_3$ in the x, y plane to (b) triangle $Q_1Q_2Q_3$ in the u, v plane.	58
5.5	Example of HER2 and H&E WSI pair containing original and resulting pathologists' annotations after semi-automatic WSI registration and annotation transformation procedure	59
5.6	Screenshot of ROIManager program with cell nuclei annotations (yellow '+' marks) made in an image tile at the highest resolution	61
5.7	Distribution of point annotations among the three types of regions (number and percentage of total annotations)	61
5.8	Defined cell nuclei classes (a) Epithelial cell (b) Leukocyte (c) Fibrocyte (d) Conglomerate (e) Fragment (f) Other cell (including blood cell in vessel) (g) Artefact.	62
5.9	Screenshot of the Object-Manager program in review mode showing labeled contours in an image tile at highest resolution	63
5.10	Distribution of cell nuclei annotations for seven classes in each of the three types of regions and total annotations at (a) $30\times$ objective magnification (b) $40\times$ objective magnification.	63
5.11	Screenshots of (a) VM Slide Explorer program and an example WSI showing rectangular regions of interest in the three types of malignancy groups (HER2 positive: yellow, HER2 negative: blue and non-tumor: green), and necrotic polygons marked inside them (red) at $0.65\times$ objective magnification (b) Object-Manager plugin and an example of image tile from non-tumor region with square annotations created at the smallest tile size, where necrotic tiles are marked with blue squares and non-necrotic in green squares at $40\times$ objective magnification.	65

5.12 (a) ROC characteristics for combined datasets using the described SVM-based method showing comparative performance between different image tile sizes (b) Distribution of labeled image tiles for necrotic and non-necrotic tissue in three types of malignancy regions and total image tiles in the final dataset for SVM-based method	67
5.13 Distribution of labeled image tiles for necrotic and non-necrotic tissue in HER2 negative tumor and non-tumor types of malignancy regions and total image tiles in the final dataset for deep learning methods	67
6.1 Schematic representation of the process of necrosis in a cell	71
6.2 Image tile pair and texture feature pictorial representation using (a) GLCM statistics (b) Gabor filter kernels.	73
6.3 Parameter selection using grid search for necrosis detection using SVM	75
6.4 Selection of discriminative thresholds for SVM classification	76
6.5 Proposed CNN architecture for necrosis detection	77
6.6 Examples of learning curves of random training rounds for necrosis detection using (a) AlexNet CNN framework (b) proposed CNN architecture.	78
6.7 Cell nuclei segmentation result example at 25× objective magnification (a) Original image (b) Processed image showing resulting cell nuclei segments. . .	81
6.8 Example of cell nuclei segmentation results of an image at different objective magnifications for visual inspection (a) 10× (b) 15× (c) 20× (d) 25× (e) 30× (f) 40×.	83
6.9 Cell nuclei segmentation performance at individual magnifications	83
6.10 Morphological feature definitions of a contour (a) Axes projections (b) Freeman code. Adapted from [Hufnagl 1984].	86
6.11 Parameter selection using grid search for multi-class SVM classification of cell nuclei	90
6.12 Parameter selection for multi-class AdaBoost classification of cell nuclei	91
6.13 Parameter selection for multi-class random forest classification of cell nuclei . .	93
6.14 Flowchart for multiresolution combination of cell nuclei segmentation	94
6.15 Cell nuclei segmentation performance after multiresolution combination	95
6.16 Results for application of automatic necrosis detection on gastric cancer WSI datasets in (a) HER2 positive tumor (b) HER2 negative tumor (c) Non-tumor. .	96
6.17 Example results of prototype application for automatic determination of tissue composition in (a) HER2 positive tumor (b) HER2 negative tumor (c) Non-tumor.	97
7.1 Representative examples of image tiles from the same WSI of the defined malignancy levels at the highest resolution for (a) HER2 positive tumor (b) HER2 negative tumor (c) Non-tumor	101

7.2	Distribution of labeled image tiles among the patients in HER2 positive tumor, HER2 negative tumor and non-tumor types of malignancy regions and total image tiles for analysis of cancer regions using traditional methods (a) at 1024×1024 pixel size (b) at 512×512 pixel size.	103
7.3	Example of H&E WSI with (a) a few annotations of HER2 positive tumor marked by expert pathologists in the corresponding HER2 WSI (b) a low magnified ($5\times$) region of agreement of most pathologists (c) example images after data augmentation at highest magnification ($40\times$).	104
7.4	Distribution of labeled image tiles among the patients in HER2 positive tumor, HER2 negative tumor and non-tumor types of malignancy regions and total image tiles for analysis of cancer regions using deep learning methods	104
7.5	Example of a HER2 negative tumor image tile, cell nuclei segments after segmentation procedure and the resulting Voronoi diagram and Delaunay triangulation for high-level state-of-the-art feature extraction	107
7.6	Schematic diagram of cell nuclei ARG	111
7.7	Uniform grid assumption for cell nuclei ARG with adaptive r	113
7.8	Example of an image tile from a larger ROI, its segmentation and representations of eight cell nuclei ARG variants (the corresponding $ncARG$ and $ncARG_{v+}$ have same appearance but difference in the corresponding vertex attributes) . . .	118
7.9	Typical example of malignancy regions overlay with their $gARG[r_F]$, $r = 50$ showing visible characteristics for (a) HER2 positive tumor (b) HER2 negative tumor (c) Non-tumor	119
7.10	AlexNet CNN architecture implemented for cancer classification in gastric cancer WSI datasets. Adapted from [Krizhevsky 2012]	126
7.11	Proposed CNN architecture for cancer classification in gastric cancer WSI . . .	128
7.12	Cascade operations in the k^{th} convolutional layer	129
7.13	Classification strategies for cancer classification (a) Single-stage classification (b) Hierarchical classification.	135
7.14	Illustrative examples of classification result (a) Original H&E WSI with pathologists' annotations for cancer classification based on IHC, and corresponding probability maps using proposed CNN architecture for (a) HER2 positive tumor (b) HER2 negative tumor (c) non-tumor at a low magnification ($0.3\times$).	137
7.15	Schematic diagram of a CBIR system	138
7.16	Illustrative experimental results of prototype CBIR application as P-R curves for categories (a) Non-tumor (b) Tumors (c) HER2 positive tumor (d) HER2 negative tumor and (e) overall result using the Manhattan distance metric. . . .	140
8.1	Schematic example of box and whisker plot	149
8.2	Overall mean and standard error of classification accuracy for necrosis detection experiments using (a) k-fold stratified shuffled split cross validation (b) Average BCA for all methods over both cross validations.	149

8.3	Overall mean and standard error of classification accuracy for cell nuclei classification experiments using (a) k-fold stratified shuffled split and (b) leave-a-sample-out cross validations (c) Average BCA for all methods over both cross validations.	151
8.4	Overall mean and standard error of classification accuracy for cancer classification experiments for GLCM statistics using (a) k-fold stratified shuffled split and (b) leave-a-patient-out cross validation (c) Average BCA for all methods over both cross validations.	157
8.5	Overall mean and standard error of classification accuracy for cancer classification experiments for Gabor filter-bank responses using (a) k-fold stratified shuffled split and (b) leave-a-patient-out cross validation (c) Average BCA for all methods over both cross validations.	158
8.6	Overall mean and standard error of classification accuracy for cancer classification experiments for LBP histograms using (a) k-fold stratified shuffled split and (b) leave-a-patient-out cross validation (c) Average BCA for all methods over both cross validations.	159
8.7	Overall mean and standard error of classification accuracy for cancer classification experiments for Varma-Zisserman textons using (a) k-fold stratified shuffled split and (b) leave-a-patient-out cross validation (c) Average BCA for all methods over both cross validations.	160
8.8	Overall mean and standard error of classification accuracy for cancer classification experiments for gray histograms using (a) k-fold stratified shuffled split and (b) leave-a-patient-out cross validation (c) Average BCA for all methods over both cross validations.	160
8.9	Overall mean and standard error of classification accuracy for cancer classification experiments for HSV histograms using (a) k-fold stratified shuffled split and (b) leave-a-patient-out cross validation (c) Average BCA for all methods over both cross validations.	161
8.10	Overall mean and standard error of classification accuracy of cancer classification experiments for RGB histograms using (a) k-fold stratified shuffled split and (b) leave-a-patient-out cross validation (c) Average BCA for all methods over both cross validations.	162
8.11	Overall mean and standard error of classification accuracy for cancer classification experiments for other color-based measurements using (a) k-fold stratified shuffled split and (b) leave-a-patient-out cross validation (c) Average BCA for all methods over both cross validations.	163
8.12	Overall mean and standard error of classification accuracy for malignancy region classification experiments for Voronoi-Delaunay method using (a) k-fold stratified shuffled split and (b) leave-a-patient-out cross validation (c) Average BCA for all methods over both cross validations.	164

8.13 Overall mean and standard error of classification accuracy for region classification experiments for only vertex attributes using (a) k-fold stratified shuffled split and (b) leave-a-patient-out cross validation (c) Average BCA for all methods over both cross validations.	165
8.14 Overall mean and standard error of classification accuracy for region classification experiments for only vertex identities using (a) k-fold stratified shuffled split and (b) leave-a-patient-out cross validation (c) Average BCA for all methods over both cross validations.	166
8.15 Overall mean and standard error of classification accuracy for malignancy region classification experiments for $gARG[r_F]$ using (a) k-fold stratified shuffled split and (b) leave-a-patient-out cross validation (c) Average BCA for all methods over both cross validations.	167
8.16 Overall mean and standard error of classification accuracy for malignancy region classification experiments for $gARG[r_A]$ using (a) k-fold stratified shuffled split and (b) leave-a-patient-out cross validation (c) Average BCA for all methods over both cross validations.	168
8.17 Overall mean and standard error of classification accuracy for malignancy region classification experiments for $nsARG[r_F]$ using (a) k-fold stratified shuffled split and (b) leave-a-patient-out cross validation (c) Average BCA for all methods over both cross validations.	168
8.18 Overall mean and standard error of classification accuracy for malignancy region classification experiments for $nsARG[r_A]$ using (a) k-fold stratified shuffled split and (b) leave-a-patient-out cross validation (c) Average BCA for all methods over both cross validations.	169
8.19 Overall mean and standard error of classification accuracy for region classification experiments for $ncARG[r_F]$ using (a) k-fold stratified shuffled split and (b) leave-a-patient-out cross validation (c) Average BCA for all methods over both cross validations.	170
8.20 Overall mean and standard error of classification accuracy for region classification experiments for $ncARG[r_A]$ using (a) k-fold stratified shuffled split and (b) leave-a-patient-out cross validation (c) Average BCA for all methods over both cross validations.	171
8.21 Overall mean and standard error of classification accuracy for region classification experiments for $ncARG_{v+}[r_F]$ using (a) k-fold stratified shuffled split and (b) leave-a-patient-out cross validation (c) Average BCA for all methods over both cross validations.	172
8.22 Overall mean and standard error of classification accuracy for region classification experiments for $ncARG_{v+}[r_A]$ using (a) k-fold stratified shuffled split and (b) leave-a-patient-out cross validation (c) Average BCA for all methods over both cross validations.	173

8.23	Overall mean and standard error of classification accuracy for cancer classification experiments for selected subset of hybrid low-level features using (a) k-fold stratified shuffled split and (b) leave-a-patient-out cross validation (c) Average BCA for all methods over both cross validations.	174
8.24	Overall mean and standard error of classification accuracy for cancer classification experiments for selected subset of hybrid low-level and high-level features using (a) k-fold stratified shuffled split and (b) leave-a-patient-out cross validation (c) Average BCA for all methods over both cross validations.	175
8.25	Overall mean and standard error of classification accuracy for region classification experiments for deep learning methods using (a) k-fold stratified shuffled split and (b) leave-a-patient-out cross validation (c) Average BCA for all methods over both cross validations.	176
8.26	Overall mean and standard error of per-class classification accuracy for all the methods over both cross validations	178
8.27	Magnified line diagram showing average BCA for the different feature groups and corresponding traditional machine learning strategies	179
8.28	Overall mean and standard error of balanced classification accuracy for traditional machine learning (using hybrid features) and deep learning methods . . .	181
9.1	Representative examples of images in three malignancy levels using HER2 stain (a) HER2 positive tumor (b) HER2 negative tumor (c) Non-tumor.	185
9.2	Heatmaps of the extracted feature vectors showing image statistics in the HER2 stain (a) GLCM statistics (b) LBP histograms (c) HSV histograms (d) RGB histograms.	185
9.3	Representative examples of images in three malignancy levels using H&E stain (a) HER2 positive tumor (b) HER2 negative tumor (c) Non-tumor.	187
9.4	Heatmaps of the extracted feature vectors showing image statistics in the HER2 stain (a) GLCM statistics (b) LBP histograms (c) HSV histograms (d) RGB histograms (e) Varma-Zisserman textons (f) Voronoi diagram and Delaunay triangulations (g) $nsARG[r_F]$ features (h) selected subset of hybrid features. .	188
A.1	Average execution time requirements of graph-based constructions for Voronoi-Delaunay and $gARG[r_F]$ graphs and variations with the number of graph vertices	211
A.2	Average execution time requirements (log-linear plot) of traditional machine learning methods and variations with number of features during training	211
A.3	Representative examples of metadata screenshots generated during (a) semi-automatic WSI registration (XML file) (b) annotations for cell nuclei segmentation evaluation (ROI file) (c) annotations and features for cell nuclei classification (XML learning sample file)	212
B.1	Color scale for values in confusion matrices	215

B.2	Box and whisker diagrams of average per-class classification accuracy for necrosis detection experiments using k-fold stratified shuffled split cross validation for (a) necrosis (b) non-necrosis.	225
B.3	Box and whisker diagrams of average per-class classification accuracy for cell nuclei classification experiments using k-fold stratified shuffled split cross validation for a) epithelial cell (b) leukocyte (c) fibrocyte (or border cell) (d) conglomerate (e) fragment (f) other cell (including blood cell in vessel) (g) artefact.	225
B.4	Box and whisker diagrams of average per-class classification accuracy of GLCM statistics for (a) HER2 positive tumor (b) HER2 negative tumor (c) Non-tumor using k-fold stratified shuffled split and (d),(e),(f) using leave-a-patient-out cross validations.	226
B.5	Box and whisker diagrams of average per-class classification accuracy of Gabor filter-bank responses for (a) HER2 positive tumor (b) HER2 negative tumor (c) Non-tumor using k-fold stratified shuffled split and (d),(e),(f) using leave-a-patient-out cross validations.	227
B.6	Box and whisker diagrams of average per-class classification accuracy of local binary patterns for (a) HER2 positive tumor (b) HER2 negative tumor (c) Non-tumor using k-fold stratified shuffled split and (d),(e),(f) using leave-a-patient-out cross validations.	227
B.7	Box and whisker diagrams of average per-class classification accuracy of Varma-Zissserman textons for (a) HER2 positive tumor (b) HER2 negative tumor (c) Non-tumor using k-fold stratified shuffled split and (d),(e),(f) using leave-a-patient-out cross validations.	228
B.8	Box and whisker diagrams of average per-class classification accuracy of gray histograms for (a) HER2 positive tumor (b) HER2 negative tumor (c) Non-tumor using k-fold stratified shuffled split and (d),(e),(f) using leave-a-patient-out cross validations.	228
B.9	Box and whisker diagrams of average per-class classification accuracy of HSV histograms for (a) HER2 positive tumor (b) HER2 negative tumor (c) Non-tumor using k-fold stratified shuffled split and (d),(e),(f) using leave-a-patient-out cross validations.	229
B.10	Box and whisker diagrams of average per-class classification accuracy of RGB histograms for (a) HER2 positive tumor (b) HER2 negative tumor (c) Non-tumor using k-fold stratified shuffled split and (d),(e),(f) using leave-a-patient-out cross validations.	229
B.11	Box and whisker diagrams of average per-class classification accuracy of other color-based measurements for (a) HER2 positive tumor (b) HER2 negative tumor (c) Non-tumor using k-fold stratified shuffled split and (d),(e),(f) using leave-a-patient-out cross validations.	230

B.12	Box and whisker diagrams of average per-class classification accuracy for the Voronoi-Delaunay method for (a) HER2 positive tumor (b) HER2 negative tumor (c) Non-tumor using k-fold stratified shuffled split and (d),(e),(f) using leave-a-patient-out cross validations.	230
B.13	Box and whisker diagrams of average per-class classification accuracy for only vertex attributes of cell nuclei ARG for (a) HER2 positive tumor (b) HER2 negative tumor (c) Non-tumor using k-fold stratified shuffled split and (d),(e),(f) using leave-a-patient-out cross validations.	231
B.14	Box and whisker diagrams of average per-class classification accuracy for only vertex identities (or tissue composition) of cell nuclei ARG for (a) HER2 positive tumor (b) HER2 negative tumor (c) Non-tumor using k-fold stratified shuffled split and (d),(e),(f) using leave-a-patient-out cross validations.	231
B.15	Box and whisker diagrams of average per-class classification accuracy for hand-crafted $gARG[r_F]$ method for (a) HER2 positive tumor (b) HER2 negative tumor (c) Non-tumor using k-fold stratified shuffled split and (d),(e),(f) using leave-a-patient-out cross validations.	232
B.16	Box and whisker diagrams of average per-class classification accuracy for hand-crafted $gARG[r_A]$ method for (a) HER2 positive tumor (b) HER2 negative tumor (c) Non-tumor using k-fold stratified shuffled split and (d),(e),(f) using leave-a-patient-out cross validations.	232
B.17	Box and whisker diagrams of average per-class classification accuracy for hand-crafted $nsARG[r_F]$ method for (a) HER2 positive tumor (b) HER2 negative tumor (c) Non-tumor using k-fold stratified shuffled split and (d),(e),(f) using leave-a-patient-out cross validations.	233
B.18	Box and whisker diagrams of average per-class classification accuracy for hand-crafted $nsARG[r_A]$ method for (a) HER2 positive tumor (b) HER2 negative tumor (c) Non-tumor using k-fold stratified shuffled split and (d),(e),(f) using leave-a-patient-out cross validations.	233
B.19	Box and whisker diagrams of average per-class classification accuracy for hand-crafted $ncARG[r_F]$ method for (a) HER2 positive tumor (b) HER2 negative tumor (c) Non-tumor using k-fold stratified shuffled split and (d),(e),(f) using leave-a-patient-out cross validations.	234
B.20	Box and whisker diagrams of average per-class classification accuracy for hand-crafted $ncARG[r_A]$ method for (a) HER2 positive tumor (b) HER2 negative tumor (c) Non-tumor using k-fold stratified shuffled split and (d),(e),(f) using leave-a-patient-out cross validations.	234
B.21	Box and whisker diagrams of average per-class classification accuracy for hand-crafted $ncARG_{v+}[r_F]$ method for (a) HER2 positive tumor (b) HER2 negative tumor (c) Non-tumor using k-fold stratified shuffled split and (d),(e),(f) using leave-a-patient-out cross validations.	235

B.22	Box and whisker diagrams of average per-class classification accuracy for hand-crafted $ncARG_{v+}[r_A]$ method for (a) HER2 positive tumor (b) HER2 negative tumor (c) Non-tumor using k-fold stratified shuffled split and (d),(e),(f) using leave-a-patient-out cross validations.	235
B.23	Box and whisker diagrams of average per-class classification accuracy of selected subset of hybrid low-level features for (a) HER2 positive tumor (b) HER2 negative tumor (c) Non-tumor using k-fold stratified shuffled split and (d),(e),(f) using leave-a-patient-out cross validations.	236
B.24	Box and whisker diagrams of average per-class classification accuracy of selected subset of hybrid low-level and high-level features for (a) HER2 positive tumor (b) HER2 negative tumor (c) Non-tumor using k-fold stratified shuffled split and (d),(e),(f) using leave-a-patient-out cross validations.	236
B.25	Box and whisker diagrams of average per-class classification accuracy for deep learning methods for (a) HER2 positive tumor (b) HER2 negative tumor (c) Non-tumor using k-fold stratified shuffled split and (d),(e),(f) using leave-a-patient-out cross validations.	237

List of Tables

2.1	GLCM statistical descriptors	16
2.2	Summary of popular algorithms for constructing Voronoi diagram, Delaunay triangulation and its subgraphs	26
5.1	Distribution of number of pathologists' annotations marked in HER2 WSI on the basis of immunohistochemical response (P: pathologist, S: slide, H+: number of HER2+ tumor annotations, H-: number of HER2- tumor annotations)	56
5.2	Number of image tiles of different sizes in the three datasets for necrosis detection using SVM-based method	65
6.1	Parameter selection in cell nuclei segmentation algorithm	81
6.2	List of object-level features computed for cell nuclei classification	88
6.3	Relevance scores assigned to classified cell nuclei objects	94
7.1	Cancer interpretation based on graphs: summary of visible characteristics and corresponding cell nuclei ARG properties	117
7.2	Results for feature selection based on correlation analysis after combination of features	122
7.3	Details of most successful empirically evaluated CNN architectures for cancer classification on gastric cancer representative datasets	127
7.4	Examples of two-dimensional outputs in the three convolutional layers of the proposed CNN architecture	131
7.5	Summary of parameter selection for the proposed CNN architecture	132
7.6	Per-class experimental results showing performance measures for the prototype CBIR application	140
8.1	Elements in the confusion matrix M	147
8.2	Scheme of comparative evaluation of studied methods for analysis of cancer regions	154
A.1	Important program modules, supporting software and computational requirements of the proposed framework	205

B.1	Average accuracy confusion matrix using k-fold stratified shuffled split cross validation for (a) SVM with discriminative thresholds (b) AlexNet CNN framework (c) Proposed CNN architecture (d) Ensemble of CNNs	216
B.2	Average accuracy confusion matrix for SVM classification method using k-fold stratified shuffled split cross validation	216
B.3	Average accuracy confusion matrix for SVM classification method using leave-a-sample-out cross validation	217
B.4	Average accuracy confusion matrix for AdaBoost ensemble classification using k-fold stratified shuffled split cross validation	217
B.5	Average accuracy confusion matrix for AdaBoost ensemble classification using leave-a-sample-out cross validation	217
B.6	Average accuracy confusion matrix for random forest classification using k-fold stratified shuffled split cross validation	217
B.7	Average accuracy confusion matrix for random forest classification using leave-a-sample-out cross validation	218
B.8	Average accuracy confusion matrix for GLCM statistics with SVM hierarchical classification (each stage) using (a) k-fold stratified shuffled split (b) leave-a-patient-out cross validation	218
B.9	Average accuracy confusion matrix for Gabor filter-bank responses with SVM hierarchical classification (each stage) using (a) k-fold stratified shuffled split (b) leave-a-patient-out cross validation	219
B.10	Average accuracy confusion matrix for local binary patterns with SVM single-stage classification using (a) k-fold stratified shuffled split (b) leave-a-patient-out cross validation	219
B.11	Average accuracy confusion matrix for Varma-Zisserman textons with SVM single-stage classification (a) k-fold stratified shuffled split (b) leave-a-patient-out cross validation	219
B.12	Average accuracy confusion matrix for gray histograms with random forests single-stage classification (a) k-fold stratified shuffled split (b) leave-a-patient-out cross validation	219
B.13	Average accuracy confusion matrix for HSV histograms with AdaBoost single-stage classification using (a) k-fold stratified shuffled split (b) leave-a-patient-out cross validation	220
B.14	Average accuracy confusion matrix for RGB histograms with SVM hierarchical classification using (a) k-fold stratified shuffled split (b) leave-a-patient-out cross validation	220
B.15	Average accuracy confusion matrix for other color-based measurements with random forests single-stage classification using (a) k-fold stratified shuffled split (b) leave-a-patient-out cross validation	220

B.16	Average accuracy confusion matrix for Voronoi-Delaunay method with random forest single-stage classification using (a) k-fold stratified shuffled split (b) leave-a-patient-out cross validation	220
B.17	Average accuracy confusion matrix for only vertex attributes (object-level features) of cell nuclei ARG with random forest hierarchical classification (each stage) using (a) k-fold stratified shuffled split (b) leave-a-patient-out cross validation	221
B.18	Average accuracy confusion matrix for only vertex identities (tissue composition) of the cell nuclei ARG with SVM single-stage classification using (a) k-fold stratified shuffled split (b) leave-a-patient-out cross validation	221
B.19	Average accuracy confusion matrix for handcrafted $gARG[r_F]$ method with SVM single-stage classification using (a) k-fold stratified shuffled split (b) leave-a-patient-out cross validation	221
B.20	Average accuracy confusion matrix for handcrafted $gARG[r_A]$ method with SVM single-stage classification using (a) k-fold stratified shuffled split (b) leave-a-patient-out cross validation	221
B.21	Average accuracy confusion matrix for handcrafted $nsARG[r_F]$ method with SVM single-stage classification using (a) k-fold stratified shuffled split (b) leave-a-patient-out cross validation	222
B.22	Average accuracy confusion matrix for handcrafted $nsARG[r_A]$ method with SVM hierarchical classification (each stage) using (a) k-fold stratified shuffled split (b) leave-a-patient-out cross validation	222
B.23	Average accuracy confusion matrix for handcrafted $ncARG[r_F]$ method with SVM single-stage classification using (a) k-fold stratified shuffled split (b) leave-a-patient-out cross validation	222
B.24	Average accuracy confusion matrix for handcrafted $ncARG[r_A]$ method with random forest hierarchical classification using (a) k-fold stratified shuffled split (b) leave-a-patient-out cross validation	222
B.25	Average accuracy confusion matrix for handcrafted $ncARG_{v+}[r_F]$ method with SVM single-stage classification using (a) k-fold stratified shuffled split (b) leave-a-patient-out cross validation	223
B.26	Average accuracy confusion matrix for handcrafted $ncARG_{v+}[r_A]$ method with SVM single-stage classification using (a) k-fold stratified shuffled split (b) leave-a-patient-out cross validation	223
B.27	Average accuracy confusion matrix for selected subset of hybrid low-level features with SVM hierarchical classification using (a) k-fold stratified shuffled split (b) leave-a-patient-out cross validation	223
B.28	Average accuracy confusion matrix for selected subset of hybrid low-level and high-level features with SVM single-stage classification using (a) k-fold stratified shuffled split (b) leave-a-patient-out cross validation	223

B.29 Average accuracy confusion matrix for AlexNet CNN framework in single-stage classification using (a) k-fold stratified shuffled split (b) leave-a-patient-out cross validation	224
B.30 Average accuracy confusion matrix for proposed CNN architecture in single-stage classification using (a) k-fold stratified shuffled split (b) leave-a-patient-out cross validation	224
B.31 Average accuracy confusion matrix for ensemble of CNNs in single-stage classification using (a) k-fold stratified shuffled split (b) leave-a-patient-out cross validation	224

List of Abbreviations

Acronyms / Abbreviations

AB-HC	AdaBoost Hierarchical Classification
AB-SC	AdaBoost Single-stage Classification
AdaBoost	Adaptive Boosting
ARG	Attributed Relational Graph
BCA	Balanced Classification Accuracy
CAD	Computer-aided Diagnosis
CBIR	Content-based Image Retrieval
CNN	Convolutional Neural Network
CPU	Central Processing Unit
ECM	Extra-cellular Matrix
gARG	Generic Attributed Relational Graph
gARG[r _A]	Generic Attributed Relational Graph and Adaptive r
gARG[r _F]	Generic Attributed Relational Graph and Fixed r
GLCM	Gray Level Co-occurrence Matrix
GPU	Graphics Processing Unit
H&E	Haematoxylin and Eosin
HER2	Human epidermal growth factor receptor 2
HSV	Hue Saturation Value
IDE	Integrated Development Environment

IHC	Immunohistochemistry
LBP	Local Binary Patterns
LMDB	Lightening Memory-mapped Database
LOO	Leave-one-out
MR8	Maximum Response 8
MST	Minimum Spanning Tree
ncARG	Nuclei-composite Attributed Relational Graph
ncARG[r _A]	Nuclei-composite Attributed Relational Graph and Adaptive r
ncARG[r _F]	Nuclei-composite Attributed Relational Graph and Fixed r
ncARG _{v+}	Nuclei-composite Attributed Relational Graph with additional vertex attributes
ncARG _{v+} [r _A]	Nuclei-composite Attributed Relational Graph with additional vertex attributes and Adaptive r
ncARG _{v+} [r _F]	Nuclei-composite Attributed Relational Graph with additional vertex attributes and Fixed r
NCCD	Nomenclature Committee on Cell Death
NNG	Nearest Neighbor Graph
nsARG	Nuclei-specific Attributed Relational Graph
nsARG[r _A]	Nuclei-specific Attributed Relational Graph and Adaptive r
nsARG[r _F]	Nuclei-specific Attributed Relational Graph and Fixed r
OCA	Overall Classification Accuracy
OOB	Out-of-bag
PDF	Probability Density Function
P-R	Precision-Recall
r _A	Adaptive r
RBF	Radial Basis Function
ReLU	Rectified Linear Unit
RF-HC	Random Forest Hierarchical Classification

r_F	Fixed r
RF-SC	Random Forest Single-stage Classification
RFS	Root Filter Set
RGB	Red Blue Green
RNG	Relative Neighborhood Graph
ROC	Receiver Operating Characteristics
ROI	Region of Interest
SDK	Software Development Kit
SVD	Singular Value Decomposition
SVM-HC	Support Vector Machine Hierarchical Classification
SVM-SC	Support Vector Machine Single-stage Classification
SVM	Support Vector Machine
TIFF	Tagged Image File Format
ViSPee	Virtual Slide Processing Environment
VSF	Virtual Slide Format
VZ	Varma Zisserman
WSI	Whole Slide Image
XML	Extensible Markup Language

Introduction

Contents

1.1	Background	1
1.2	Study Objectives	2
1.3	Motivations	3
1.4	Contributions	5
1.4.1	Scientific Contributions	5
1.4.2	Application Areas	6
1.5	Challenges in Study	7
1.6	Organization of Thesis	8

1.1 Background

Histological images are the magnified images of tissues of an organism's body. These are a special kind of medical images, which contain various complex structures and the underlying semantic knowledge pertaining to different biological conditions of the organism. *Histopathology* is the field of examining tissue specimen by medical specialists (pathologists) to determine the presence and extent of abnormal conditions, especially tumors and cancers. Generally, a pathological test is recommended after observing physiological symptoms of a subject, and performed by acquiring histopathological specimens from the human body through biopsy or surgical resection [Weidner 2009]. The tissue specimens are mounted on glass slides and treated with detailed preparation procedure before viewing [Rolls 2011]. This includes a staining process with special chemicals in order to improve contrast and visibility of particular histological and cellular structures as required for observation. Conventional viewing and examination of glass slides is performed using diverse microscopy techniques [Rochow 1994].

Digital pathology as a sub-branch of pathology involves the use of digitized histological images for observation and analysis. Glass slides are first scanned using sophisticated equipment

such as advanced microscope cameras or whole slide scanners, as a result of which high resolution whole slide images (WSI) are obtained. These images are stored on local computers or remote servers, observed on screens using viewing programs, and archived into large databases for future references. They provide a comprehensive view of the tissue architecture, allowing the pathologists to form opinions about the underlying disease conditions. Several benefits of digital pathology have been identified, the most apparent being convenient slide navigation, efficient data handling, parallel viewing at distinct locations (telepathology), parallel viewing of different stains and positions, reduction in glass archives and glass transportation, and simplicity to handle pathologists' annotations [Sucaet 2014].

Medical image analysis techniques are being developed to automate the process of extracting meaningful information from whole slide images for clinical routine practice and research in digital pathology. Majority of the research in this direction has been aimed towards developing computer-aided diagnostic methods in order to automatically predict tissue regions of interest (ROI) as healthy or cancerous, benign or malignant, or one of the tumor grades or types. Less critical applications are also being explored that can potentially improve the current state of technology in this field, for instance, content-based image retrieval of interest regions from whole slide image data, archiving and bio-banking of large databases, marker quantification and detecting malignant changes over time, thereby, contributing directly or indirectly towards diagnosis, prognosis, education and research in biology and medicine. The developed computer-based tools and techniques provide assistance to pathologists by reducing their time inputs, manual efforts and subjective variations. Nevertheless, pathologists' experience and expertise are most valuable and essential for the development of these image processing and analysis software, as well as for the validation of obtained results.

1.2 Study Objectives

In this study, the prime objective is recognized as the development of a fully-automated systematic framework to perform medical image analysis of tissue regions in whole slide images of gastric cancer. The sub-objectives of the research are described as the following.

1. The comprehensive research goal is to explore suitable medical image analysis algorithms, leading to efficient knowledge description and representation of meaningful visual information for automatic categorization and retrieval in histopathological whole slide images of gastric cancer. This demands the study and modifications of existing methods, and the proposal of novel approaches with theoretical descriptions and experimental evidence for further enhancement in current state of technology.
2. In this work, a prime objective is the analysis of cancer-affected tissue regions using the H&E stain based on HER2 immunohistochemical response in gastric cancer whole slide images. H&E stained images need to be thoroughly analyzed in order to determine whether these are sufficient for cancer characterization on the basis of HER2 immunohistochemistry, and to establish if any visible correlations exist between the two stains. Thus, computer-based image

analysis of gastric cancer in H&E stain requires to be pursued in depth through traditional methods as well as deep learning methods in digital histopathology.

3. During the research tasks, related goals have been discovered, namely, computer-based exclusion of necrotic tissue regions and effective cell nuclei segmentation. For necrosis detection, an adequate and timely solution needs to be devised. Also, the H&E stain has the property to distinctly stain the cell nuclei in the tissue specimen, so, for the automatic isolation of cell nuclei, a segmentation algorithm is required to be extensively evaluated. This involves determining the working magnification and image resolution suitable for optimum image analysis. Further, a multiresolution approach needs to be explored for enhancement of segmentation results compared to individual magnifications.
4. Computerized applications corresponding to the above defined problems can greatly assist professionals in routine and research, thus, a part of this study is dedicated towards contribution of application-specific software tools. This includes demonstrating example applications such as computer-aided diagnosis, automatic necrosis detection, tissue composition determination and content-based image retrieval in H&E stained gastric cancer whole slide images.

1.3 Motivations

In the current scenario, pathologists mostly examine collections of large-sized whole slide images (or glass slides) using visual inspection, which is a tedious and time-consuming process that requires to be automated. Therefore, the significance of histological image analysis has been recognized by the scientific community to a great extent worldwide. However, compared to other types of medical images, research on histological images is somewhat limited. This is because histological images have typical characteristics which set them apart from other types of images. These include diverse magnifications, complex appearances, large image resolutions, highly specialized acquisition process, specific staining methods and the corresponding differences in semantic interpretations. Furthermore, there is always a possibility of human errors in analyzing histological images due to high subjectivity and observer variations during interpretation owing to a high visual similarity of such images, which may subsequently affect the selection of an appropriate treatment method. Most visual observations vary from person to person and time to time, leading to inter- and intra-observer variability and ambiguities in the decision-making process. Hence, there is clearly a need for developing computer-aided methods to introduce objectivity, which is a strong motivation behind this work.

Gastric cancer is the fourth most common cancer and the third most common cause of cancer-related deaths in the world. According to the WHO, gastric cancer accounts for around 800,000 deaths worldwide annually [Rugge 2015]. Computer-aided analysis of gastric cancer histological images is still in its early stages of development and an actively emerging topic of research. In literature, a few studies have been performed in the direction of gastric tissue image

analysis in digital histopathology ¹. However, to the best of the author's knowledge, no previous study focuses on the specified research objectives as described in this work. This motivated the author to proceed in the direction of medical image analysis in gastric cancer histopathology.

Haematoxylin and eosin (H&E) stain is routinely used in histopathological examinations as it provides a detailed view of tissue components, is easy to apply and less costly. In contrast, HER2 immunohistochemical staining is not commonly applied in laboratory practice and involves higher costs. For this study, both HER2 and H&E stained sections were available but H&E stain has been analyzed because of its wider usability and lower preparation costs. When HER2 sections are observed by pathologists using optical microscopy techniques, cancer categories are mostly visually distinguishable. However, while observing H&E stained tissue sections, pathologists require a greater time and effort to identify the corresponding malignancy levels due to very subtle differences, hence, immunostaining is often suggested. One of the main motivations of this study is to discriminate between cancer types using computer-based image analysis in the H&E stained tissue regions based on HER2 immunohistochemistry, with the assumption that minor deviations in the texture, color, morphological, neighborhood and architectural properties of tissues that may be difficult and laborious to recognize barely with the human eye, can allow sophisticated computerized methods to make desired decisions.

One of the most promising approaches for histological image analysis is the use of graph-based techniques [Sharma 2015a]. Graphs are effective and flexible representation structures and have lately been of major interest to the computer vision and image analysis fields due to their expressive ability to model topological and relational information between image components. Moreover, histological image data is visually observed and interpreted by pathologists by considering the architectural characteristics, neighborhood relationships and spatial arrangements between tissue components, and graphs have proved to be able to quantitatively represent the visual information cues acquired and processed by pathologists from histological images. Hence, in the thesis, the author is motivated to explore fully automated handcrafted graph-based methods for computerized analysis of histopathological images.

Another motivation is to explore the potential of deep learning methods in digital histopathology. Most importantly, deep learning has recently gained popularity in various research domains and achieved breakthrough results. But deep learning has not been explored so far for H&E stained histopathological whole slide images of gastric cancer for the specified goals. Particularly, the deep convolutional neural networks aim towards replacing hand-engineering of image descriptions in traditional methods with end-to-end learning algorithms. This work attempts to further investigate this hypothesis and the overall prospective capabilities of deep learning for the mentioned problems.

¹The terms *histology*, *pathology* and *histopathology* are interchangeably used throughout the thesis, referring to the same underlying concept.

1.4 Contributions

The study contributes towards the scientific domain of histological image analysis, and can be used for diverse application areas in the field of digital histopathology, in the following ways.

1.4.1 Scientific Contributions

The proposed systematic framework for medical image analysis of gastric cancer images in digital histopathology offers novel and constructive ideas, methods, experimental results, observations and conclusions, which can be valuable for the scientific community. The salient scientific contributions of this work are highlighted as follows.

One of the most prominent scientific contributions includes the design, implementation and variation in a novel graph-based method called *cell nuclei attributed relational graph* for effective representation and subsequent classification of histological regions in gastric cancer whole slide images. The graph-based approach attempts to capture and quantify the visual characteristics of individual cell nuclei, their spatial interactions with neighbors and the global tissue architecture. Eight variants of the proposed cell nuclei attributed relational graph are conceived, engineered and computationally analyzed. The hand-engineered graph-based features are quantitatively evaluated with comparisons among themselves and several well-known state-of-the-art feature extraction methods in digital histopathology.

An original self-designed *deep convolutional neural network architecture* is contributed after a thorough empirical investigation, to attain the defined objectives of medical image analysis of gastric cancer in digital histopathology. The deep learning method follows a supervised approach for classifying image regions, and its performance is quantitatively evaluated, suggesting comparable results to handcrafted features and traditional machine learning methods. The developed architecture is studied along with its relative performance to AlexNet framework [Krizhevsky 2012] which is extremely successful in general object categorization. An ensemble of the two networks is further explored to reduce individual variations.

The detailed experimental results confirm the hypothesis that a fair visual correspondence exists between H&E stain and HER2 immunohistochemical stain, thereby suggesting that in general, automatic image analysis is possible in the more routinely used H&E stain based on HER2 immunohistochemical response. Computer-based methods have been able to achieve favorable results, sometimes even better than humans, by considering the subtle details which are usually not clearly distinguishable by the human visual system.

The other scientific contributions worth mentioning include the development of suitable approaches for necrosis detection, cell nuclei classification, multiresolution combination of visual information for segmentation enhancement, and a comprehensive comparative scheme for the purpose of performance evaluation.

1.4.2 Application Areas

The developed methods can potentially be adapted to the following application areas for practical usability. These have been briefly introduced below, and demonstrated in detail in Section 6.5 and Section 7.5 respectively.

1.4.2.1 Automatic Necrosis Detection

While examining the whole slide gastric cancer image datasets, medical experts have discovered that some of the images contain small necrotic areas. These regions must be excluded before proceeding with the image analysis of cancer regions, as these act as noisy areas and interfere with the available useful information. As a result, an automatic necrosis detection application is developed for this purpose. This method is performance and time efficient, and can be potentially utilized for detecting necrosis in other tissue types.

1.4.2.2 Automatic Determination of Histological Tissue Composition

The process for multiresolution segmentation enhancement includes an automatic cell nuclei classification stage. This stage can be used independently for determining the types of cell nuclei in tissue regions, and assist in revealing histological composition in gastric cancer for heterogeneous datasets. Such tissue composition measurements can possibly assist pathologists in computer-aided diagnosis by providing a basis for automatic differentiation between tumor and non-tumor compartments of the tissue to resolve the cancer type, grade or extent. So, the tissue compositions determined in this way have been used later for the analysis of cancer regions in this work.

1.4.2.3 Computer-aided Diagnosis: Immunohistochemistry-based Cancer Classification

Immunohistochemical staining is used in practice by pathologists to determine extent of malignancy, however, it is laborious to visually discriminate the corresponding malignancy levels in the more commonly used H&E stain. This computer-based application attempts to solve the problem that may seem manually tedious by classifying unlabeled data in large-sized whole slide images. It can assist pathologists in diagnosis and prognosis, reducing preparation and inspection times as well as inter and intra-observer variabilities.

1.4.2.4 Content-based Image Retrieval

Suitable hand-engineered representations of tissue images can be further utilized for content-based image retrieval (CBIR) of desirable image regions from the database having visual and semantic contents similar to a query image. Based on this premise, an introductory CBIR application has been developed and tested for the available gastric cancer datasets in this work.

1.5 Challenges in Study

As discussed before, computer-based analysis of histological images is being actively explored worldwide. However, complete dissemination of computerized methods in practical or clinical environments is still a prospective challenge, mainly due to lower reliability compared to pathologists' verdicts. Pathologists' experience and expertise have been acquired through a specialized long-term training process, and they diagnose diverse cases on day-to-day basis. Along with assessment of visual appearance of tissues, pathologists also consider additional clinical information such as patient history and etiological agents, which makes the decision-making process more complicated and difficult to formalize. Also, the image analysis software developed till date is specific to certain pathological disorders, and not able to handle exceptional cases. In general, the extent of usability and deployment of image analysis applications in a clinical setting is a challenge for the technical developers. Nevertheless, these can be used as bases for further scientific research to optimally achieve the desired objectives. Additional considerations before computer-based systems are adopted into routine diagnostics include the size of the programs, ease of use, robustness, time requirements and data storage required for successful application. Strict regulations and financial issues add to the challenges, making their routine adoption difficult. The regulatory standards and guidelines are expected to become more distinct and unambiguous in future advancements in the field of digital pathology. However, even at present, such computer-based methods can assist medical professionals by automating several less serious tasks, *e.g.* providing additional opinions to pathologists, bio-banking, archiving, content-based image retrieval and so on.

The image acquisition process is not ideal in most cases, hence the quality of the acquired digital images is also a current question of discussion among pathologists worldwide. During the process of image acquisition, several problems may occur such as non-uniform cutting and staining of tissue specimens, touching or overlapping of samples, unclear separation, existence of stain artifacts and inhomogeneity in the tissues. As a result, segmentation of the tissue in order to identify exact boundaries of its components may become complex and suboptimal. This effect is also observed in this research, and was overcome by exploring different ways (such as re-scanning slides with different scanners and designing a multiresolution combination approach) to improve the results of a suitable segmentation algorithm. The methods have been tailored for the specific whole slide image datasets of gastric cancer in order to address the issues related to a non-ideal acquisition process.

Medical image interpretation is a complex and poorly understood process. It requires a consistently active communication channel between the computer vision researchers and the physicians for proper understanding of the images, so that it is ensured that the research is streamlined in the correct direction and is able to deliver as promised. This interaction requirement proves a bottleneck mainly due to strict time schedules on both sides. Creating or reviewing of ground truth data is performed by pathologists which is manually demanding and time-consuming in high resolution images, so acquiring medical expert knowledge is one of

the most challenging aspects of this study. Considering the complexity involved into acquiring and annotating histological image data, our method is currently limited to experimentation with the described gastric cancer whole slide image datasets. Moreover, reproducibility of developed methods is demanding due to the existing variabilities at several levels. For the same tissue type, these include biological heterogeneity, varying tissue thickness, different scanner properties and stain variations during the acquisition process. Most of these factors are considered in this research and also require to be addressed for future extensions. The usability of developed methods can further be extended for multiple tissue types, however, the challenge will be enhanced and is a future prospect of this study.

In general, there do not exist common public benchmark databases for many typical pathological problems that are highly specific to certain populations owing to several factors like geographical location, lifestyle, gender and biological agents. The study and analysis of the patterns, causes and effects of health and disease conditions in a defined population are subject of a separate branch of science called *Epidemiology* [Morabia 2013]. As a result, availability of large-scale disease-specific data is limited by such factors. Moreover, the situation is complicated by legal restrictions such as access permissions and ethical issues for usage and distribution of human data in the existing datasets or their expansion. Access to most histopathological datasets is restricted largely to the organizations holding their license, and patient data is either pseudonymized and (or) proper permission is sought for research purpose. The author and associated research groups have acquired and analyzed the available whole slide image datasets according to the ethical and usage guidelines prescribed by the source of the gastric cancer histopathological data [Warneke 2013].

1.6 Organization of Thesis

The thesis is organized as follows. Chapter 2 presents a theoretical background about the various methods explored in digital histopathology, including the domain description, widely known state-of-the-art feature extraction techniques and machine learning methods, which are studied and applied for comparative evaluations in the next chapters. Chapter 3 emphasizes the related work in literature for the investigated methods in the field of digital histopathology. Chapter 4 describes a brief overview of the proposed scientific framework showing the process pipeline of the conducted experiments. Chapters 5, 6 and 7 explain in detail, the design, implementation, analysis and implication of the explored methods in the three main stages of the experimental framework. In Chapter 8, the quantitative evaluation of performance of the proposed techniques along with comparative analysis with the established methods demonstrates the accomplishments and scrutinizes the shortcomings and challenges in experimentation. Chapter 9 discusses the statistical aspects in datasets in reference to the stains and its consequences. Chapter 10 summarizes the studies by concluding the findings, suggesting the important prospective directions of research and providing useful recommendations in the field of digital histopathology.

Theoretical Background

Contents

2.1	Introduction	9
2.2	Domain Description	10
2.2.1	Overview of Digital Histopathology	10
2.2.2	Introduction to Gastric Histopathology	12
2.3	Feature Extraction in Digital Histopathology	14
2.3.1	Low-level (Pixel-based) Methods	15
2.3.2	Object-level Methods	21
2.3.3	High-level (Architectural) Methods	22
2.4	Machine Learning in Digital Histopathology	33
2.4.1	Support Vector Machines	34
2.4.2	AdaBoost Ensemble Learning	34
2.4.3	Random Forests	34
2.4.4	Deep Convolutional Neural Networks	35
2.5	Summary	36

2.1 Introduction

In this chapter, firstly the important principles of digital histopathology are described, including an overview of the preparation and image acquisition procedures and a background of the gastric tissue. Next, the most widely applied feature extraction and image description methods in digital histopathology are introduced. A part of these methods describing the state-of-the-art high-level (architectural) feature extraction in Section 2.3.3 were published by the author in a review article [Sharma 2015a]. Finally, the machine learning algorithms explored in the study are briefly discussed.

2.2 Domain Description

2.2.1 Overview of Digital Histopathology

2.2.1.1 Introduction

Histology, a compound of the Greek words *histo*-tissue and *logos*-study, is the branch of science comprising the study of microscopic structures and functions of tissues of plants and animals. Pathology is another compound of the Greek words *patho*-suffering and *logos*-study, relating to the causal study of diseases. The term *histopathology* is derived from the fusion of the branches histology and pathology, and defined as the study of microscopic diseased tissues, where pathologists examine tissue specimens and provide diagnostic inferences based on their professional medical knowledge, experience and observations. Histopathology lays down the scientific foundation for disease diagnosis, prognosis, clinical decision-making, research and education in medicine and biology.

2.2.1.2 Tissue Specimen Preparation

Microscopic analysis of tissue specimens requires their preparation using a complex sample preparation procedure for subsequent observation of tissue components [Hunter 1972], [Rolls 2011]. A schematic overview of the general process of tissue specimen preparation is illustrated in Figure 2.1. Initially, a tissue specimen is extracted from the corresponding body part of the subject. The tissue specimen submitted for pathological analysis can be broadly classified into two main categories, namely, biopsy and surgical resection.

1. **Biopsy:** A biopsy [Zerbino 1993] is a small tissue sample acquired mostly for the purpose of rendering a definitive diagnosis.
2. **Surgical resection:** This type of tissue specimen [Tjandra 2006] is obtained by the therapeutic surgical removal of the whole disease-affected tissue region or organ(s). Such a specimen is usually acquired in order to provide definitive treatment of a disease when the diagnosis is already known or strongly suspected.

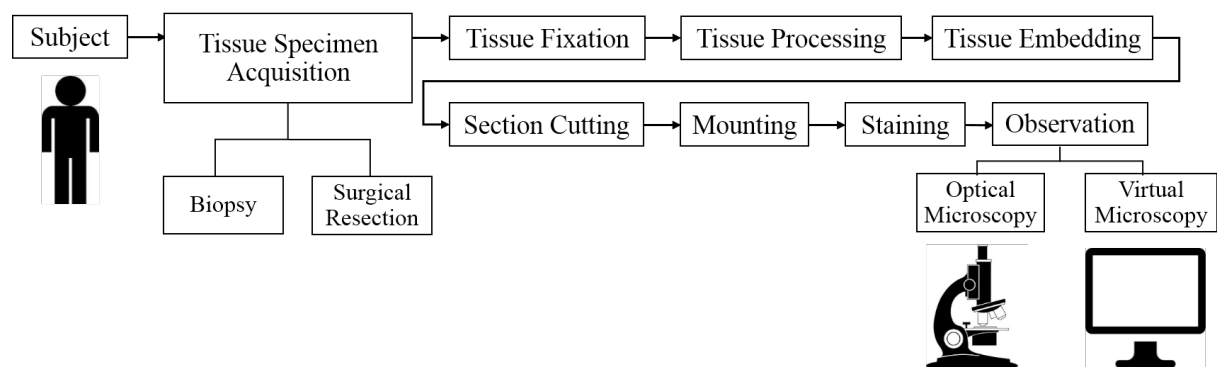


Fig. 2.1 A schematic overview of tissue specimen preparation

After acquiring the tissue specimen from the organism's body, it is required to be chemically fixed to prevent decay and preserve the constituent cells. The most popular fixing agent is formaldehyde, usually in the form of a phosphate-buffered solution, often referred as *formalin*. Following the fixation, the tissue specimens are processed in automated instruments called *tissue processors*, which allow the specimens to be infiltrated with a sequence of different solvents finishing in molten paraffin wax to provide support during cutting. Embedding procedure results in a tissue block ready for section cutting. Sections are cut on a precision instrument called *microtome* using extremely fine steel blades at a thickness of 3 - 5 μm , ensuring that only a single layer of cells makes up the section. This step produces very thin, high quality sections called *slices*. The slices are then mounted on glass slides and appropriately stained in order to improve contrast of specific tissue components, demonstrating normal and abnormal structures. A professional visually inspects the sections under an optical microscope. Moreover, a set of digital images called whole slide images can be acquired to be further processed or analyzed in digital pathology.

2.2.1.3 Virtual Microscopy and Whole Slide Imaging

Recent advancements in technology have led to the development of a set of modern methods called *virtual microscopy* [Weinstein 2009] including whole slide imaging, that incorporate the use of technical equipment such as whole slide scanners and high-performance computers to facilitate digitization of glass slides. The resulting whole slide images (WSI) usually have Gigapixel resolutions. Virtual microscopy methods also aid in digital viewing at several magnifications, storage and transmission of histological WSI over computer networks. Hence, virtual microscopy can be seen as an integration of optical microscopy and digital technologies, and an advanced alternative to examining glass slides under the light microscope. The general virtual microscopy pipeline including whole slide imaging is shown in Figure 2.2. The procedure for histological image analysis usually starts with the acquisition and labeling of the digital whole slide images using virtual microscopy techniques.

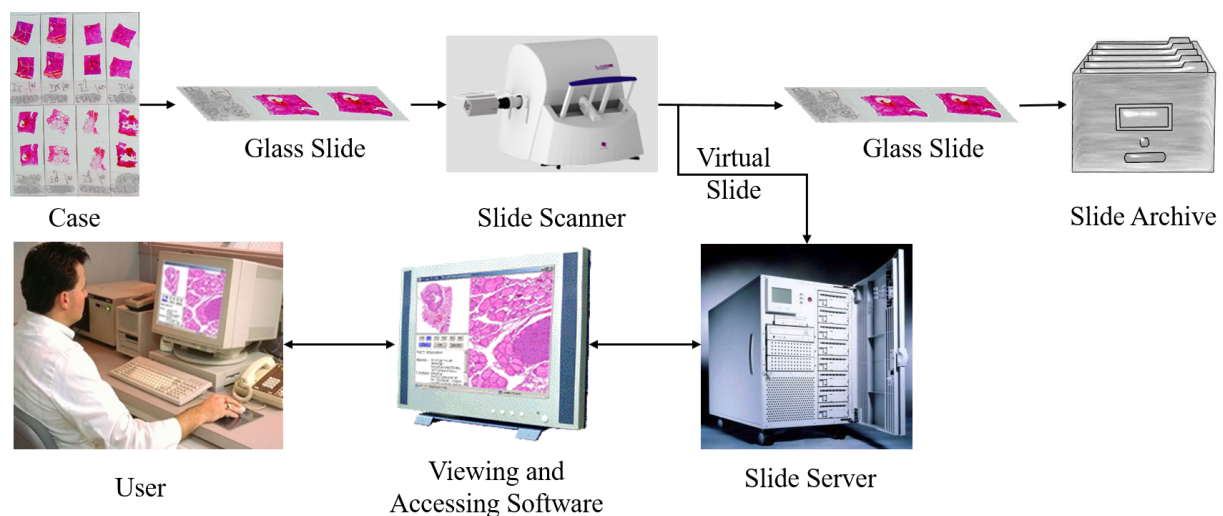


Fig. 2.2 Pipeline of virtual microscopy. Adapted from [Saeger 2009].

The virtual microscope functions in similar way to optical microscope. Magnification and resolution are the two main characteristics to be considered in this regard. *Magnification* is the ability to make the objects appear larger, and measured in terms of magnifying power (represented by a number followed by \times), which denotes how many times the object is magnified. The conventional notation in optical microscopy for compound microscopes uses only the magnification provided by the objective lens, *i.e.* varying with each instrument, and usually a fixed magnification of the eye piece *i.e.* $10\times$. For example, a magnification of $a\times$ means that the optical lens magnifies the object a times, followed by a 10 times magnification of eye piece, leading to a final effective $10a$ magnification of the object. Such notation is extended to virtual microscopes, and also considered in this work. *Resolution* is the ability to distinguish between two points and determines the amount of details that can be observed. It measured in terms of resolving power as the smallest distance between objects that can be distinguished. For the human eye, it is around 0.1 mm. For digitized microscopic images, it is usually measured in $\mu m/pixel$ representing the distance resolved by each pixel, thus, smaller the quantity, higher is the resolving power. This notation has been used later in the thesis. Usually, the WSI scanner resolution is sufficient to present all relevant magnifications on the monitor, and the size or scale of the physiological structures to be viewed in the slide determines the choice of working magnification of histological images to be analyzed.

2.2.2 Introduction to Gastric Histopathology

2.2.2.1 Anatomy of the Stomach

The stomach is a muscular organ situated in the left upper abdominal side of the body that secretes acids and enzymes for digesting food. Anatomically, the stomach is divided into four sections [Moore 2011], each of which has different types of cells and functions. The sections are *Cardia* where the contents of the esophagus empty into the stomach, *Fundus* formed by the upper curvature, *Body or Corpus* is the main central region and *Pylorus* is the lower section that empties the contents into the small intestine *via* duodenum. The stomach wall is folded into *Rugae* to provide higher absorption area. Histologically, the wall consists of four layers of lining [Sorenson 2008] as the following.

1. *Mucosa*: It consists of the epithelium and the lamina propria (loose connective tissue), with a thin smooth muscle layer called muscularis mucosa which separates it from the next layer.
2. *Submucosa*: It is the next layer after mucosa and comprises of the fibrous connective tissue.
3. *Muscularis Externa*: It comprises of the muscular layers which lie beneath the submucosa. It is unique from other organs of the gastrointestinal tract and consists of three layers, namely, the inner oblique layer, the middle circular layer and the outer longitudinal layer.
4. *Serosa*: It consists of layers of the connective tissue continuous with the peritoneum.

The columnar epithelium of the mucosa forms invaginations called gastric pits [Sorenson 2008]. The lamina propria contains gastric glands opening into the bases of the gastric pits, which

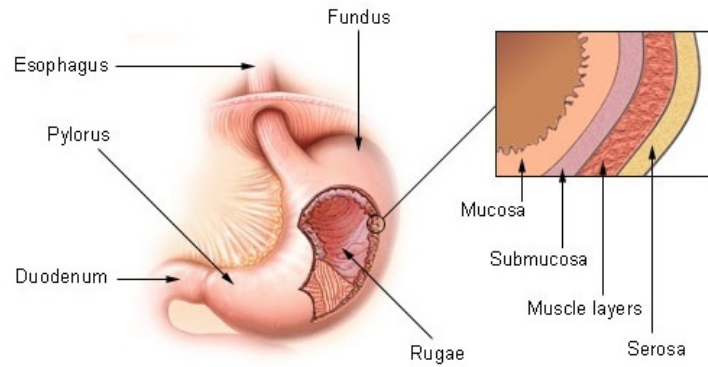


Fig. 2.3 Pictorial representation of the stomach. Source: [Wikipedia 2017]

produce and secrete gastric juices. Different types of cells are found at different parts of these glands. A pictorial representation of stomach is shown in Figure 2.3.

2.2.2.2 Gastric Carcinoma

Gastric cancer is a leading cancer type that can be subdivided into different gastric malignancies. Among these, gastric adenocarcinoma is the most common gastric malignancy (90-95%), followed by lymphomas (1-5%), gastrointestinal stromal tumor (2%), carcinoids (1%), adenoacanthomas (1%) and squamous cell carcinomas (1%) [DeVita 2010]. Histopathologically, gastric carcinoma is classified according to Lauren classification [Lauren 1965] into intestinal type and diffuse type.

Gastric adenocarcinoma is the stomach malignancy studied in this work. It is defined as a malignant epithelial tumor that originates from glandular epithelial tissue in the mucosa layer of the stomach, and aggressively spreads by infiltrating the submucosa and muscularis externa of the gastric wall. Gastric adenocarcinoma can be subclassified according to histological description of the predominant structures as tubular, papillary, mucinous and poorly cohesive [Bosman 2010].

2.2.2.3 Associated Staining Methods

Staining is a technique that can be used to better visualize cells and other tissue components under a microscope. By using different stains, one can preferentially stain certain cellular components, such as a nucleus or a cell wall, or the entire cells. Some commonly used stains for this purpose are Haematoxylin & Eosin (H&E), Feulgen, Schiff, Wright, Ki-67 protein, Toluidine blue and immunohistochemical stains [Pawlina 2006]. The author describes two of the many staining techniques, namely, HER2 immunohistochemical stain and H&E stain, because these two stains have been applied in the whole slide image data of the subsequent experiments.

Immunohistochemical (IHC) staining (or immunostaining) is a group of staining methods used in histopathological diagnostics of cancer [Ramos-Vara 2014]. An example of immunostaining is the HER2 staining. The human epidermal growth factor receptor 2 (Her2) gene is a proto-oncogene whose high amplification causes a protein overexpression in cell membrane

of a malignant cell, leading to abnormal cell division and growth [Chua 2012]. It has been most widely studied in breast cancers. Neu is a protein which is encoded by the Her2 gene in humans. Her2/neu (or simply HER2) has been recently introduced as a predictive biomarker for the treatment of gastric cancer with trastuzumab. Trastuzumab is an antibody targeting HER2 and is applied in combination with chemotherapy for the treatment of HER2 positive advanced gastric cancer [Warneke 2013]. The HER2 status is assessed by pathologists using tumor tissue obtained by biopsy or surgical resection and immunohistochemistry along with in-situ hybridization. A gastric cancer is HER2 positive when a strong circumferential, lateral or basolateral immunostaining is shown by $\geq 10\%$ of the tumor cells in the tissue, or when weak to moderate circumferential, lateral or basolateral immunostaining is shown by $\geq 10\%$ of the tumor cells in combination with HER2 gene amplification, also called the *10% cut-off rule* [Behrens 2015].

Hematoxylin-Eosin (H&E) staining is an example of staining method that has been used by pathologists for over a hundred years because it provides a detailed view of the tissue by clearly staining the tissue components. Hematoxylin stains the cell nucleus and other acidic structures, such as RNA-rich portions of the cytoplasm and the matrix of hyaline cartilage with blue color. In contrast, eosin stains the cytoplasm and collagen with pink color. Hence, the image analysis of H&E stained images would mainly emphasize on cell nuclei, cytoplasm and other stained tissue component characteristics. H&E stain is the most widely used stain in histological diagnosis and is often a basis for primary diagnosis in the field of histopathology.

As stated earlier, HER2 immunohistochemical staining is a more complex staining procedure involving higher preparation costs, and is still not common in laboratory practice. Haematoxylin and eosin (H&E) stain, in contrast, is routinely used in histological examinations due to clear staining of cell nuclei, easy application, lower preparation costs and wider usability [Bancroft 2008]. However, pathologists find difficulty in visually differentiating between corresponding malignancy levels in H&E stain, the reason that H&E is used initially and HER2 immunohistochemical staining is subsequently recommended to reveal the details of malignancy. Due to its advantages, the H&E stain has been preferably analyzed in this work.

2.3 Feature Extraction in Digital Histopathology

In computer vision and image processing, a *feature* can be defined as a set of values derived from the higher dimensional image data that can be useful and informative for a given task, for example, image recognition, object detection and content-based image retrieval. Features are required to be appropriately engineered for sophisticated image analysis such as machine learning and classification. The process of computing features from the image is known as *feature extraction*, and considered as a dimensionality reduction, because the salient characteristics in the image are represented by a lower dimensional feature vector. The systematic procedure to make image data more accessible, leading to a relevant computer interpretation is referred as *image representation*, and the process of quantification as *image description*. In the field of

digital histopathology, the established and frequently studied handcrafted feature extraction and image description methods are broadly divided into three groups, namely, low-level (pixel-based) methods, object-level methods and high-level (architectural) methods [Gurcan 2009].

In this section, the three groups (and contained methods) considered as state-of-the-art for extracting suitable features from tissue images in digital histopathology are briefly introduced. A subset of these methods has been applied in our experiments as a baseline for quantitative comparisons with the proposed methods and also during the image pre-analysis stage.

2.3.1 Low-level (Pixel-based) Methods

These methods incorporate detection of features through individual pixels in the image, and do not include morphological or architectural properties of tissue components. Research in this area mainly includes approaches for extracting meaningful information in the form of texture, intensity, color and frequency-based features directly from the image pixels. In literature, these methods have been applied on various types of histological images for several applications like cancer classification, grading, tissue classification and content-based retrieval, discussed in detail in Section 3.2.1.

2.3.1.1 Texture Descriptors

Texture is the most frequently used tissue characteristic in histopathological image analysis. This is primarily because the working magnification in histological WSI ($5\times$ to $40\times$ objective magnification) makes it possible to distinguish the variations in histological components purely on the basis of textural differences. Some of the most popular texture descriptors are explained as follows.

Gray Level Co-occurrence Matrix Statistics

The commonly used features in this group consist of statistics [Haralick 1973] computed after deriving a gray level co-occurrence matrix (GLCM) for each image. The GLCM is used for a series of second order statistical calculations, measuring the relationship between the gray levels of neighboring pixels in the original image. The GLCM is denoted by G , with dimension $N_g \times N_g$, where N_g is the number of gray levels in the image. Each element of G is computed by counting the number of times $N_{i,j}$ a pixel with value i is adjacent to a pixel with value j and then dividing the entire matrix by the total number of such comparisons. Each entry is considered as the probability that a pixel with value i will be found adjacent to a pixel of value j . $p[i, j]$ is given in equation 2.1 as the probability of co-occurrence of pixels with values i and j .

$$p[i, j] = \frac{N_{i,j}}{\sum_{j=0}^{N_g-1} \sum_{i=0}^{N_g-1} N_{i,j}} \quad (2.1)$$

The common statistical features extracted from the GLCM are defined in Table 2.1 [Haralick 1973].

Table 2.1 *GLCM statistical descriptors*

Statistical quantity	Expression	Remarks
Angular Second Moment	$\sum_i \sum_j [p(i, j)]^2$	Also called Energy or Uniformity
Contrast	$\sum_{n=0}^{N_g-1} n^2 \sum_{i=1}^{N_g} \sum_{j=1}^{N_g} p(i, j)$	$n = i - j $
Correlation	$\frac{\sum_i \sum_j (ij)p(i, j) - \mu_x \mu_y}{\sigma_x \sigma_y}$	μ_x and μ_y are mean values σ_x and σ_y are standard deviations of partial PDFs of x, y
Entropy	$-\sum_i \sum_j p(i, j) \log[p(i, j)]$	–
Inverse Difference Moment	$\sum_i \sum_j \frac{p(i, j)}{1 + (i - j)^2}$	Similar to (or referred as) Homogeneity
Sum of Squares: Variance	$\sum_i \sum_j (i - \mu)^2 p(i, j)$	–
Sum Average	$\sum_{i=2}^{2N_g} i p_{x+y}(i)$	x and y are coordinates of an entry in G and $p_{x+y}(i)$ is the probability of the co-occurrence of coordinates $x + y$
Sum Entropy	$-\sum_{i=2}^{2N_g} p_{x+y}(i) \log[p_{x+y}(i)]$	–
Sum Variance	$\sum_{i=2}^{2N_g} (i - SE)^2 p_{x+y}(i)$	SE refers to sum entropy
Difference Entropy	$-\sum_{i=0}^{N_g-1} p_{x-y}(i) \log[p_{x-y}(i)]$	–
Difference Variance	$\sum_{i=0}^{N_g-1} i^2 p_{x-y}(i)$	–
First Information Measure	$\frac{HXY - HXY1}{\max(HX, HY)}$	HXY is Entropy, HX and HY are entropies of partial PDFs p_x, p_y and $HXY1 = -\sum_i \sum_j p(i, j) \log[p_x(i)p_y(j)]$
Second Information Measure	$\sqrt{1 - e^{[-2(HXY2 - HXY)]}}$	$HXY2 = -\sum_i \sum_j p_x(i)p_y(j) \log[p_x(i)p_y(j)]$
Maximal Correlation Coefficient	$\sqrt{\text{second largest eigenvalue of } Q}$	$Q(i, j) = \sum_k \frac{p(i, k)p(j, k)}{p_x(i)p_y(k)}$

Gabor Filter-bank Responses

Gabor filters are considered good approximations of the receptive fields of simple cells in primary visual cortex of mammalian brain [Marčelja 1980]. [Daugman 1985]. Gabor filter-bank features can be selected for simulating the human visual detection abilities, as frequency and orientation representations of Gabor filter are similar to response of human visual system. A two-dimensional Gabor filter is a band-pass filter with impulse response $h(x, y)$ in the spatial domain as the multiplication of two functions, a complex sinusoidal carrier $s(x, y)$ and a Gaussian envelope $g_r(x, y)$ and denoted by equation 2.2 [Bovik 1990], [Movellan 2002].

$$h(x, y) = s(x, y) \cdot g_r(x, y) \quad (2.2)$$

The sinusoidal carrier in spatial domain in the Cartesian coordinate system is shown in equation 2.3, where U and V are spatial frequency and ϕ is the phase of the sinusoid.

$$s(x, y) = \exp(j(2\pi(Ux + Vy) + \phi)) \quad (2.3)$$

It can be inferred from above that, $s(x, y)$ can be represented as equation 2.4 in the polar coordinate system, where $F = \sqrt{U^2 + V^2}$ is the magnitude and $\omega_0 = \tan^{-1}(V/U)$ is the direction.

$$s(x, y) = \exp(j(2\pi F(x\cos\omega_0 + y\sin\omega_0) + \phi)) \quad (2.4)$$

The Gaussian envelope in spatial domain is expanded in equation 2.5, where K is the scale of the magnitude, σ_x and σ_y are asymmetric spread of the Gaussian, x_0 and y_0 are locations of peaks, r specifies rotational operation, and θ is the rotation angle.

$$g_r(x, y) = K \exp \left(-\frac{1}{2} \left(\frac{(x - x_0)_r^2}{\sigma_x^2} + \frac{(y - y_0)_r^2}{\sigma_y^2} \right) \right) \quad (2.5)$$

$$(x - x_0)_r = (x - x_0)\cos\theta + (y - y_0)\sin\theta \quad (2.6)$$

$$(y - y_0)_r = -(x - x_0)\sin\theta + (y - y_0)\cos\theta \quad (2.7)$$

Example of a Gabor filter-bank consisting of twelve two-dimensional Gabor filters with three frequencies and four orientations is shown in Figure 2.4. When an image is filtered with a set of filter-banks at different frequencies, orientations and standard deviation combinations, their filter responses are considered as powerful texture discriminators [Fogel 1989]. The Gabor filter bank responses can be used as effective texture features in histopathological image analysis.

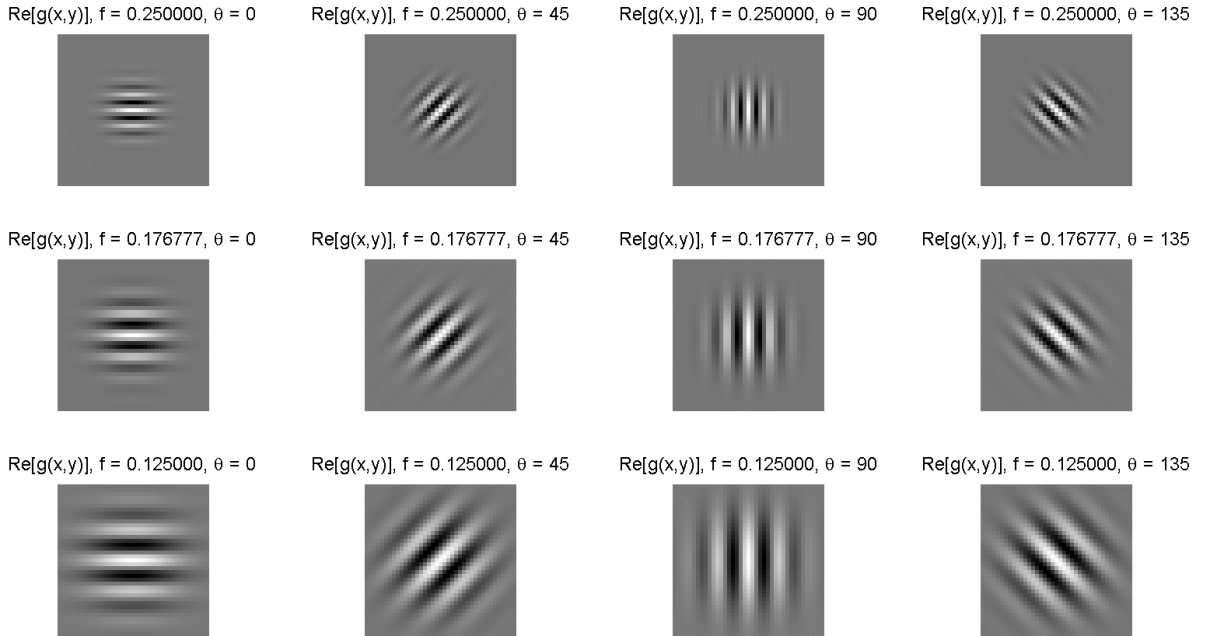


Fig. 2.4 Example of Gabor filter-bank with varying frequencies and orientations

Local Binary Patterns

The process of generating Local Binary Patterns (LBP) labels the pixels of an image by thresholding the neighborhood of each pixel and considers the result as a binary number [Ojala 1996]. A LBP histogram can be computed and found effective to represent textures. The LBP texture operator has recently become popular approach for texture classification because of high distinguishing potential and computational simplicity, and has been applied in several applications such as face recognition [Ahonen 2006] and moving object detection [Heikkilä 2006].

The LBP operator first divides the image into windows of smaller size, and compares the intensity of each center pixel c to its p -nearest neighbors, generating a 0 or 1 as the result of each comparison. This operation gives a p -digit binary number corresponding to the pixel, which is then converted to an equivalent decimal value called the LBP code. The resulting LBP image $L_{c,p}(x, y)$ is shown in equation 2.8, where I_c is the intensity of the considered pixel and s is a sign function mapping to 1 if the difference is greater than or equal to 0, and 0 otherwise. An example of the process of generation of the LBP code is illustrated in Figure 2.5.

$$L_{c,p}(x, y) = \sum_{k=0}^{p-1} 2^k s(I_k - I_c) \quad (2.8)$$

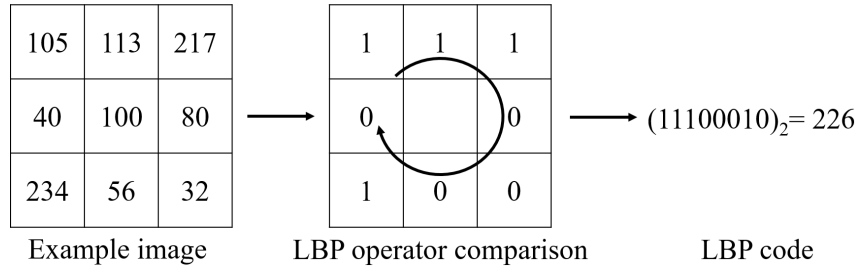


Fig. 2.5 The LBP code generated from an example image

The LBP image $L_{c,p}(x, y)$ can be obtained, and its histogram for i^{th} LBP code with total $N = 2^p$ codes is defined in equation 2.9.

$$H_i = \sum_{x,y} f\{L_{c,p}(x, y) = i\}, \quad i = 0, 1, \dots, N - 1 \quad (2.9)$$

where, $f\{P\}$ is a conditional operator which gives 1 where predicate P is true, else 0. The classical LBP operator uses eight nearest neighbors. As a result, for each window, a histogram of the frequency of each number occurring can be computed as a 256-dimensional feature vector, and can be optionally normalized and concatenated over the image. Several extensions of LBP have also been explored [Mäenpää 2003], such as the use of different sizes of neighborhood [Ojala 2002].

Varma-Zisserman Textons

Varma-Zisserman texture classification approach [Varma 2002], [Varma 2005] is also a popular method for extracting texture features in digital histopathology. It makes use of maximum response 8 (MR8) filters to generate filter responses of images selected randomly from each class. The MR8 filter bank consists of 38 filters derived from a common Root Filter Set (RFS), but only eight filter responses are used, as the filters are at multiple orientations but their outputs are reduced by considering the maximum filter response across all orientations for rotational invariance. The filters used in the Varma-Zisserman approach from the RFS are edge filters at 3 scales with 6 orientations at each scale, bar filters at the same 3 scales with 6 orientations at each scale, a Gaussian and a Laplacian of Gaussian both with $\sigma=10$ which are rotationally symmetric. The RFS filter bank generated using the open-source MATLAB filter implementation [Varma 2002] is shown in Figure 2.6.

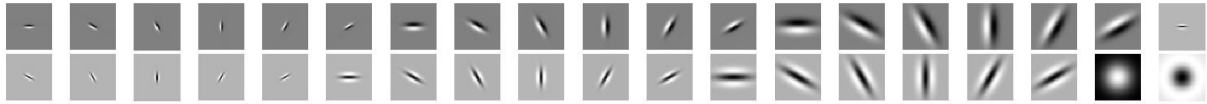


Fig. 2.6 The RFS filter-bank to obtain MR8 filter responses in the Varma-Zisserman approach. Constructed using the open-source MATLAB filter implementation in [Varma 2002]

Filter responses of sample images are aggregated and texton cluster centers are then collected into a texton dictionary using the K-means clustering algorithm [Hartigan 1979]. For each image, its filter responses are also generated and labeled with the texton closest to it in the filter response space. A histogram of textons showing the frequency of occurrence of each texton in the image is computed as the feature vector representing its textural characteristics.

2.3.1.2 Color and Intensity Descriptors

Color and intensity based features play a significant role in histopathological image analysis since different requirement-specific staining techniques are applied to stain the tissue sections in order to highlight specific components. Approaches in various color spaces [Tkalčič 2003], e.g. RGB, HSV and $L^*a^*b^*$ can be utilized to extract features for particular regions of interest. A few commonly used color and intensity-based features in digital histopathology are described below.

Intensity Histogram

Intensity or gray level histogram is the simple histogram of pixel intensity values. It contains the number of pixels in an image at each of the different intensity values found in that image. For an 8-bit grayscale image there are 256 different possible intensities, and so the histogram will consist of 256 bins showing the distribution of pixels among those grayscale values. For an $M \times N$ sized image $I(x, y)$ with total L intensity levels, intensity histogram for i^{th} intensity

level can be defined as equation 2.10.

$$H_i = \frac{1}{MN} \sum_{x,y} f\{I(x,y) = i\}, \quad i = 0, 1, \dots, L-1 \quad (2.10)$$

RGB Histogram

This histogram can be considered as the extension of intensity histogram for color images in the RGB color space. Either individual intensity histograms of red, green and blue channels can be considered, or a three-dimensional histogram can be produced, with the three axes representing the red, blue and green color channels. The component (or channel) histograms can be concatenated to generate a single feature vector. This histogram represents the distribution of colors in the image.

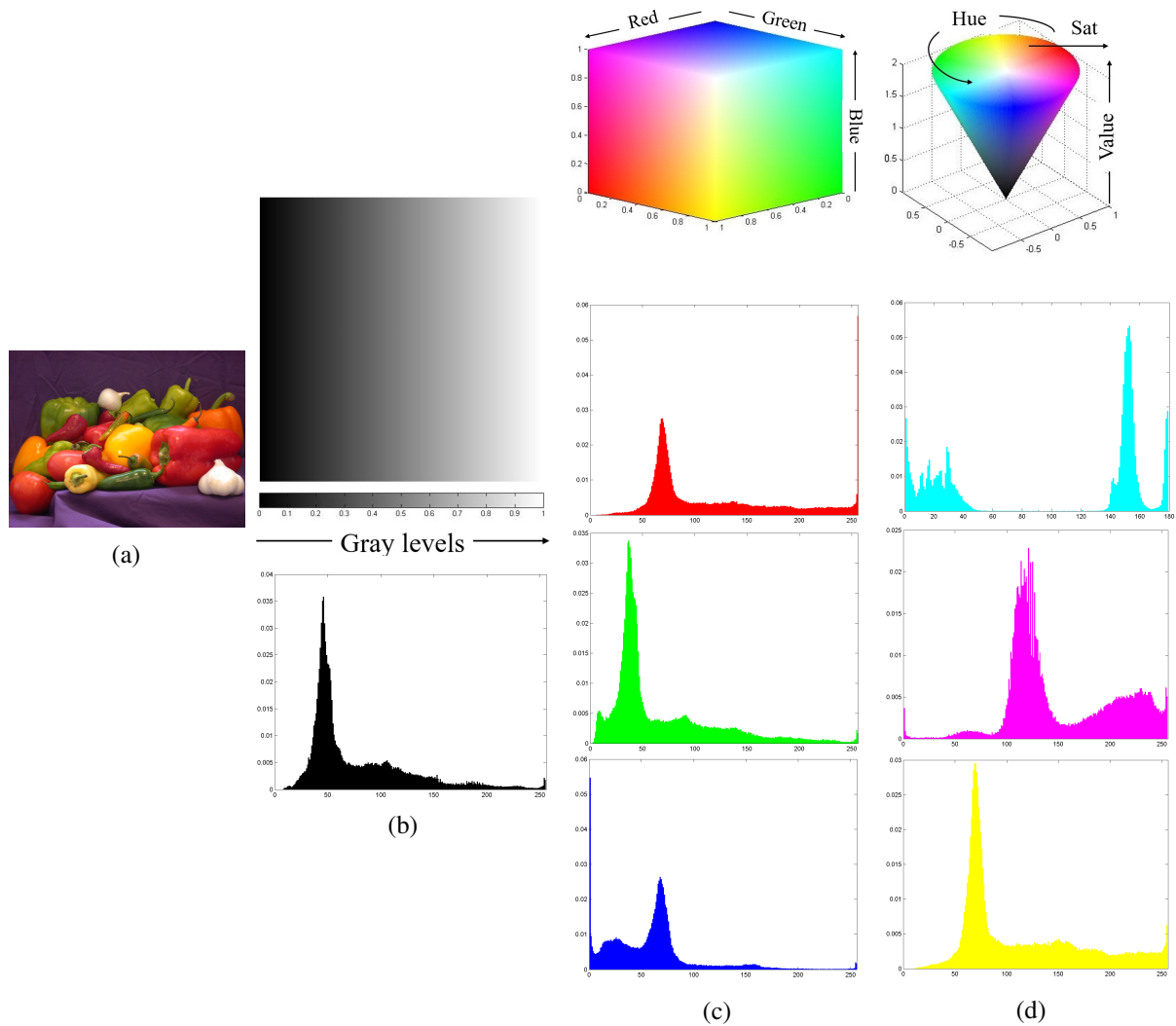


Fig. 2.7 Example showing (a) an image (b) gray model and image gray histogram (c) RGB color model and image component histograms (d) HSV color model and image component histograms (generated in MATLAB)

HSV Histogram

HSV stands for hue, saturation and value, and is also alternatively called HSB (B: brightness). It is a common cylindrical-coordinate representation of points in the RGB color model. Hue is the characteristic of a visual sensation by virtue of which a stimulus is described as similar to one of the perceived colors, namely, red, yellow, green and blue, or their combinations. Saturation represents the degree of colorfulness of a stimulus relative to its own brightness. Value (or brightness) is the measure of a visual sensation according to which a stimulus is considered to emit certain degree of light. Transformation of an image from RGB to HSV color space can reveal additional details from the different color components. These color models and an example image with the respective intensity, RGB and HSV histograms are shown in Figure 2.7.

2.3.1.3 Frequency-based Descriptors

There are additional feature extraction methods, including the measurements in the frequency domain, however, these are outside the scope of this study. Although the frequency-based features have not been widely explored in histology, they have been of interest in a few studies. The feature space can be expanded to other domains beyond the spatial domain (where the features of previous sections are computed). Basic image transforms such as Fourier transform [Gonzales 2009] and Wavelet transform [Antonini 1992] can be efficiently calculated from the raw image, and a set of the most representative coefficients are selected as image descriptors. A few more image transforms include the Chebyshev transform [Mukundan 2000] where a set of orthogonal polynomials are applied to approximate a smooth function, the Radon transform [Fiddy 1985] projecting pixel intensities onto a radial line and the Curvelet transform [Candes 2000] which is a two-dimensional extension of wavelet transform at multiple scales and angles. Feature extraction algorithms may also be applied to the transformed image, further expanding the representations in the alternate domain. The combinations of different image transforms have also been introduced as image descriptors. For instance, image transforms and their compositions for image classification in fluorescence microscopy images and other image types have been explored in [Orlov 2008].

2.3.2 Object-level Methods

Features based on image objects are also used frequently in digital histopathology. Pathologists usually describe object characteristics in terms of their structural appearance, where an object is defined as a connected group of pixels satisfying a similarity criterion, for instance, cell nucleus, vessel or gland [Gurcan 2009]. Fundamentally, object-level analysis methods depend on an underlying segmentation mechanism that detects the required objects in the image. Therefore, this research direction makes use of image analysis techniques for description and representation of constituent image objects in the regions of interest.

Object-level features can be categorized into one of four groups, namely, morphological, radiometric and densitometric, texture, and chromatin-specific [Gurcan 2009]. While the ra-

diometric and densitometric, texture, and chromatin-specific features are low-level features extracted from local neighborhoods, morphological measures are the most predominant and distinct object-level features. The common features in this group include area, perimeter, contour properties, bounding ellipsoid properties, convex hull properties, and lengths and orientations of the axes through the centroid [Hufnagl 1984]. Many of these cell nuclei morphological metrics have been applied in the proposed methods, hence, defined later in the thesis.

2.3.3 High-level (Architectural) Methods

The state-of-the-art pixel-based methods may be inadequate for solving certain problems in histological image analysis since they do not incorporate high-level topological information, neighborhood relationships and tissue architecture, which are essential characteristics for distinguishing between healthy and abnormal tissues. The main objective of this group of methods is to model the spatial interactions and arrangements between histological objects. Segmentation is a prerequisite for majority of these methods. Graph-based techniques are a subset of the high-level methods, involving extraction of discriminative structural features from tissues by first constructing representative graphs on tissue objects. This study explores such graph-based methods for analyzing histopathological images.

Graph is a data structure consisting of a finite set of points called nodes (or vertices) and a set of edges (or arcs) which link the vertices with each other based on predefined criteria. Mathematically, a graph is a tuple $G(V, E)$ containing the set of vertices V and edges $E \subseteq V \times V$. An edge in E connects two vertices in V . A graph is planar if it can be embedded in the plane, in other words, it can be depicted on the plane such that its edges do not cross one another except only at their endpoints, otherwise it is a non-planar graph. A graph is undirected if the edges have no orientation. A directed graph (digraph) is the one where each edge has a direction associated with it, connecting an ordered pair of vertices. The following sections discuss some commonly used graph-based methods in digital histopathology [Sharma 2015a].

2.3.3.1 Voronoi Diagram, Delaunay Triangulation and Related Graphs

The first formal definitions of Dirichlet tessellation and Voronoi diagram were proposed by Dirichlet [Dirichlet 1850] and Voronoi [Voronoi 1907] respectively. Let $V = \{v_1, v_2, \dots, v_n\}$ be a set of n points (or vertices) in a plane and d denotes distance between two given points. It is assumed that no three points are collinear and no four points are co-circular. The planar graphs Voronoi diagram, Delaunay triangulation and related graphs are defined using these definitions and assumptions. These graphs are included in the larger group called *proximity graphs*, also called neighborhood graphs, in which two vertices are linked by an edge if and only specific geometric requirements are satisfied by the vertices [Toussaint 1991]. Proximity graphs are defined with reference to various metrics, however Euclidean metric is the most commonly used.

Voronoi Diagram

Let two points (or sites) v_i and v_j be connected by the line segment $\overline{v_i v_j}$ and its perpendicular bisector $B(v_i, v_j)$ divides the plane between the two sites into two half planes. The half plane of site v_i with respect to v_j is denoted by $H(v_i, v_j)$ and contains the set of points given by,

$$H(v_i, v_j) = \{x \mid d(v_i, x) < d(v_j, x)\} \quad (2.11)$$

The Voronoi region (or Voronoi polygon) $V_P(v_i)$ for the site v_i is given by equation 2.12. $V_P(v_i)$ is a convex polygon which may be unbounded.

$$V_P(v_i) = \cap_{j=1, j \neq i}^n H(v_i, v_j) \quad (2.12)$$

Hence, $V_P(v_i)$ contains all such points in the plane closer to site v_i than to any other site v_j . The Voronoi diagram of the set of V sites is obtained by using equation 2.13.

$$VD(V) = \cup_{i=1}^n V_P(v_i) \quad (2.13)$$

Voronoi diagram performs a nearest site proximity partitioning of the plane. It is illustrated in Figure 2.8. Using planarity and Euler's formula, the Voronoi polygons always follow the conditions given in equations 2.14 [Rozenberg 1993].

$$\begin{aligned} \text{Number of Voronoi polygons} &= n \\ \text{Number of polygon edges} &\leq 3n - 6 \\ \text{Number of polygon vertices} &\leq 2n - 5 \end{aligned} \quad (2.14)$$

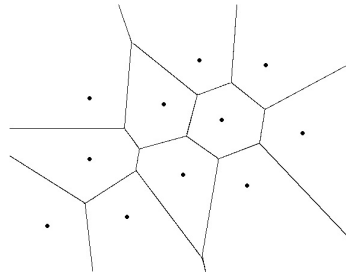


Fig. 2.8 Voronoi Diagram of a set of random points

Delaunay Triangulation

The Delaunay triangulation, also called Delaunay graph, was first defined by Delaunay in 1934 [Delaunay 1934]. It is obtained by connecting pair of points v_i and v_j in the plane such that the triangle formed by joining three non-collinear points with one side as $\overline{v_i v_j}$ is a Delaunay triangle, and is enclosed within a circumcircle with no other point $v_k \in V - \{v_i, v_j\}$ inside this circle, called *empty circle* property. Delaunay triangulation is depicted in Figure 2.9. Delaunay triangulation is the dual of Voronoi diagram, as the centroids of Voronoi polygons

correspond to the vertices in Delaunay triangulation. The duality property is also shown in Figure 2.9, and as a result of this property, the conditions in equation 2.14 are also satisfied by Delaunay triangulation with the following modification given in equation 2.15 [Rozenberg 1993]. Subgraphs of Delaunay triangulation can be generated from Delaunay graph and explained next.

$$\begin{aligned} \text{Number of Delaunay vertices} &= n \\ \text{Number of Delaunay edges} &\leq 3n - 6 \\ \text{Number of Delaunay triangles} &\leq 2n - 5 \end{aligned} \quad (2.15)$$

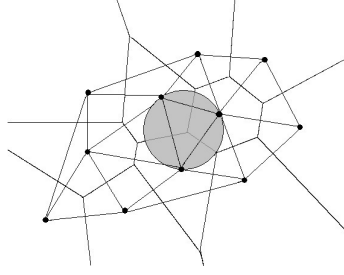


Fig. 2.9 Delaunay Triangulation corresponding to the Voronoi diagram in Figure 2.8 showing empty circle and duality properties

Gabriel Graph

Gabriel graph [Gabriel 1969] is a subgraph of Delaunay graph and an edge exists between vertices v_i and v_j if they are least square adjacent, *i.e.* if for all other vertices $v_k \in V - \{v_i, v_j\}$ the condition given in equation 2.16 is satisfied [Matula 1980].

$$d^2(v_i, v_j) < d^2(v_i, v_k) + d^2(v_j, v_k) \quad (2.16)$$

Gabriel graph can be derived from Delaunay graph by retaining all the edges of the graph such that each edge is the diameter of a circle centered on the point halfway between its endpoints that has empty circle property. Formally, line segment $\overline{v_i v_j}$ is a Gabriel graph edge for all points $v_k \in V - \{v_i, v_j\}$ if the circle with diameter $\overline{v_i v_j}$ does not contain v_k . It is depicted in Figure 2.10.

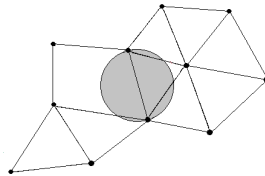


Fig. 2.10 Gabriel Graph corresponding to Delaunay graph in Figure 2.9 showing empty circle property

Relative Neighborhood Graph

Relative neighborhood graph (RNG) is a subgraph of the Delaunay graph and Gabriel graph. In this graph, two vertices are connected by an edge if there is no other vertex closer to both than

they are to each other [Toussaint 1980]. An edge exists between vertices v_i and v_j if for all other vertices $v_k \in V - \{v_i, v_j\}$ the condition in equation 2.17 is satisfied.

$$d(v_i, v_j) \leq \max\{d(v_i, v_k), d(v_j, v_k)\} \quad (2.17)$$

In other words, $\overline{v_i v_j}$ is an RNG edge, if for all points $v_k \in V - \{v_i, v_j\}$, $\overline{v_i v_j}$ is not the longest edge of triangle (v_i, v_j, v_k) . It can be derived from Gabriel graph by retaining only those edges $\overline{v_i v_j}$ for which $\text{lune}(v_i, v_j)$ is empty, where $\text{lune}(v_i, v_j)$ is the intersection of two circles centered at v_i and v_j , each with radius $\overline{v_i v_j}$. RNG corresponding to the given set of points along with the empty lune property is represented in Figure 2.11.

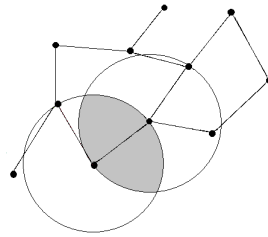


Fig. 2.11 Relative Neighborhood Graph corresponding to Delaunay graph in Figure 2.9 showing empty lune property

Euclidean Minimum Spanning Tree

A tree is obtained from an undirected graph by eliminating cycles, such that any two vertices are connected by exactly one path. A spanning tree of a graph is a tree including all its vertices. A minimum spanning tree (MST) [Graham 1985] is the spanning tree whose sum of edge weights is less than or equal to the sum of edge weights of every other spanning tree. It is called Euclidean minimum spanning tree (EMST) when each edge weight is the Euclidean distance between the two vertex-coordinates [Czumaj 2008]. EMST can be derived from the Relative Neighborhood Graph such that edge $\overline{v_i v_j}$ is retained which is not the longest edge of a cycle in the RNG. It is shown in Figure 2.12.

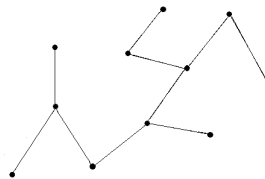


Fig. 2.12 Euclidean minimum spanning tree corresponding to Delaunay graph in Figure 2.9

Nearest Neighbor Graph

Nearest Neighbor Graph [Preparata 1985] can be defined as a directed graph where $\overline{v_i v_j}$ is a directed edge from v_i to v_j if for all vertices $v_k \in V - \{v_i, v_j\}$, the condition in equation 2.18 is satisfied. It is depicted in Figure 2.13 for the given set of points.

$$d(v_i, v_j) \leq d(v_i, v_k) \quad (2.18)$$

Nearest Neighbor Graph, when considered as an undirected graph is a subgraph of Delaunay triangulation and can be obtained from Euclidean minimum spanning tree by retaining the edge for each vertex to the closest neighbors.

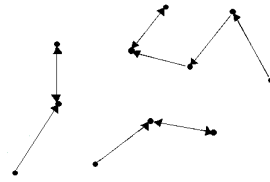


Fig. 2.13 *Nearest Neighbor Graph corresponding to Delaunay graph in Figure 2.9*

Several algorithms exist in literature for generation of Voronoi diagram, Delaunay triangulation and its subgraphs. A summary of the popular algorithms, with their associated complexities is given in Table 2.2, where a graph with n vertices are considered. Many authors have also proposed time-efficient versions of the existing algorithms with improvement in complexities.

Table 2.2 *Summary of popular algorithms for constructing Voronoi diagram, Delaunay triangulation and its subgraphs*

Graph type	Associated algorithms	Time complexity
Voronoi diagram	Fortune's sweepline algorithm [Fortune 1987] Divide and conquer algorithm [Shamos 1975] Incremental algorithm [Green 1978] Naïve algorithm [Okabe 2009] Lloyd's algorithm [Lloyd 1982]	$O(n \log(n))$ $O(n \log(n))$ $O(n^2)$ $O(n^2 \log(n))$ $O(n \log(n))$
Delaunay triangulation	Bowyer-Watson algorithm [Bowyer 1981], [Watson 1981] Lawson's flip algorithm [Lawson 1977] Lifting or projection algorithm [Zimmer 2005]	$O(n \log(n))$ to $O(n^2)$ $O(n^2)$ $O(n \log(n))$
Gabriel graph	From Delaunay graph [Matula 1980]	$O(n)$
Relative neighborhood graph	From Delaunay graph [Lingas 1994]	$O(n)$
Euclidean minimum spanning tree	From Delaunay graph [Eppstein 1999]	$O(n \log(n))$
Nearest neighbor graph	From Delaunay graph [Edelsbrunner 1987]	$O(n \log(n))$

Ulam Tree

Ulam Tree is also derived from the Voronoi diagram. It is a mathematical object growing in space and time according to certain rules [Ulam 1966]. It is generated from the Delaunay graph in such a way that the tree branches traverse only those Voronoi polygons which are not traversed by any other branch of the tree. Marginal polygons are not considered during traversal. However, there is not sufficient theoretical explanation or experimental evidence for this derivative in previous literature. It is depicted in Figure 2.14 (with start point S).

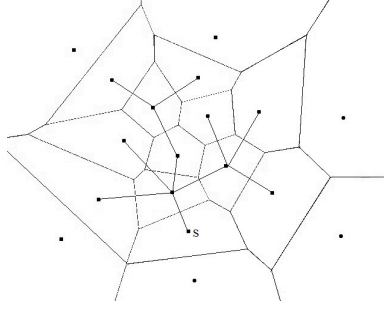


Fig. 2.14 A Voronoi diagram and its corresponding Ulam Tree

2.3.3.2 β -Skeleton

In [Kirkpatrick 1985], Kirkpatrick and Radke have defined a parameterized group of proximity graphs known as β skeletons, β being the parameter. The neighborhood of vertices v_i and v_j is defined by $N_{i,j}(\beta)$ for a fixed β . The β skeleton is an undirected graph, where two vertices v_i and v_j are connected by an edge if and only if no other point is located in $N_{i,j}(\beta)$. A sphere centered at point p and radius r is denoted as $S(p, r)$.

For $0 < \beta \leq 1$, vertices v_i and v_j are β -neighbors if and only if the neighborhood defined by the intersection of the two spheres each of radius $d(v_i, v_j)/2\beta$ which pass through points v_i and v_j in their boundary, contains no point $v_k \in V - \{v_i, v_j\}$. In this case, the lune-based and circle-based neighborhoods are identical. For $1 < \beta < \infty$, there are two definitions for β skeletons, as follows.

Lune-based β Skeleton

Points v_i and v_j are lune-based β -neighbors if and only if the lune $N_{i,j}(\beta)$ defined by the intersection of the spheres in equation 2.19 contains no other point $v_k \in V - \{v_i, v_j\}$.

$$N_{i,j}(\beta) = S\left(\left(1 - \frac{\beta}{2}\right)v_i + \frac{\beta}{2}v_j, \frac{\beta}{2}d(v_i, v_j)\right) \cap S\left(\left(1 - \frac{\beta}{2}\right)v_j + \frac{\beta}{2}v_i, \frac{\beta}{2}d(v_i, v_j)\right) \quad (2.19)$$

Circle-based β Skeleton

Points v_i and v_j are circle-based β -neighbors if and only if the neighborhood defined by the union of the two spheres each of radius $\frac{\beta}{2}d(v_i, v_j)$ passing through points v_i and v_j in their boundary, contains no point $v_k \in V - \{v_i, v_j\}$.

β neighborhoods with diverse values of β are shown in Figure 2.15. It can be seen that β -skeletons contain both relative neighborhood graphs and Gabriel graphs as special cases. Gabriel graph can be defined as lune-based and circle-based 1-skeleton ($\beta = 1$) and the relative neighborhood graph as lune-based 2-skeleton ($\beta = 2$) [Jaromczyk 1992].

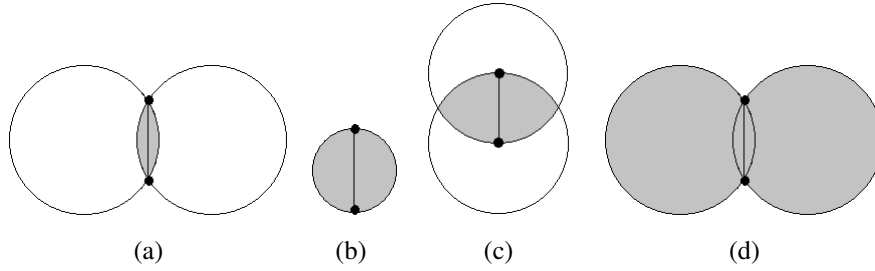


Fig. 2.15 β neighborhoods (shaded regions) with different values of β . (a) Lune-based and circle-based $\beta = 0.5$ (b) Lune-based and circle-based $\beta = 1$ (c) Lune-based $\beta = 2$ (d) Circle-based $\beta = 2$

2.3.3.3 Johnson-Mehl Tessellation

Visualization of growth models is essential in many technical processes, including histological changes. For representing such phenomena, spatial patterns obtained from simple growth processes can be used. The Johnson-Mehl model [Johnson 1939] is defined as a Poisson-Voronoi growth model showing growth of particle aggregates, where a Poisson point process is applied to generate nuclei asynchronously, and the nuclei grow at the same radial speed [Anton 2009]. The i^{th} generator $P_i = (\mathbf{p}_i, t_i)$ is defined by a planar position vector \mathbf{p}_i and associated appearance time t_i . The Johnson-Mehl tessellation can be viewed as comparable to a dynamic version of an additively weighted Voronoi diagram [Anton 1998], in which the weights indicate the associated appearance times of particles in \mathbb{R}^2 [Okabe 2009]. In stochastic geometry, the Johnsons-Mehl model is constructed to measure arbitrarily distributed geometrical properties.

The generation of Johnson-Mehl tessellation is explained in [Anton 2009]. After appearance of a new nucleus the tessellation changes, as the incoming nucleus is inserted with a new Voronoi region and the neighboring Voronoi polygons are changed. The sizes of the associated spheres of nuclei are increased by the growth proportional to the time interval between the previous appearance and current one ($t_i - t_j$). This type of spatial growth uses a Poisson point process, defined according to two cases of radial speed in [Anton 2009], namely, time homogeneous Poisson point process and time inhomogeneous Poisson point process. An example of Johnson-mehl tessellation for a random set of points at two time instants is shown in Figure 2.16.

2.3.3.4 O'Callaghan Neighborhood Graph

An alternate definition of the neighborhood of a point manifested by O'Callaghan neighborhood graphs in [O'Callaghan 1975] has been applied to histopathological images. For the histological context, O'Callaghan neighborhood graphs have been discussed in detail in [Kayser 1987], where vertices are defined as *structures*. Two types of constraints apply to the neighborhood, namely distance constraint and direction constraint. Two structures are neighbors if they are located within a certain distance (distance constraint), and not hidden behind other points classified as neighbors (direction constraint). It is shown in Figure 2.17, where v_1, v_2, v_3 are neighbors of v_i ,

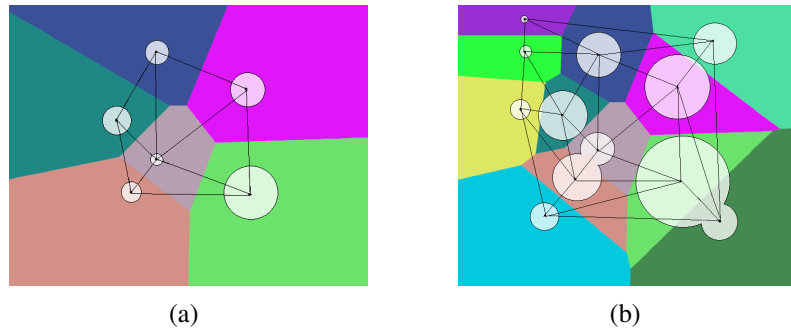


Fig. 2.16 Johnson-Mehl tessellation for a set of random points (a) growth of particles at $t = t_1$ (b) growth of particles at $t = t_1 + 50$

however v_4 is not a neighbor of v_i . All distances d_{i1} to d_{i3} are below the upper distance threshold. Direction constraint is not fulfilled by v_4 , as v_4 is hidden behind v_2 .

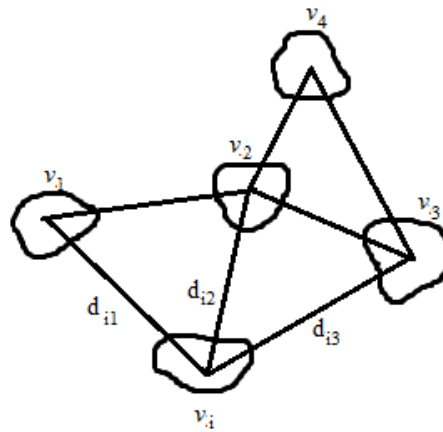


Fig. 2.17 O'Callaghan neighborhood graph

2.3.3.5 Cell Graph

As the name suggests, cell graphs are formed by considering cells or cell clusters as vertices, and relationships between them as edges using certain predefined linking rules. Cell graphs are non-planar graphs as crossing of edges are allowed. The authors introducing cell graphs initially define the linking probabilities using the Waxman model [Gunduz 2004]. However, in subsequent works like [Bilgin 2007], three different variations of cell graphs are defined. These are depicted in Figure 2.14 using available histological images. Figure 2.18(a) and Figure 2.18(b) are created with 400 times magnification, whereas Figure 2.18(c) with 100 times magnification for clarity.

Simple Cell Graph

In simple cell graph, an edge exists between given two vertices v_i and v_j if and only if the Euclidean distance d_e between the vertices is less than a predefined threshold D , which is defined

according to the characteristics of the tissue architecture.

$$d_e(v_i, v_j) < D \quad (2.20)$$

Probabilistic Cell Graph

In probabilistic cell graph, edges are assigned to vertices depending on the distance between vertices, by using a probability function which can be defined according to the tissue being analyzed. One such probability function used in literature is given by:

$$P(v_i, v_j) = d_e(v_i, v_j)^{-\alpha} \quad (2.21)$$

where v_i and v_j are two vertices and d_e denotes the Euclidean distance between them. α is an experimental parameter which determines the rate at which probability of a link decreases with the increase in distance between vertices, hence, controls the graph density. The edges are defined by a binary relation E on V for all pair of vertices v_i and v_j , such that $E = \{(v_i, v_j) : P(v_i, v_j) > r\}$, r being a real number between 0 and 1. With such construction, the vertices closer to each other are more likely to be linked compared to vertices farther away, however it is not necessary that a link will be formed even if the distance between vertices is small, and will depend on the parameter α .

Hierarchical Cell Graph

Hierarchical graph is generated by connecting smaller subgraphs representing cell clusters together to form the larger graph on the image. The clusters are identified as cell groups exceeding a threshold of the number of cells in a given region, defined by placing a grid on top of the image. Hierarchical graphs are built using these clusters as vertices, where each vertex represents a simple cell graph on the cells included in the cluster. The same linking rules are applied as for probabilistic cell graphs.

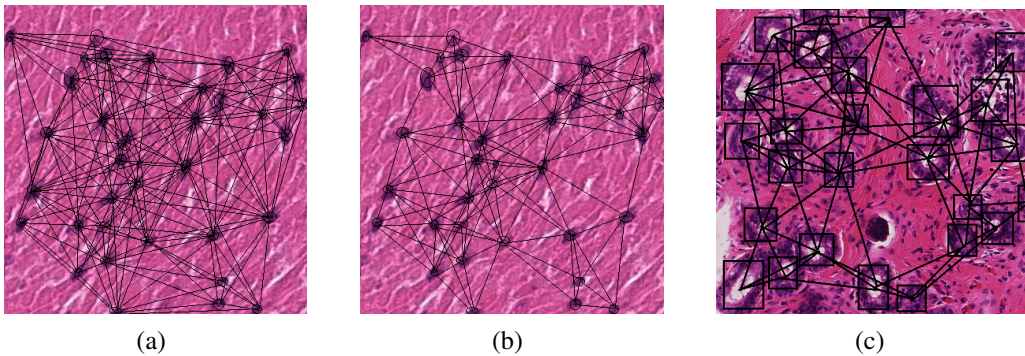


Fig. 2.18 Representative cell graphs for small regions of H&E stained images (a) Simple cell graph with $D = 200$ for gastric tissue region (b) Probabilistic cell graph with $\alpha = 2$ for gastric tissue region (c) Hierarchical cell graph for breast tissue where each vertex represents a cell cluster with a corresponding simple cell graph built on the constituent cells (denoted by rectangle)

2.3.3.6 Attributed Relational Graph

A graph G is an Attributed relational graph (ARG) when its vertices and the edges contain associated attributes. The vertex attributes for v_i are denoted as a vector $\mathbf{a}_i = [a_i^{(k)}]$, ($k = 1, 2, 3, \dots, K$), where K is the number of vertex attributes, and the edge attributes (or weights) for e_j by the vector denoted as $\mathbf{b}_j = [b_j^{(m)}]$, ($m = 1, 2, 3, \dots, M$), where M is the number of edge attributes [Sharma 2012]. In Figure 2.19, ARG with vertex attribute vectors \mathbf{a}_i , $i = 1, 2, 3$ and edge attribute vectors \mathbf{b}_j , $j = 1, 2, 3$ is shown. Vertex attributes represent object properties like size, position, shape and color whereas edge attributes define relationships between vertices such as the distance, common boundary and dissimilarity between objects. Linking between vertices may be based on the selected neighborhood condition (e.g. O’Callaghan, proximity graph).

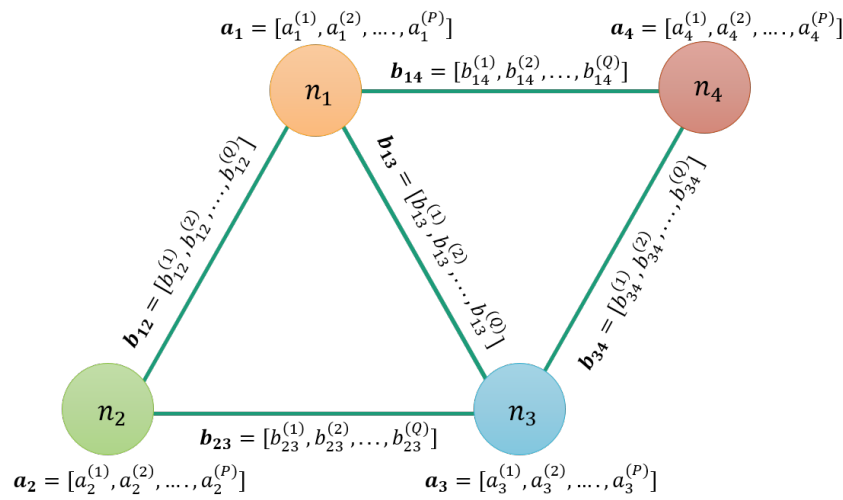


Fig. 2.19 Attributed Relational Graph

2.3.3.7 Global Graph Features

When the graphs have been constructed for images using the above described methods, the next step in image analysis is to extract appropriate features from these graphs. One way of achieving this is to define a set of global graph measures. There are many graph-based metrics and measures which can be derived from graph representations. Definitions of several measures can be found in [Wallis 2010]. Some of the most commonly used global graph features in image analysis are described as follows:

1. **Graph size:** Graph size represents the span and extent of a graph, and can be measured by counting the total number of vertices, total number of edges and total number of trees in the graph.
2. **Degree:** Degree of a vertex is defined as the number of neighbors of a vertex, or number of vertices linked to the given vertex by edges. Average degree D_{av} of graph G with n vertices and degree D_i of the i^{th} vertex v_i is calculated as below, and suitable for representing the

relative density or sparsity of the graph.

$$D_{av} = \frac{1}{n} \sum_{i=1}^n D_i \quad (2.22)$$

3. **Clustering coefficient:** Clustering coefficient is the measure of the extent to which vertices in a graph have a tendency to cluster together. For a given vertex v_i , let the neighboring vertices be contained in the neighborhood set N_i defined as,

$$N_i = \{v_j \in V \mid v_j \neq v_i, (v_i, v_j) \in E\} \quad (2.23)$$

For an undirected graph, clustering coefficient C_i for vertex v_i is computed by,

$$C_i = \frac{2|e_{jk}|}{D_i(D_i - 1)} : v_j, v_k \in N_i; e_{jk} \in E \quad (2.24)$$

where, D_i is the degree of vertex v_i and $|e_{jk}|$ is the total number of edges between the neighbors of v_i . It can be noted that $0 \leq C_i \leq 1$. The average clustering coefficient C_{av} for the graph can be calculated as,

$$C_{av} = \frac{1}{n} \sum_{i=1}^n C_i \quad (2.25)$$

4. **Eccentricity:** The eccentricity ϵ_i of a graph vertex v_i is the maximum distance between v_i and any other vertex v_j in V . It can be considered as the measure of how far a vertex is from the vertex most distant from it in the graph. The vertices of a disconnected graph are said to have infinite eccentricity [West 2000]. The minimum graph eccentricity is called the graph radius and the maximum eccentricity of graph is called the graph diameter. Average eccentricity of the graph can be calculated as,

$$\epsilon_{av} = \frac{1}{n} \sum_{i=1}^n \epsilon_i \quad (2.26)$$

5. **Path length:** A path in a graph or tree is a finite or infinite sequence of edges which connects a sequence of vertices. Edges can be part of a path only once. The minimum, average and maximum path lengths of the graph can be calculated. Let $d(v_i, v_j)$, where $v_i, v_j \in V$ denote the shortest distance between v_i and v_j , assuming that $d(v_i, v_j) = 0$ if v_i cannot be reached from v_j . Average path length l_{av} for an undirected graph can be defined as [Tomassini 2010],

$$l_{av} = \frac{2}{n(n-1)} \sum_{i \neq j} d(v_i, v_j) \quad (2.27)$$

6. **Connected Component Ratio:** A connected component (or just component) of an undirected graph is a subgraph in which any two vertices are connected to each other by paths, and which is not connected to additional vertices in the supergraph. The connected component ratio for a connected component in the supergraph is defined in equation 2.28, where n_{vc} is the

number of vertices in the given connected component and n is the total number of vertices in the supergraph.

$$CCR = \frac{n_{vc}}{n} \quad (2.28)$$

7. **Cyclomatic number:** A cycle is a path where the start vertex is also the end vertex. Cyclomatic number or circuit rank γ of a graph is the minimum number of edges which must be removed in order to eliminate all cycles from the graph [Volkman 1996]. Considering a graph with n vertices, n_e edges and n_c connected components, γ is given by,

$$\gamma = n_e - n + n_c \quad (2.29)$$

8. **Statistical descriptors:** For Voronoi diagrams and Delaunay triangulations, there can be certain measures such as area and perimeter associated with the individual Voronoi polygons and Delaunay triangles respectively. For ARGs, there can be several attributes associated with vertices and edges. Statistical descriptors can be used for globally describing these attributes of individual vertices and edges. These include mean, standard deviation, minimum to maximum ratio, disorder, skewness, kurtosis and higher-order descriptors. For these graphs, co-occurrence matrix can also be constructed and statistical features [Haralick 1973] can also be extracted from them. These statistical descriptors are capable to represent several diverse properties of entire graphs and their variations quantitatively.
9. **Spectral descriptors:** Spectral features include properties extracted from the matrices used to describe the graph, such as adjacency matrix and Laplacian matrix, and their study is called spectral graph theory [Brouwer 2011]. Some of the frequently used features include eigenvalues, spectral radius, eigen exponent, Cheeger constant, eigenmode perimeter, eigenmode volume and so on [Wilson 2005], [Luo 2003]. These spectral features have the ability to indicate various fundamental properties of the graph.

2.4 Machine Learning in Digital Histopathology

Machine learning is the branch of computer science which includes development of algorithms using which computers can learn from empirical data without the requirement of explicit programming. The spectrum of machine learning methods ranges from artificial neural networks to decision trees, ensemble learning and most recently deep learning. Many machine learning methods have been employed for several image analysis applications, *e.g.* semantic segmentation, region classification, cell nuclei classification and content-based retrieval in the field of digital histopathology. Four popular machine learning methods have been explored in this work for the specified image analysis tasks on the gastric cancer datasets, namely, support vector machines, AdaBoost ensemble learning, random forests as traditional methods and convolutional neural networks as deep learning methods. These four methods are briefly introduced in this section, and described later in the respective sections of the thesis.

2.4.1 Support Vector Machines

Support vector machines (SVM) is a group of non-probabilistic supervised learning methods. An SVM is a discriminative classifier formally defined by a separating hyperplane, *i.e.*, given labeled training data consisting of handcrafted features, the algorithm outputs an optimal hyperplane which categorizes new incoming examples. The learning phase generates non-linear classifiers using *maximum margin hyperplanes* after applying the kernel trick [Boser 1992] in order to discriminate between non-linearly separable data. The current standard SVM algorithm was proposed in [Cortes 1995], however, its variations have been successfully applied to diverse classification and regression analysis tasks in pattern recognition and computer vision [Byun 2002]. A few drawbacks of SVM include lower robustness to irrelevant descriptors leading to the requirement of feature selection [Chen 2006], and also the overhead of exhaustive parameter selection of multiple parameters. Details of the SVM algorithm applied in this work is given in Section 6.2.4.1.

2.4.2 AdaBoost Ensemble Learning

Ensemble learning [Rokach 2010] is a set of machine learning methods which use a collection of classifiers or learners and obtain a joint classification decision by combining their predictions. AdaBoost [Freund 1995] or adaptive boosting is a well-known ensemble learning approach that allows adding a sequence of weak learners to the algorithm until a desired low training error is achieved. The strong learner is assembled through a weighted majority voting scheme on the weak learners. AdaBoost is the most widely used form of boosting, as a boosting algorithm like AdaBoost involving a combination of classifiers is proposed as a new direction for improving the performance of individual classifiers like kNN, Naive Bayes and SVM [Kotsiantis 2007]. AdaBoost is also found to be easier to implement than other maximal margin classifiers like SVM. These reasons give AdaBoost an edge over other traditional machine learning approaches for supervised classification, and has been explored in the research. However, AdaBoost also suffers from some limitations such as high sensitivity towards noise and outliers, and high computational requirements due to an ensemble of classifiers. The details of the specific AdaBoost algorithm is provided in Section 6.4.2.2.

2.4.3 Random Forests

Random forest [Breiman 2001] is also a popular ensemble learning method which constructs many decision trees during the training phase. The final class is voted by the individual trees during classification. Like the SVM and AdaBoost methods, training a random forest also requires prior computation of suitable handcrafted features from the corresponding data to represent their characteristics. Random forests improve the classification by individual decision trees which are weak learners through ensemble learning, and is considered as good as AdaBoost in terms of prediction power, even sometimes better by its authors. Furthermore, it is faster than

bagging and boosting, simple to implement and interpret, and can be easily parallelized. It is also described relatively robust to outliers and redundant variables, and is able to efficiently handle large databases and missing data. Therefore, random forests have been selected as one of the traditional machine learning methods, details of which are described in Section 7.4.1.3.

2.4.4 Deep Convolutional Neural Networks

Deep learning is a subset of a larger group of machine learning methods, which comprises of algorithms with hierarchical processing layers performing non-linear transformations to represent and learn data characteristics effectively [LeCun 2015]. Deep learning methods are currently being explored in various fields such as computer vision, audio and speech processing, natural language processing, information retrieval and bioinformatics [Deng 2014]. Several deep learning architectures, for example, convolutional neural networks, deep belief networks and recurrent neural networks have been introduced [Deng 2012], and have reported to achieve state-of-the-art results in a number of tasks. The goodness of data representation notably affects the performance of machine learning algorithms. Specifically, the existence of large-scale data has been recognized as a prerequisite for the success of majority of deep learning applications, leading to a convergence between the fields of deep learning and big data analytics [Najafabadi 2015].

One of the most frequently used methods of deep learning for two-dimensional data is the deep *convolutional neural networks* (CNN). These networks consist of interconnections emulating the arrangements in visual cortex in animals, like individual neurons are organized in a manner to respond to the overlapped tessellations comprising the visual field [Hinton 2005]. The principle transformations in deep CNNs mainly consist of combinations of convolutional, pooling and fully connected layers, and their parameters are trained using backpropagation through these layers [Ciregan 2012]. Deep learning using convolutional neural networks does not necessarily require prior computation of handcrafted features, and directly processes raw input images to compute self-derived and learned features during the training process. In recent years, CNNs have achieved breakthrough performance due to the availability of large-scale training data and huge parallelization using graphics processing units (GPU) to speed up the application *i.e.* training and deployment process [Strigl 2010]. This has led to development of a number of time-efficient GPU-based frameworks [Jia 2014] for deep learning which can be applied for various image analysis tasks. All these recent advancements have led to the choice of deep learning to explore their potential for image analysis in digital histopathology.

The above machine learning methods have been selected due to mutually distinct and unique properties, for instance, traditional machine learning versus deep learning, single classifiers versus ensemble of classifiers, probabilistic versus deterministic algorithms, and different paradigms involved in learning such as margin maximization, boosting, decision trees or neural networks, that can facilitate an exhaustive comparison, along with corresponding observations and conclusions in the undertaken study.

2.5 Summary

This chapter provides the description of related medical concepts in the histopathological domain, theoretical background of the state-of-the-art and widely known image analysis techniques including feature extraction and machine learning in digital pathology. The useful methods related to this work have been introduced and defined in the chapter, unless otherwise described later in the corresponding sections of the thesis.

Related Work in Digital Histopathology

Contents

3.1	Introduction	37
3.2	Feature Extraction in Digital Histopathology	38
3.2.1	Low-level (Pixel-based) Methods	38
3.2.2	Object-level Methods	39
3.2.3	High-level (Architectural) Methods	40
3.3	Machine Learning in Digital Histopathology	43
3.3.1	Support Vector Machines	43
3.3.2	AdaBoost Ensemble Learning	44
3.3.3	Random Forests	44
3.3.4	Deep Convolutional Neural Networks	44
3.4	Image Analysis in Gastric Cancer	45
3.5	Summary	46

3.1 Introduction

This chapter discusses the recent studies in literature that are related to the work demonstrated in this research, for image analysis in the field of digital histopathology. These are organized in the sequence starting with the feature extraction methods followed by the machine learning methods and finally the research related to analysis of gastric cancer. A section of the literature review, especially relating to the feature extraction methods, was published by the author in a review article [[Sharma 2015a](#)] during her studies.

3.2 Feature Extraction in Digital Histopathology

3.2.1 Low-level (Pixel-based) Methods

Previous literature in this area mainly includes approaches for extracting information in the form of textural, color and intensity, and frequency-based features, for example, GLCM statistics, Gabor filter-banks, LBP histograms, Varma-Zisserman textons, color and intensity measurements. These methods have been applied on several types of histological images for applications such as cancer classification, tumor grading, tissue categorization and content-based image retrieval.

Texture has been recognized as a highly significant group of low-level features to characterize histological images. GLCM texture features are extracted in [Walker 1994] to determine if cervical cell nuclei are normal or abnormal. Texture analysis using optical density, GLCM and gray level run-lengths is used in [Hamilton 1997] to classify colorectal mucosa images into normal and adenomatous (dysplastic). Studies described in [Shuttleworth 2002a] and [Shuttleworth 2002b] use texture features from co-occurrence matrices in distinct color spaces for classifying colon cancer images. In [Zhao 2005], Gabor filter-bank based texture features are applied to classify histological images into different tissue types depending on their organs or parts of body. Another study using Gabor filters is presented in [Yu 2008] for analyzing semantic content in gastrointestinal tract histological images. GLCM along with Gabor filter-bank responses are used for feature extraction in [Thomas 2008], a prototype method to transform the appearance of irregular tissues to frieze-like patterns in one dimension. Local binary pattern texture features are applied for automated identification of epithelium and stroma in tissue microarrays of colorectal cancer in [Linder 2012]. Local binary patterns are also explored for stromal area removal in Ki-67 stained histological images in [Alomari 2012]. Co-occurrence features along with local binary patterns are used for classifying neuroblastoma images in [Sertel 2009]. Varma-Zisserman texton approach has been used in histological image analysis tasks and performed well such as in [Khurd 2010] for automatic Gleason grading of H&E images of prostate cancer. Several filter-banks such as Gabor and Maximum Response filters used in the Varma-Zisserman approach have also been applied in a similar texton histogram approach for classification in breast cancer tissue microarray samples in [Yang 2009].

Color and intensity-based features have been widely explored in digital pathology. Early works include the computation of gray-level features in [Wong 1983] for parallel classification of muscle tissue images. More recently, analysis of stain components has been performed using color-based SIFT features in H&E stained skin biopsy [Díaz 2010]. Histological specimens stained with H&E are segmented in original color space using a supervised color normalization and classification scheme in [Kothari 2011]. In additional related works, color and intensity features are used in conjunction with the other low-level feature groups (mentioned later under hybrid approaches).

Frequency-based features include the deployment of spectral bands, Fourier transforms and wavelets among others. Applications utilizing multiple spectral bands for analysis in-

clude [Roula 2002] and [Rajpoot 2003]. Wavelets have also been proved as powerful low-level features *e.g.* in [De Wouwer 2000], [Jafari-Khouzani 2003] and [Qureshi 2008] for breast cancer, prostate cancer and meningioma classification respectively.

Hybrid methods with different combinations of low-level features have also been explored in the past. Integration of texture and intensity-based features has been popularly used in different applications. For example, GLCM features are studied in combination with intensity-based features in [Esgiar 1998] for classifying colon mucosa as normal or cancerous. In [Doyle 2006], texture features obtained from co-occurrence matrices and Gabor filters are used with intensity-based features for detecting prostatic adenocarcinoma. Other approaches using combination of texture and intensity-based features include [Wiltgen 2003] and [Kong 2009] for analysing melanoma and neuroblastoma respectively. A work combining frequency-based features with texture properties can be found in [Orlov 2008]. A hybrid low-level method proposed as a part of this research in [Sharma 2016] combines texture and color features.

Content-based retrieval of histological images has been investigated in previous literature using low-level methods. A work for retrieval of skin cancer images using texture features is described in [Chung 2001]. A CBIR system for retrieving visually similar and pathologically relevant images from a database of cervicographic specimens has been proposed in [Xue 2007], where color features, texture features, size of lesion and spatial location features are first extracted from the image regions, and distance measures are defined to establish similarity to a query region marked by the user. In [Caicedo 2011], the authors perform retrieval of basal-cell carcinoma images using an example image as query, where a set of global texture and intensity features is applied, followed by a kernel-based semantic image annotation strategy, facilitating automatic annotation of images in large collections. A semi-automatic method using a combination of low-level features, namely, color, intensity, orientation and texture from tissue images is proposed in [Romo 2011] for automatic selection of regions of interest for further diagnosis. A multi-tiered CBIR system for microscopic images of several tissues (follicular lymphoma slides with three sub-types and neuroblastoma slides with four sub-types) is described in [Akakin 2012], where the authors extract color and texture features and perform slide-level image retrieval using multi-image query.

3.2.2 Object-level Methods

This group of methods predominantly includes the extraction of shape-based or morphological features of image objects. These features have been extensively studied in histological image analysis. The dissertation [Hufnagl 1984] explores the mathematical modeling of cell nuclei in histological images using object-level metrics, which has built the premise for an essential part of this research. Moreover, in [Hamilton 1987], H&E stained colorectal epithelium images are analyzed using morphological features followed by discriminant analysis that classifies images as normal and malignant. In [Hamilton 1994], cytometric features based on cell morphology are used for classifying fine needle aspiration cytology breast specimens into benign and malignant by constructing Bayesian belief networks. Anderson et al. [Anderson 1997]

use morphological features of glandular tissue components (ducts and lumens) to discriminate breast tissue images into ductal hyperplasia and ductal carcinoma in situ. In [Farjam 2007], morphological features are extracted to classify prostate biopsy samples into benign and malignant. An application called FatFind [Goode 2007] considers the property of perfect round shape of adipocytes for their automatic counting. Other morphological studies in histology include [Deligdisch 1993], [Thiran 1996] and [Nedzved 2007] for analyzing ovarian cancer, lungs and digestive tract, and multiple kinds of tumors respectively.

Several studies also exhibit a fusion of the low-level and object-level methods. Papanicolaou smears of cervical epithelial lesions are analyzed using texture, color and shape features in [Stenkvis 1987]. An application of CBIR with histopathological images using color and morphological characteristics and property concept frame representation based on fuzzy logic is described in [Jaulent 2000]. A combination of texture and morphological features is explored in [Spyridonos 2001] for bladder carcinoma, and morphological and colorimetric features are combined in [Zhou 2002] for the analysis of lung cancer images. Tissue composition in prostate neoplasia using GLCM-based texture and morphology is determined in [Diamond 2004]. A multiresolution approach developed as a part of this work also uses object-level texture, intensity and morphological features for automatic classification of cell nuclei of gastric cancer tissue in [Sharma 2015b].

3.2.3 High-level (Architectural) Methods

Graphs constitute an interesting area of research for analysis of histological image data, due to their ability to efficiently represent tissue architecture in suitable ways. High-level or architectural feature extraction methods are being increasingly developed to characterize tissues using spatial distribution and neighborhood properties by constructing representative graphs and calculating a set of appropriate local and global graph-based metrics from respective image descriptions. As stated before, a comprehensive literature survey of graph-based methods for image analysis in digital histopathology performed as part of this research is presented in [Sharma 2015a], discussing the progress of exciting developments and applications in this field.

Graph-based methods for histological images were initially explored in the direction of syntactic analysis. Syntactic methods based on neighborhood conditions in pattern recognition were introduced in [Fu 1974]. Following this, a relatively new technique was introduced in diagnostic quantitative pathology called *syntactic structure analysis*, which was believed to provide quantitative information on tissue architecture [Van Diest 1995]. The earliest work in this direction is presented by Prewitt et al. [Prewitt 1978] for characterizing epithelial tissues of urinary bladder mucosa, where measurements are performed using an interactive digital image processing and decision-making system called PEEP/DECIDE/GRAPH [Prewitt 1977] to automate diagnostic procedures. Features are extracted from haematoxylin stained cells representing intensity, texture, shape, differentiation and structure, and from tissue regions quantifying spatial distributions and tissue organization. Graphs and grammars in histology are formally defined in [Prewitt 1979], where a graph-theoretic model and syntactic pattern recognition are introduced for tissues using

two random spatial structures, namely, Dirichlet tessellation and Delaunay tessellation as the unique topological planar dual graph of the Dirichlet tessellation. These structures are used to obtain a unique, invertible, relational and attributed graph representation of histological sections. The work also suggests that Dirichlet's domain can be extended to a more general Johnsons-Mehl domain. Another early application of the technique is found in [Sanfeliu 1981] for analyzing muscle tissue patterns using a distance measure between graphs. The syntactic structure analysis approach though theoretically powerful, has limited usage in practical applications of the present times due to its extensive computational requirements, which was also experimentally observed by scientists later. Subsequently, with development of digital pathology techniques, some more efficient graph-based methods were analyzed.

The commonly used graph-theoretic methods in histopathology are Voronoi diagrams, Delaunay triangulations and related graphs, cell graphs and attributed relational graphs. It has been stated in [Gurcan 2009] that a total of approximately 150 spatial-relational features have been extracted from all graph structures for histological images.

The most abundantly applied graph-theoretic techniques for representation and analysis of histological images till date are the Voronoi diagram, Delaunay graph and related graphs. Voronoi diagrams and Dirichlet domains have been explored in [Darro 1993] for growth assessment of degrees of differentiation in terms of cell population dynamics in Feulgen stained colorectal neoplastic cell colonies growing on histological slides, by studying the structure of clones in 11 different media. The authors in [Keenan 2000] describe a graph-based approach for analysis of H&E stained images of Cervical Intraepithelial Neoplasia, where Delaunay triangulation mesh is first computed on a region and 18 quantitative features are extracted. Discriminant analysis is applied to categorize image regions into normal or one of the three cancer grades. Minimum spanning tree has been used abundantly for quantitative representation of tissue architecture. For instance, H&E stained tissue sections of colorectal adenomas are analyzed in [Meijer 1992] using MST features to differentiate between three grades of dysplasia. In [vanDiest 1992], the authors analyze invasive breast cancer tissue images using 10 syntactic structure features extracted from corresponding MST of each H&E stained breast tissue region. A method for detecting lymphocytic infiltration in HER2 positive breast cancer images is described in [Basavanahally 2010] using features derived from Voronoi diagrams, Delaunay triangulations and minimum spanning trees.

Cell graphs have been introduced and extensively studied in [Gunduz 2004], [Demir 2004], [Demir 2005a], [Demir 2005b] and [Gunduz-Demir 2007] for analysis of H&E stained brain cancer (glioma) biopsy samples. In [Bilgin 2007], the authors use hierarchical cell graphs to model H&E stained breast tissue images. H&E stained histopathological images of bone tissue samples are analyzed using advanced version of cell graphs called *ECM-aware cell graphs* in [Bilgin 2010]. In the cell graph methods, the authors have used low magnification (80-100 times) tissue images to construct representative graphs, global graph metrics (*e.g.* average degree, clustering coefficient, average eccentricity, eigenvalues of graph) are extracted, and the images are classified with the help of neural networks or SVM classification methods. The cell graph

representation is a flexible method for histological images due to advantages over other geometric graph-based techniques, especially the use of fewer geometrical constraints for cancer tissue modeling. However, it has limitations when highly magnified histological images need to be analyzed, and are addressed later in the thesis.

Attributed relational graphs (ARG) are emerging as an appealing topic of interest for researchers in this field. Attributed minimum spanning trees are explored for representing Feulgen stained soft tissue tumors (malignant fibrous histiocyoma, fibrosarcoma, rhabdomyosarcoma, osteosarcoma and Askin tumor) in [Kayser 1991], where a basic graph is first constructed according to the neighborhood condition of O'Callaghan [O'Callaghan 1975]. Nuclear features related to morphometry and DNA-content are attributed to the vertices, and the differences between features of connected vertices are attributed to the corresponding edges. Resulting MST is decomposed into clusters using a suitable decomposition function on the edges, and clusters of distinct nuclear orientation are detected. Then a cluster tree is constructed by defining the geometric center of a cluster as a new vertex and by computing the neighborhood of the cluster vertices. Another study employing regional adjacency graphs using the ARG principle for content-based retrieval of H&E stained breast tumor tissue images is given in [Sharma 2012]. The authors have represented segmented tissue images using ARG, where nodes are centroids of tissue regions having attributes as a label from the segmentation and morphological features of the associated segment. Edges connect neighboring segments and are characterized by edge attributes which are Euclidean distance and perimeter of the common boundary between connecting nodes. A* based graph matching algorithm is employed to retrieve the image regions most similar to a given query region. In the study [Arslan 2013], attribute relational graphs are applied for model-based segmentation of cell nuclei in fluorescence microscopy images of hepatocellular carcinoma, where primitives are defined and the ARG is constructed on these on the basis of a set of rules. In [Sharma 2016], a part of this work has been described, with introduction of the cell nuclei attributed relational graphs for automatic classification of H&E stained gastric cancer regions into malignancy levels based on HER2 immunohistochemistry. In [Sharma 2017a], the work is extended using a comparative study of developed variations in the proposed graph-based method.

Other less popular graph-based methods include O' Callaghan neighborhood graphs [Kayser 1986], [Kayser 1988], cell cluster graphs [Ali 2013] and cell webs [Ficsor 2008]. Hybrid methods combining architectural information with low-level features have also been introduced in literature quite early such as [Kayser 1989] and [Kayser 1990]. A related work is [Weyn 1999], where Feulgen stained tissue sections of malignant mesothelioma, hyperplastic mesothelium and pulmonary adenocarcinoma are analyzed where a total of 82 features including texture features, morphometric features, densitometry features and graph-based features based on Voronoi diagram, Gabriel graph and MST are extracted from histological sections. Another recent study is found in [Doyle 2008], where H&E stained breast biopsy images are studied. A total of 3400 textural and architectural features are extracted. Texture features include gray level statistical features, Haralick features and Gabor features, and architectural features include

Voronoi diagram, Delaunay graph and MST-based features. In [Chekkoury 2012], H&E stained breast biopsy images are classified into cancerous and non-cancerous using morphometric, Fourier-based, Hessian-based, Varma-Zisserman textons and graph-based features.

Summarizing the evolution of graphs for image description in digital histopathology, the earliest methods range from the mathematically complex but less efficient tessellations like Dirichlet and Johnson Mehl tessellations, followed by studies on proximity graphs, especially Delaunay graphs and their subgraphs. Subsequently, graphs based on alternative neighborhood conditions like O’Callaghan and zone of influence were explored, promising an edge over the Voronoi-based graphs. Meanwhile, there was also an increased interest in syntactic pattern recognition methods deriving quantitative information from graph representations, mostly utilizing minimum spanning trees. However, with further developments in virtual microscopy techniques, a paradigm shift has occurred from the constrained (or geometric) methods utilizing proximity graphs to more relaxed (or flexible) application-specific graphs, for example cell graphs and attributed relational graphs applied on different types of tissues. Many authors have also used a combination of the low-level methods and high-level graph-based methods for analysis tasks. It can be concluded that graph-based techniques comprise an important direction in the field of histopathological image analysis, and can provide the basis for developing various automatic applications and tools to retrieve and classify tissue sections in a reliable way.

3.3 Machine Learning in Digital Histopathology

In this section, prominent literature in digital histopathology that incorporate image analysis using the four selected machine learning methods, namely, support vector machines, AdaBoost ensemble learning, random forests and deep convolutional neural networks are reported.

3.3.1 Support Vector Machines

Support vector machines have been explored extensively for supervised learning tasks in histological image analysis. Different SVM kernels such as linear, Gaussian and polynomial have been compared for optimizing SVM classification performance of hyper-spectral colon tissue [Rajpoot 2004]. Diagnosis of prostate cancer and Gleason grading in H&E stained histological images has been attempted using object-based features and supervised learning methods including SVM [Tabesh 2007]. Another work is described in [Altunbay 2010] using Delaunay graph representation followed by SVM to classify H&E stained colon tissue images into normal, low-grade and high-grade cancer. A pre-analysis component of this research involves the automatic detection of necrotic areas in H&E stained gastric cancer images, performed using textural features with SVM learning using discriminative thresholds is explained in [Sharma 2015c].

3.3.2 AdaBoost Ensemble Learning

AdaBoost ensemble learning is explored in [Doyle 2006] for classifying H&E prostate adenocarcinoma images by extracting texture and intensity features and training based on AdaBoost and decision tree classifiers. In [Yang 2009], breast cancer tissue microarray samples are classified using texture features, along with comparison of classical AdaBoost with its variations such as gentle AdaBoost, LogitBoost and other classifiers like Bayesian classifiers and SVM. Prostate cancer is also analyzed in a multiresolution approach by describing texture features of tissue images with AdaBoost ensemble learning in [Doyle 2012]. Automatic detection and grading in prostate cancer histopathological images is performed in [Gorelick 2013] using two stages of AdaBoost classification algorithm. Related to this work, cell nuclei classification in H&E stained gastric cancer images is studied using morphological characteristics and AdaBoost classifiers in [Sharma 2015b]. Additionally, gastric carcinoma images are classified on the basis of immunohistochemical response using a range of low-level and graph-based features along with their combinations followed by AdaBoost ensemble learning in [Sharma 2016].

3.3.3 Random Forests

Handcrafted image features can be used in conjunction with random forests in a supervised learning approach for various classification and detection tasks in digital histopathology. An example includes [DiFranco 2011], where co-occurrence color texture features and random forest feature selection and ensemble learning are applied for classification of prostate carcinoma in histological sections from radical prostatectomy. In [Sommer 2012], hierarchical learning workflow is proposed using random forest classifiers for automated mitosis detection in breast cancer images. Neutrophils are identified in H&E stained histological images by first constructing Voronoi diagrams followed by random forest binary classification in [Wang 2014]. In [Sharma 2017b], random forests are applied as the traditional machine learning method for comparative evaluation with deep learning, and in [Sharma 2017a], random forests hierarchical classification is performed for image analysis in H&E stained gastric cancer WSI.

3.3.4 Deep Convolutional Neural Networks

Deep learning using convolutional neural networks (CNN) is presently receiving tremendous attention in diverse fields of image analysis [LeCun 2015]. Earliest applications of deep convolutional neural networks include handwritten character recognition [LeCun 1989], [Lecun 1998] and general object categorization [Krizhevsky 2012], that are now expanding in different fields of image analysis.

Several recent advances have been proposed by scientists in the field of digital histopathology relating to deep learning methods. The introduction of deep learning in this area was enabled initially by application-specific challenges where large-scale labeled datasets have been made available publicly, for instance, the mitosis detection challenge in breast cancer [Veta 2015] and

gland segmentation challenge in colon histology images [Sirinukunwattana 2017]. In literature, a few examples of widely known studies of digital pathology using deep convolutional neural networks are as follows. The work [Ciresan 2013] shows highly favorable result for mitosis detection in breast cancer images using deep neural networks. A CNN method is proposed in [Cruz-Roa 2014] for classification of invasive ductal carcinoma in breast cancer WSI with results superior to handcrafted features. U-nets have recently become very popular for biomedical image segmentation [Ronneberger 2015]. These are fully convolutional neural networks, composed of a convolution block followed by a deconvolution block to generate segmented images as the output. The approach was initially used for segmenting neuronal structures in electron microscopy stacks, and was further applied to transmitted light microscopy images to perform cell tracking, depicting most superior results in the ISBI cell tracking challenge [ISBI 2016]. A patch-based convolutional neural network approach is examined in [Hou 2015], for discriminating glioma and non-small-cell lung carcinoma into respective subtypes in histological WSI. A comparative study of deep CNN architectures with handcrafted methods for classification of stromal and epithelial histological images of breast cancer and colorectal cancer is given in [Xu 2016]. The study [Sirinukunwattana 2016] describes a spatially constrained convolutional neural network for cell nuclei detection and classification in colon cancer images. Fully convolutional networks have been developed in [Chen 2016a], and this method won the MICCAI gland segmentation challenge [Sirinukunwattana 2017]. More research in this area includes [Spanhol 2016] for classification in breast cancer histopathological images, and [Litjens 2016] for prostate cancer classification in biopsy specimens and breast cancer metastasis detection in sentinel lymph nodes.

A part of this research has been presented in [Sharma 2017b] to contribute to this field by analysis of gastric cancer WSI, where a novel CNN architecture is proposed, quantitatively compared with handcrafted features followed by traditional machine learning and also with the AlexNet deep convolutional framework.

3.4 Image Analysis in Gastric Cancer

Medical image analysis of gastric cancer has not been extensively explored in previous literature of digital histopathology. Recent studies comprising of related work illustrate application-specific techniques for automatically analyzing the gastric tissue. For instance, gastric atrophy is quantitatively analyzed in H&E stained sections using syntactic structure methods in [Zaitoun 1998]. Image regions depicting normal mucosa, gastritis and adenocarcinoma in H&E stained histological sections are distinguished by use of cytometric measurements in [Ficsor 2006]. The ability of image processing methods to allow differentiation between normal and cancerous tissues in gastric cancer using confocal endomicroscopy has been strengthened in [Kakeji 2006]. A semi-supervised approach for detection and diagnosis of gastric cancer is described in [Cosatto 2013] using multiple instance learning in H&E stained tissue images. A multiresolution approach introduced in this work to improve cell nuclei segmentation of gastric cancer is explained in [Sharma 2015b]. The author has also explored necrosis detection in gastric carcinoma using

textural features and SVM-based classification in [Sharma 2015c]. The work in [Sharma 2016] involves graph-based analysis of H&E stained gastric carcinoma WSI based on their HER2 immunohistochemistry, with a comparative study with variations in [Sharma 2017a]. Additionally, the problem of cancer classification and necrosis detection are further addressed using deep convolutional neural networks in [Sharma 2017b].

3.5 Summary

This chapter mentions the related work in past and recent literature regarding medical image analysis in digital histopathology, especially, in the domains of feature extraction and image description methods, machine learning approaches and histopathological image analysis of gastric cancer.

Overview of Research Methodology

Contents

4.1	Introduction	47
4.2	Stage 1: Preparation of Materials	47
4.3	Stage 2: Image Pre-analysis	49
4.4	Stage 3: Analysis of Cancer Regions	49
4.5	Summary	50

4.1 Introduction

In this chapter, an overview of the workflow of the proposed research methodology and conducted experiments is presented, which allows the readers to attain a clear understanding of the following chapters. The schematic overview of the experimental pipeline is demonstrated in Figure 4.1. It consists of three main stages (or phases) named as preparation of materials, image pre-analysis and analysis of cancer regions. These are briefly described in the next sections and extensively discussed in the following chapters.

4.2 Stage 1: Preparation of Materials

In the first stage, histopathological whole slide image data is acquired and processed for subsequent image analysis tasks. It begins with the precise procedures for acquisition of tissue specimens from relevant sources, followed by the scanning process to generate digitized whole slide images. The next step includes image data labeling by expert pathologists for different modules of the framework, in order to create the *gold standard* or *ground truth* datasets. This stage also comprises intermediate steps such as visual quality assessment and whole slide image registration required to generate consistent image databases necessary for the succeeding analysis experiments.

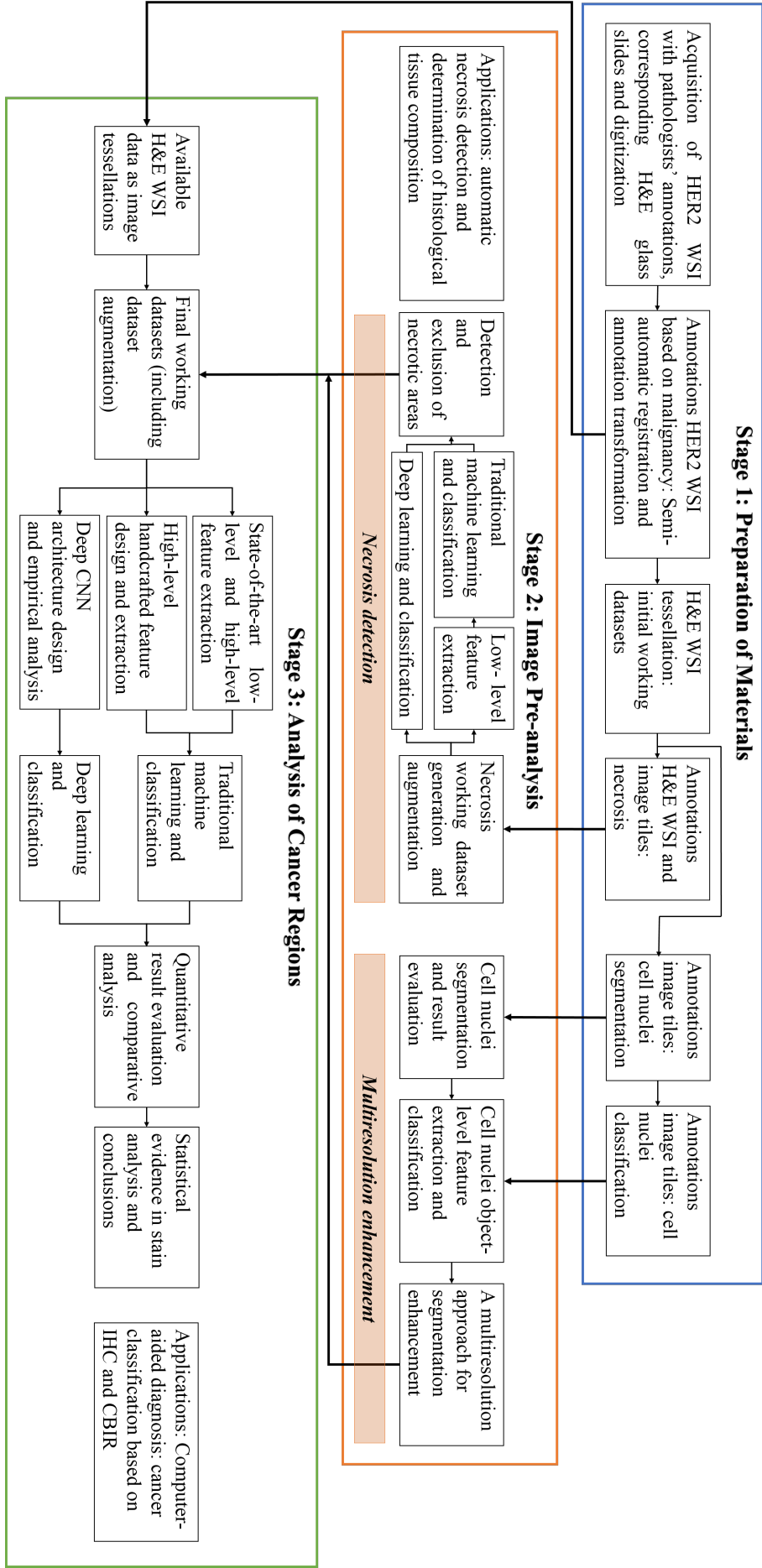


Fig. 4.1 Schematic overview of the experimental pipeline

4.3 Stage 2: Image Pre-analysis

As the name suggests, this stage incorporates certain image processing and analysis tasks which are a prerequisite before proceeding with the primary research objectives of this study. The necessity of these tasks arises out of thorough examination of the acquired whole slide image data, and the evaluation of classical image processing steps followed by refinements according to the particular nature of the histopathological images of gastric cancer.

A few challenges are observed during this stage, which are required to be addressed by suitable methods. These are as follows.

1. Necrosis was identified by expert pathologists in a small part of the acquired whole slide image data. So, these necrotic areas are required to be detected and excluded from the final working datasets before continuing with the analysis of cancer.
2. The working magnification is needed to be finalized in order to achieve optimal results for cell nuclei segmentation and subsequent image analysis tasks. This requires an elaborate multiresolution evaluation and may potentially lead to combination of information.

These challenges have led to the development of two main directions in the pre-analysis stage in the experimental pipeline, namely:

1. *Necrosis detection* using traditional machine learning and deep learning methods.
2. *Multiresolution enhancement* of image segmentation, which also includes developing a computerized cell nuclei classification procedure.

Outcomes of this stage are the procurement of histological image datasets after excluding necrotic areas, determination of the working magnification for further image analysis tasks, procedure for optimal isolation and categorization of cell nuclei, and the proposal of two applications in digital histopathology, namely, automatic necrosis detection and automatic determination of tissue composition in whole slide images of gastric cancer.

4.4 Stage 3: Analysis of Cancer Regions

In this stage, analysis of cancer regions is performed in H&E stained gastric cancer whole slide images based on the malignancy groups defined by the corresponding HER2 immuno-histochemical response. This is accomplished using a wide range of methods including novel analysis approaches, comparative evaluation with existing methods as well as modifications in the classical methods.

It starts with the generation of working datasets for region-based analysis of gastric cancer. This is followed by the design and implementation of handcrafted features and performance comparison with the state-of-the-art low-level and high-level features in the traditional analysis route. Additionally, deep learning is explored by design and empirical analysis of CNNs and the proposal of a suitable CNN architecture. The traditional machine learning and deep learning

methods are quantitatively and comparatively assessed using a comprehensive performance evaluation scheme for the described cancer classification problem. At the end, visual and statistical analysis of the image data and extracted features for the two staining methods is performed to conclude the experimental findings.

Outcomes of this stage are the accomplishment of research objectives for image analysis in H&E stained gastric cancer whole slide images in digital histopathology, and contribution to two additional applications, namely, computer-aided diagnosis consisting of cancer classification based on immunohistochemistry, and content-based image retrieval of regions of interest from WSI data.

4.5 Summary

In this chapter, the author has introduced and briefly described the three stages of the research framework, namely, preparation of materials, image pre-analysis and analysis of cancer regions.

Stage 1: Preparation of Materials

Contents

5.1	Introduction	51
5.2	Whole Slide Image Acquisition	52
5.2.1	Specimen Properties	52
5.2.2	Scanner Details	52
5.3	Annotations: HER2 whole slide images	54
5.3.1	Labeling Process	54
5.3.2	Semi-automatic WSI Registration	55
5.3.3	Annotation Transformation	57
5.4	Initial Working Datasets	59
5.4.1	Annotations for Cell Nuclei Segmentation Evaluation	60
5.4.2	Annotations for Cell Nuclei Classification	60
5.5	Annotations for Necrosis Detection	64
5.5.1	Datasets for SVM-based Method	66
5.5.2	Datasets for Deep Learning Methods	66
5.6	Summary	66

5.1 Introduction

This chapter describes in detail the procedures followed to prepare the materials required for experiments in histopathological image analysis. It includes the process of whole slide image acquisition, expert labeling for generation of ground truth including transfer of labeled information between the two histological stains, and organization of working datasets for subsequent image analysis tasks comprising this study.

5.2 Whole Slide Image Acquisition

5.2.1 Specimen Properties

Initially, twelve HER2 immunohistochemically stained and corresponding H&E stained surgical specimens are selected from clinical and oncological studies on 454 cases of gastric carcinoma [Warneke 2011], [Warneke 2013]. In the related previous studies, information from patients with gastric cancer was retrieved from archives of the Institute of Pathology, University Hospital Kiel, Kiel, Germany and their study cohort comprised of 454 all Caucasian patients undergone either total or partial gastrectomy for adenocarcinoma of the stomach or esophago-gastric junction between 1997 and 2009. The retrieved paraffin blocks were used to obtain 454 whole tissue sections. These sections are stained using the HER2 immunohistochemical method.

The prior selection of twelve whole slide sections followed by digitization was originally performed for the study described in [Behrens 2015], and the digitized WSI data along with the corresponding H&E glass slides were shared for this research. Hence, the selection process to retrieve the specified tissue sections was performed remotely in the Department of Pathology, Christian-Albrechts-University, Kiel, Germany by a group of medical experts and the author did not participate or effect the choice of the selected twelve tissue specimens.

To summarize, twelve tissue specimens have been acquired from proximal or distal parts of the stomach. These specimens belong to twelve patients with one specimen per patient. Tissue specimen acquisition is followed by specimen preparation procedure similar to the process described in Section 2.2.1.2, and glass slides are obtained using two staining methods, namely, HER2 immunohistochemical staining and haematoxylin and eosin staining. The glass slides were first prepared using HER2 immunohistochemical staining and non-consecutive sections from the same tissue block were later stained with H&E stain. The approximate thickness of the tissue sections is 4 μm . The underlying cancer grades based on HER2 immunohistochemical response of the tissue sections are identified by expert pathologists as 0, 1+, 2+, 3+, based upon immunostaining grades analogous to the rules described by Rüschoff et al [Rüschoff 2010].

5.2.2 Scanner Details

To digitize the acquired glass slides, two whole slide image scanners are employed. Initially, HER2 and H&E stained glass slides were scanned using a *Leica SCN400* microscopic whole-slide scanner at its maximum, nominally 400 times magnification with pixel resolution 0.26 μm per pixel at 40 \times objective magnification with quadratic pixels. Images were exported from the scanner system into files of Leica SCN format, which is a multi-image, pyramidal, multiresolution 64-bit TIFF format. This scanning process was conducted in Christian-Albrechts-University, Kiel, Germany and resulting digital whole slide images were shared electronically for the research.

However, on careful visual observation of the acquired H&E WSI by the author, pathologists and field experts along with the initial evaluation of cell nuclei image segmentation results, a few

issues were discovered as follows. Firstly, when a tissue section is scanned using a single focal plane, the WSI scanner focuses on one two-dimensional layer in the region of interest, however, as the tissue section is three-dimensional, all the observed cells are not simultaneously focused and result in unfocused neighboring cells, unclear boundaries and reduced sharpness in the digital whole slide image. The problem is enhanced by non-uniform illumination during the image acquisition. Due to these effects, the automatic segregation of cells is a tedious process leading to unsatisfactory segmentation results, as observed experimentally. Therefore, a re-scanning of the original H&E glass slides was recommended and later implemented using a more sophisticated scanning technique, as described below.

To overcome the challenges in the original H&E WSI data, eleven H&E glass slides were transported from Christian-Albrechts-University, Kiel, Germany to Charité University Hospital, Berlin, Germany. One corresponding H&E glass slide could not be physically acquired. The glass slides were re-scanned using *3DHistech Panoramic-250* whole-slide scanner, which has a higher resolution of $0.22\ \mu\text{m}$ per pixel at $40\times$ objective magnification with quadratic pixels. Additionally, it provides an extended depth of field [Bradley 2005] using an advanced digital image processing technique called *Focus Stacking* which combines multiple images taken at different focus distances. Hence, the resulting single in-focus whole slide image has a higher visual quality compared to the whole slide image generated using a single focal plane. The average size of each histopathological whole slide image is 13.65 Gigapixels. So, these WSI are high resolution with high sharpness and good visual quality. Further, the H&E WSI database can be called *heterogeneous* with variations in malignancy levels, stain intensities and inter-patient biological characteristics.

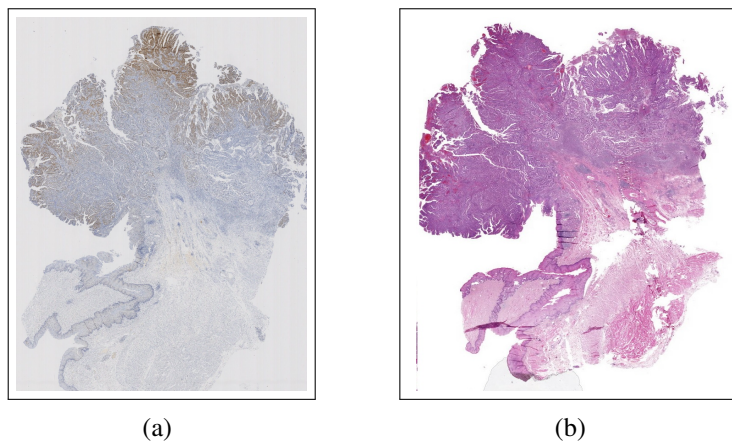


Fig. 5.1 Examples of corresponding sections in (a) HER2 and (b) H&E stains.

In conclusion, eleven H&E WSI acquired with the help of the second whole-slide scanner are used for further experiments. HER2 stained WSI image data is used later for statistical comparison in Chapter 9, otherwise only the pathologists' marked annotations of HER2 stained WSI are utilized for generating the ground truth and creating working databases in H&E stained WSI after registration and annotation transformation procedures. An example of a WSI pair in HER2 and H&E stain used in these studies is demonstrated in Figure 5.1.

5.3 Annotations: HER2 whole slide images

5.3.1 Labeling Process

HER2 expression and gene amplification in the described HER2 whole slide sections is analyzed in [Behrens 2015]. During this procedure, each HER2 immunohistochemically stained WSI is marked by ten expert pathologists with polygon annotations following a *10% cut-off rule* based on HER2 immunohistochemical response of the tissue. The pathologists have remotely marked annotations with the help of a virtual microscopy software [Behrens 2015]. A screenshot of the program is depicted in Figure 5.2. It mainly consists of a viewer application to display whole slide images with options for user interaction and assistance tools. The pathologists' annotations generated in this process are provided along with each HER2 WSI for this research.

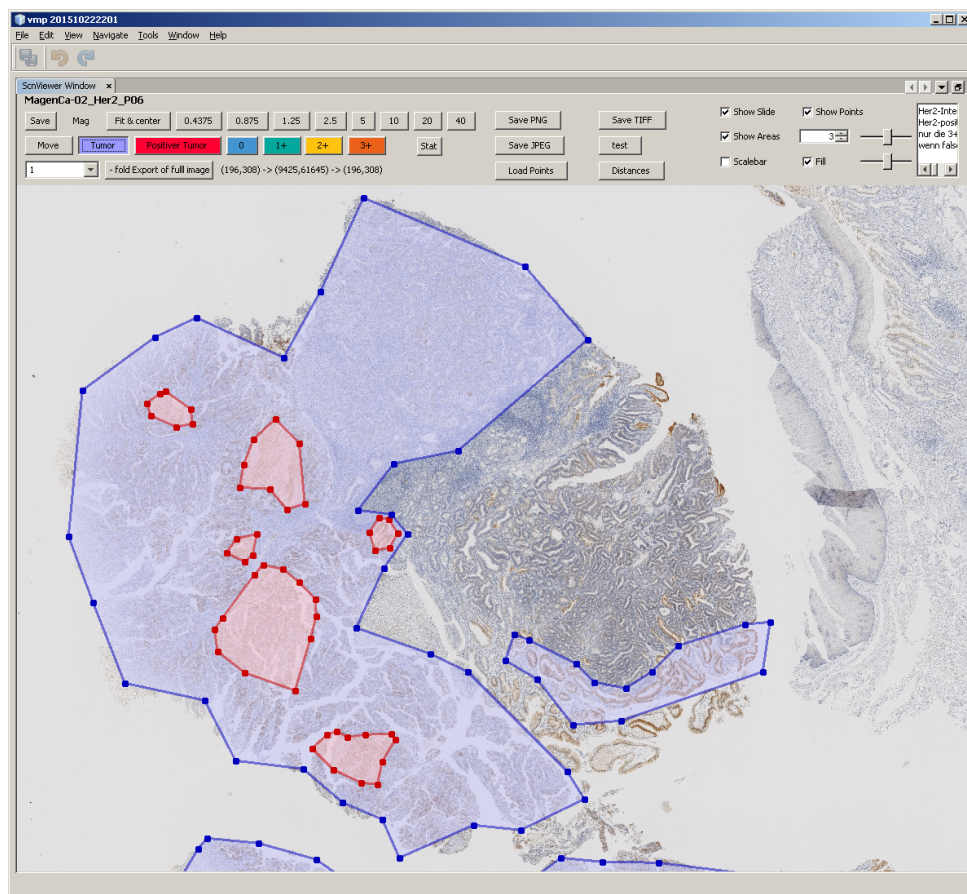


Fig. 5.2 Screenshot of virtual microscopy program used by pathologists to create annotations in HER2 WSI based on immunohistochemical response

The HER2 positive tumor areas are marked according to the 10% cut-off rule [Behrens 2015], which consist of gastric cancer grades 2+ and 3+ and are mostly visually distinguishable from the remaining tissue due to distinct brown colored staining. Additionally, HER2 negative areas are also marked, which consist of gastric cancer grades 0 and 1+ and don't show a distinct stain response, however, these are morphologically identified as tumor areas by the pathologists. It can be noted that HER2 positive areas define a higher degree of malignancy, whereas

HER2 negative areas denote a lower malignancy level. Inter- and intra-observer variability and pathologists' behavior for observing the HER2 stained WSI dataset has been discussed in detail in [Behrens 2015]. The average number of pixels per annotation for HER2 positive tumor are 231.7×10^6 and for HER2 negative tumor are 1279.0×10^6 . The distribution of number of HER2 positive and HER2 negative annotations for each expert pathologist and each slide is shown in Table 5.1.

The HER2 WSI are converted from the scanner format into Virtual Slide Format (VSF) which is a compressed image format consisting of a folder with several files to support a faster access on single whole slide image. The corresponding annotations are converted to VMSM files which are XML-formatted metafiles containing metadata about the slides including the pathologists' annotations. These data formats are suitable for accessing whole slide image data and related annotations with the help of *VMscope* software support [VMscope 2010b]. In order to create the ground truth using H&E stained WSI, a semi-automatic registration and annotation transformation procedure is first applied to transform these polygon annotations from each HER2 WSI to the corresponding H&E WSI. This procedure is described in the following sections.

5.3.2 Semi-automatic WSI Registration

Image registration is a widely used digital image processing method of aligning two or more images of the same scene. This process involves designating one image as the reference (also called the reference image or the fixed image), and applying geometric transformations to the other image(s) so that they align with the reference [Zitova 2003]. In this work, whole slide image registration is required to transform annotations of pathologists from HER2 WSI into corresponding annotations in H&E WSI for further experiments, as H&E is the working stain for reasons explained earlier. Spatial differences in position and orientation between corresponding whole tissue sections in two stains are not negligible, hence, WSI pairs are first semi-automatically registered to generate spatial correspondence data between pairs of HER2 and H&E whole slide images.

The Virtual Slide Processing Environment (*ViSPEe*: details in Appendix A.4) was applied for creating spatial correspondences between pairs of whole slide images of the same tissue in different stains. The program helps the user to manually generate a set of control points in the WSI pair, allowing visual selection of common features in each image to map to the same pixel location. It also creates Delaunay triangulation using the selected control points as seeds, which is utilized during annotation transformation procedure. The triangulation edges are called *smart links* between control points in the same WSI, and this feature also aids the user to visualize the relative positions of control points in order to create good distribution indicated by equal sized Delaunay triangulations to cover the entire WSI. The output of the process is an XML file containing correspondences in the form of pixel positions and Delaunay triangulation information, for each pair of whole slide images. Screenshot of the *ViSPEe* program for semi-automatic registration in an example WSI pair along with the selected control points and Delaunay links is shown in Figure 5.3.

Table 5.1 Distribution of number of pathologists' annotations marked in HER2 WSI on the basis of immunohistochemical response (P: pathologist, S: slide, H+: number of HER2+ tumor annotations, H-: number of HER2- tumor annotations)

	P1		P2		P3		P4		P5		P6		P7		P8		P9		P10		Annotations per S	
	H+	H-	H+	H-	H+	H-	H+	H-	H+	H-	H+	H-	H+	H-	H+	H-	H+	H-	H+	H-	H+	H-
S1	5	6	5	9	1	2	1	5	3	4	5	10	3	7	4	5	3	4	2	4	32	56
S2	6	10	5	10	4	7	2	4	6	10	17	22	6	10	4	8	15	18	10	14	75	113
S3	3	6	5	7	2	3	13	14	3	4	15	24	3	7	3	5	2	5	4	10	53	85
S4	1	4	1	2	2	5	1	2	1	2	2	3	3	4	1	2	16	18	1	2	29	44
S5	3	4	3	4	1	2	2	3	1	2	3	4	1	2	2	4	4	5	3	4	23	34
S6	2	4	3	4	1	2	2	3	1	2	1	2	1	3	1	2	1	2	3	5	16	29
S7	3	5	5	7	4	5	2	3	2	3	5	7	5	6	5	6	4	5	4	5	39	52
S8	3	4	3	8	2	3	2	3	1	2	10	11	3	4	2	3	1	2	3	4	30	44
S9	2	3	2	5	2	3	4	5	2	3	3	7	4	5	2	3	2	3	6	10	29	47
S10	1	4	2	3	1	2	2	3	3	4	1	2	1	6	1	2	3	4	1	2	16	32
S11	3	15	1	2	1	2	2	3	1	2	5	6	4	5	1	2	1	2	1	16	20	55
Annotations per P	32	65	35	61	21	36	33	48	24	38	67	98	34	59	26	42	52	68	38	76	362	591

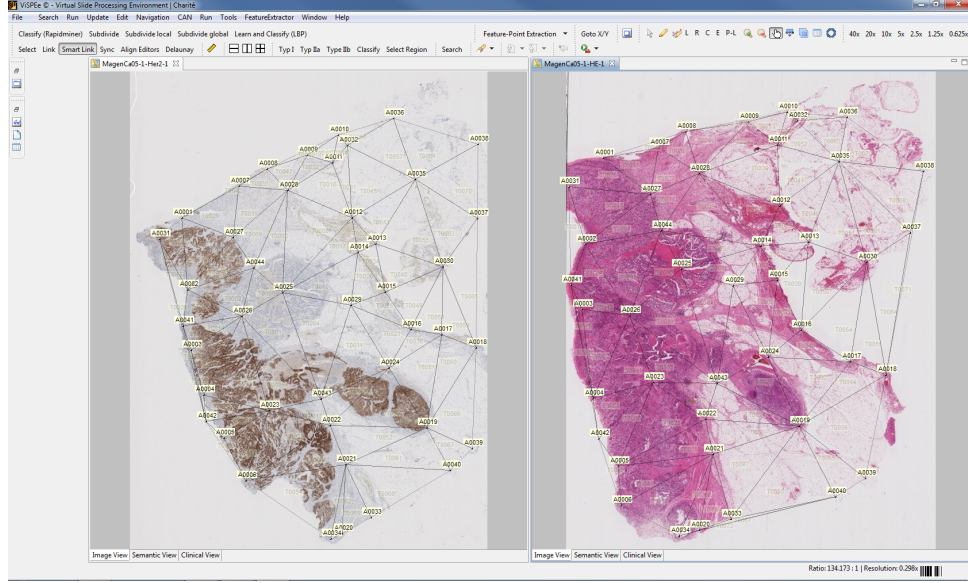


Fig. 5.3 Screenshot of ViSPee program for semi-automatic registration in one HER2 and H&E WSI pair

5.3.3 Annotation Transformation

Using the stored WSI registration information, source coordinates of polygon annotations of pathologists in HER2 WSI are transformed into destination coordinates in corresponding H&E WSI. The annotation transformation method requires a mapping $f : \mathbb{R}^2 \rightarrow \mathbb{R}^2$ for bivariate two-valued data. It is achieved using two transformation algorithms depending on the position of each polygon vertex in the reference polygon annotation of HER2 WSI with respect to the Delaunay triangulation created during the semi-automatic registration. A local affine transformation is applied for all the polygon points inside the convex hull of the control point cloud, and global rigid transformation is applied for all points lying on or outside the convex hull of control points, as explained below.

5.3.3.1 Interpolation using Local Affine Transformation

If the source point lies inside one of the Delaunay triangles, it is mapped to its destination point using an affine transformation function [Amidror 2002]. Suppose the source point P for which we want to estimate the mapping is situated at $\mathbf{x} = (x, y)$, inside the Delaunay triangle with vertices P_1 at $\mathbf{x}_1 = (x_1, y_1)$, P_2 at $\mathbf{x}_2 = (x_2, y_2)$ and P_3 at $\mathbf{x}_3 = (x_3, y_3)$ of the reference dataset. The points and associated vectors can be represented in affine coordinate system as shown in Figure 5.4a. From the figure, it can be observed that,

$$\begin{aligned} \mathbf{x} &= \mathbf{x}_1 + \mathbf{x}' \\ &= \mathbf{x}_1 + a_2(\mathbf{x}_2 - \mathbf{x}_1) + a_3(\mathbf{x}_3 - \mathbf{x}_1) \end{aligned} \quad (5.1)$$

where a_2, a_3 are the coordinates of the point \mathbf{x} in the new coordinate system such that $0 \leq (a_2, a_3) \leq 1$. Expanding equation 5.1, we get,

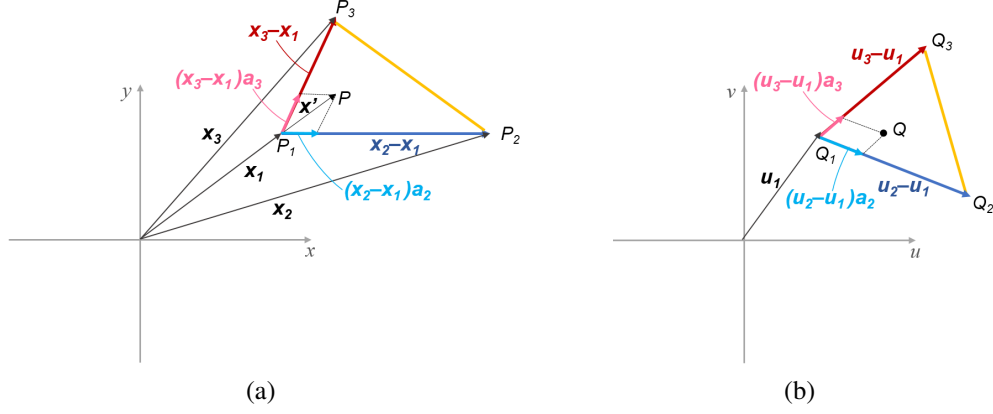


Fig. 5.4 Mapping by linear triangular interpolation of (a) triangle $P_1P_2P_3$ in the x, y plane to (b) triangle $Q_1Q_2Q_3$ in the u, v plane.

$$\begin{aligned} \begin{pmatrix} x \\ y \end{pmatrix} &= \begin{pmatrix} x_1 \\ y_1 \end{pmatrix} + \begin{pmatrix} x_2 - x_1 \\ y_2 - y_1 \end{pmatrix} a_2 + \begin{pmatrix} x_3 - x_1 \\ y_3 - y_1 \end{pmatrix} a_3 \\ &= \begin{pmatrix} x_1 \\ y_1 \end{pmatrix} + \begin{pmatrix} x_2 - x_1 & x_3 - x_1 \\ y_2 - y_1 & y_3 - y_1 \end{pmatrix} \begin{pmatrix} a_2 \\ a_3 \end{pmatrix} \end{aligned} \quad (5.2)$$

Using the above equation, (a_2, a_3) can be obtained as

$$\begin{pmatrix} a_2 \\ a_3 \end{pmatrix} = \begin{pmatrix} x_2 - x_1 & x_3 - x_1 \\ y_2 - y_1 & y_3 - y_1 \end{pmatrix}^{-1} \begin{pmatrix} x - x_1 \\ y - y_1 \end{pmatrix} \quad (5.3)$$

Let the mapping of the three Delaunay vertices in the destination dataset be represented by $\mathbf{u}_1 = (u_1, v_1)$, $\mathbf{u}_2 = (u_2, v_2)$ and $\mathbf{u}_3 = (u_3, v_3)$ in the u, v plane for the source points \mathbf{x}_1 , \mathbf{x}_2 and \mathbf{x}_3 respectively. Using linear triangulation interpolation, the mapping of point $\mathbf{x} = (x, y)$ will be a point $\mathbf{u} = (u, v)$ inside the destination Delaunay triangle, that has the same relative coordinates a_2, a_3 with respect to the new coordinate system. This new point and co-ordinate system is shown in Figure 5.4b. Using analogy with equations 5.1 and 5.2, we get,

$$\mathbf{u} = \mathbf{u}_1 + a_2(\mathbf{u}_2 - \mathbf{u}_1) + a_3(\mathbf{u}_3 - \mathbf{u}_1) \quad (5.4)$$

On expanding equation 5.4 and substituting from equation 5.3, the final value of (u, v) is obtained as,

$$\begin{pmatrix} u \\ v \end{pmatrix} = \begin{pmatrix} u_1 \\ v_1 \end{pmatrix} + \begin{pmatrix} u_2 - u_1 & u_3 - u_1 \\ v_2 - v_1 & v_3 - v_1 \end{pmatrix} \begin{pmatrix} x_2 - x_1 & x_3 - x_1 \\ y_2 - y_1 & y_3 - y_1 \end{pmatrix}^{-1} \begin{pmatrix} x - x_1 \\ y - y_1 \end{pmatrix} \quad (5.5)$$

The computation is repeated for all source polygon points lying inside the Delaunay triangulations generated during semi-automatic registration, in order to obtain the transformed destination polygon points.

5.3.3.2 Interpolation using Global Rigid Transformation

When a source point is not situated inside any of the Delaunay triangles in the reference dataset, it is mapped globally with the help of rigid transformations using Singular Value Decomposition (SVD) method [Sorkine 2009]. A very few polygon points are lying outside the boundary of the tissue section and are situated out of the convex hull of the control point cloud (or outside the Delaunay triangles). These points are interpolated using the global rigid transformation method, however, it is scarcely used compared to the local affine transformation method.

Example of a WSI pair with original and resulting polygon annotations after semi-automatic registration and annotation transformation procedure are shown in Figure 5.5. Red polygons represent marked HER2 positive areas, and blue polygons represent marked HER2 negative areas.

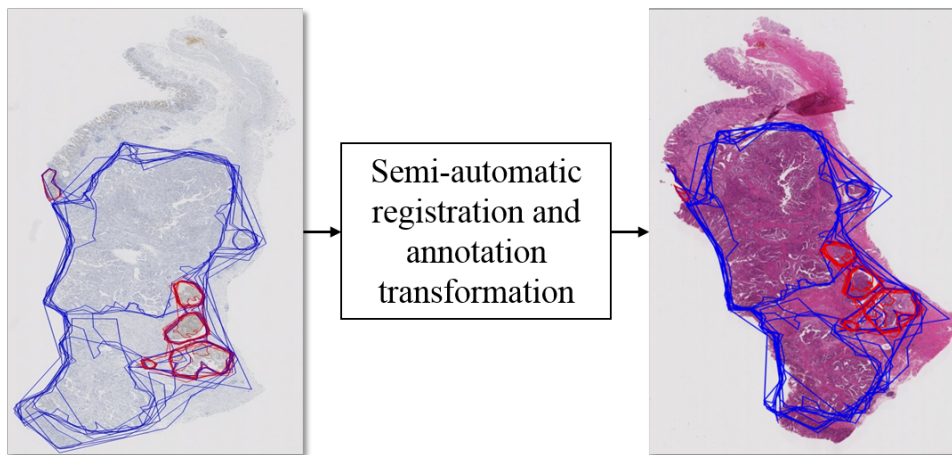


Fig. 5.5 Example of HER2 and H&E WSI pair containing original and resulting pathologists' annotations after semi-automatic WSI registration and annotation transformation procedure

5.4 Initial Working Datasets

Initially, smaller working datasets are generated using a fraction of the available H&E WSI data to perform image pre-analysis experiments in the second stage of the research framework, namely, cell nuclei segmentation and evaluation, cell nuclei classification, multiresolution enhancement and determination of tissue composition (partly in necrosis detection). The reason to generate initial working datasets is to lower the time requirements for creating the ground truth consisting of manual annotations by expert pathologists, and also because the involved tasks mainly require labeling of cell nuclei as data instances, which can be obtained in sufficient quantities using fewer large-sized WSI. Hence, five H&E stained WSI specimens are considered from the entire available WSI data for this purpose. Regions of interest are first selected for generating image

tessellations using the polygon annotations of pathologists, and consist of three main types of regions based on malignancy, namely, HER2 positive tumor, HER2 negative tumor and non-tumor. Then, these regions of interest are tessellated at different objective magnifications ranging from $10\times$ to $40\times$ to generate *non-overlapping* image tiles, with the tiles at highest magnification of size 1024×1024 pixels. As pathologists' annotations are not completely overlapping, the selected tissue areas are enclosed in maximum annotations, ensuring agreement of most pathologists to minimize inter-observer variability.

From the generated image tessellations, five image tiles are selected from each type of malignancy region for each WSI. A total of $5(\text{number of tiles per region type})\times 3(\text{number of region types})\times 5(\text{number of WSI}) = 75$ image tiles comprise the initial working dataset, where each malignancy type is represented by 25 tiles. This step results in evenly distributed initial working datasets with one-third (33.33%) representation of each type of malignancy. The tiles are selected such that they contain variation in stain intensity and malignancy levels, introducing heterogeneity in the resulting image datasets. Using these initial working datasets, two main labeling tasks are performed *i.e.* creating annotations for cell nuclei segmentation evaluation and cell nuclei classification, which are later used for multiresolution enhancement and determining tissue composition. Additionally, a part of these datasets are included for the labeling of necrotic regions (as necrosis was initially discovered by expert pathologist in this dataset). *CognitionMaster* [Wienert 2013], an object-oriented analysis framework (details in Appendix A.4), is applied for user interaction with the tissue images for the two cell nuclei labeling tasks.

5.4.1 Annotations for Cell Nuclei Segmentation Evaluation

For quantitatively evaluating the results of cell nuclei segmentation algorithm, non-overlapping tiles containing unique cell nuclei are required, which are generated in the initial working datasets. Haematoxylin-stained cell nuclei in the 75 image tiles are manually labeled using point annotations, first on an individual basis and then presented to an expert pathologist for validation. The labeling is performed using *ROIManager*, a utility plugin in the *CognitionMaster* program. The point annotations are carefully created by marking the centroid of each cell nucleus. Each image tile is annotated at highest resolution *i.e.* size 1024×1024 pixels. The resulting annotation information of each image tile is saved as a text file, analyzed later during quantitative evaluation of cell nuclei segmentation and enhancement. A screenshot of the plugin used for labeling process, with yellow '+' marks as the point annotations, is shown in Figure 5.6. A total of 25,257 manual point annotations are obtained for the 75 image tiles in the initial working datasets. Their distribution among the three types of regions is shown in Figure 5.7.

5.4.2 Annotations for Cell Nuclei Classification

Non-overlapping image tiles containing unique cell nuclei are also required to obtain reference data for cell nuclei classification, hence, initial working datasets are used for creating the ground

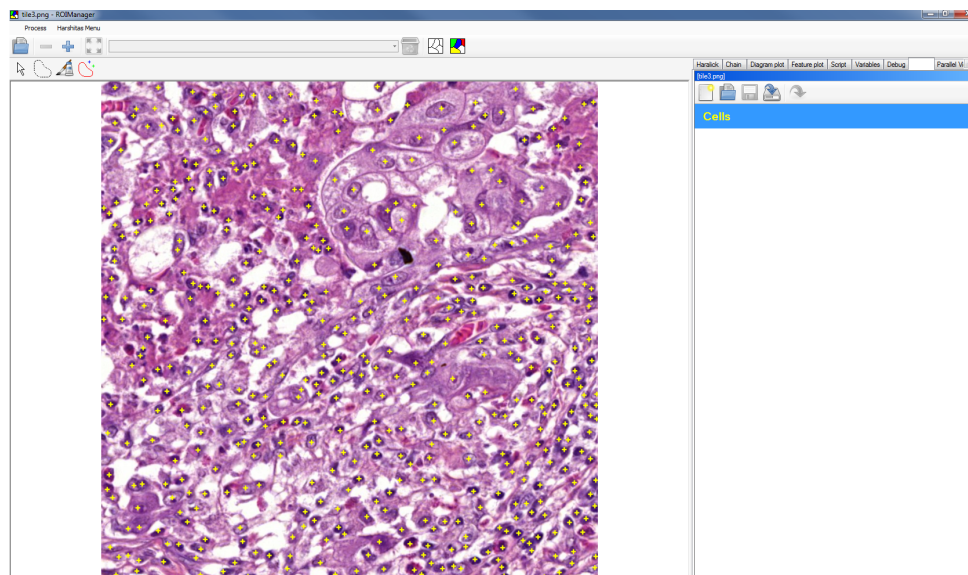


Fig. 5.6 Screenshot of ROIManager program with cell nuclei annotations (yellow '+' marks) made in an image tile at the highest resolution

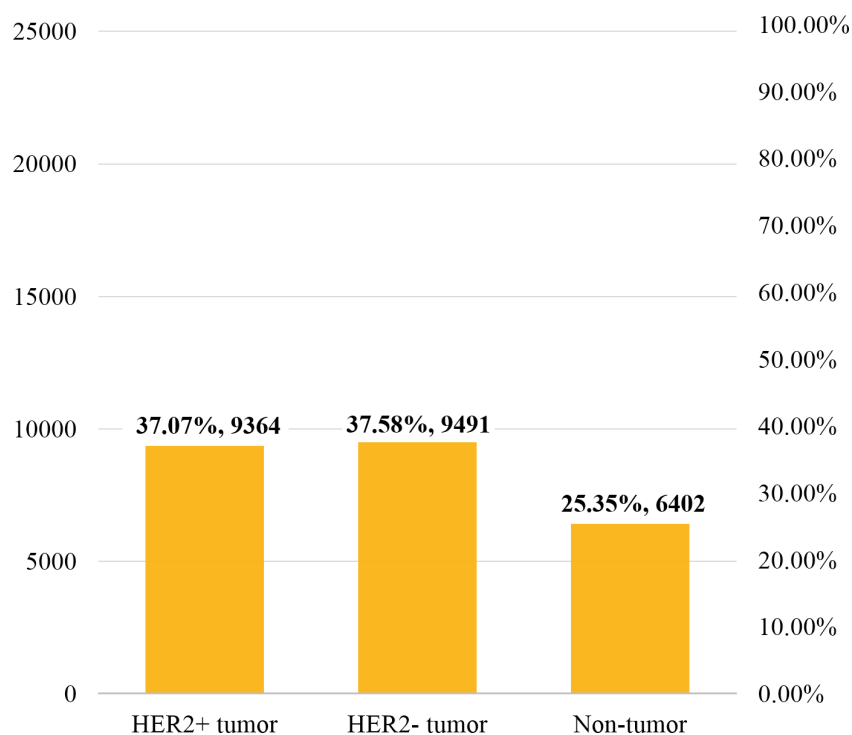


Fig. 5.7 Distribution of point annotations among the three types of regions (number and percentage of total annotations)

truth data for this purpose. The inputs for this step are cell nuclei contours obtained the result of automatic segmentation of the tissue images. This step involves the manual labeling of previously segmented cell nuclei contours with one of the seven cell nuclei types or classes. The seven cell nuclei classes have been suggested by expert pathologists and are named as *epithelial cell*, *leukocyte*, *fibrocyte (or border cell)*, *conglomerate*, *fragment*, *other cell (including blood cell in vessel)* and *artefact*. Figure 5.8 illustrates the seven defined cell nuclei classes.

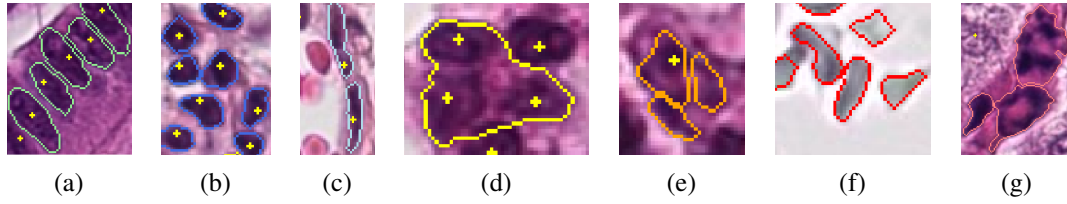


Fig. 5.8 Defined cell nuclei classes (a) Epithelial cell (b) Leukocyte (c) Fibrocyte (d) Conglomerate (e) Fragment (f) Other cell (including blood cell in vessel) (g) Artefact.

The defined cell nuclei classes are explained as follows. **Epithelial cells** are either elliptical or round shaped, typically larger than leukocytes with a distinct texture and characteristic of tumor-affected areas. **Leukocytes** are deeply stained round cells. **Fibrocytes** are long shaped cells, which are generally found as border cells of vessels, and may also include smooth muscle cells and fragments of these types of cells. **Conglomerates** consist of segments containing more than one cell and badly segmented cells such as a parts of multiple cells. **Fragments**, as the name suggests, are the segments which are only parts of a cell nucleus. **Other cells** include the cells in the tissue which are not stained by haematoxylin, *e.g.* blood cells inside a vessel, hence, not marked by point annotations. **Artefacts** are segments *i.e.* not a cell nucleus but may appear to be like one, also not marked by point annotations.

Nearly one-half of the segmented image data from the initial working datasets is manually labeled for creating ground truth for cell nuclei classification. This includes a total of 33 image tiles from the five H&E WSI at $40\times$ objective magnification. The same image tiles are also labeled at $30\times$ objective magnification (768×768 pixels), as this magnification shows potential of good segmentation results which can be utilized in multiresolution combination for segmentation enhancement (details in Section 6.3). This has resulted in manual generation of cell nuclei annotations in a total of 66 image tiles.

Cell nuclei segments are labeled first by the author, which are later reviewed by an expert pathologist. A plugin called *Object-Manager* is used in conjunction with *CognitionMaster* for the manual labeling of cell nuclei segments. This program can be operated by the user in two modes, namely, ‘classify mode’ and ‘review mode’. In the classify mode, each contour is labeled for the first time and the resulting annotations are saved along with sample contour information in an XML file, of the same name as the image tile, called *learning sample file* (LSF). In the review mode, classes of previously labeled contours can be modified, and the corresponding LSF is updated. A screenshot of the plugin in review mode, consisting of already labeled contours (different colors depicting different cell nuclei classes) in an image tile at $40\times$ objective magnification is shown in Figure 5.9. A total of 17,516 cell nuclei segments in $30\times$ and 25,848 cell nuclei segments in $40\times$ have been annotated as one of the seven categories using the described method. The distribution of the manual annotations in various classes of the cell nuclei segments in the three type of regions, and total annotations in each type of region is demonstrated in Figure 5.10(a) for $30\times$ and Figure 5.10(b) for $40\times$ objective magnification, respectively.

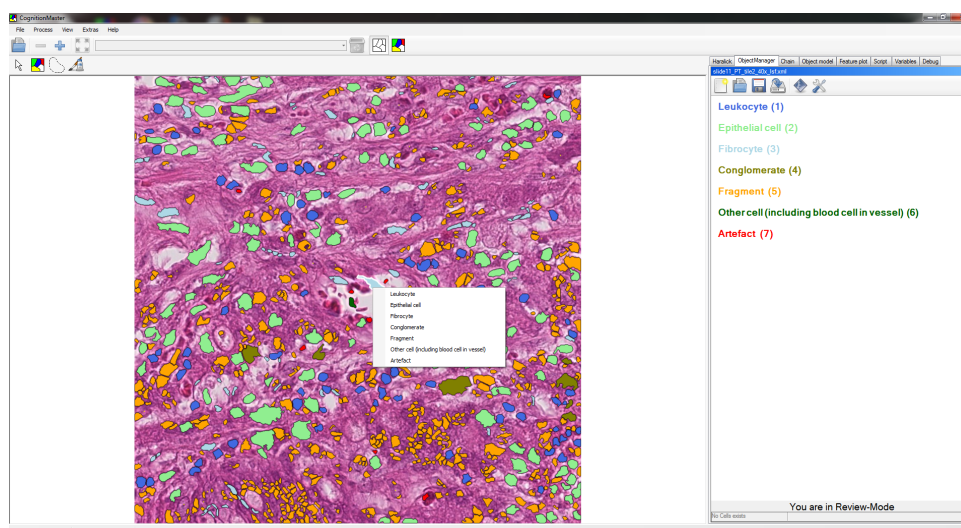


Fig. 5.9 Screenshot of the Object-Manager program in review mode showing labeled contours in an image tile at highest resolution

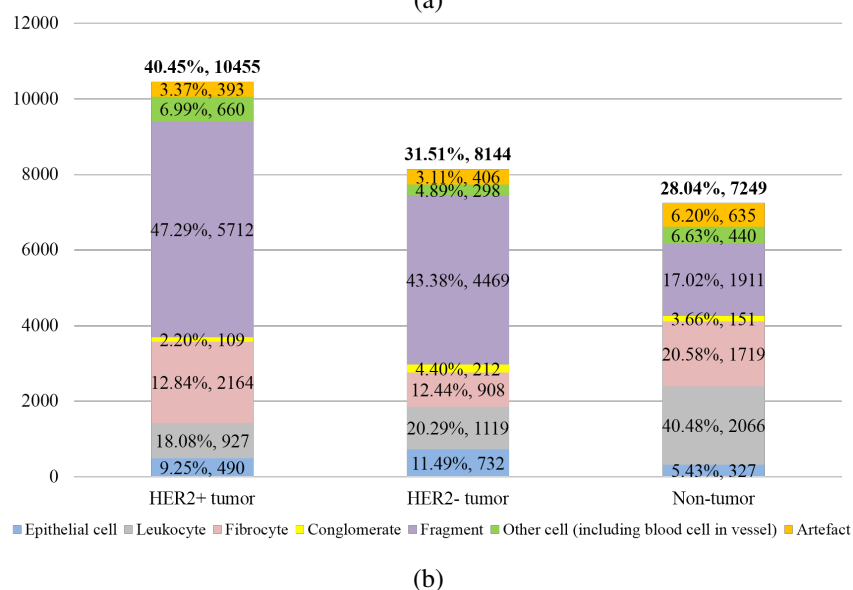
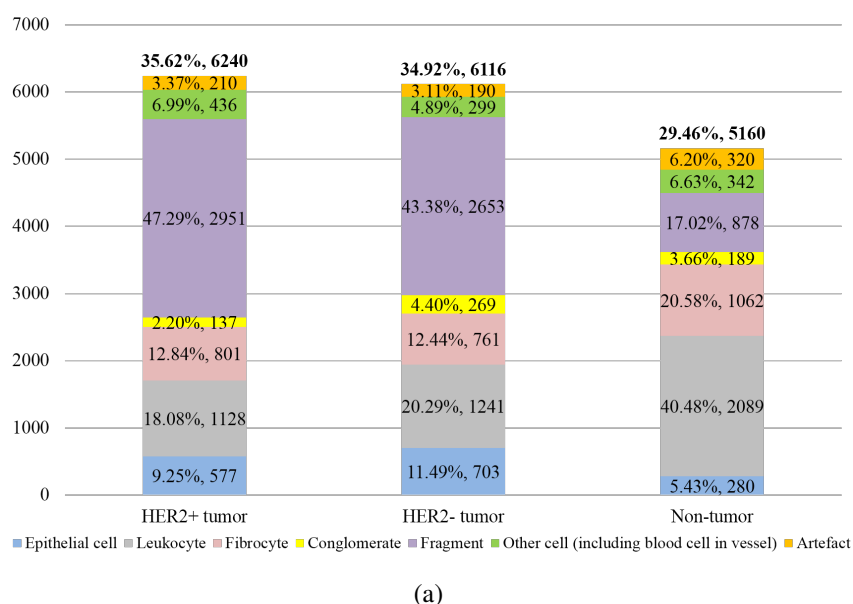
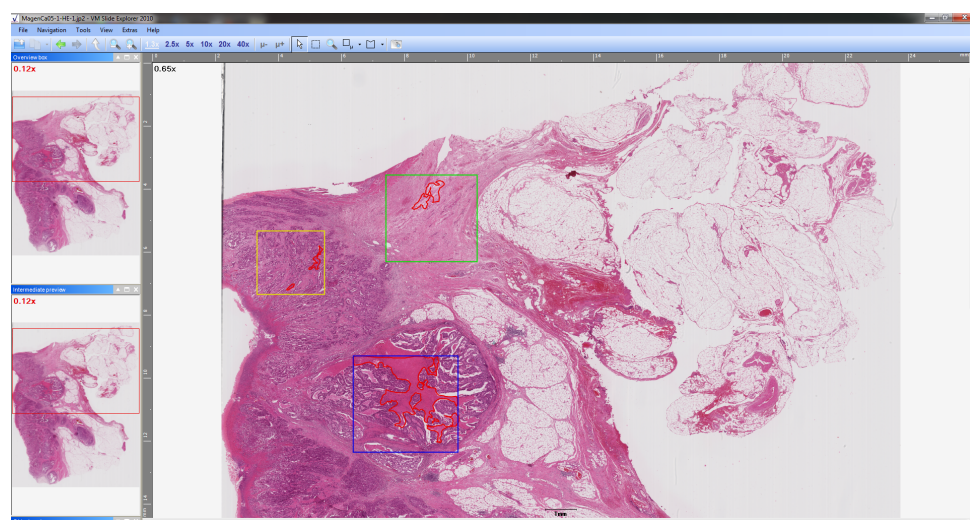


Fig. 5.10 Distribution of cell nuclei annotations for seven classes in each of the three types of regions and total annotations at (a) 30 \times objective magnification (b) 40 \times objective magnification.

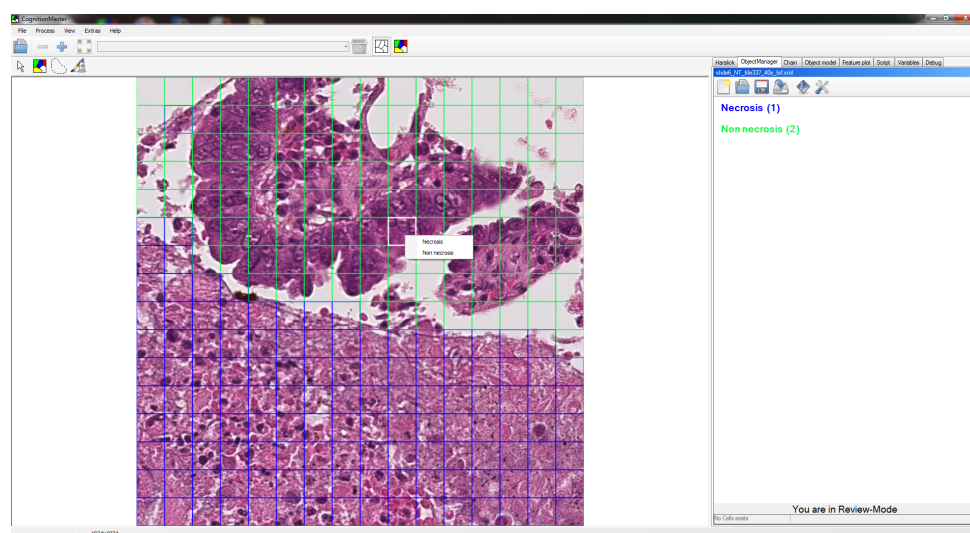
5.5 Annotations for Necrosis Detection

One expert pathologist has observed and marked necrotic areas to create the ground truth for necrosis detection in two distinct ways, namely, at a low magnification and at the highest magnification. This labeled information is then used to generate three datasets for the initial experiments consisting of non-overlapping image tiles of different sizes. The three initial working datasets for necrosis detection are explained as follows.

1. **Annotations in WSI - splitting approach:** To generate the first dataset, rectangular regions of interest (ROI) are selected in two whole slide images according to three groups, namely, HER2 positive tumor, HER2 negative tumor and non-tumor, depending on the associated HER2 polygon annotations. All the WSI with rectangular ROIs are presented to an expert pathologist, who has marked polygon annotations of necrotic areas at the WSI level after visual assessment in predefined regions of interest. On visual inspection, it is found that the marked areas are larger and more prominent in two WSI for the given rectangular ROIs. These larger necrotic areas and surrounding non-necrotic areas are split into non-overlapping image tiles of specific sizes at the highest magnification. Hence, this dataset has been generated using a *splitting approach*, and includes well-defined images of the two types. Average number of pixels per polygon annotation labeling necrosis in the WSI are 48.01×10^6 . The rectangular regions of interest and polygon necrotic annotations are marked using the VMscope program *VM Slide Explorer* [VMscope 2010b]. A screenshot of the *VM Slide Explorer* program and an example WSI showing rectangular regions of interest in the three types of malignancy groups (HER2 positive: yellow, HER2 negative: blue and non-tumor: green), and necrotic polygons marked inside them (red) at low magnification ($0.65\times$) is depicted in Figure 5.11(a).
2. **Annotations in image tiles - merging approach:** The second dataset contains image tiles from four whole slide images, and are a subset of the initial working dataset of size 1024×1024 belonging to the three malignancy groups. The pathologist has labeled the image tiles at the highest magnification *i.e.* at $40\times$, with each square annotation is made at the smallest size of 64×64 pixels. Labels are mapped from smaller to larger tiles using the information at the smallest size. Class of a larger sized image tile assigned as the one with maximum instances of constituent smaller image tiles. Hence, this dataset is created using a *merging approach* and possesses a higher visual complexity. *Object-Manager* plugin of the *Cognitionmaster* software (explained in Section 5.4.2) is applied to create annotations for this purpose. A screenshot of the *Object-Manager* plugin and an example of image tile from a non-tumor region with square annotations created at the smallest tile size is shown in Figure 5.11(b), where necrotic tiles are marked in blue squares and non-necrotic in green squares.
3. **Combined dataset:** The third dataset contains the combination of first and second datasets by blending the ground truths created by expert pathologist using two different approaches. It is considered comprehensive as it contains heterogeneity in terms of visual distinction of image tiles with respect to stain intensity, malignancy levels and biological variations, required in order to develop a potentially robust necrosis detection system. Image tile sizes and number of image tiles of each size in the respective datasets are shown in Table 5.2.



(a)



(b)

Fig. 5.11 Screenshots of (a) VM Slide Explorer program and an example WSI showing rectangular regions of interest in the three types of malignancy groups (HER2 positive: yellow, HER2 negative: blue and non-tumor: green), and necrotic polygons marked inside them (red) at $0.65\times$ objective magnification (b) Object-Manager plugin and an example of image tile from non-tumor region with square annotations created at the smallest tile size, where necrotic tiles are marked with blue squares and non-necrotic in green squares at $40\times$ objective magnification.

Table 5.2 Number of image tiles of different sizes in the three datasets for necrosis detection using SVM-based method

Image tile size	Using splitting approach	Using merging approach	In combined dataset
64×64	7680	3790	11470
128×128	1920	1039	2959
256×256	480	291	771
512×512	120	78	198
1024×1024	30	18	48

5.5.1 Datasets for SVM-based Method

The combined dataset of image size of 512×512 pixels is used as the final working database for necrosis detection using traditional machine learning with SVM, with a total of 198 non-overlapping image tiles. It must be specified that the selection of this size is due to the following reasons. Firstly, when the specified classification methods (described in Section 6.2) are applied on the combined dataset specified in Table 5.2, most favorable results are achieved using image sizes 512×512 and 64×64 . These observations have been discussed in detail in [Sharma 2015c]. The ROC curves obtained for the combined dataset is shown in Figure 5.12(a), clearly suggesting that the choice of the two sizes for image tessellations. As a larger field of view is always significant and preferable for analysis of histological images, the larger size of 512×512 pixels is considered. Another reason is the uniformity of image size for all methods, including deep learning. The distribution of the image tiles among the three types of malignancy regions is shown in Figure 5.12(b). It can be observed that the image tiles are generated mostly from the HER2 negative tumor and non-tumor malignancy regions, as smaller necrotic areas have been recognized by expert pathologist in the HER2 positive tumor regions.

5.5.2 Datasets for Deep Learning Methods

The final working database for necrosis detection using deep learning methods also consists of image tiles of the most convenient and performance optimized pixel size *i.e.* 512×512 pixels. Further, the size is consistent with the deep learning method applied for the analysis of cancer regions. The dataset has been obtained using data augmentation methods on the existing non-overlapping images, (detail in Section 7.2). Using this process, 47,128 image tiles are generated and their distribution among the three types of malignancy regions is shown in in Figure 5.13. Data augmentation has been performed using the initial image tiles of size 1024×1024 to generate nearly uniform datasets with images belonging to HER2 negative tumor and non-tumor malignancy regions, due to smaller necrotic areas in the HER2 positive tumor regions not sufficient to create image tiles at the given pixel size.

5.6 Summary

This chapter provides a detailed description of the first stage of the research framework, *i.e.*, preparation of materials. In this stage, WSI data has been acquired from relevant sources, labeled with the help of domain knowledge of medical experts, and processed and organized into workable databases for subsequent image analysis experiments. It begins with the WSI acquisition from gastric cancer tissue specimens, followed by expert labeling of polygon annotations based on the immunohistochemical response in HER2 stained WSI, semi-automatic registration and annotation transformation to the H&E stained WSI, and generation of the initial working datasets. Ground truth is obtained as reference data in the form of point and contour annotations for cell nuclei segmentation and classification procedures respectively of the image pre-analysis stage. This stage also includes the process of labeling necrosis in the gastric cancer tissue for the necrosis detection task.

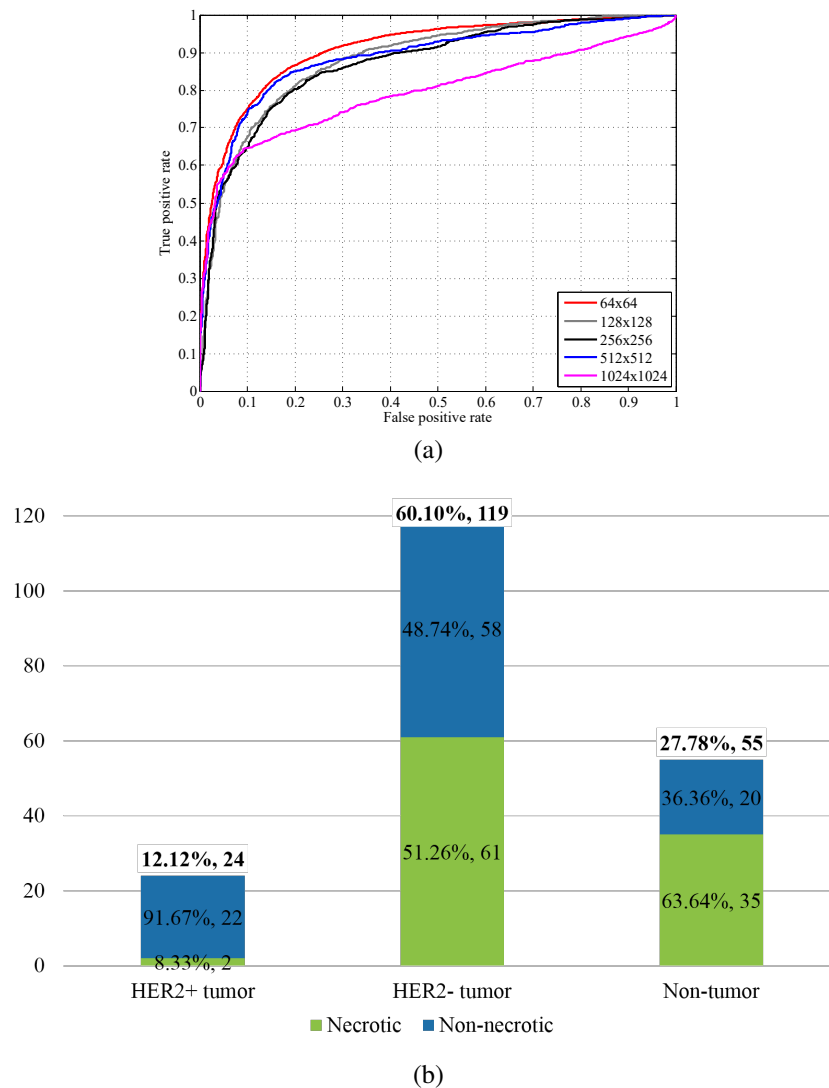


Fig. 5.12 (a) ROC characteristics for combined datasets using the described SVM-based method showing comparative performance between different image tile sizes (b) Distribution of labeled image tiles for necrotic and non-necrotic tissue in three types of malignancy regions and total image tiles in the final dataset for SVM-based method

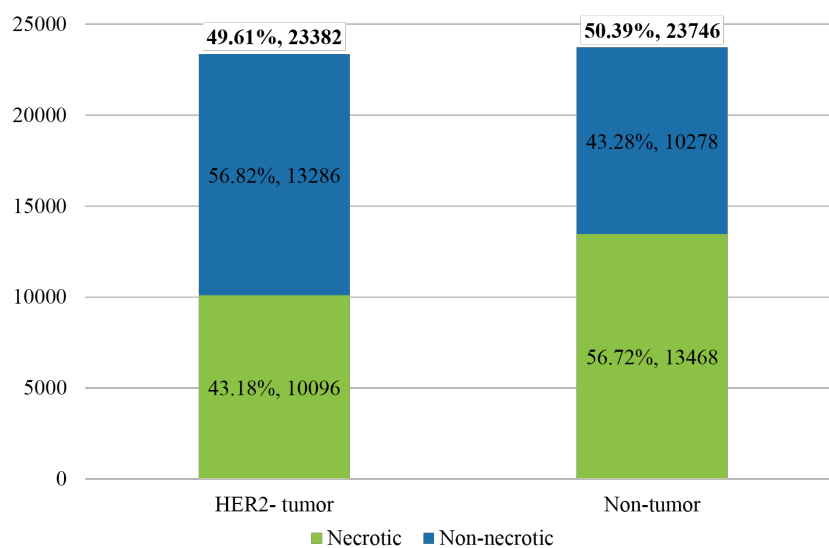


Fig. 5.13 Distribution of labeled image tiles for necrotic and non-necrotic tissue in HER2 negative tumor and non-tumor types of malignancy regions and total image tiles in the final dataset for deep learning methods

Stage 2: Image Pre-analysis

Contents

6.1	Introduction	69
6.2	Necrosis Detection	70
6.2.1	Background	70
6.2.2	Motivation	71
6.2.3	Feature Extraction	72
6.2.4	Machine Learning	73
6.3	Cell Nuclei Segmentation and Evaluation	79
6.3.1	Segmentation Algorithm	79
6.3.2	Segmentation Evaluation	82
6.4	Multiresolution Segmentation Enhancement	84
6.4.1	Background	84
6.4.2	Cell Nuclei Classification	85
6.4.3	Multiresolution Combination	93
6.5	Applications	95
6.5.1	Automatic Necrosis Detection	95
6.5.2	Automatic Determination of Histological Tissue Composition	96
6.6	Summary	98

6.1 Introduction

In general, image preprocessing includes methods for image quality improvement, for example, image contrast enhancement and brightness correction, such that the resulting images are more suitable for subsequent image processing and analysis. The gastric cancer WSI specimens have been acquired in good quality using highly advanced digitization equipment. Consequently,

improvement of image quality is not considered essential and is outside the scope of this study. However, certain tasks can provide additional information from the images aiding further analysis, for instance, image registration and segmentation, and can be grouped into the image pre-analysis stage. The image pre-analysis stage in the described process pipeline is highly specialized towards digital histopathology, and includes task-specific methods such as necrosis detection, cell nuclei segmentation and multiresolution segmentation enhancement of the tissue images of gastric cancer WSI datasets. Among these, image registration has been explained in detail in Section 5.3 due to its high significance in the preparation of materials.

Relating to the necrosis detection methods described in this chapter, a preliminary feasibility study using textural features and SVM machine learning to establish the choice of appropriate patch sizes and demonstrate a suitable classification performance was published in [Sharma 2015c]. A comparative examination of deep learning methods and traditional random forests machine learning for this classification problem form the partial contents of [Sharma 2017b]. Further, the multiresolution segmentation enhancement approach was introduced in [Sharma 2015b] using AdaBoost ensemble learning, but the methodology described in this work is an extension of the initial study with required modifications (details in Section 6.4.2).

6.2 Necrosis Detection

As necrosis was visually detected in small parts of gastric cancer WSI, the first pre-analysis step is to exclude the necrotic tissue regions using a computer-based method. This step is comparable to *noise removal* in general image pre-analysis, however, the ‘noise’ here refers to entire necrotic areas in the WSI which are undesirable for further histopathological image analysis procedures. Hence, necrosis can be considered analogous to, but different from image noise which consists of unwanted signals adding spurious information on the original image.

6.2.1 Background

Necrosis or necrotic cell death, is premature cell death which may occur due to several factors such as infection, heat, mechanical injury or chemicals. According to the Nomenclature Committee on Cell Death (NCCD), necrosis is largely defined in a negative fashion as cell death which lacks apoptotic or autophagic features [Kroemer 2009]. It is morphologically characterized by cellular changes, *e.g.* gain in cell volume, swelling of cell organelles, rupture of plasma membrane, moderate chromatin condensation, followed by disappearance of intracellular contents [Kroemer 2009]. In [Majno 1995], necrosis is defined as drastic and visible changes which appear in the tissue after cell death, which can be “*hinted at by ordinary histological techniques*”. The process of necrosis in a cell is illustrated in Figure 6.1.

Necrosis occurring in tissues afflicted with malignancies is known as *tumor necrosis*. Tumor necrosis is a usual feature of solid tumors caused from chronic ischemic injury owing to rapid tumor growth [Swinson 2002], and can be classified as focal, moderate or extensive, depending

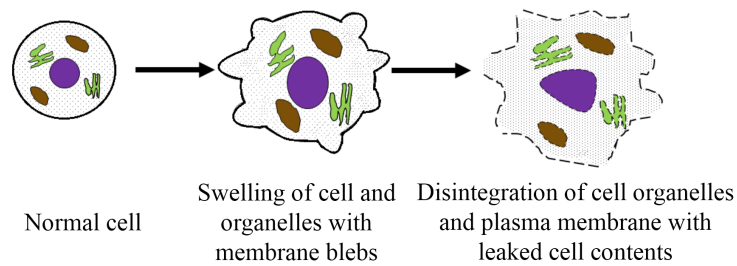


Fig. 6.1 Schematic representation of the process of necrosis in a cell

on its occupancy in the tumor area [Pollheimer 2010]. Focal necrosis has a central appearance as the tumor is depleted of oxygen supply, a condition called tumor hypoxia, in the central portion of tumor. There are usually living cancer cells in neighborhood that can be further observed by pathologists or computer-based analysis tools for diagnostic purposes and treatment planning.

In literature, necrosis has been usually studied at the cytological level by considering the cellular morphology and biochemical processes. For instance, a review about necrosis and apoptosis detection and discrimination, presenting a selection of techniques using electron and fluorescence microscopy and different cell markers is presented in [Krysko 2008]. At the histological level, tumor necrosis has been found to be an important prognostic factor. The prognostic value of tumor necrosis in histological specimens has been quantitatively analyzed with respect to biological variables in [Leek 1999], [Muro-Cacho 2000], [Swinson 2002], [Sengupta 2005], [Langner 2006], and [Pollheimer 2010] for breast carcinoma, gastrointestinal stromal cancer, lung cancer, renal carcinoma, upper urinary tract carcinoma, and colorectal cancer respectively. These studies indicate that the properties of tumor necrosis can be correlated with the type and extent of the respective tumors, and show significant impact on tumor prognosis.

A related work for classifying histological image signatures of Glioblastoma images into necrosis, apoptosis and viable regions has been performed in [Le 2012], but their method uses a small set of training images (50 patches of size 80×80 pixels), and performs classification by reconstruction from subspace analysis, which the authors state is “usually slow” and takes many hours to learn small images, hence, not suitable for datasets containing a large number of images. In this work, the author has implemented a fast and accurate computer-based method performing dedicated task of necrosis detection using image analysis techniques for heterogeneous whole slide images of gastric cancer, which can be potentially generalized for different types of tissues.

6.2.2 Motivation

Automatic detection of necrosis in histological images is an interesting problem of digital histopathology that needs to be addressed. It may be more difficult to reach diagnostic conclusions by observing a WSI containing necrotic regions, and manual identification of necrosis using visual inspection can be a time-consuming task for large-sized whole slide images. Automatic necrosis detection can provide useful information about the the type and extent of malignancy and help in formation of prognosis, as higher necrosis may promote tumor growth and consequently

lead to a lower possibility of survival [Vakkila 2004]. In some cancer patients, (neoadjuvant) chemotherapy is followed by surgery and histological diagnosis, where determining the extent of necrosis can prove useful. Moreover, the detected necrotic areas can be excluded in order to carefully analyze the remaining living tissue. Therefore, necrosis detection can constitute a preparatory stage that is helpful to pathologists and subsequent analysis tools for more precise disease observation and characterization.

The prime motivation behind the step in the described experimental pipeline is, while observing gastric cancer tissue whole slide images, few necrotic regions have been identified visually by expert pathologist. Detection and subsequent exclusion of necrotic regions has been recognized as a necessary task for computer-based analysis to succeed, otherwise necrosis will distort the perceived information and will interfere in the decision-making process. Hence, these regions are required to be excluded from the WSI datasets before performing more sophisticated cancer analysis tasks on the remaining living tissue. So, automatic necrosis detection method has been developed as a pre-analysis step which solves the problem in a timely manner.

6.2.3 Feature Extraction

When observed in histological slides, tumor necrosis has a patch-like appearance with contracted or disappeared cell nuclei in the extracellular space, so it can be visually distinguished from remaining living tissue on the basis of these typical and discernible textural characteristics. Thus, for this study, it is hypothesized that automatic pattern recognition methods using texture information can be suitably applied for detecting necrotic areas in histopathological images of gastric cancer. Low-level state-of-the-art textural image features are extracted from image patches to efficiently represent the visual characteristics of necrotic and non-necrotic areas in the tissue, primarily because these areas can be differentiated on the basis of their textural appearance without the knowledge of the underlying tissue architecture.

A combination of gray level co-occurrence matrix features and Gabor filter-bank texture features is extracted from the histological image tiles comprising the necrosis detection datasets, each at $40\times$ objective magnification and size 512×512 pixels. This texture-based description is used in order to achieve an optimum balance of algorithm efficiency and simplicity, as necrosis detection is a pre-analysis step which requires to be performed using a quick and precise implementation. The computed textural characteristics are defined as follows.

1. **Gray level co-occurrence matrix features:** The gray level co-occurrence matrix descriptors include the 14 statistical features explained in detail in Section 2.3.1.1. These features belong to the state-of-the-art texture features used frequently in digital histopathology.
2. **Gabor filter-bank features:** Gabor filter-banks are selected to detect necrosis for simulating the human visual detection abilities, due to reasons discussed in Section 2.3.1.1. Image tiles are filtered using the real parts of Gabor filter kernels of 16 different frequency, orientation and standard deviation combinations. The first and second order statistics of the filter responses

are derived, based on least squared error for simplicity, and a total number of 32 Gabor filter-bank features is used to represent image texture.

An example of a pair of necrotic and non-necrotic image tile and a subset of the extracted textural features using GLCM statistics and Gabor filter-banks are pictorially depicted in Figure 6.2. The pictorial representations show apparent differences in the texture characteristics of the two categories, ensuring that the extracted features are potentially powerful texture discriminators that can be utilized in combination with traditional machine learning for efficient automatic necrosis detection in gastric cancer histopathological whole slide images.

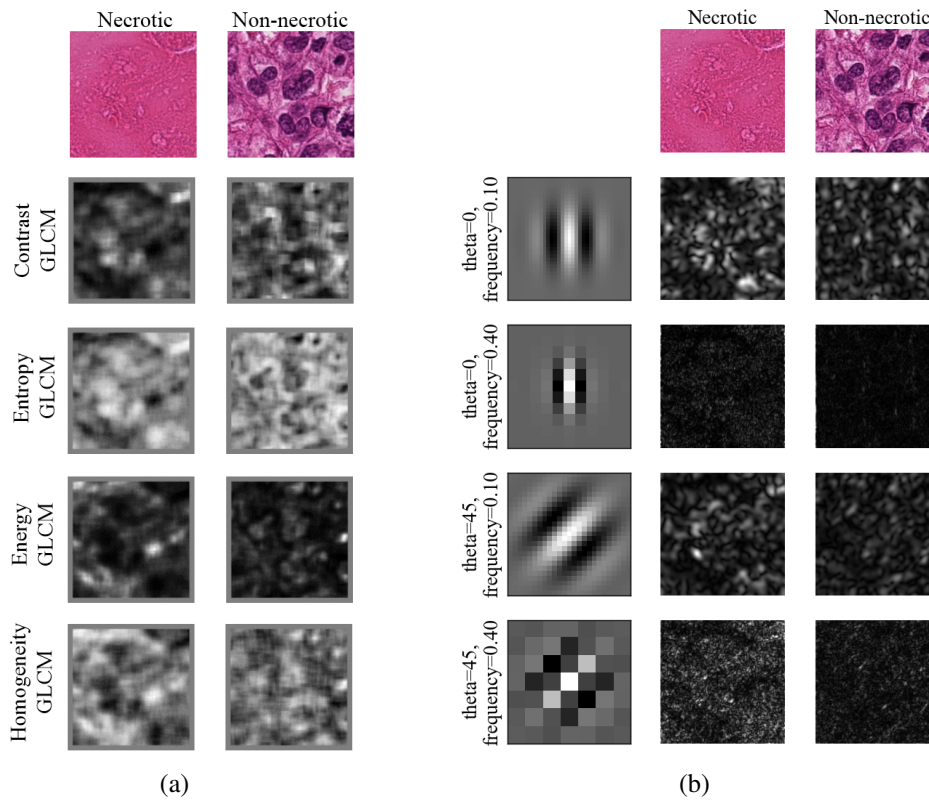


Fig. 6.2 Image tile pair and texture feature pictorial representation using (a) GLCM statistics (b) Gabor filter kernels.

6.2.4 Machine Learning

After the extraction of a set of 46 textural features from each image tile, traditional machine learning is performed using a supervised learning method, namely, support vector machines followed by discriminative thresholding. Later, deep learning methods are also explored for the problem, *i.e.*, AlexNet CNN framework, proposed CNN architecture and ensemble of the two CNNs. The details of machine learning algorithm are described as follows. Using the training datasets and the method with optimal performance, necrotic areas are detected in regions of interest in the whole slide images and excluded prior to the generation of final working datasets for analysis of cancer regions in the third stage of research framework.

6.2.4.1 Support Vector Machines with Discriminative Thresholds

Support vector machines (SVM) is a group of non-probabilistic supervised learning methods that were first introduced by Boser, Guyon and Vapnik in [Boser 1992]. Like any machine learning method, the goal of SVM is to generate a model from the training data that predicts the target values of test data given its features or attributes. Support vector machines aim at finding an optimal separating hyperplane using the training data to classify the incoming instances.

The classical SVM method is originally defined for binary classification. Initially, given a training set $(x_1, y_1), (x_2, y_2) \dots (x_m, y_m)$ of m previously seen attribute-label pairs such that attributes $x \in \mathbb{R}^n$ and class labels $y \in \{\pm 1\}$, where $\{x \in X, y \in Y\}$, the mapping $X \rightarrow Y$ needs to be learned using a classifier $y = f(x, \alpha)$, where α represents the parameters of the classifier function. The easiest approach is the use of linear classifiers, with the decision function given by,

$$f(x) = w \cdot x + b \quad (6.1)$$

where w represents the weights and b is the offset (or bias). However, for complex datasets consisting of a nonlinear feature space, linear classifiers may not be optimal. Hence, the method is elevated to use of nonlinear classifiers which produce maximum margin hyperplanes [Boser 1992], [Cortes 1995]. This is performed by first pre-analysis the data by mapping it to a richer, nonlinear feature space *i.e.* $x \rightarrow \phi(x)$ and the decision function now becomes

$$f(x) = w \cdot \phi(x) + b \quad (6.2)$$

In the nonlinear case, the SVM requires the solution to the following quadratic optimization problem [Boser 1992], [Cortes 1995],

Minimize:

$$\frac{1}{2} ||w||^2 + C \sum_{i=1}^m \xi_i$$

subject to:

$$y_i(w \cdot \phi(x_i) + b) \geq 1 - \xi_i, \quad \xi_i \geq 0 \quad (6.3)$$

However, the dimensionality of $\phi(x)$ can be very high, thereby making it difficult for representation in memory and solving it. For this purpose, a **kernel trick** is applied. The *Representer theorem* [Kimeldorf 1970] shows that (for SVM as a special case), w can be represented as a linear combination,

$$w = \sum_{i=1}^m \alpha_i \phi(x_i) \quad (6.4)$$

Hence, instead of optimizing w , α_i can be optimized. After substituting w , the decision rule now becomes:

$$f(x) = w \cdot \phi(x) + b \quad (6.5)$$

$$= \sum_{i=1}^m \alpha_i \phi(x_i) \phi(x) + b \quad (6.6)$$

$$= \sum_{i=1}^m \alpha_i K(x_i, x) + b \quad (6.7)$$

where, $K(x_i, x) = \phi(x_i) \phi(x)$ represents the kernel function. Several kernels have been proposed by researchers over the past few years, however, the most popular ones are, for instance, linear, sigmoid, Gaussian and polynomial kernels [Scholkopf 2001].

Gaussian or radial basis function (RBF) [Vert 2004] is used as the kernel function for the necrosis detection task and is defined below as,

$$K(x_i, x) = e^{-\gamma(\|x_i - x\|^2)}, \gamma > 0 \quad (6.8)$$

where $\gamma = \frac{1}{2\sigma^2}$ and $\|x_i - x\|$ denotes Euclidean distance between the attributes. The choice of RBF kernels is suitable for the necrosis datasets as they can non-linearly map samples into a higher dimensional space. The two experimental parameters required to be pre-selected for SVM training and classification are c and γ [Chapelle 2002], where $c > 0$ controls the trade-off between margin maximization and error minimization and γ is the kernel parameter respectively. Parameter selection of optimum parameters c and γ is performed using a grid search with k -fold cross validation ($k = 5$) [Chang 2011], where all (c, γ) parameter combinations from a specified grid of parameter values are exhaustively applied, and the one with the best cross validation accuracy is selected. The result of grid search for the given training samples is depicted in Figure 6.3, where optimal parameters $c = 8$ and $\gamma = 0.5$ are applied for further analysis.

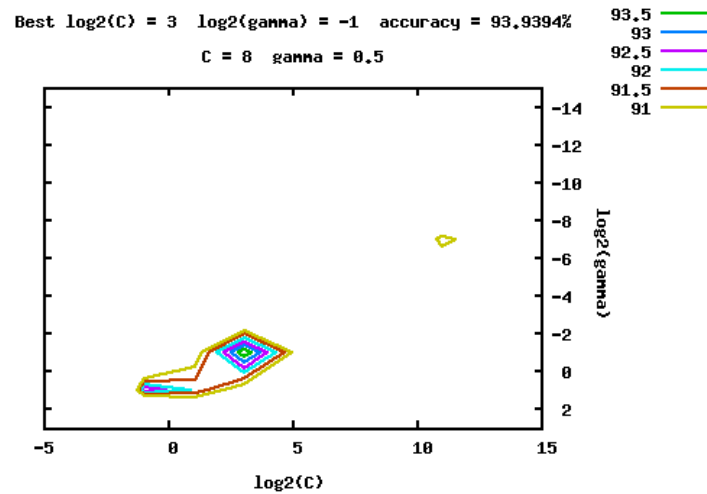


Fig. 6.3 Parameter selection using grid search for necrosis detection using SVM

During the classification process, probability estimates are also generated along with each prediction [Wu 2004], [Chang 2011]. The predicted label is obtained as the one with the largest value of probability estimate. A detailed observation of the predicted labels and probability estimates on the training dataset confirms that there are some patches with borderline probability estimates classified as the opposite class. Hence, thresholds on probability estimates are introduced as additional experimental parameters, known as *discriminative thresholds*. These parameters aim to reduce confusion of the learning method, where a higher confusion to classify into a given category is associated with lower probability estimate values. Based on the discriminative thresholds and probability estimate value of each instance, final class labels are assigned according to the following:

$$L_i = \begin{cases} y_{ij}, & \text{if } P(y_{ij}|X) \geq T_j \\ 1 - y_{ij}, & \text{otherwise.} \end{cases} \quad (6.9)$$

where L_i is the final class label of instance i with original class labels $y_{ij} \in \{0, 1\}$ of class j given observation X . T_j denotes the threshold on class j . Selection of discriminative thresholds is performed using an iterative refinement procedure to achieve optimum classification performance on the training data using a k-fold cross validation ($k = 5$) in the range $[0.5, 1]$ at discrete intervals of 0.05. The threshold search is shown in Figure 6.4 as a color scale representation using blue to red for lower to higher values of classification error rates. The values with least classification error rate can be observed for $T_1 = 0.55$ and $T_2 = 0.6$. This method provides flexibility in classification of doubtful cases and takes into consideration the heterogeneous nature of the whole slide images.

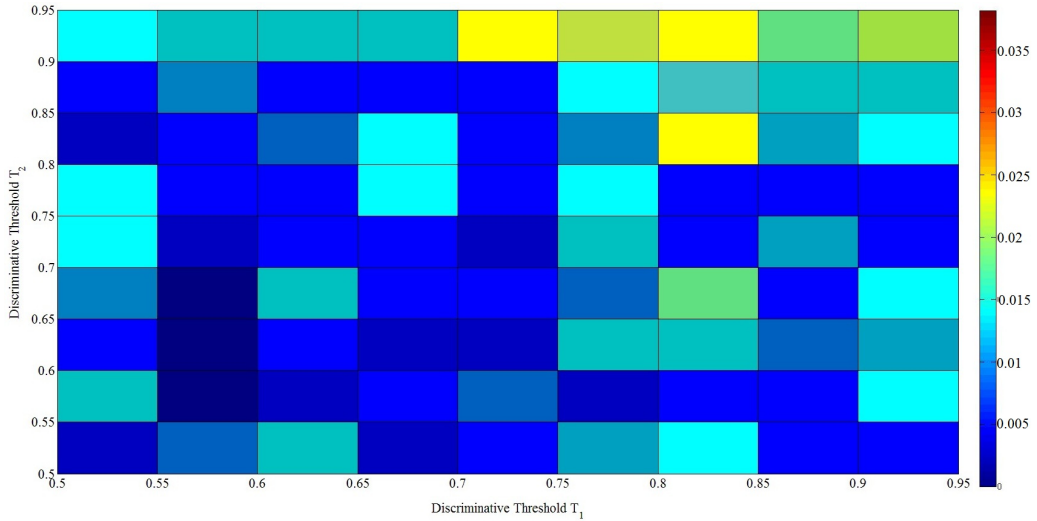


Fig. 6.4 Selection of discriminative thresholds for SVM classification

After observing the experimental results on final working datasets from Section 8.2, it can be seen that using textural features and support vector machines with discriminative thresholds, a

satisfactory value of average overall and balanced classification accuracy is achieved as **87.66%**, suggesting that the described classes are easily distinguished by the classifiers.

Due to a satisfactory cross validation performance using the above described features and machine learning method for necrosis detection in gastric cancer WSI, more complex traditional machine learning methods using ensemble learning are not further explored. It should be emphasized that, as this stage still being the pre-analysis stage, the primary goal is to perform the required tasks using a fast and accurate algorithm, which has already been achieved. However, convolutional neural networks are additionally explored due to the novelty of the research problem with respect to deep learning methods. Additionally, it is performed to examine the generalizability of the proposed CNN architecture, that was originally designed for the analysis of cancer regions but can theoretically achieve accurate classification by directly using the image knowledge without any prior requirement of handcrafted features. Hence, subsequent experiments are performed for necrosis detection using deep convolutional neural networks, which are described in detail in the following section.

6.2.4.2 Deep Learning

Deep learning methods have been explored for detecting necrosis in gastric cancer histopathological images. A significant interest has been shown lately in the image analysis field for deep convolutional neural networks, due to their ability to replace the requirement of hand-engineered image descriptions with direct processing, representation and learning from raw images. This property has been exploited for necrosis detection without any prior extraction of handcrafted features. The widely known AlexNet CNN [Krizhevsky 2012] and a self-designed CNN architecture have been trained from scratch and deployed to study the performance of convolutional neural networks for the necrosis detection problem. An ensemble of the two CNNs is also tested for this purpose. Due to requirement of large-scale datasets in deep learning, data augmentation is performed (discussed in Section 7.2.2). The proposed CNN architecture applied to necrosis detection is shown in Figure 6.5. The selection of experimental parameters, as well as training and deployment have been performed using an elaborate process, described later in Section 7.4.2. Examples of learning curves are demonstrated in Figure 6.6 for randomly selected training round using the AlexNet CNN framework and the proposed CNN architecture.

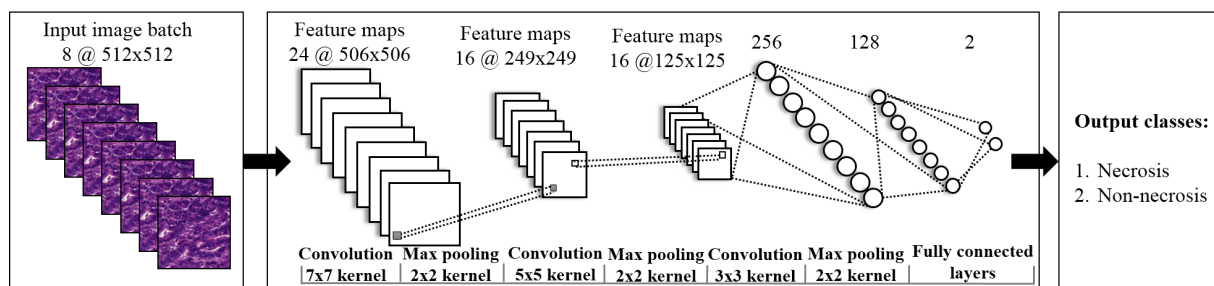


Fig. 6.5 Proposed CNN architecture for necrosis detection

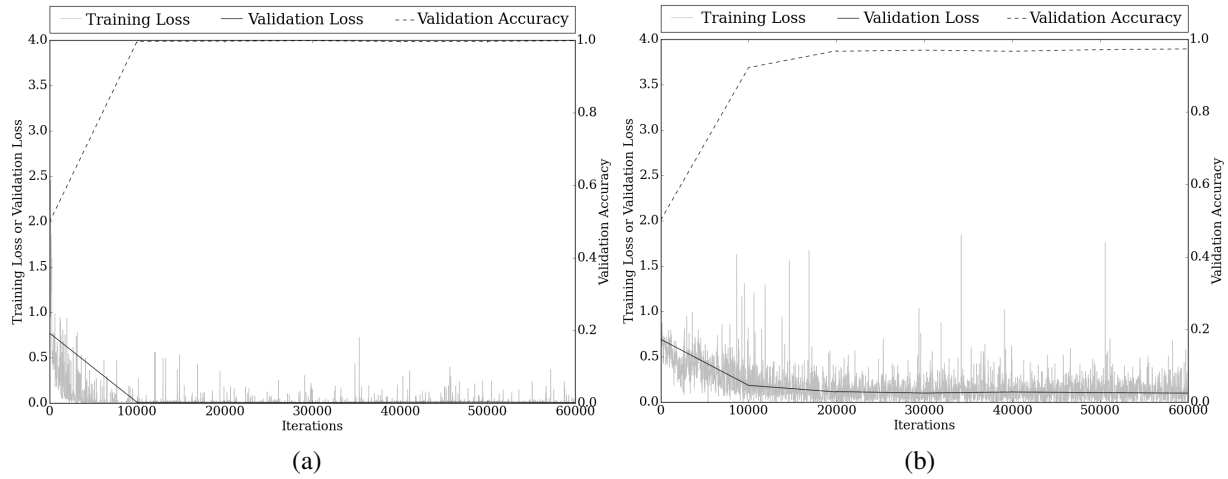


Fig. 6.6 Examples of learning curves of random training rounds for necrosis detection using (a) AlexNet CNN framework (b) proposed CNN architecture.

As observed in the experimental results of Section 8.2, cross validation yields an overall and balanced classification accuracy for AlexNet CNN method as **99.71%**. For the proposed CNN architecture, these are calculated as **93.56%**. An ensemble of the two CNNs provides improvement in performance compared to the individual CNNs, with balanced classification accuracy as **99.89%**. A high cross validation accuracy with lower false positives highlights a successful classification of necrotic and non-necrotic tissue in gastric cancer histopathological images using deep learning methods.

There is a minor difference in detection rates of the explored machine learning methods, which shows that necrosis detection is a fairly simple classification problem compared to the cancer analysis problem considered later, due to a distinct appearance characteristics of necrotic versus non-necrotic or living areas. One drawback of the described analysis is, due to a limited amount of training data consisting of visual information from four WSI, a possibility of misclassification may exist among the regions of WSI from unknown patients, that are not learned by the classifiers. This can be attributed to the heterogeneity among tissue characteristics of individual patients, owing to biological variance in a population. In the worst case, necrotic regions which need to be excluded from further examination are misclassified as non-necrotic, and the final goal may not be completely achieved. To prevent this effect during the creation of the final working datasets comprising information of the seven non-labeled WSI for the analysis of cancer regions, it is atleast safe to exclude the automatically classified necrotic regions from further analysis. Additionally, a brief visual inspection of the non-necrotic classified regions follows the automatic necrotic detection in order to ensure that the false negative necrotic regions are not included in the final working datasets for analysis of cancer regions. The detailed experimental results and observations for these methods are discussed in Section 8.2 and associated computational requirements are summarized in Appendix A.5.

6.3 Cell Nuclei Segmentation and Evaluation

After exclusion of necrotic regions, the next step is to isolate the cell nuclei in the H&E stained tissue images using appropriate segmentation methods. Segmentation is a crucial and challenging step in most histological image analysis problems, and performance of subsequent tasks like feature extraction and classification largely depends on the results of the segmentation algorithm. It is difficult mainly due to the complex appearance of cells due to unclear cell boundaries and overlaps, hence, manual and automatic segregation of cell nuclei is a tedious process. A recently developed and advanced computerized cell nuclei segmentation algorithm is used for this purpose, and its segmentation results are utilized to perform analysis of cancer regions in the subsequent stages of the research work.

6.3.1 Segmentation Algorithm

The cell nuclei segmentation algorithm [Wienert 2012] explored in the study is a minimum-model method as it requires minimal apriori information to detect contours of cell nuclei in H&E stained tissue images. It is a fully-automatic approach and has been applied on the gastric cancer image tiles to discern the constituting cell nuclei. The cell nuclei segmentation algorithm consisting of six main steps is defined as follows.

The algorithm starts with the detection of all possible closed contours regardless of their size, shape or intensity in the grayscale-transformed image using a gradient-based technique. It uses a conventional contour tracing approach for binary images [Hufnagl 1983], extended for the use of grayscale images. Each row is a one-dimensional function of intensity $I(x)$ and contour pixels are recognized as those at which gradient between pair of neighbors is maximum and these pixel positions are stored. Subsequently, 8-connected neighbors of the stored positions are scanned to find all possible contours. A valid contour is the one where start and end pixels are the same. This step results in multiple overlapping contours, which are evaluated based on gradient features of the input image in the second step.

In the next step, contour evaluation is performed by computing a metric called *Contour Value* to select contours that best represent image objects. It is defined as the combination of two contour-based features, namely, mean contour gradient that measures the relative importance of objects in a group of overlapping objects, and gradient fit that determines which of several alternative contours best represents a certain object. For this purpose, firstly a Sobel operated image S is computed by convolving the image I with 3×3 Sobel filter operators G_x and G_y . This is given as,

$$|S| = \sqrt{(I * G_x)^2 + (I * G_y)^2} \quad (6.10)$$

where, Sobel filter operators G_x and G_y are [Gonzales 2009],

$$G_x = \begin{pmatrix} -1 & 0 & 1 \\ -2 & 0 & 2 \\ -1 & 0 & 1 \end{pmatrix} \quad G_y = \begin{pmatrix} -1 & -2 & -1 \\ 0 & 0 & 0 \\ 1 & 2 & 1 \end{pmatrix} \quad (6.11)$$

From the Sobel image $|S|$, the mean contour gradient MG_i for the i^{th} contour C_i , where p_{ij} represents the j^{th} contour pixel, is calculated by,

$$MG_i = \frac{\sum_j |S(p_{ij})|}{|C_i|} \quad (6.12)$$

Also, gradient fit GF_i which represents the relative number of contour pixels which have local maxima of gradient is measured as,

$$GF_i = \frac{\sum_j p_{ij}^{max}}{|C_i|} \quad (6.13)$$

where,

$$p_{ij}^{max} = \begin{cases} 1, & \text{if } \max\{|S(p_{ab})|\} = |S(p_{ij})| \quad \forall (a, b) \text{ s.t. } x_{j-1} \leq a \leq x_{j+1} \wedge y_{j-1} \leq b \leq y_{j+1} \\ 0, & \text{otherwise} \end{cases} \quad (6.14)$$

Contour value CV_i will then be computed as

$$CV_i = MG_i \cdot GF_i \quad (6.15)$$

In the third step, a non-overlapping segmentation is generated where the enclosed areas of the ranked contours are labeled in a two-dimensional map of the same size as the original image in descending order of contour values, and blocking overwriting of already assigned pixels.

Subsequently, the segmentation produced is improved using a novel contour optimization method involving elimination of non-compact pixels which do not belong to the object using Manhattan metric. A distance d is defined for testing the compactness of object pixels, and a pixel p_i with $d_i < d$ is compact if connected to another pixel p_j with $d_j = d$ over $d - d_i$ edges. Pixels connected over more than $d - d_i$ edges are removed, and this results in compact objects.

An optional cluster separation step involves detection and separation of concave objects by removing pixels around a cutting line passing through the concavity. Finally, cell nuclei are detected by assessing nucleus-specific Hematoxylin within each contour area. Hematoxylin strength is found by color deconvolution [Ruifrok 2001] and all objects with intensity greater than a predefined threshold [Otsu 1979] are selected as cell nuclei objects.

A set of optimal parameters for cell nuclei segmentation in H&E images of gastric cancer is selected after visual inspection of results for a wide range of involved algorithm parameters. The definitions and optimal values of parameters are shown in Table 6.1.

Table 6.1 *Parameter selection in cell nuclei segmentation algorithm*

Name of parameter	Definition	Optimum value for gastric cancer H&E images
minContourLength	Minimum length of acceptable contours	50
maxContourLength	Maximum length of acceptable contours	500
distance	d defined for testing compactness of object pixels	3
mindepthabs	Minimum absolute depth of concavity during concave object separation	0
mindepthrel	Minimum relative depth of concavity during concave object separation	0.3
recursive	Boolean value whether the concave object separation step should be repeated until no concave objects are found	TRUE

A segmentation map is a data structure that is used to store spatial positions of image objects after segmentation. It encapsulates a two-dimensional array of unsigned integers that has the size of the input image, where the unsigned integers are the identifiers of the corresponding segment of a given pixel. For each image tile, its corresponding segmentation map is generated and stored. Cell nuclei segmentation of gastric cancer image tiles yields a result similar to the one shown in Figure 6.7. The yellow outlines are the resulting contours of segments obtained after the cell nuclei segmentation is executed.

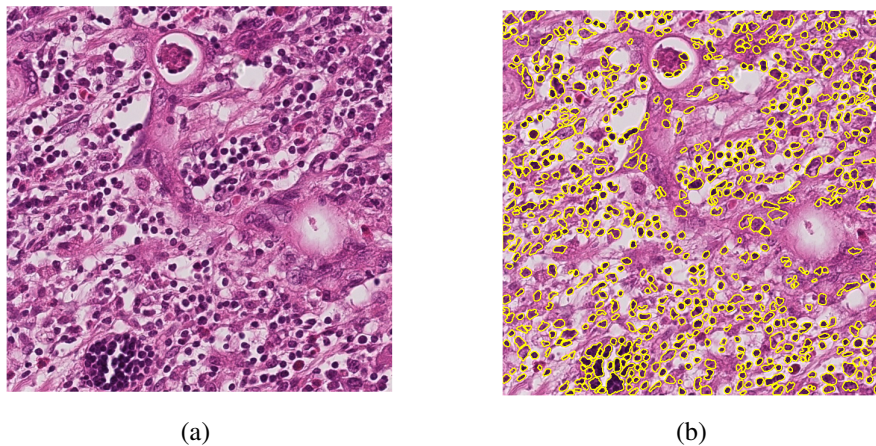


Fig. 6.7 Cell nuclei segmentation result example at 25 \times objective magnification (a) Original image (b) Processed image showing resulting cell nuclei segments.

6.3.2 Segmentation Evaluation

The initial working dataset consisting of 75 image tiles is used in order to evaluate the performance of the cell nuclei segmentation algorithm at different magnifications. The cell nuclei are first manually located with point annotations at highest resolution ($40\times$) to create the ground truth which is validated by an expert pathologist (details in Section 5.4.1).

It is intuitive that highest amount of useful information should be contained in higher magnifications, but it is also important to observe the ratio between the relevant and irrelevant information and how to differentiate between them. So different resolutions are simultaneously considered and visual and quantitative assessment is performed to verify this assumption. Each image tile is first automatically segmented with the help of the described cell nuclei segmentation algorithm at the selected magnifications of $10\times$, $15\times$, $20\times$, $25\times$, $30\times$ and $40\times$, hence, a total of 450 segmented image tiles are generated. Following segmentation, the results are quantitatively compared against the ground truth information containing positions of cell nuclei as point annotations. The total number of point annotations N_p are defined as,

$$N_p = \sum_{i=1}^{N_{im40\times}} n_{pi} \quad (6.16)$$

where n_{pi} represents the number of point annotations in the i^{th} image tile and $N_{im40\times}$ is the total number of image tiles at $40\times$ magnification. Also, for a given magnification, the total number of segments obtained N_s are denoted by,

$$N_s = \sum_{i=1}^{N_{imx}} n_{si} \quad (6.17)$$

where n_{si} represents the number of segments found in the i^{th} image tile and N_{imx} is the total number of image tiles at magnification X. The following quantities are measured and compared for each of the six magnifications.

1. The *percentage of correctly detected cell nuclei* can be calculated as the fraction of all the point annotations which are enclosed inside a cell nuclei segment, represented as,

$$P_d = \left(\frac{1}{N_p} \sum_{i=1}^{N_{imx}} \sum_{j=1}^{n_{pi}} f\{p_j(x, y) \in S_i\} \right) \times 100 \quad (6.18)$$

where $p_j(x, y)$ denotes the j^{th} point annotation in the i^{th} image tile containing a set S_i of two-dimensional spatial point coordinates lying inside or on all the detected segments. $f\{P\}$ is a conditional operator which gives 1 when P is true, and 0 other wise. Hence, a cell nucleus is considered as detected if its associated point annotation is inside or on any of the detected segments, so it is marked as a cell nucleus in the ground truth data. Since the segments are non-overlapping, each cell nucleus will be detected only once at maximum.

2. The *percentage of segments not cell nuclei*, denoting the amount of oversegmentation, can be calculated as the fraction of the total number of segments which do not contain any point annotation *i.e.*,

$$P_{os} = \left(\frac{1}{N_s} \sum_{i=1}^{N_{imx}} \sum_{j=1}^{n_{si}} f\{n_p(s_j) = 0\} \right) \times 100 \quad (6.19)$$

where where $n_p(s_j)$ represents the number of point annotations lying inside or on the j^{th} segment s_j , and the summation term denotes the total number of segments not representing actual cell nuclei in the images. This percentage denotes the segments representing false positive instances, which mostly constitute smaller fragments of cell nuclei and other tissue components.

3. The *percentage of conglomerates* is computed by finding all those segments which contain more than one point annotation as a fraction of the total number of segments, and represents the number of cell nuclei clusters, *i.e.*, segments containing more than one cell nucleus.

$$P_c = \left(\frac{1}{N_s} \sum_{i=1}^{N_{imx}} \sum_{j=1}^{n_{si}} f\{n_p(s_j) > 1\} \right) \times 100 \quad (6.20)$$

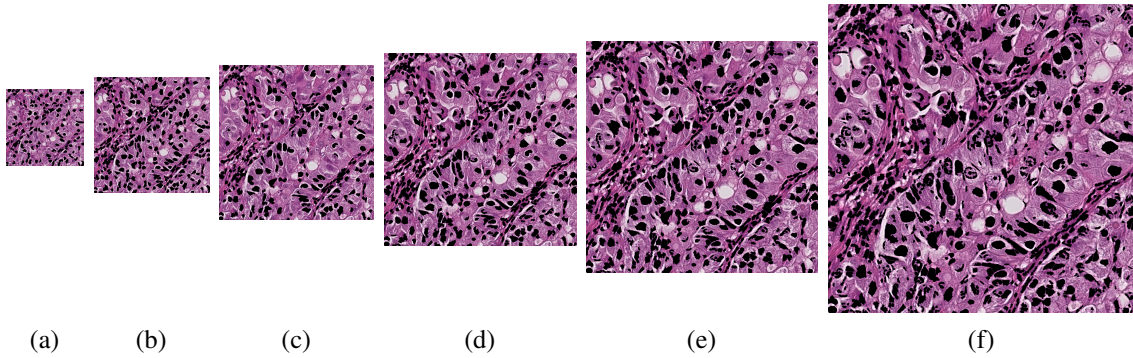


Fig. 6.8 Example of cell nuclei segmentation results of an image at different objective magnifications for visual inspection (a) 10× (b) 15× (c) 20× (d) 25× (e) 30× (f) 40×.

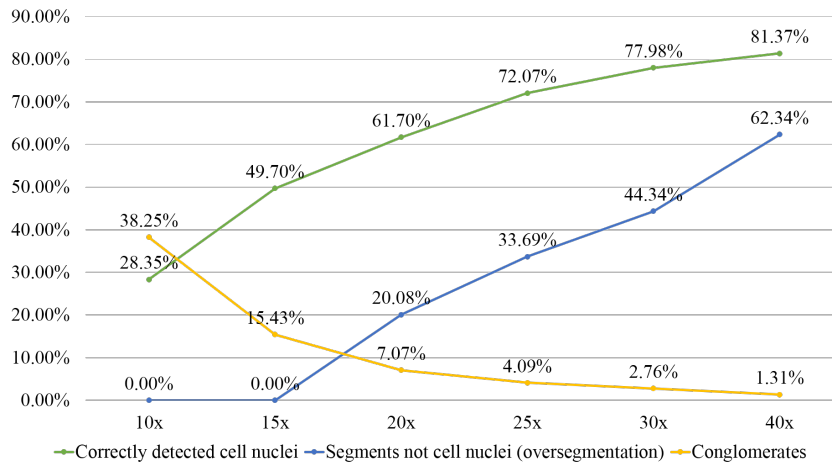


Fig. 6.9 Cell nuclei segmentation performance at individual magnifications

Example of cell nuclei segmentation results of an image at different objective magnifications for visual inspection is depicted in Figure 6.8. The overall result showing the three performance quantities at each magnification can be summarized in Figure 6.9. From the outcome of the described coarse-to-fine segmentation analysis using visual and quantitative assessment, it is observed that the percentage of correctly detected cell nuclei increases with magnification. Also, oversegmentation (segments not cell nuclei) increases with magnification. No oversegmentation for $10\times$ and $15\times$ magnifications indicates that the number of annotated cell nuclei are equal to or greater than the number of segments, which means that each segment corresponds to one or more cell nuclei point annotations. The number of conglomerates decrease with increase in magnification. In order to preserve the finer details in the tissue images and capture maximum number of correctly detected cell nuclei, one choice is the highest magnification $40\times$. However, to deal with the problem of oversegmentation caused due to nuclear fragmentation evident mostly in $40\times$, it may be combined with a lower magnification to achieve more accurate results.

For identifying which of the lower magnifications is suitable for combining with $40\times$ objective magnification, a pairwise comparison is performed between segmentation results of $40\times$ and each of the lower magnifications. The goal is to find the percentage of correctly segmented nuclei in lower magnifications which can additionally contribute to the total correctly segmented nuclei with $40\times$, and is calculated as P_{adx} by,

$$P_{adx} = \left(\frac{N_{dx} - N_{dc}}{N_p} \right) \times 100 \quad (6.21)$$

where N_{dx} denotes the number of correctly detected cell nuclei found at magnification X and N_{dc} is the number of correctly segmented nuclei common in both magnifications X and $40\times$. It is found almost equal ($\approx 5\%$) for magnifications between $15\times$ to $30\times$. It is lower for $10\times$ ($\approx 4\%$). The other factor used for deciding the other magnification is clustering which leads to formation of conglomerates. It is already seen that clustering decreases with magnification, thus, in order to minimize it, the next magnification with minimum conglomerates (2.8%) *i.e.* $30\times$ is selected. Hence, after evaluating the results of nuclei segmentation algorithm it is concluded that segmentation information at $30\times$ and $40\times$ objective magnifications will be utilized for segmentation enhancement.

6.4 Multiresolution Segmentation Enhancement

6.4.1 Background

As observed after the detailed evaluation of segmentation results, it is clear that visual information contained in more than one resolutions needs to be combined to achieve desired results for further analysis of cancer regions. Therefore, an appropriate improvement in this direction can be the use of multiresolution methods. In previous literature, there have been a few studies relating to multiresolution analysis in digital histopathology. For instance, a multiresolution texture

analysis technique is used in [Shuttleworth 2002b] for classifying colon cancer images that focuses on varying distances of texture co-occurrence matrix instead of spatial resolution, and does not include isolation of nuclei before classification. A multiresolution approach is also reported in [Kong 2009] for neuroblastoma images where higher magnification is considered only when the results from lower magnifications are unsatisfactory, using a feedback loop involving pathologists. Another work related to cell classification reports a multi-class and two-stage categorization of Wright stained WBCs [Ramesh 2012]. In the first stage, cells are classified as ones with segmented and non-segmented nucleus, and in the second stage as one of the five subtypes. In the author's paper [Sharma 2015b], a part of the work including cell nuclei segmentation evaluation, cell nuclei classification (using AdaBoost ensemble learning) and resulting multiresolution segmentation enhancement have been published.

6.4.2 Cell Nuclei Classification

Combining information at the level of segmentation itself is a non-trivial task, due to which automatic classification of cell nuclei segments is required to be performed. This work emphasizes on object-level classification of cell nuclei segments in gastric cancer images based on their morphological, texture, and color and intensity characteristics. The seven cell nuclei classes have been suggested by expert pathologists and are named as epithelial cell, leukocyte, fibrocyte (or border cell), conglomerate, fragment, other cell (including blood cell in vessel) and artefact (details in Section 5.4.2).

Initial working datasets are used to generate ground truth data and performance evaluation. As explained in Section 5.4.2, 33 tiles were initially selected at $30\times$ and $40\times$ objective magnifications to generate the ground truth annotations, however, after performing necrosis detection, labeled segments from 24 tiles at $40\times$ magnification have been retained for the final training sample as remaining tiles contain total or partial necrotic regions. A balanced dataset has been created for training purpose containing equal number *i.e.* 400 samples per category, depending on the relative availability of training samples of each class in the labeled data, hence, a total of 2,800 samples are employed for training the classifiers using the supervised traditional machine learning methods.

In [Sharma 2015b], two different classification approaches have been explored. In the first approach, *single-stage classification* of segments is performed into eight predefined classes. However, in later experiments, a modification *i.e.* combination of clusters and badly segmented cells as conglomerates has been introduced. In the second approach, *hierarchical classification* is executed, where, in first stage, segments are classified into three broad classes, namely, compact objects (including epithelial nuclei, leukocytes, fibrocytes/border cells, other nuclei and nuclei fragments), conglomerates (including clusters of nuclei and badly segmented nuclei) and artefacts, and in the second stage, each class objects are further classified into one of the respective subclasses. Hierarchical classification is avoided in the described experiments because this approach does not result in any significant improvement as noted in [Sharma 2015b], hence, unwanted addition to the complexity of the algorithm is prevented. However, the explored

traditional machine learning methods are expanded from just AdaBoost in [Sharma 2015b] to SVM and random forests for further investigations. Another modification is the exclusion of $30\times$ labeled samples in the set of training samples, again after the initial observations in [Sharma 2015b], suggesting lower classification rates. Hence, for quantitative evaluation and classification of unknown objects, only labeled samples at the highest resolution are used in the training dataset to reduce time and space requirements of the proposed method.

6.4.2.1 Feature Extraction

In this step, object-level feature extraction is performed where numerical features are computed on the cell nuclei segments. A feature vector of 32 cell nuclei features based on morphology, texture, color and intensity are extracted from segments, as described below. A listing of the 32 cell nuclei features with their mathematical descriptions is summarized in Table 6.4.2.1.

1. **Morphological features:** Shape or morphological properties are used by pathologists to identify or distinguish between different types of nuclei components. The following morphological features are applied, namely, Object Pixels, Minimum Distance to Tessellation Border, Pixels at Layer Border, Maximum Distance to Border, Aspect Ratio of Bounding Ellipsoid, Minor Axis of Bounding Ellipsoid, Major Axis of Bounding Ellipsoid, Angle of Bounding Ellipsoid, Form Factor of Contour, Convexity of Contour, Length of Contour, Area of Contour, Form Factor of Convex Hull, Length of Convex Hull, Area of Convex Hull [Hufnagl 1984], Feret, Minimal Radius of Enclosing Centered Circle, Maximal Radius of Enclosed Centered Circle, Roundness and Form Factor [Zerbe 2008]. For the subset of feature definitions from [Hufnagl 1984], the axes projections and Freeman code is given in Figure 6.10.

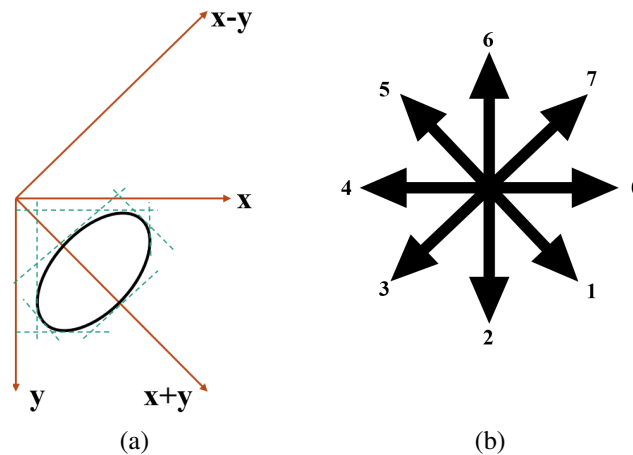


Fig. 6.10 Morphological feature definitions of a contour (a) Axes projections (b) Freeman code. Adapted from [Hufnagl 1984].

In addition, the fractal dimension of an object is also used as a morphological feature. It is a ratio determining the complexity of a given object or pattern by comparing the changes

in the pattern to the scale of measurement. It was first used for the idea of self-similarity of objects in [Mandelbrot 1967], [Mandelbrot 1983]. Fractal dimension has been recognised as an important tool for cancer analysis [Baish 2000]. For example, it has been proved in [Simeonov 2006], that fractal dimension can play a role in differentiating between tumor cells in mammary glands in cytological specimens. For a Euclidean space \mathbb{R}^n , fractal dimension is measured as *Minkowski dimension* or *box-counting dimension* [Falconer 2004]. It is calculated by using the box-counting algorithm, where a uniform grid is considered on the object with each box of side ϵ . The number of boxes covering the object will be $N(\epsilon)$, as this number will grow as $(1/\epsilon)^d$ when $\epsilon \rightarrow 0$, *i.e.* when the box size is reduce by making the grid finer. Hence,

$$N(\epsilon) = \lim_{\epsilon \rightarrow 0} \left(\frac{1}{\epsilon} \right)^d \quad (6.22)$$

$$d = \lim_{\epsilon \rightarrow 0} \frac{\log N(\epsilon)}{-\log \epsilon} \quad (6.23)$$

The fractal dimension d has been estimated in this work in two steps, by first counting the number of filled boxes $N(\epsilon)$ in a cell nuclei segment at each corresponding ϵ in a predefined range, and then solving linear regression using least-squares method [Intriligator 1978].

2. **Color and intensity features:** Color and intensity features are important for histological images due to the specific stains used. The features used to characterize the intensity of segments in this work include Mean Intensity, Mean Intensity on Contour, Standard Deviation of Intensity, Standard Deviation of Intensity on Contour [Hufnagl 1984], Contour Value and Gradient Fit [Wienert 2012]. Mean Chromaticity is also calculated, where chromaticity for i^{th} RGB pixel \mathbf{p}_i is defined as the minimum euclidean distance d_{min} between pixel RGB value and points on the diagonal where each point \mathbf{p}_d is defined by RGB value $R = G = B = x, x \in \{0, 1, ..255\}$. In other words, it is the minimum euclidean distance of a pixel to a grey pixel value.

$$d_{min}(\mathbf{p}_i, \mathbf{p}_d) = \min(\|\mathbf{p}_i - \mathbf{p}_d\|) \quad \mathbf{p}_i, \mathbf{p}_d \in \mathbb{R}^3 \quad (6.24)$$

3. **Texture features:** Texture is also a widely used characteristic in histological image analysis, and varies with different tissue components. Four features based on gray level co-occurrence matrix [Haralick 1973], namely contrast, entropy, energy and homogeneity have been selected to represent the visual characteristics of cell nuclei segments. Contrast denotes variation in pixel intensities, entropy is a statistical measure of randomness, energy measures repetitiveness of texture patterns and homogeneity quantifies extent of similarity in the image. For high computational efficiency in the image pre-analysis stage, these features are considered to be a representative subset of the set of GLCM features.

Table 6.2 List of object-level features computed for cell nuclei classification

Feature name	Feature definition	Remarks
Minimum Distance to Tesselation Border	$\text{Min}\{d_i \text{ from border of image}\}$	d_i : distance of i^{th} pixel of the object to image border
Pixels at Layer Border	Number of pixels in the object at the border of the image	
Maximum Distance to Border	$\text{Max}\{d_i \text{ from border of image}\}$	
Object Pixels	$ S_j $	S_j is the set of two-dimensional point coordinates constituting the j^{th} object
Minor Axis of Bounding Ellipsoid	Length of minor axis of circumscribing ellipse	
Major Axis of Bounding Ellipsoid	Length of major axis of circumscribing ellipse	
Aspect Ratio of Bounding Ellipsoid	$\frac{1000(\text{Major Axis of Bounding Ellipsoid})}{\text{Minor Axis of Bounding Ellipsoid}}$	
Angle of Bounding Ellipsoid	Angle between major axis of circumscribing ellipse and the horizontal plane	
Length of Contour	Numerical contour length + $(\sqrt{2} - 1)SKP$	SKP: Number of contour points with a skewed (odd) Freeman direction
Form Factor of Contour	$\frac{10(\text{Length of Contour})^2}{\text{Geometric area of segment}}$	
Convexity of Contour	$\frac{100 \cdot 2(D_x + D_y + D_m + D_p)}{3(\text{Numerical contour length}) + SKP}$	Numerical contour length is the number of points in the contour of the object and D_y, D_x, D_m, D_p are the total number of contour points in Freeman directions except in $\mathbf{x}, \mathbf{y}, \mathbf{x+y}$ and $\mathbf{x-y}$ axes respectively.
Area of Contour	Number of pixels inside contour (excluding contour pixels)	
Length of Convex Hull	Number of pixels at boundary of convex hull	Convex hull is the convex envelope of S_j is the smallest convex set of points that contains S_j .
Form Factor of Convex Hull	$\frac{10(\text{Length of Convex Hull})^2}{\text{Geometric area of convex hull}}$	
Area of Convex Hull	Number of pixels in convex hull	
Feret	$\text{Max}\{\text{Euclidean distance between two points in an object}\}$	
Minimal Radius of Enclosing Centered Circle	Radius of smallest circle enclosing the object	
Maximal Radius of Enclosed Centered Circle	Radius of largest circle enclosed by the object	
Roundness	$\frac{4(\text{Area of Contour})}{\pi(\text{Feret})^2}$	
Form Factor	$\frac{4\pi(\text{Area of Contour})}{(\text{Numerical contour length})^2}$	
Fractal Dimension	$d = \lim_{\epsilon \rightarrow 0} \frac{\log N(\epsilon)}{-\log \epsilon}$	ϵ is the length of box side, $N(\epsilon)$ is the number of boxes covering the object in box-counting algorithm
Mean Intensity	$\frac{1}{n} \sum_{i=1}^n I_i$	I_i is the intensity of the i^{th} pixel and n is the number of pixels in the object

Mean Intensity on Contour	$\frac{1}{n_c} \sum_{i=1}^{n_c} I_i$	I_i is the intensity of the i^{th} pixel and n_c is the number of pixels in the contour of the object
StdDev Intensity	$\sqrt{\frac{1}{n} \sum_{i=1}^n (I_i - \mu)^2}$	μ is Mean Intensity of object
StdDev Intensity on Contour	$\sqrt{\frac{1}{n_c} \sum_{i=1}^{n_c} (I_i - \mu_c)^2}$	μ_c is Mean Intensity on Contour of object
Mean Chromaticity	$\frac{1}{n} \sum_{i=1}^n d_{i,min}, d_{i,min}(\mathbf{p}_i, \mathbf{p}_d) = \min(\ \mathbf{p}_i - \mathbf{p}_d\)$	Minimum euclidean distance $d_{i,min}$ for i^{th} RGB pixel \mathbf{p}_i in segment where \mathbf{p}_d is a diagonal RGB point where $p_d(R) = p_d(G) = p_d(B)$.
Gradient Fit	$\frac{\sum_j p_{ij}^{max}}{ C_i }$	Details in Section 6.3.1
Contour Value	$MG_i \cdot GF_i$	MG_i is mean gradient of i^{th} object, details in Section 6.3.1
Contrast GLCM	$\sum_{i=0}^{N-1} \sum_{j=0}^{N-1} i - j ^2 p(i, j)$	$p(i, j)$ is the (i, j) value of GLCM and N is the number of gray levels in the image
Entropy GLCM	$\sum_{i=0}^{N-1} \sum_{j=0}^{N-1} p(i, j) \ln(p(i, j))$	
Energy GLCM	$\sum_{i=0}^{N-1} \sum_{j=0}^{N-1} [p(i, j)]^2$	
Homogeneity GLCM	$\sum_{i=0}^{N-1} \sum_{j=0}^{N-1} \frac{p(i, j)}{1 + i - j }$	

6.4.2.2 Machine Learning

After the feature extraction process, each of the cell nuclei objects are represented by a set of 32 derived object-level measures. The cell nuclei balanced dataset consisting of learning samples is used in a supervised approach. Three machine learning methods, namely, support vector machines, AdaBoost and random forests are comparatively analyzed for the classification of cell nuclei objects into one of the seven predefined classes. Cross validation methods such as k-fold stratified shuffled split and leave-a-sample-out are applied to quantitatively evaluate the classification performance of the various machine learning methods. The calculated performance metrics, results and observations for each method are discussed in detail in Section 8.3. Using the selected learning samples and the machine learning method with optimal performance, unknown cell nuclei objects can be automatically classified into one of the seven different classes.

Support Vector Machines

Training and deployment of SVM is similar to the approach elaborated in Section 6.2.4 using Gaussian kernels. However, in this case, the SVMs are trained for a multi-class classification problem instead of a binary one, which is implemented using the *one-against-one* approach [Knerr 1990], [Chang 2011]. Discriminative thresholds are avoided to prevent complexity due to multiple classes. The parameter selection process for SVM, as previously discussed, is used to determine two experimental parameters c and γ for optimum training. Parameter selection is performed using grid search on the training samples with k-fold cross validation ($k = 5$), and the pictorial result of the grid search is depicted in Figure 6.11. Optimal parameters $c = 512$ and $\gamma = 0.125$ are selected and applied for cross validation using the entire training dataset.

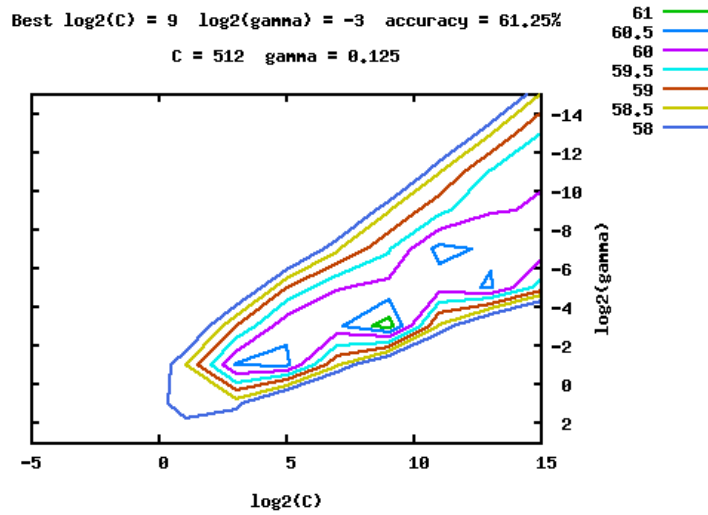


Fig. 6.11 Parameter selection using grid search for multi-class SVM classification of cell nuclei

After quantitative assessment of classification performance, the average overall cross validation accuracy achieved is **60.39%** and average balanced cross validation accuracy is **60.41%** using support vector machines as the machine learning method.

AdaBoost Ensemble Learning

A multi-class AdaBoost classification algorithm has been developed for cell nuclei classification task. AdaBoost [Freund 1995] or Adaptive Boosting is the most widely used form of boosting, and involves *ensemble learning* where a collection of component classifiers or learners is used and a joint decision is taken by combining their predictions. AdaBoost allows adding a sequence of weak learners to the algorithm, until a desired low training error is achieved. Each weak learner consists of decision stumps. The ensemble of weak learners can be defined as $h_k(\mathbf{x})$, $k = 1, 2, \dots, K_{max}$. K_{max} is an experimentally determined parameter of the algorithm, and as expected, higher the value of K_{max} , higher is the time complexity of the algorithm. The strong learner is assembled from all weak learners through a weighted majority voting scheme. The binary AdaBoost algorithm [Bishop 2006] is explained as follows.

Input: Consider N input vectors $\mathbf{x}_1 \dots \mathbf{x}_N$ comprising the training data.

Output: Binary variables $y_n \in \{-1, +1\} \forall n \in \{1, 2, \dots, N\}$

Steps:

1. Initialize weights by setting: $w_n^{(1)} = 1/N \forall n \in \{1, 2, \dots, N\}$
2. For $k = 1 \dots K_{max}$:
 - (a) Train classifier $h_k(\mathbf{x})$ on training data by minimizing the weighted error function given as:

$$E_k = \sum_{n=1}^N w_n^{(k)} \times \begin{cases} 1 & h_k(\mathbf{x}_n) \neq y_n \\ 0 & \text{otherwise} \end{cases}$$

(b) Evaluate: $\epsilon_k = \frac{E_k}{\sum_{n=1}^N w_n^{(k)}}$

(c) Evaluate: $\alpha_k = \ln \left[\frac{1-\epsilon_k}{\epsilon_k} \right]$

(d) Update weights as:

$$w_n^{(k+1)} = w_n^{(k)} \times \begin{cases} e^{\alpha_k} & h_k(\mathbf{x}_n) \neq y_n \\ 1 & \text{otherwise} \end{cases}$$

3. Final vote of the test point \mathbf{x} is given by: $H(\mathbf{x}) = \text{sgn}[\sum_{k=1}^{K_{max}} \alpha_k h_k(\mathbf{x})]$

The binary classification algorithm is extended to multi-class by considering the final vote as the class with majority votes of weighted binary learners. In general, boosting involves the sequential minimization of an exponential error function given as [Bishop 2006],

$$E = \sum_{n=1}^N w_n^{(k)} e^{-\frac{1}{2} y_n \alpha_k h_k(\mathbf{x}_n)} \quad (6.25)$$

The parameter K_{max} has been experimentally determined by using an exhaustive search that calculates classification error on randomly selected parts of the training data, for each value of K_{max} in a predefined range until no further improvement is observed. Figure 6.12 shows the result of the parameter search for $K_{max} \in [1, 50]$. It can be observed that the global minima of classification error is obtained for $K_{max} = 7$ after which classification error increases non-uniformly with less significant drops at other values like 10 and 14. To minimize the computational requirements and optimize classification accuracy, $K_{max} = 7$ is selected for AdaBoost ensemble learning. After quantitative evaluation of the multi-class classification performance, the average overall cross validation accuracy achieved is **63.38%** and average balanced cross validation accuracy is **63.89%** using AdaBoost ensemble learning as the machine learning method.

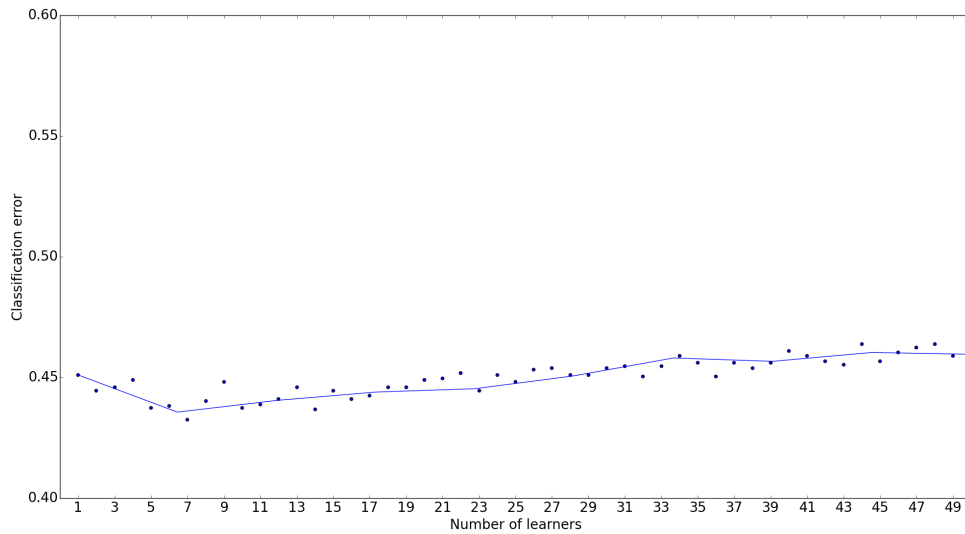


Fig. 6.12 Parameter selection for multi-class AdaBoost classification of cell nuclei

Random Forests

The latest successful traditional machine learning method for supervised classification after prior computation of suitable handcrafted features is the random forest. A decision tree or classification tree [Rokach 2014] is a predictive model that grows by using a series of binary splits called recursive splitting (or partitioning), until terminal nodes called *leaves* are reached, which are the decision nodes and dominated by one of the classes in the classification process. Decision trees have several advantages, such as easy interpretation, ignoring redundant variables, handling of missing data and good performance with large datasets. Several ensemble classifiers are constructed with more than one decision trees, for instance, random forests, bagging trees, boosted trees, and rotation forests.

Random forests operate by constructing multiple decision trees during the training phase, and the classification result is the mode of the classes predicted by constituent decision trees [Breiman 2001]. During the training phase, the random forest grows many classification trees, where each constituent decision tree of the random forest grows by selecting a random sample m out of M input variables (or features) at each node and splitting the node using best split on these variables. During the classification phase, the input feature vector is passed through each of the decision trees in the random forest. Each decision tree predicts a class, and the class with the majority votes is the final result. In this work, instead of majority voting of individual decision trees, their probabilistic classification scores p_k are averaged [Pedregosa 2011] to obtain class probabilities P for an input x and number of trees N_t , as below.

$$P(x) = \frac{1}{N_t} \sum_{k=1}^{N_t} p_k(x) \quad (6.26)$$

Random forests are superior to decision trees as they prevent overfitting and also enhance the estimation ability of individual trees. The variance among decision trees is also decreased due to the averaging operation.

The random forest algorithm requires a prior selection of two experimental parameters, namely, number of trees N_t and number of features m to consider for best split. Examples of frequently used values of m are \sqrt{M} and $\log_2 M$. For the described classification experiments, these are estimated internally, during the run using *bootstrap aggregation*, where each new decision tree is fit from a bootstrap sample of the training observations $z_i = (x_i, y_i)$. An out-of-bag (OOB) error is computed as the average error for each sample z_i using predictions from the trees that do not contain z_i in their respective bootstrap sample [Friedman 2001]. The OOB error is computed for each value of N_t in a predefined range and each number of features *i.e.* square root, \log_2 or all of the total features to approximate a suitable value of N_t at which the error stabilizes. Figure 6.13 shows the result of the parameter search for $N_t \in [1, 1500]$, where minimum and more stable values of OOB error are seen for square root of features for $N_t \geq 1000$. In order to reduce the trade-off between time complexity and classification accuracy, lowest possible number of trees, *i.e.* $N_t = 1000$ with features as square root for best split, are

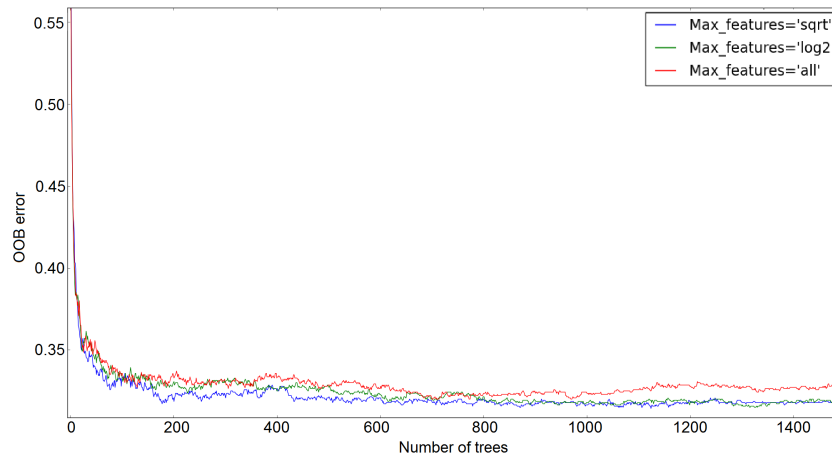


Fig. 6.13 *Parameter selection for multi-class random forest classification of cell nuclei*

selected for further experiments in cell nuclei classification using random forests. Quantitative evaluation of classification performance of random forests shows the average overall cross validation accuracy as **64.01%** and average balanced cross validation accuracy as **63.99%**.

To summarize, it is observed that both the ensemble learning probabilistic methods have a superior classification performance over support vector machines for the classification problem of multiple categories of cell nuclei objects in the histopathological images. The performance of the described methods is comparable to each other, with an overall precedence of random forests as a result of highest average classification accuracies and a greater number of diagnostically relevant classes achieving better detection rates. Due to a satisfactory performance and higher computational efficiency, random forests machine learning method is selected for cell nuclei classification in order to accomplish multiresolution enhancement of segmentation, and determination of tissue composition in histological images. The overall achieved classification accuracies are reasonable, however, not exceptionally superior, the prime reason being a difficult multi-class problem at hand to classify the cell nuclei segments into one of the seven defined classes, some of which do not show consistent behavior in visual appearance across the tissue images, especially the tumor-affected epithelial cells. The detailed results, observations and implications of cell nuclei classification step are discussed in Section 8.3. Also, the computational complexity and associated implementation details of the described attribute calculation and machine learning methods are summarized in Appendix A.5.

6.4.3 Multiresolution Combination

To combine the segmentation results of different magnifications, the procedure depicted in Figure 6.14 is used. Firstly, the lower magnification ($30\times$) contours are upscaled using nearest neighbor interpolation [Gonzales 2009] to obtain cell nuclei objects and respective set of 32 cell nuclei features at the highest resolution. Next, the learning samples at the highest resolution are used in a training procedure to learn random forest classifiers, which are then deployed for

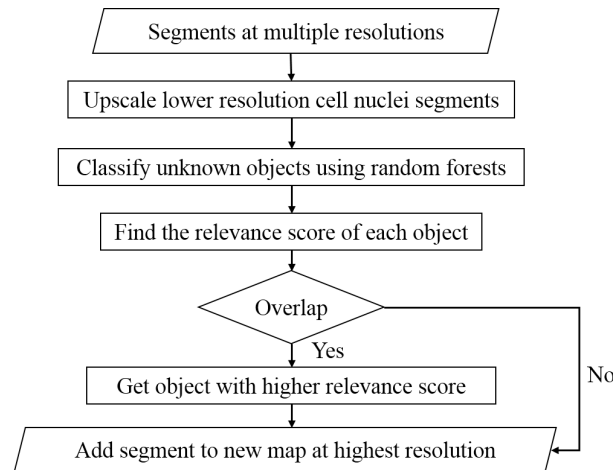


Fig. 6.14 Flowchart for multiresolution combination of cell nuclei segmentation

Table 6.3 Relevance scores assigned to classified cell nuclei objects

Type of objects	30×	40×
All cells	200	150
Conglomerates	100	100
Fragments	80	50
Artefacts	0	0

classification of unknown cell nuclei objects from the two resolutions. Following the automatic classification, a relevance score is assigned to each classified cell nuclei object depending on its predicted class. This scoring is relative, depending on which information is visually more significant for diagnostic purpose and for subsequent analysis of cancer regions, as given in Table 6.3. The cell nuclei and fragments at 30× magnification are given a higher relevance than the corresponding segments at 40×, due to a better visual quality at 30× and to reduce oversegmentation which is mainly found at 40× objective magnification. No relevance is given to artefacts and they are completely eliminated from the combined results as desired.

A new segmentation map is created at the highest resolution where cell nuclei segments from individual constituent magnifications are added according to their visual importance. Hence, by using the classification results of segments at different magnifications, a more accurate combined segmentation result is obtained, containing more useful information than individual ones. The approach can potentially lead to enhancement of cell nuclei segmentation results from individual magnifications.

The outcome of multiresolution combination is quantitatively evaluated using the method described in Section 6.3.2. The overall combination result is summarized in Figure 6.15. From the quantitative and visual evaluation of the segmentation results after multiresolution combination, it can be observed that these are superior to individual magnifications 30× and 40×, with desired characteristics of both individual magnifications. Correctly detected cell nuclei segments are almost maintained to the percentage in 40× magnification which has the highest quantity. Oversegmentation *i.e.* the percentage of segments not cell nuclei is reduced from 40×

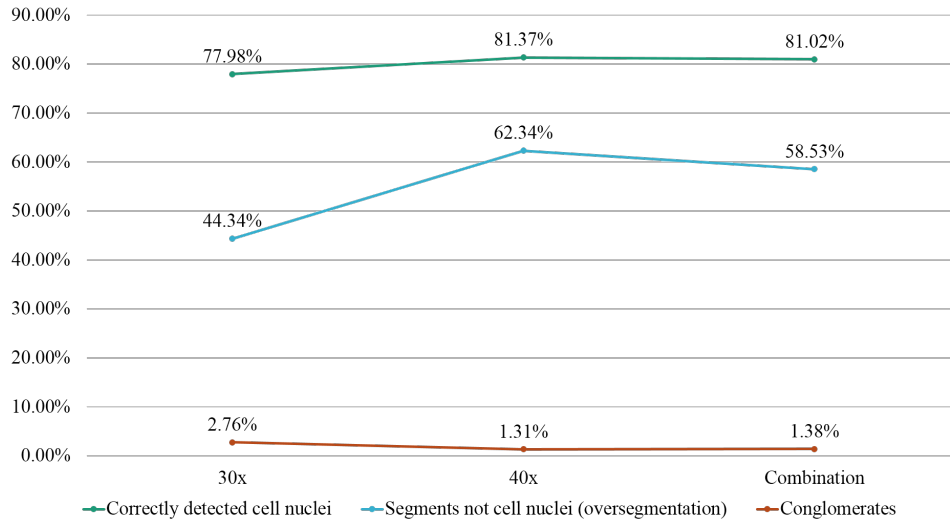


Fig. 6.15 Cell nuclei segmentation performance after multiresolution combination

magnification by $\approx 4\%$, and clustering *i.e.* the percentage of conglomerates is also reduced from 30 \times magnification by $\approx 1.5\%$. By visual assessment of combined results, we find that some cell nuclei fragments of 40 \times have been replaced with corresponding whole cell nuclei of 30 \times . Similarly, some conglomerates of 30 \times are replaced with individual nuclei of 40 \times . Hence, the multiresolution segmentation result combines the characteristics at individual magnifications.

However, one effect to be carefully analyzed is the slight reduction in the number of correctly detected cell nuclei compared to 40 \times magnification, as theoretically, should be higher than individual magnifications to realize the true advantage of multiresolution approach by combining their individual contributions. A prime reason for this observation is misclassification during the cell nuclei classification stage, which leads to a few higher precedence segments being replaced by other segments (*e.g.* fragments) that are misclassified, resulting in decline in the number of correctly detected nuclei. So the combined result is closer to 40 \times , but a better version due to lower oversegmentation. Hence, it can be concluded that the multiresolution combination has caused an enhancement of segmentation properties in histopathological images of gastric cancer and a more meaningful segmentation has been achieved, containing diagnostically relevant visual information compared to individual magnifications.

6.5 Applications

6.5.1 Automatic Necrosis Detection

The primary requirement of the pre-analysis step is a high algorithmic performance and efficiency with lower time and space requirements, which has been fulfilled using textural features and SVM classification with discriminative thresholds (details of computational requirements in Appendix A.5). Hence, this method is retained and applied for classification of unknown

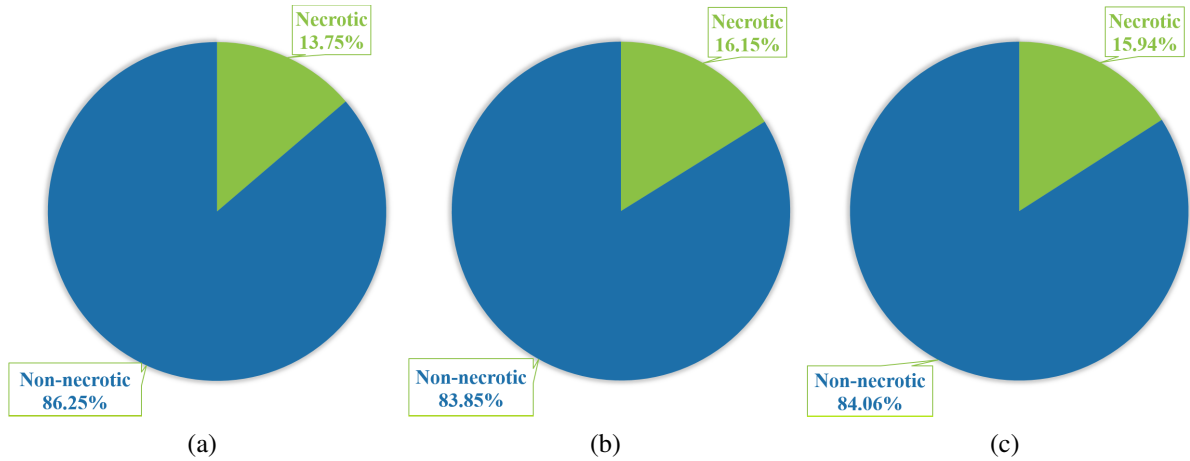


Fig. 6.16 Results for application of automatic necrosis detection on gastric cancer WSI datasets in (a) HER2 positive tumor (b) HER2 negative tumor (c) Non-tumor.

histological regions as necrotic or non-necrotic, using the labeled images from the training datasets. For the studied WSI datasets, the result of necrosis detection after applying the automatic process is shown in Figure 6.16. As expected, the percentage of necrotic areas is found lower than non-necrotic areas in the three malignancy regions, which coincides with pathologists' observations during creation of ground truth. It also corresponds with the underlying diagnostic hypothesis according to which the tumor is in the early stages, hence, necrosis is not widespread in the tissue. The resulting necrosis classified regions are excluded from subsequent image analysis tasks in the research pipeline.

Therefore, necrosis detection can be a useful preparatory step for pathologists and automatic analysis tools for accurate disease observation and characterization. The proposed computer-based method is a promising tool to detect necrosis in heterogeneous haematoxylin and eosin stained whole slide images with variation in malignancy levels, stain intensity and biological characteristics among patients. The experimental results on the gastric cancer WSI datasets are encouraging and show robustness of the method to varying visual appearances. The novel proposed necrosis detection application can act as a standalone program or as a pre-analysis step in digital histopathology.

6.5.2 Automatic Determination of Histological Tissue Composition

Histological composition of tissues for diagnostic purpose is currently determined by pathologists using visual inspection in routine and research, which is a tedious and time-consuming process. The cell nuclei classification step can be used as an automatic application to determine histological composition of gastric cancer tissues. Such tissue composition analysis can potentially assist pathologists in computer-assisted diagnosis of gastric cancer. The method also provides a basis for automatic differentiation between tumor and non-tumor compartments of the tissue and determination of cancer type, grade or extent (also considered for this work in Chapter 7). In this

section, a prototype application for automatic determination of histological tissue composition is described and its preliminary results are illustrated using a pilot dataset, which can be extended to larger datasets consisting of non-overlapping tiles from larger regions of interest in the H&E stained WSI.

Cell nuclei classification has been comparatively analyzed for three machine learning methods in Section 6.4.2, and random forests are selected for classifying unknown cell nuclei samples. The initial working dataset consisting of 75 image tiles is first used to create an enhanced segmentation result using the multiresolution combination approach, including automatic classification of unlabeled cell nuclei segments using random forest classifiers. The combined result is quantitatively analyzed to determine the proportion of each cell category in the three types of malignancy regions, namely, HER2 positive tumor, HER2 negative tumor and non-tumor. This information can provide diagnostically relevant cues for subsequent visual or computational analysis.

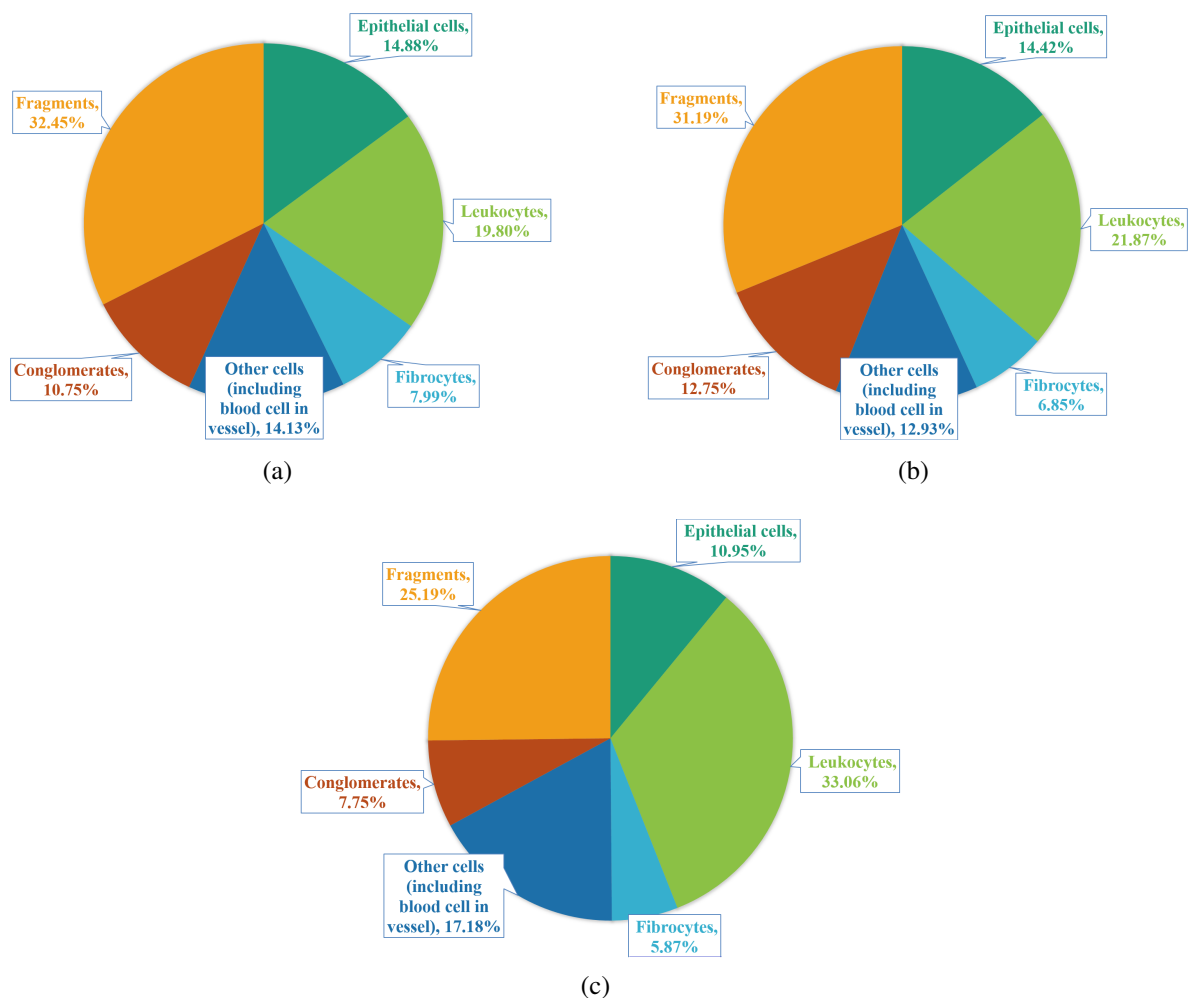


Fig. 6.17 Example results of prototype application for automatic determination of tissue composition in (a) HER2 positive tumor (b) HER2 negative tumor (c) Non-tumor.

The measured tissue composition for the three types of regions is shown in Figure 6.17. It can be observed that the two type of tumors have similar tissue compositions, which are visibly different from non-tumor. HER2 positive areas consist of the highest ratio of cell nuclei fragments, epithelial cells and fibrocytes and lowest percentage of leukocytes compared to other two types of regions, as expected. Also, HER2 negative regions have intermediate properties in terms of percentage of epithelial cells, leukocytes, fibrocytes and fragments, and consist of largest amount of conglomerates among the three malignancy types. Non-tumor, in contrast, has the maximum proportion of leukocytes and lower quantities of cell nuclei fragments, epithelial cells, fibrocytes and conglomerates.

Therefore, the automatically determined tissue composition resonates with the tissue composition expected pathologically in the corresponding malignancy areas. The results are also in synchronization with the actual composition of the labeled data comprising part of initial working datasets at $40\times$ magnification, shown in Figure 5.10(b). A few differences arise due to the error in the classification of cell nuclei, and multiresolution combination, leading to lower fragment and conglomerate percentages in the resulting segmentation.

6.6 Summary

In this chapter, the second stage of the proposed framework, including the necessary image pre-analysis methods, is defined. The stage incorporates the essential steps for processing the current image datasets as required for favorable image analysis of gastric cancer. Firstly automatic detection and exclusion of necrotic areas is performed such that the final datasets consist of non-necrotic regions. Moreover, a cell nuclei segmentation algorithm is applied, evaluated and segmentation results are enhanced using a novel multiresolution combination approach. This stage also contributes towards two applications, namely, automatic necrosis detection and automatic determination of histological tissue composition in microscopic images using supervised learning and classification methods. The result of this stage is attainment of necrosis-free histopathological image data consisting of additional information from the isolated cell nuclei segments for subsequent analysis of cancer regions.

Stage 3: Analysis of Cancer Regions

Contents

7.1	Introduction	100
7.2	Ground Truth Dataset Generation	101
7.2.1	Datasets for Traditional Machine Learning Methods	101
7.2.2	Datasets for Deep Learning Methods	102
7.3	Image Description Methods	105
7.3.1	Low-level State-of-the-art Features	105
7.3.2	High-level State-of-the-art Features	107
7.3.3	High-level Handcrafted Features	108
7.3.4	Combination of Features	121
7.4	Machine Learning Methods	123
7.4.1	Traditional Methods	123
7.4.2	Deep Convolutional Neural Networks	125
7.4.3	Classification Strategies	135
7.5	Applications	136
7.5.1	Computer-aided Diagnosis: Cancer Classification based on Immuno-histochemistry	136
7.5.2	Content-based Image Retrieval	137
7.6	Summary	141

7.1 Introduction

In this chapter, systematic procedures designed and implemented for analyzing histopathological whole slide images of gastric cancer are described in detail. Two research directions are pursued, namely, traditional methods and deep learning methods for efficient image representation and automatic classification of malignancy levels in H&E whole slide images of gastric cancer tissues, based on their HER2 immunohistochemical response. Most essentially, among the traditional methods, a novel graph-based technique is handcrafted for knowledge discovery from tissue regions, followed by traditional machine learning algorithms. Deep learning using convolutional neural networks is also explored by empirical analysis, formulation, realization and investigation of appropriate network architecture. In each direction, comparisons are established using quantitative evaluations with the state-of-the-art traditional methods and deep learning methods.

The image analysis tasks are performed in the more widely used haematoxylin and eosin stain using the malignancy information from the immunohistochemical counterparts. Pathologists in routine practice visually navigate and inspect glass slides or WSI to identify and analyze abnormalities in the immunohistochemical stain as malignancy levels are mostly visually distinguishable. However, visual discrimination of the corresponding malignancy levels in the more commonly used H&E stain will be a more prolonged and tedious process. Moreover, even in the more advanced immunohistochemical stain, inter-and intra-observer variability has been observed and formally studied for the described gastric cancer datasets in [Behrens 2015], suggesting subjectivity in the visual examination process. This study attempts to address the above challenges by developing automated image analysis techniques for computer-based classification in the H&E stained tissue regions that can enable automatic analysis to make desired decisions, but may be difficult to visually discern in H&E stain.

The primary goal is to enable speedy and accurate computer-aided characterization from the routinely used H&E stained slides without the requirement of immunohistochemical staining. Analysis of cancer regions can potentially assist pathologists in understanding of the disease in the more common H&E stain, thereby, reducing costs and efforts in preparation and inspection, and observer variability. A few results obtained on smaller representative datasets after applying part of the methods described in the chapter, consisting of the initial proposal of the high-level handcrafted features and comparison with low-level and high-level state-of-the-art methods with AdaBoost ensemble learning has been published by the author in [Sharma 2016]. Similarly, the section of the chapter discussing image analysis of gastric cancer regions using deep learning and relative evaluation with low-level state-of-the-art features followed by traditional machine learning with random forests has been explored by the author in detail in [Sharma 2017b]. Lastly, the design and implementation of hand-engineered graph-based variants explained in this chapter, and subsequent random forests hierarchical classification have been discussed [Sharma 2017a]. In this work, the author has attempted to present a comprehensive analysis by understanding, extending, modifying and comparing the developed methods for optimal performance, and

demonstrating the all-inclusive outcomes in depth, using the available WSI datasets of gastric cancer.

7.2 Ground Truth Dataset Generation

The gastric cancer WSI datasets used in this study consist of 11 digitized whole slide images of surgical gastric tissue sections, prepared using HER2 immunohistochemical techniques and H&E stain, described in detail in Section 5.2. The pathologists have marked polygon annotations using specialized computer software, where annotations consist of HER2 positive areas according to *10% cut-off rule* and HER2 negative areas morphologically identified as tumor, hence, pathologists used their expert knowledge to visually discriminate between the two tumor classes in the HER2 immunohistochemical stain. Details of labeling annotations in HER2 WSI datasets are provided in Section 5.3. Therefore, the HER2 positive areas define a high degree of malignancy, whereas HER2 negative areas denote a lower malignancy level.

7.2.1 Datasets for Traditional Machine Learning Methods

To create the ground truth dataset, H&E WSI are tessellated at highest resolution ($40\times$ objective magnification), with each image tile of size 1024×1024 pixels. Non-overlapping image tiles are selected from areas inside polygon annotations considering regions of agreement of most pathologists, and labeled as HER2 positive tumor or HER2 negative tumor, depending on the corresponding region in the HER2 WSI. Non-tumor category comprises of image tiles tessellated from regions lying outside both types of polygon annotations. To generate the final working datasets, necrotic areas are excluded using automatic necrosis detection procedure described in detail in Section 6.2, and visual inspection follows to further eliminate any falsely classified necrotic image tile and tiles containing more empty areas compared to tissue. Hence, the experiments are performed on large ROIs containing vast context information at highest magnification. A representative example of image tiles from each category is shown in Figure 7.1.

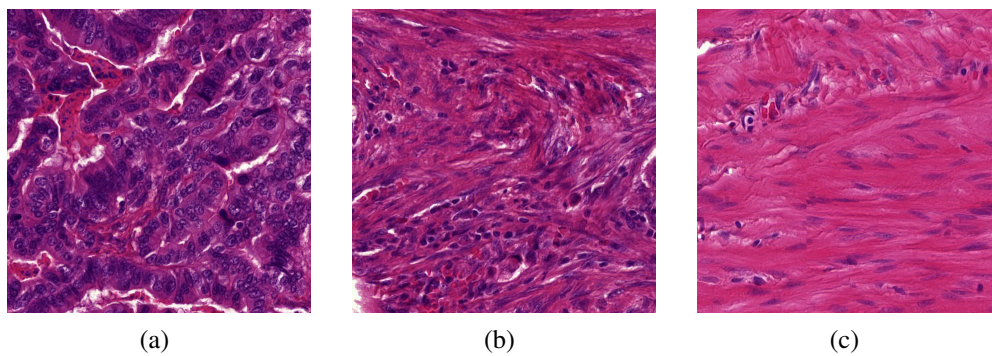


Fig. 7.1 Representative examples of image tiles from the same WSI of the defined malignancy levels at the highest resolution for (a) HER2 positive tumor (b) HER2 negative tumor (c) Non-tumor

A total of 795 labeled image tiles have been generated according to the immunohistochemical response of the tissue to conduct experiments for exploring different image descriptions followed by traditional machine learning methods. The distributions of the image tiles showing the number and proportion of the image tiles for individual patients (denoted by S) in each of the three type of regions is illustrated in Figure 7.2(a). The number of image tiles in these datasets is not as high as the datasets used for deep learning, because traditional machine learning methods do not require large-scale datasets compared to deep learning methods to optimize predictive performance. Furthermore, the selected tiles are large-sized and non-overlapping, and obtained from regions of agreements of most pathologists to create an accurate ground truth, which are limited by inter-observer variability in annotations. Another reason for the smaller number is to achieve balance among the three types of classes, especially because HER2 positive class has smaller annotated areas compared to the HER2 negative tumor class.

A subset of experiments using traditional methods described in the thesis also examine the performance using a dataset comprising of smaller image tiles of 512×512 pixels, with a total of 3,153 images. This is created as it is an interesting observation to determine if there is a deterioration in the system performance by using smaller sized image tiles (observations mentioned briefly later). This dataset has also been utilized for the CBIR prototype application explained in Section 7.5.2. The distributions of the image tiles of the smaller size showing the number and proportion of the image tiles in this dataset for individual patients (denoted by S) in each of the three type of regions is illustrated in Figure 7.2(b).

7.2.2 Datasets for Deep Learning Methods

In the studies related to deep learning methods, dataset augmentation is recognized as a prior requirement to expand a comparatively smaller amount of labeled WSI data, owing to many of the recent successes of deep learning algorithms with massive datasets [Deng 2014]. Standard data augmentation methods are used to generate large number of image tiles, each of size 512×512 pixels at highest magnification, from the WSI regions annotated by expert pathologists. The prime reason for the use of 512×512 image size, which is quite reasonable for the existing deep learning methods, is the memory limitations of the GPU during practical implementation of convolutional neural networks, as current capabilities of the specified single GPU are not sufficient for the study of deep architectures with larger sized images (*e.g.* 1024×1024). An advantage of using the smaller image size is larger datasets are generated which can potentially lead to a more reliable evaluation of classification performance for deep learning methods.

The data augmentation methods include overlapping tiles by a factor of 0.3, and affine transformations such as rotations with 10 degree intervals, reflection, rotation after reflection and shear by a factor of 0.1. Brightness, contrast and intensity adjustments, and significant geometric deformations are not applied in order to preserve the salient texture, color and morphological properties of the original H&E stained tissue images. For the analysis of cancer regions, a total of 21,000 images from each WSI are generated, thus, a large-scale balanced dataset containing 231,000 images is obtained. The size of datasets depends on the availability of corresponding

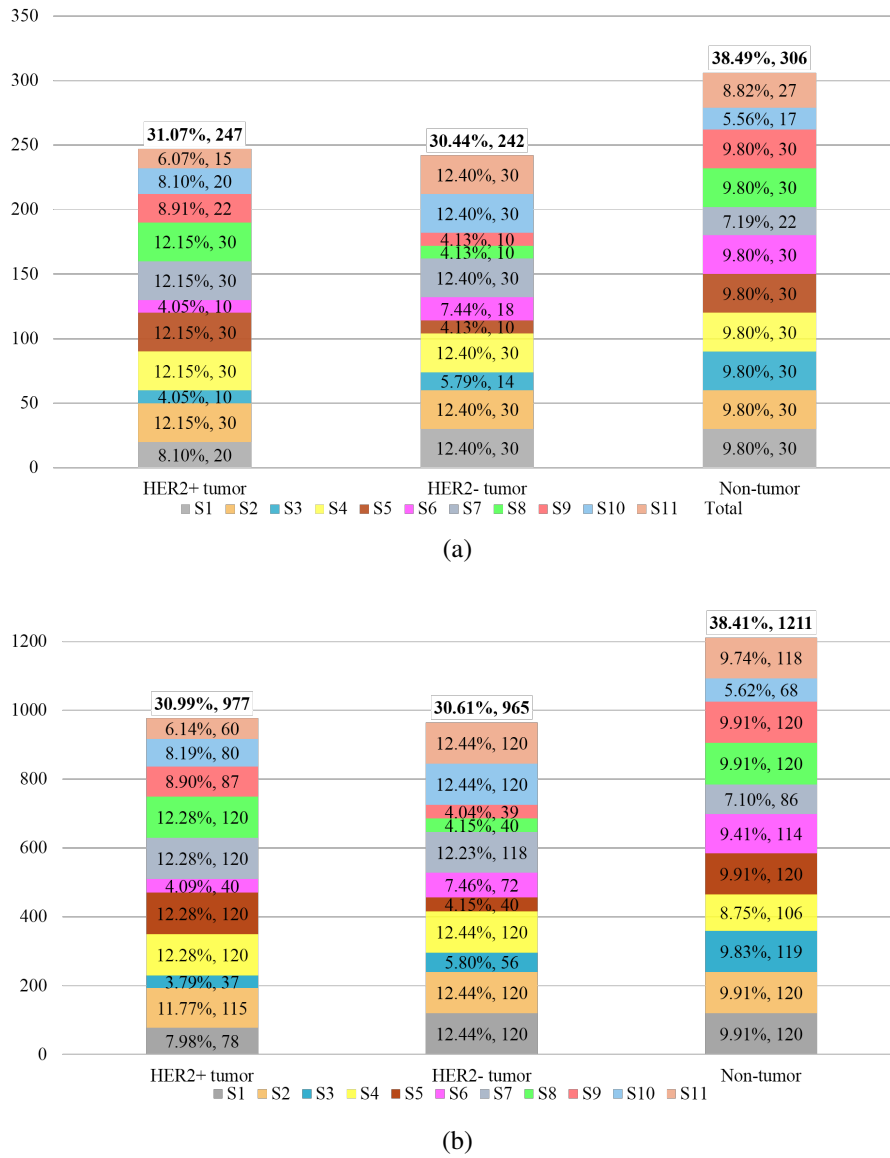


Fig. 7.2 Distribution of labeled image tiles among the patients in HER2 positive tumor, HER2 negative tumor and non-tumor types of malignancy regions and total image tiles for analysis of cancer regions using traditional methods (a) at 1024×1024 pixel size (b) at 512×512 pixel size.

ground truth marked by medical experts. To speed up the reading process during training of convolutional neural networks, image datasets are first converted into lightning memory-mapped databases (LMDB) [Chu 2011] storing images and corresponding labels. For the described experiments, data augmentation is performed offline to organize the databases among different cross validation rounds and to use them for comparative analysis using two CNN architectures. However, it is possible to perform this step online for speedup, as achieved in many previous works such as [Krizhevsky 2012]. Using the same technique, a dataset containing a total of 47,128 images has been generated for necrosis detection tasks of Section 6.2. The process of dataset generation and augmentation is demonstrated in Figure 7.3, where an H&E WSI with five annotations of HER2 positive tumor marked by expert pathologists (one color per pathologist) in

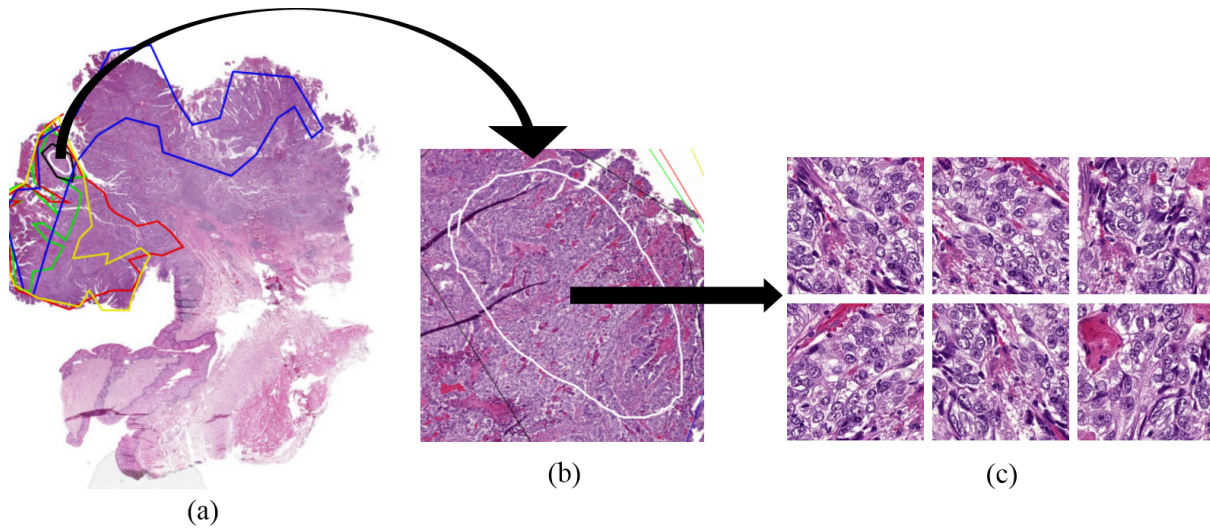


Fig. 7.3 Example of H&E WSI with (a) a few annotations of HER2 positive tumor marked by expert pathologists in the corresponding HER2 WSI (b) a low magnified ($5\times$) region of agreement of most pathologists (c) example images after data augmentation at highest magnification ($40\times$).

the corresponding HER2 slide, is shown in Figure 7.3(a), a low magnified region ($5\times$) containing areas of agreement of most pathologists for the given class in Figure 7.3(b), and a few image tiles after data augmentation in Figure 7.3(c).

The distributions of the image tiles in each of the three type of regions for individual patients (denoted by S) are given in Figure 7.4, showing completely balanced large-scale datasets for exploring deep learning methods. Only the deep learning methods are studied in conjunction to the large-scaled datasets described in this section, mainly in order to reduce space and time complexity of the entire framework. Hence, smaller datasets described in Section 7.2.1 are utilized for image analysis using the traditional machine learning methods.

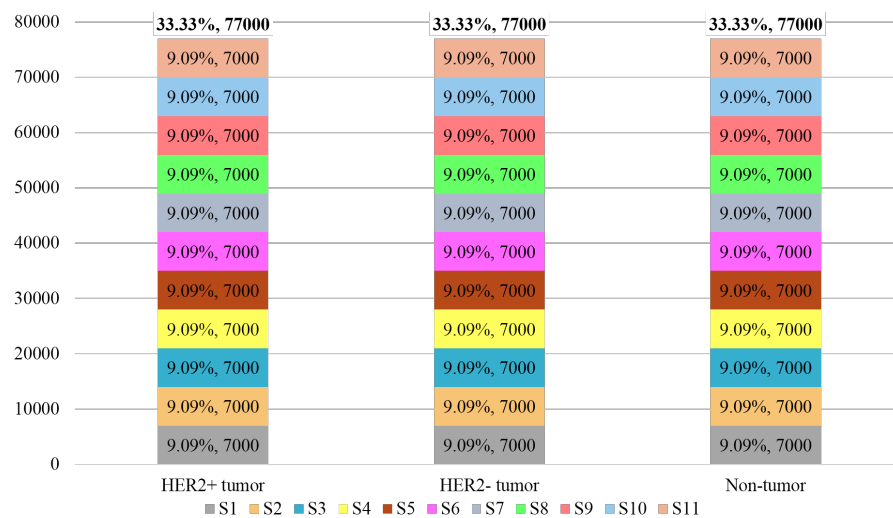


Fig. 7.4 Distribution of labeled image tiles among the patients in HER2 positive tumor, HER2 negative tumor and non-tumor types of malignancy regions and total image tiles for analysis of cancer regions using deep learning methods

7.3 Image Description Methods

7.3.1 Low-level State-of-the-art Features

The low-level methods refer to the group of image description methods applied directly on individual image pixels. Research in this area mainly includes approaches for extracting information in the form of texture, intensity, color, and frequency-based features. In literature, these methods have been applied on various types of histological images for several applications like cancer classification, grading, tissue classification and content-based retrieval. The most frequently used state-of-the-art low-level methods in digital histopathology have been described in detail in Section 2.3.1. The studied low-level feature groups include texture descriptors such as gray level co-occurrence matrix statistics, Gabor filter-bank responses, local binary patterns, Varma-Zisserman textons and color and intensity descriptors such as gray histograms, RGB histograms, HSV histograms and other color-based measurements. These specific feature groups have been explored for analysis of cancer regions in gastric carcinoma histological datasets for the purpose of comparative evaluations with the high-level state-of-the-art and handcrafted features in conjunction with traditional machine learning methods, along with the deep learning methods. The computation of the low-level state-of-the-art image descriptors utilized in experiments is described in the following sections, however, their theoretical background and mathematical details are provided earlier in Section 2.3.1.

7.3.1.1 Texture Descriptors

Texture-based image descriptors have been very frequently used in digital histopathology since many years, due to their ability to quantify the visual appearance of images on the basis of pixel quantities in local neighborhoods, without requiring the knowledge of high-level tissue architecture. In this work, the following state-of-the-art texture-based features have been extracted from gastric cancer tissue images.

GLCM Statistics

These are computed as the 14 texture features described using the same procedures as for necrosis detection in Section 6.2.3, after deriving a gray level co-occurrence matrix for each image tile. The extracted GLCM statistics are Angular Second Momentum, Contrast, Correlation, Sum of Squares, Inverse Difference Moment, Sum Average, Sum Variance, Sum Entropy, Entropy, Difference Variance, Difference Entropy, First Information Measure, Second Information Measure and Maximal Correlation Coefficient.

Gabor Filter-bank Responses

Corresponding to the textural feature extraction step of the necrosis detection process in Section 6.2.3, real parts of Gabor filter kernels of 16 different frequency, orientation and standard

deviation combinations are applied to filter each image in the datasets and a set of 32 Gabor filter-bank features is obtained.

Local Binary Patterns

In the experiments, LBP histograms are generated by considering 8-nearest neighbors of each pixel, hence, LBP codes are generated in the range 0-255 depending on the intensity of each center pixel. From the LBP-coded image, 256-bin LBP histograms are computed and normalized to obtain feature vectors of 256 values.

Varma-Zisserman Textons

Using the Varma-Zisserman (VZ) texton approach, maximum response 8 filters are applied to generate filter responses of images selected randomly from each class. A set of 50 texton cluster centers is selected to represent the texture corresponding to each malignancy level and included in the texton dictionary using the K-means clustering algorithm. A histogram of VZ textons of length 150 is used as the feature vector for each labeled image.

7.3.1.2 Color and Intensity Descriptors

Color and intensity are crucial characteristics in histological image analysis due to specific staining of different tissue components. In the cancer analysis experiments, RGB color space is used. Again, the following color and intensity descriptors are theoretically described in Section [2.3.1](#).

Gray Histograms

Gray level histograms are computed from each image tile to give rise to 256-length feature vectors. These features represent the intensity distributions in gray-level images.

Color Histograms

The computed color histograms include HSV histograms and RGB histograms from each image tile to yield 692- and 768-length feature vectors respectively. These histograms can accurately depict the color distributions of different channels in H&E stained histological images using HSV and RGB color spaces respectively.

Other Color-based Measurements

Other color-based measurements include six quantities namely mean red, mean green, mean blue, mean intensity, standard deviation of intensity and mean chromaticity (defined in Table [6.4.2.1](#)). Some features in this feature group have been introduced as a subset of object-level features for cell nuclei classification, and extended to image regions.

7.3.2 High-level State-of-the-art Features

The well-known graph-based method in histopathological image analysis is known as *proximity graph*, also sometimes called neighborhood graph. Proximity graphs are planar graphs consisting of vertices linked by an edge if and only if specific geometric requirements are satisfied by the vertices [Toussaint 1991]. Proximity graphs are defined with reference to various metrics, however the Euclidean metric is the most commonly used. Voronoi diagram and its dual Delaunay triangulation are two types of proximity graphs widely studied in digital pathology [Sharma 2015a], hence, these can be considered as state-of-the-art for high-level feature extraction. Voronoi diagram [Voronoi 1907] creates a nearest site proximity partitioning of a plane consisting of a set of points (sites) into Voronoi polygons. Delaunay triangulation [Delaunay 1934] (also called Delaunay graph) is the dual of Voronoi diagram. A number of related graphs have been introduced in literature, and explained in Section 2.3.3.1.

In order to facilitate a quantitative comparison of the handcrafted high-level graph-based features with high-level state-of-the-art features, the author has described a set of feature computations using the most common Voronoi-Delaunay representation. Other proximity graphs discussed in Section 2.3.3.1 are not studied in the comparative analysis, as most of these are related or derived from the Voronoi diagram and Delaunay triangulation, hence, will not have significant additional contribution to the originally existing graph information. In the subsequent experiments, a Voronoi diagram is constructed with the corresponding Delaunay triangulation for each image, where centroids of each of the cell nuclei segments are considered as the set of proximity points in the Euclidean plane. Global graph features namely, moments of vertex degree, clustering coefficient, eccentricity, area, perimeter and number of faces in corresponding Voronoi polygon, mean area and mean perimeter of neighboring triangles for each vertex and Euclidean distances between neighboring Delaunay vertices, along with size-based graph features are extracted from each Voronoi-Delaunay representation, leading to a feature vector of 61 features per image (Global graph features are discussed in detail in Section 7.3.3.4). Example of an image tile from the cancer analysis datasets depicting a HER2 negative image region with corresponding cell nuclei segments after performing cell nuclei segmentation procedure, and the resulting Voronoi diagram and associated Delaunay triangulation are illustrated in Figure 7.5.

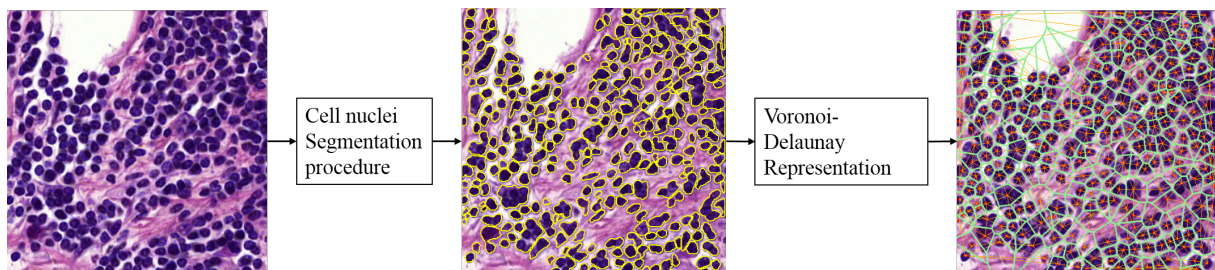


Fig. 7.5 Example of a HER2 negative tumor image tile, cell nuclei segments after segmentation procedure and the resulting Voronoi diagram and Delaunay triangulation for high-level state-of-the-art feature extraction

However, it should be emphasized that for describing the tissue images in the complex histological gastric carcinoma image dataset, Voronoi-Delaunay graphs have certain limitations, such as a geometrically constrained nature, planar properties and only one connected component. Some of these limitations are also noted by authors of cell graphs [Bilgin 2010]. For instance, Voronoi-Delaunay method only expresses geometric relations between vertices, where the vertices are simply point coordinates, and it also imposes certain geometric constraints during construction. Furthermore, since the Delaunay triangulation is planar, it consists of edges only between adjacent vertices and captures spatial relationships within a small vicinity, however, no evidence justifies such a restriction in architecture of tissues. Also, Delaunay graphs are always a single connected component, hence, cannot be effectively used for representing sparse tissues. Moreover, the length of shortest path in Delaunay triangulation is always three and does not depend on spatial distribution in the tissue. Due to these reasons, features derived from Voronoi and Delaunay representations are not considered sufficient to discriminate between the malignancy levels in histopathological regions of gastric cancer, and this statement will also be validated by the experimental results in next sections.

7.3.3 High-level Handcrafted Features

The main objective of this group of methods is to model the spatial relationships and architectural properties between histopathological tissue components using innovative handcrafted techniques. To outline the procedure, cell nuclei are first isolated at high magnification using the automatic cell nuclei segmentation strategy, followed by construction of cell nuclei attributed relational graphs of the tissue regions. These graphs represent tissue architecture comprehensively, as they contain information about cell nuclei characteristics as vertex attributes, along with knowledge of neighborhood in the form of edge linking and edge attributes. Global graph measurements are derived and used to discriminate between the different types of malignancy levels in H&E images based on HER2 immunohistochemistry. The developed graph-based techniques can be included in the category of high-level (architectural) feature extraction methods, as they involve extraction of suitable topological features from tissue images by first constructing representative graphs on the underlying tissue objects.

7.3.3.1 Motivation

Despite of several widely known low-level state-of-the-art methods, the author firmly believes, and this phenomenon has also been observed in past studies for histopathological image analysis, that most of the described pixel-based methods are not self-sufficient to effectively represent the malignancy properties in histological images, since they do not incorporate other significant characteristics such as architectural information and neighborhood relationships of the tissue components. Therefore, the author has proposed a set of solutions using high-level image description involving hand-engineered image analysis techniques for modeling and quantifying tissue components and relations between them. For this purpose, a novel graph-based method

is developed for graph construction in histopathological image analysis called *cell nuclei attributed relational graphs* (cell nuclei ARG), which can be potentially designed to represent the knowledge in tissue regions in an effective way. In the proposed approach, numerical attributes are assigned to the graph vertices *i.e.* cell nuclei and graph edges, representing morphological and spatial neighborhood properties of the cell nuclei in the tissue. The author is motivated to explore the strengths and weaknesses of the cell nuclei ARG method due to several reasons.

To the best of author's knowledge, mostly the graph-theoretic techniques explored for histological image analysis have been applied on standard sized individual regions of smaller dimensions and low resolutions. With the rapid advancement in hardware and software equipment in the field of digital pathology in the past few years, glass slides are now available as high resolution digital images *i.e.* WSI ready to be analyzed. The histopathological community has observed the need of developing analysis techniques applicable to WSI datasets, and not only to some specific regions. The previous solutions using graphs are computationally expensive to be applied on such high resolution WSI. Another limitation of many of the earlier graph-based approaches is their semi-automatic nature, as cells or other interesting objects have been mostly interactively marked with their boundaries before graph construction, which also makes the process time consuming. For some other graph-based methods, the corresponding tessellations provide cell boundaries which may not coincide with the actual cell boundaries. Sometimes simpler and more general segmentation algorithms like thresholding and k-means algorithm have been applied, leading to blob-like appearance of cells and cell clusters which can clearly affect the accuracy of subsequent analysis. Hence, there is a requirement to include automatic segmentation for isolating the different tissue components before the use of graphs.

Moreover, the resultant cell nuclei ARGs are topologically different from proximity graphs, as these are generated using non-planar linking of vertices in a region of interest. The cell nuclei ARGs are more appropriate graph-based method for the given classification problem, as they contain more comprehensive information about the tissue architecture due to vertex and edge attribute assignment, compared to proximity graph which only expresses the geometric relation between a given set of Euclidean points as vertices. Also, as mentioned earlier, proximity graphs have geometric constraints so they can only capture spatial relationships between cell nuclei within a small vicinity, whereas cell nuclei ARGs are constructed depending on the spatial arrangements of the cell nuclei, hence, can be used for representing dense or scattered tissue components. Furthermore, unlike the Delaunay triangulation, cell nuclei ARGs can contain more than one connected component. These properties provide a greater flexibility to cell nuclei ARGs for modeling diverse types of tissues based on the tissue topology.

Cell graphs [Gunduz 2004] have been an inspiration for the commencement of the novel graph-based handcrafting procedure. Cell graphs are described by their authors as more suitable for representing cancer tissue characteristics compared to proximity graphs due to the above advantages, especially due to less geometrical constraints for cancer tissue modeling. This has provided a higher clarity to the author in general, as the proposed cell nuclei ARG and its variants have been conceptualized with fewer linking constraints keeping the heterogeneous

structural properties of the tumor in consideration. However, the cell nuclei ARG are distinct from cell graphs, as they have been developed as more detailed graph constructions at higher image resolutions and tissue magnifications by computing visual characteristics of individual cell nuclei (instead of single points), their neighborhoods and relational information between neighbors, using different edge linking rules and additional edge information. The overall global graph feature extraction is also distinct from the cell graph method.

Therefore, the author investigates the power of handcrafted graph-based approaches for the classification of gastric carcinoma, and also examines the performance with respect to the most widely used state-of-the-art low-level and high-level image description methods. Variations in the cell nuclei attributed relational graphs are also extensively studied and comparatively analyzed for effective knowledge description and classification in H&E stained whole slide images of gastric cancer. The comparative evaluation between the different tiers of feature extraction methods can emphasize whether the primary basis of the author's hypothesis is true, and is an interesting aspect of research in this direction further leading to hybrid approaches to combine the salient visual information.

7.3.3.2 Cell Nuclei Attributed Relational Graphs

A graph is a mathematical structure consisting of a finite set of points (or objects) known as vertices that are linked to each other by a set of edges using a predefined criteria. Attributed Relational Graph (ARG) is a graph $G(V, E, A, B)$ consisting of vertices and edges along with their associated attributes [Sharma 2012]. The vertex attributes for a vertex $v_i \in V$ are included in a vector $\mathbf{a}_i = [a_i^{(p)}]$, ($p = 1, 2, 3, \dots, P$, $\mathbf{a}_i \in A$), where P is the length of vector \mathbf{a}_i , and the edge attributes for edge $e_{ij} \in E$ between vertices v_i and $v_j \in V$ are described by vector $\mathbf{b}_{ij} = [b_{ij}^{(q)}]$, ($q = 1, 2, 3, \dots, Q$, $\mathbf{b}_{ij} \in B$), where Q is the length of vector \mathbf{b}_{ij} . In the proposed method, cell nuclei ARGs are constructed by considering the cell nuclei segments as graph vertices, which are connected by the following linking rule. A vertex v_j is included in the neighborhood $N_{i,j}$ of vertex v_i if and only if it is located inside a disc of radius r centered at the vertex centroid, which is referred as the *maximum spanning edge length*. In other words, a vertex v_j is included the neighborhood $N_{i,j}$ of vertex v_i if and only if the following linking rule is satisfied.

$$d_e(i, j) \leq r \quad (7.1)$$

where $d_e(i, j)$ denotes the Euclidean distance between the centroids of vertex v_i and v_j . The *concave object separation* step of cell nuclei segmentation procedure [Wienert 2012] ensures that the resultant cell nuclei segments are mainly convex, however, they may consist of smaller concave subregions. Thus, centroids are mostly located inside or on the cell nuclei segments. The resulting cell nuclei ARGs are undirected and non-planar. The schematic diagram of cell nuclei ARG is shown in Figure 7.6. After identifying the vertices and linking the edges between neighboring vertices, attribute vectors are assigned to each vertex and each edge, and defined as follows.

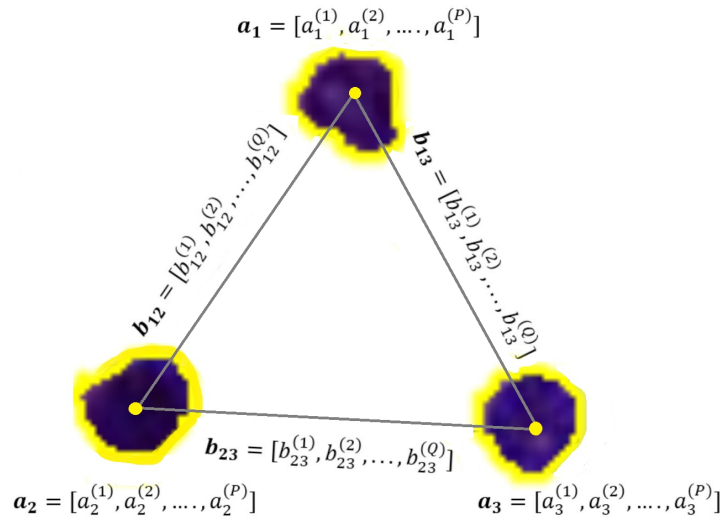


Fig. 7.6 Schematic diagram of cell nuclei ARG

1. **Vertex attributes:** An attribute vector \mathbf{a}_i of 29 attributes are assigned to each cell nuclei ARG vertex $v_i \in V$. The two subsets of attributes associated with the vertices are:
 - (a) **Object-level attributes:** The properties of individual cell nuclei are important indicators in detecting malignant changes in tissues. A set of 25 object-level attributes is used for representing the visual characteristics of each cell nuclei segment. These include morphological or shape-based attributes of cell nuclei and characteristics of their contour intensities. The shape-based attributes are Object Pixels, Minimum Distance to Tessellation Border, Pixels at Layer Border, Maximum Distance to Border, Aspect Ratio of Bounding Ellipsoid, Minor Axis of Bounding Ellipsoid, Major Axis of Bounding Ellipsoid, Angle of Bounding Ellipsoid, Form Factor of Contour, Convexity of Contour, Length of Contour, Area of Contour, Form Factor of Convex Hull, Length of Convex Hull, Area of Convex Hull, Feret, Minimal Radius of Enclosing Centered Circle, Maximal Radius of Enclosed Centered Circle, Roundness and Form Factor. Additionally, fractal dimension of each contour is also computed using the box counting algorithm for Minkowski's dimension [Falconer 2004]. Attributes based on contour intensities are Mean Intensity on Contour, Standard Deviation of Intensity on Contour, Contour Value and Gradient Fit. These have been explained in Table 6.4.2.1, and are a subset of the cell nuclei segment features applied for cell nuclei classification.
 - (b) **Neighborhood attributes:** These include the neighborhood characteristics of each vertex and include vertex degree, vertex clustering coefficient, vertex eccentricity and mean euclidean distance of a vertex to its neighbors. Vertex degree is the number of number of vertices linked to it by edges. Vertex clustering coefficient is the measure of the extent of the tendency of a vertex to cluster with other vertices. Vertex eccentricity is the measure of how far a vertex is from the most distant vertex from it in the graph. Their numerical definitions are presented in Section 2.3.3.7. The measure of mean distance of a vertex to

all its neighbors is indicative of cell nuclei density of a region and inversely proportional to it, as a higher value represents cell nuclei scattered more sparsely and a smaller value denotes a cluster of closely spaced cell nuclei. These neighborhood attributes collectively quantify the interactions between cell nuclei with their neighbors, thereby characterizing the spatial relationships between tissue components in the overall tissue architecture.

2. **Edge attributes:** The relationships between linked vertices are described using edge attributes. For an edge $e_{ij} \in E$ between vertices $\{v_i, v_j\} \in V$ such that $v_j \in N_{i,j}$, the edge attribute vector \mathbf{b}_{ij} consisting of 26 attributes is computed. It includes the euclidean edge length, where centroids of vertices v_i and v_j are given as (x_i, y_i) and (x_j, y_j) respectively, and the difference between the corresponding object-level vertex attributes as shown in the following equations.

$$b_{ij}^1 = \sqrt{(x_i - x_j)^2 + (y_i - y_j)^2} \quad (7.2)$$

$$b_{ij}^k = |a_i^{(k-1)} - a_j^{(k-1)}|, \quad k = 2, 3, \dots, 25 \quad (7.3)$$

An additional edge attribute is orientation, defined as the gradient of the edge line connecting the two vertices.

$$b_{ij}^{26} = \frac{y_i - y_j}{x_i - x_j} \quad (7.4)$$

This attribute is only indicative of the underlying open, closed or dense arrangements of cells and its usage and significance is explored later while computing the global graph features.

7.3.3.3 The Variants of Cell Nuclei Attributed Relational Graphs

The basic construction principles described above have been utilized to generate eight different kinds or variants of cell nuclei ARG, representing diversity in the contained knowledge of the corresponding image regions and their malignancy levels. The two variables used for designing the variants of cell nuclei ARGs are linking rule to determine r and ARG construction method based on vertex information.

Linking Rule: Determining r

There are two different ways in which the maximum spanning edge length r has been assigned. Different types of cell nuclei ARGs are generated, with one of the variables being the linking rule used for determining r . The linking rules are as follows.

1. **Fixed r (r_F):** In this type of linking rule, the maximum spanning edge length r is set to a fixed value that is independent of the cell nuclei density of each image. In other words, r is empirically predetermined for an effective representation of spatial arrangements of cell nuclei in the corresponding image regions, and kept constant for all types of regions. An advantage of using fixed r (denoted by r_F) irrespective of image or region type, is that an accurately estimated and sufficiently large value of r_F ensures that the cell nuclei segments are connected to all their ‘closer’ neighbors included in a group of cells in the tissue, leading

to formation of connected components which represent the histological architecture and spatial relationships within the tissue in an effective manner. Also, an increase in maximum spanning edge length r_F means a higher number of edges and associated edge attributes, leading to denser cell nuclei ARGs, and vice-versa. The value of r_F is decided after visual and experimental investigation for a set of different values while construction of the cell nuclei ARG.

2. **Adaptive r (r_A):** A cell nuclei ARG with adaptive maximum spanning edge length r is constructed using a dynamic value of r (denoted by r_A) to perform the edge linking in each cell nuclei ARG individually. In this case, the edge linking parameter r_A is not kept constant for all regions, and computed for each region of interest depending on the number of cell nuclei in the image, ensuring that these are connected to their closer neighbors.

For computation of r_A , an imaginary regular grid consisting of equal sized rectangular grid cells is placed on each image tile and it is assumed that the cell nuclei segments are uniformly spaced such that each grid cell consists of a cell nuclei centroid at its center. The maximum spanning edge length r_A is now set to the maximum distance between the given cell nuclei centroid and its 8-neighbors. The uniform grid assumption is depicted in Figure 7.7. Let the length of the image tile be represented by l and its breadth be represented by b in

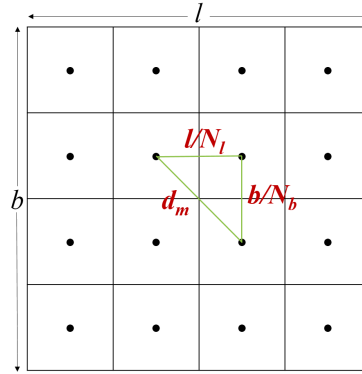


Fig. 7.7 Uniform grid assumption for cell nuclei ARG with adaptive r

pixels. For a regular grid placed on the image tile with each grid cell consisting of a cell nuclei centroid, there are N_l number of grid cells in each row and N_b number of grid cells in each column. Horizontal distance between two consecutive grid centers will be l/N_l and vertical distance will be b/N_b . Hence, the maximum distance of a given grid cell center (coinciding with a cell nuclei centroid) with its 8-neighbors will be the diagonal distance given by,

$$d_m = \sqrt{\left(\frac{l}{N_l}\right)^2 + \left(\frac{b}{N_b}\right)^2} \quad (7.5)$$

For a square image tile, $l = b$ and $N_l = N_b = N$. The distance now becomes,

$$d_m = \sqrt{\frac{2l^2}{N^2}} = \frac{\sqrt{2}l}{\sqrt{N}} \quad (7.6)$$

where, N^2 defines the total number of grid cells in a square grid. Also, according to the assumption, each grid cell consists of a cell nuclei centroid, so the value N^2 is the total number of vertices N_v in the corresponding cell nuclei ARG. To obtain the maximum spanning edge length r_A for an image tile, d_m is rounded off to the nearest higher integer.

The main advantage of the dynamic linking approach is the elimination of the prerequisite of empirically fixing r as it is automatically calculated for each cell nuclei ARG during graph construction. This calculation of a dynamic r_A is more flexible depending on the type of malignancy, and increases the probability that a cell nuclei segment is connected to its nearest neighbors irrespective of their actual distance in the tissue. The edge linking method used in this graph can be considered similar to proximity graphs due to the imposed geometrical constraints. However, it is less constrained compared to the Voronoi-Delaunay representations because the computation is performed for determining the maximum spanning edge length instead of actual positions of the cell nuclei centroids, so all the edges having a value less than the calculated r_A will still be present in the resulting cell nuclei ARG.

For the computation of r_A as above, a uniform distribution assumption considering equally spaced cell nuclei segments is not always indicative of the tumor tissue because in the real case, cells are not uniformly spaced and can lead to multiple connected components. However, it should be specified that the ideal case assumption is not determining actual edge lengths but only putting a threshold to determine the connectivity between tissue components. The approach can result in multiple connected components representing cell neighborhoods in tissues, and the true tissue architecture.

One limitation of the adaptive r technique described above, similar to Voronoi-Delaunay representations, it is constrained by a dynamic value of r , so in some cases the graph may not fully represent the global arrangements of cell nuclei in the tissue. Nevertheless, it will be interesting to observe the behavior of dynamically and automatically calculated r_A without the prerequisite of estimating it, and compare the performance of the two edge linking approaches.

A grid-based approach for cell graphs is described in [Demir 2004], [Demir 2005a], [Bilgin 2007], [Gunduz-Demir 2007]. However, it is different from the linking method described above, as in that approach, in order to construct cell graphs, grids of specific sizes (*e.g.* 4, 8, 10 pixels) are considered and the probability values for existence of a cell node is determined by averaging their binary pixel values inside a grid cell during the node identification of cells or cell clusters. In the above described approach, an imaginary grid has been used only to calculate the maximum spanning edge length during the edge linking phase, where grid size is not constant and depends on the number of cell nuclei using the uniform distribution assumption.

On comparing the computational complexities of the two linking approaches, if the empirical analysis part is ignored, fixed r method has lower costs as it doesn't require the r calculation step for each image. On the other hand, accuracy of the assumption of r affects the subsequent classification performance, hence, should be carefully determined using experimental evidence, which increases the overheads related to fixed r method.

ARG Construction Method based on Vertex Information

The second variable to develop variants of cell nuclei ARG is the construction method to include and represent vertex information. There are four ways in which this has been accomplished.

1. **Generic cell nuclei ARG construction (gARG):** As the name suggests, a generic cell nuclei ARG is constructed by considering all the cell nuclei segments comprising the image tile as its vertices, irrespective of their individual classes or types. The resulting cell nuclei ARG considers all the vertex and edge attributes as described in Section 7.3.3.2.
2. **Nuclei-specific ARG construction (nsARG):** A nuclei-specific ARG is constructed by considering all the cell nuclei of one type (or class) in question. The class for each cell nuclei segment is automatically computed using the cell nuclei classification step described in Section 6.4.2. For each image tile, one nuclei-specific ARG is used to represent the spatial arrangements of each of the following cell nuclei classes, namely, epithelial cells, leukocytes, fibrocytes and conglomerates. The other three types of cell nuclei segments, namely, fragments, other cells and artefacts have not been considered in constructing nuclei-specific ARG due to insignificant impact on the malignancy characteristics of tissue regions. Hence, a maximum of four nuclei-specific ARGs are generated for each region of interest. The vertex attributes of nuclei-specific ARGs do not include the morphological and contour intensity-based measures because these have been previously utilized during cell nuclei classification stage as a subset of the cell nuclei features. Hence, such a construction emphasizes purely on the neighborhood relationships among cell nuclei of the same type in a given ROI.
3. **Nuclei-composite ARG construction (ncARG):** A nuclei-composite ARG takes into account all the cell nuclei segments in an image, along with their individual identities. Here also, the class for each cell nuclei segment is automatically determined as one of the seven types using the cell nuclei classification step described in Section 6.4.2, however, there is only a single nuclei-composite ARG generated for each image. The vertex attributes of nuclei-composite ARGs do not include the morphological and contour intensity-based measures because of the same reason described for nuclei-specific ARG, but there is an additional vertex attribute denoting the category of each cell nuclei segment. Another difference from the nuclei specific construction is the consideration of different types of cell nuclei in the same ARG, emphasizing on the inter-class interactions in the neighborhood.
4. **Nuclei-composite ARG construction with additional vertex attributes (ncARG_{v+}):** This is constructed similar to nuclei-composite ARG, however, it also takes into account the morphological and contour intensity-based vertex attributes along with the class of each vertex. This can be seen as a repetition using the same information of vertex attributes twice, once during cell nuclei classification and once during graph-based region analysis, however, it may be argued as a method for information recycling in order to enrich the knowledge from the tissue image.

In terms of space and time complexity, the generic cell nuclei ARG construction has the lowest requirements due to its simplistic and straightforward implementation. Nuclei-composite ARG

construction has intermediate requirements due to additional step of cell nuclei classification prior to graph construction. For the construction with additional vertex attributes, space requirement is higher than the former counterpart. Nuclei-specific ARG construction is the most computationally intensive as it consumes around four times more execution time due to multiple graphs generated for each image and more space when the corresponding graphs need to be explicitly saved, also requiring additional time due to the intermediate cell nuclei classification step.

Eight Resulting Cell Nuclei ARG Variants

Based on the combinations of the above two variables, the eight different versions (or variants) of the cell nuclei ARG developed and explored in the subsequent experiments, described as follows. The cell nuclei ARG variants are assigned the respective acronyms to easily refer to these in the following text.

1. **gARG**[r_F]: The generic cell nuclei ARG and fixed r variant is the most elementary and intuitive type, and has also been extensively explored in [Sharma 2016]. For the described experiments, r is empirically determined and then fixed prior to graph construction.
2. **gARG**[r_A]: A generic cell nuclei ARG with adaptive r is constructed by using the generic construction method and the adaptive edge linking method to compute individual values of r for each image.
3. **nsARG**[r_F]: Nuclei-specific ARG and fixed r uses the nuclei-specific ARG construction method such that values of $r_k, k = \{1, 2, 3, 4\}$ are empirically determined for the four types of cell nuclei. It is intuitive that these values will be greater than r determined for generic cell nuclei ARG, as the latter considers more closely spaced cell nuclei segments of all types instead of one specific type of cell nuclei segments.
4. **nsARG**[r_A]: Nuclei-specific ARG with adaptive r is constructed for each category of cell nuclei separately and maximum spanning edge length r is dynamically calculated for a region of interest considering only the particular cell nuclei category of the corresponding nuclei-specific ARG.
5. **ncARG**[r_F]: Nuclei-composite ARG and fixed r uses the nuclei-composite ARG construction method using all the cell nuclei segments with identities, and a predetermined value of r for all the images.
6. **ncARG**[r_A]: Nuclei-composite ARG and adaptive r is constructed using all the cell nuclei segments with their identities, and r is dynamically calculated for each image.
7. **ncARG_{v+}**[r_F]: Nuclei-composite ARG with additional vertex attributes and fixed r are obtained by considering all cell nuclei segments, their vertex identities, along with shape and contour intensity-based vertex attributes and fixing the value of r .
8. **ncARG_{v+}**[r_A]: Nuclei-composite ARG with additional vertex attributes and adaptive r use the same construction method as above, though edge linking is performed using the adaptive r linking method.

An example of a smaller image tile from a larger region (same image tile shown in Figure 7.5), corresponding segmentation and the cell nuclei ARG representations of the eight types of variants are shown in Figure 7.8.

In Table 7.1, visible characteristics in H&E stained images for the three malignancy types on the basis of HER2 immunohistochemistry, and the corresponding architectural properties of the resulting cell nuclei ARG are summarized. Figure 7.9 shows typical examples of each type of region from the same WSI, overlay with the constructed $gARG[r_F]$ where $r = 50$ and illustrates majority of the properties listed in the table. These properties, along with other subtle changes not clearly visible to human eyes, are extracted using the various self-designed handcrafted features of the different attributed relational graphs to facilitate an accurate automatic differentiation of the types of cancer regions.

Table 7.1 *Cancer interpretation based on graphs: summary of visible characteristics and corresponding cell nuclei ARG properties*

Type of malignancy region (based on HER2 IHC)	Visible characteristics of region	Architectural properties of resulting cell nuclei ARG
HER2 positive tumor	Cells are mostly large, long and oval	Large vertex size and distinct morphological properties
	Cellular density is very high	Dense graphs
	Intercellular distance is very less	Highly clustered graphs
	Cell boundaries are overlapping	Many vertices are large conglomerates
	Cells have a clustered appearance	Few large and dense connected components
	Cells are either randomly accumulated or arrange in dense patterns	Random or globular arrangements in the graph
HER2 negative tumor	Partly, the areas contain large cells and partly contain small cells	Intermediate vertex sizes and morphological properties
	Cellular density is medium or high	Medium or high graph density
	Intercellular distance is medium	Intermediate clustering in graphs
	Some cell boundaries are not clear	Many vertices are small conglomerates
	Cells can be scattered or clustered	Intermediate connected components
	Cells just started to arrange in closed patterns by forming a single ring	Distinct closed patterns (including single layers) in the graph
Non-tumor	Cells are usually small and round	Small vertex size and distinct morphological properties
	Density of cells is medium or low	Sparse graphs
	Intercellular distance is highest	Scattered graphs over whole image
	Cells are distinct and boundaries are not overlapping	More number of isolated vertices or lower vertex degrees
	Cells are scattered in connective tissue	Many connected components, each of small size
	No distinct arrangements or grouping of cells	Fewer closed patterns and more open structures in the graph

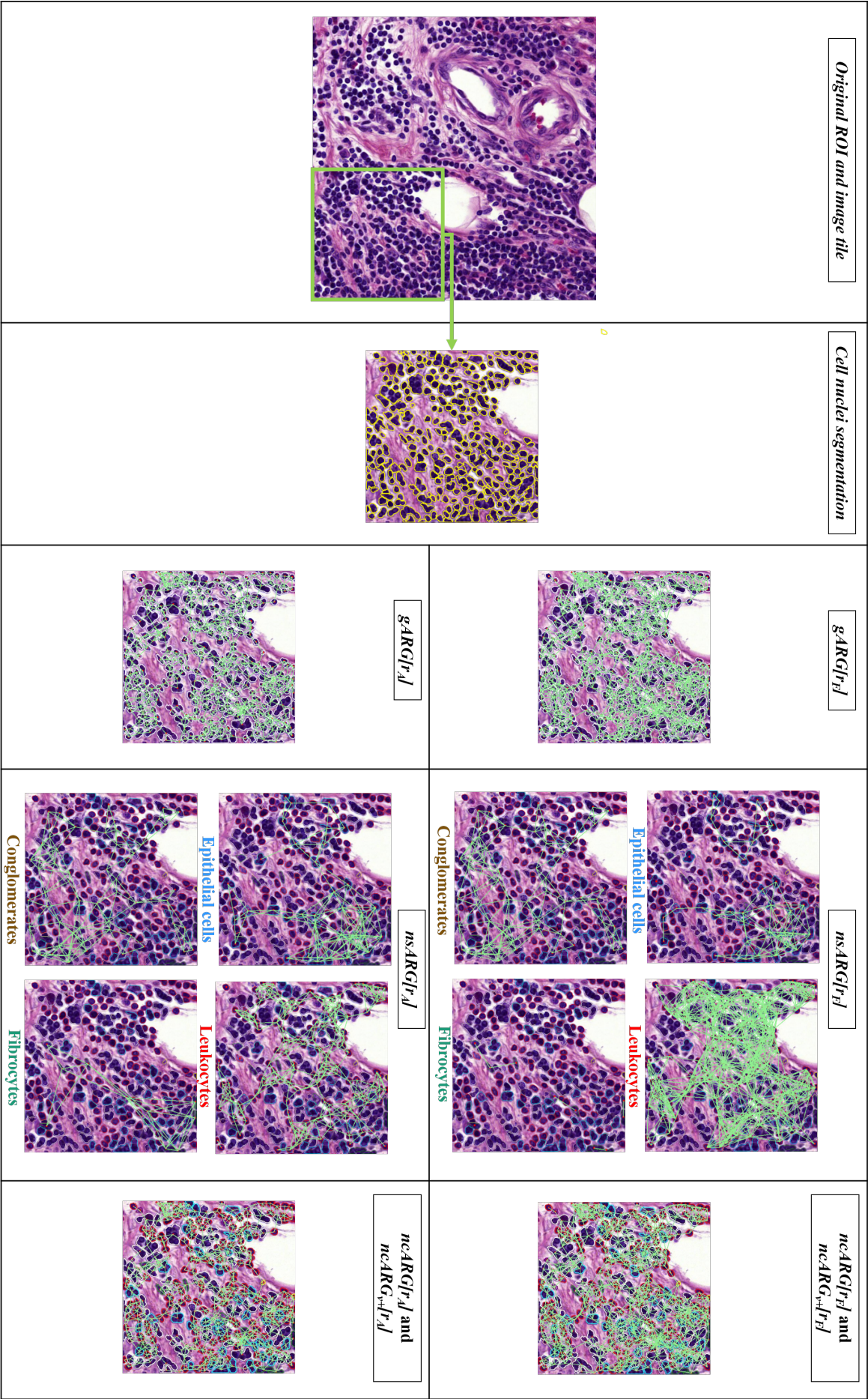


Fig. 7.8 Example of an image tile from a larger ROI, its segmentation and representations of eight cell nuclei ARG variants (the corresponding ncARG and ncARG_u have same appearance but difference in the corresponding vertex attributes)

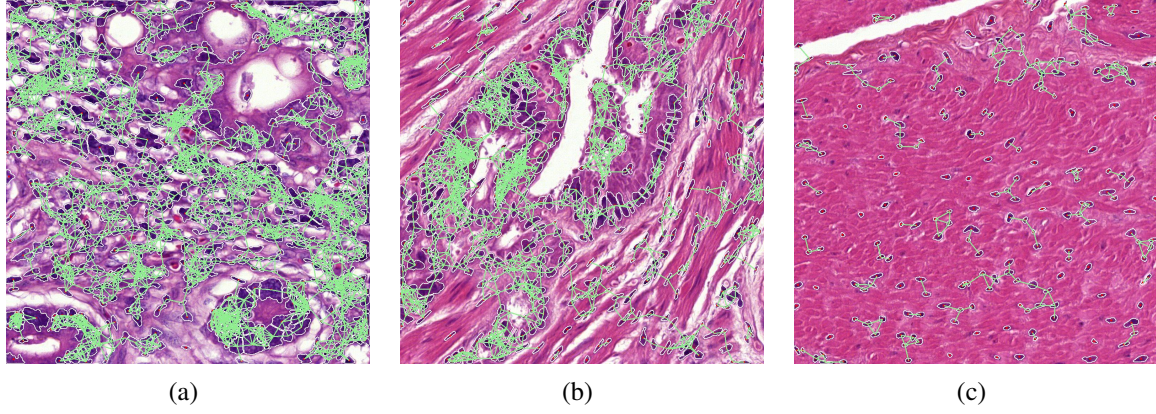


Fig. 7.9 Typical example of malignancy regions overlay with their $gARG[r_F]$, $r = 50$ showing visible characteristics for (a) HER2 positive tumor (b) HER2 negative tumor (c) Non-tumor

7.3.3.4 Global Graph Features

After design and construction of the cell nuclei ARGs for each region of interest, a method is required to extract global quantitative information describing the overall characteristics of each graph. The applied method uses averaging techniques on vertex and edge attributes giving rise to moment-based features, along with additional computation of measurements based on graph size, explained as follows.

Moment-based Features

These features include first, second and higher order moments of the vertex and edge attributes of each cell nuclei ARG. They are calculated as mean, variance, skewness and kurtosis. Minimum and maximum values of the attributes in the cell nuclei ARG are also included in this feature set. Their definitions are shown as follows, where $a_i^{(p)}$ represents the p^{th} vertex attribute for the i^{th} vertex. Similar computations are also made for each edge attribute $e_{ij}^{(q)}$ in the cell nuclei ARG containing N_v vertices and N_e edges.

Mean:

$$\mu\{a^{(p)}\} = \frac{1}{N_v} \sum_{i=1}^{N_v} a_i^{(p)} \quad (7.7)$$

Variance:

$$\sigma^2\{a^{(p)}\} = \frac{1}{(N_v - 1)} \sum_{i=1}^{N_v} (a_i^{(p)} - \mu\{a^{(p)}\})^2 \quad (7.8)$$

Skewness:

$$s\{a^{(p)}\} = \frac{1}{\sigma^3(N_v - 1)} \sum_{i=1}^{N_v} (a_i^{(p)} - \mu\{a^{(p)}\})^3 \quad (7.9)$$

Kurtosis:

$$\kappa\{a^{(p)}\} = \frac{1}{\sigma^4(N_v - 1)} \sum_{i=1}^{N_v} (a_i^{(p)} - \mu\{a^{(p)}\})^4 \quad (7.10)$$

Maximum:

$$m_1\{a^{(p)}\} = \max_{1 \leq i \leq N_v} \{a_i^{(p)}\} \quad (7.11)$$

Minimum:

$$m_2\{a^{(p)}\} = \min_{1 \leq i \leq N_v} \{a_i^{(p)}\} \quad (7.12)$$

For the orientation edge attribute, moment-based features defined above are not calculated. Instead, one feature called *total orientation* is obtained as the sum of the orientations of all the edges in the cell nuclei ARG. The feature will vary with rotation for open structures, but the author wants to specify that it is mainly introduced in order to differentiate between single closed layer (or ring) of cells found mostly in HER2 negative tumor which will have lower value of this feature, compared to dense group of cells in HER2 positive tumor which will have intermediate and higher values of this feature. For non-tumor, due to scattered and more isolated cells, even when the structures are open, the value will tend to remain lower compared to the values for HER2 positive tumor. Hence, the feature can differentiate between a few distinct patterns in the cell nuclei ARG, such as isolated cells or small open structures (non-tumor), closed circular arrangements in single layer (HER2 negative tumor) and dense groups of cells (HER2 positive tumor).

In addition, the two nuclei-composite ARG constructions, namely, $ncARG$ and $ncARG_{v+}$, use seven additional global graph features related to vertex identities. As vertex identity is a vertex attribute in nuclei-composite ARGs, computation of the total number of vertices belonging to each cell nuclei category in the ARG give rise to seven numeric values which are included in the global graph features.

Size-based Features

These features include number of vertices, number of edges, number of connected components, cyclomatic number, graph density, graph irregularity and number of triangles in each cell nuclei ARG. A connected component of an undirected graph is a subgraph containing vertices connected to each other by paths, but it is not connected to any other additional vertex in the supergraph. Connected components are computed using a depth-first search procedure [Hopcroft 1973]. A cycle is a path where the start vertex is same as the end vertex. Cyclomatic number (also called circuit rank) γ_G of a graph G is the minimum number of edges which must be removed in order to eliminate all cycles from G [Volkman 1996]. Considering a graph with N_v vertices, N_e edges and N_c connected components, γ_G is given by,

$$\gamma_G = N_e - N_v + N_c \quad (7.13)$$

The density of a graph G , denoted as D_G measures ratio of edges in E compared to the maximum possible number of edges between vertices in V . Graph density is calculated using [Coleman 1983],

$$D_G = \frac{2N_e}{N_v(N_v - 1)} \quad (7.14)$$

Graph irregularity [Albertson 1997] I_G denotes the imbalance of all edges and is calculated by,

$$I_G = \sum_{e_{ij} \in E} |D_i - D_j| \quad (7.15)$$

where D_i represents the degree of vertex v_i and $|D_i - D_j|$ is defined as the imbalance of edge e_{ij} . Number of triangles in the graph is measured by [Alon 1997].

$$\Delta_G = \frac{tr(A^3)}{6} \quad (7.16)$$

where A represents the adjacency matrix of the graph and tr represents trace of a matrix.

As a result, from each $gARG$ construction, a feature vector of 332 global graph features are extracted. The number of features extracted from individual $nsARG$ construction vary for each image depending on the existence of a particular type of cell nuclei and (or) links between them, yielding upto a maximum of four graphs and 188 features per graph. Thus, from the combined set of four nuclei-specific ARGs, a total of 752 (188*4) global graph features are extracted for each image. The missing global features in a small fraction of samples are substituted to zero for simplicity and to further emphasize on the missing feature values of certain cell nuclei types (especially in the non-tumor class). Furthermore, for the $ncARG$ construction, a set of 195 (188+7) features, and for the $ncARG_{v+}$ construction, 339 (332+7) global graph features are extracted from each image.

7.3.4 Combination of Features

In order to achieve optimum classification performance, a hybrid approach is used to combine the information contributed by different feature groups, thereafter relevant information is selected. Firstly, a hybrid low-level feature set is generated by combining specific low-level features. Furthermore, the extracted high-level cell nuclei ARG features from its effective variants and low-level features are also concatenated to obtain a richer hybrid feature set for each image region.

To overcome the curse of dimensionality [Hughes 1968] due to a very large number of features, each combination of features is reduced to a smaller dimension using a feature selection process. Feature selection is defined as the process of selecting a subset of relevant features that can be utilized in subsequent machine learning and classification [Guyon 2003]. The primary reason for using a feature selection for the cancer classification problem in conjunction with dimensionality reduction, is that the image data may contain features that are either redundant or irrelevant, that may be removed without incurring loss of meaningful information or influencing the classification performance significantly.

The feature selection approach applied on the combinations of feature vectors for analysis of cancer regions simply makes use of correlation analysis. The procedure applied evaluates the degree of association between each pair of features x and y using the Pearson's correlation function [Pearson 1895], [Fisher 1921]. Pearson's correlation coefficient for features

$x = \{x_1, x_2, \dots, x_n\}$ and $y = \{y_1, y_2, \dots, y_n\}$ for n instances with means \bar{x} and \bar{y} and standard deviations s_x and s_y is shown in equation 7.17 as,

$$r_{xy} = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{(n-1)s_x s_y} \quad (7.17)$$

where,

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i \quad (7.18)$$

and

$$s_x = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2} \quad (7.19)$$

For m features, this procedure generates an $m \times m$ correlation matrix, and highly correlated features ($r_{xy} \geq 0.9$) are discarded. Removing the correlated attributes can avoid redundancy during prediction as these attributes usually show similar effect on the classifiers. Such a selected subset of features also leads to a higher time and space efficiency. The procedure is applied on the set of hybrid low-level features and the set of hybrid low-level and high-level features to obtain the respective selected subsets of features.

The feature groups initially considered for combination and lengths of feature vectors before and after performing correlation analysis are provided in Table 7.2. Generally, the individual feature groups represent unique information in the image. Specifically, among the low-level features, RGB histograms are included from the three types of color histograms due to most satisfactory classification performance during experiments, and because other color histograms have long lengths of the corresponding feature vectors but may contain potentially correlated color information. Also, among the high-level features, $ncARG_{v+}[r_F]$ variant of the cell nuclei ARG is selected as it contains composite knowledge about the underlying graphs.

Table 7.2 Results for feature selection based on correlation analysis after combination of features

Combination type	Feature groups considered for combination	Original length of feature vectors	Length of feature vectors after feature selection	Correlated features removed as percentage of total from respective group
Hybrid low-level features	GLCM statistics, Gabor filter-bank response, LBP histograms, Varma-Zisserman textons, RGB histograms and other color-based measurements	1225	351	35.71%, 84.38%, 33.20%, 34.67%, 91.15%, 83.33%
Hybrid low-level and high-level features	GLCM statistics, Gabor filter-bank response, LBP histograms, Varma-Zisserman textons, RGB histograms, other color-based measurements and $ncARG_{v+}[r_F]$ global graph features	1564	534	35.71%, 84.38%, 33.20%, 34.67%, 91.15%, 83.33%, 46.31%

As observed in Table 7.2, the highest percentage ($\geq 50\%$) of correlated features have been removed from RGB histograms, Gabor filter-bank features and other color-based measurements. Therefore, there are a few discriminative but correlated features (*e.g.* RGB histograms) that are excluded after feature selection, which is the probable explanation of hybrid approaches performing similar to, or sometimes slightly worse than the best performing individual feature extraction methods, as observed later in the experimental results. Furthermore, since it was noted in the studies in [Sharma 2016], the performance difference between hybrid features without feature selection, and after correlation analysis based feature selection is very low and insignificant ($\approx 2\%$), hence, only the subset of hybrid features selected after correlation analysis have been considered in this study to lower the computational requirements.

Dedicated feature selection methods, also called *embedded methods*, use a specific learning algorithm to perform feature selection and classification simultaneously, *e.g.* simulated annealing and genetic algorithm [Hamon 2013]. The author wants to specify that such complex methods are not considered in the experiments, mainly due to the reason that features should not be specifically selected for a particular classification algorithm as multiple traditional machine learning methods are simultaneously evaluated. Hence, an independent method is required, and correlation-based feature selection is a generic approach that simply removes the highly correlated features without giving preference to any classification algorithm. Additionally, due to lower computational requirements, it economically integrates with all the other steps of the process pipeline.

7.4 Machine Learning Methods

7.4.1 Traditional Methods

After the elaborate procedure of representing histopathological images using a wide range of low-level and high-level descriptors, the next logical step is to select suitable machine learning methods to facilitate automatic classification of the H&E stained tissue image tiles into appropriate category, based on the malignancy levels which are defined by the corresponding HER2 immunohistochemical stain. The selected traditional machine learning methods include support vector machines, AdaBoost ensemble learning and random forests for the classification experiments, due to their wide acceptability and success in various fields of pattern recognition, specifically in the domain of image analysis in digital histopathology. Their individual benefits and limitations, leading to their choice for subsequent research are provided in Section 2.4.

7.4.1.1 Support Vector Machines

The first widely known traditional machine learning method applied for cancer classification after feature extraction is support vector machines. The algorithmic details of the support vector machines learning and classification are provided in Section 6.2.4, and subsequent extension to multi-class classification in Section 6.4.2.2. For classification of cancer regions based on

malignancy, a similar approach using support vector machines is deployed. Each image is represented by a set of features from one of the low-level or high-level graph-based feature groups, or a combination of features after feature selection. Gaussian or radial basis function kernels, showing a desirable classification performance in necrosis detection and cell nuclei classification are also employed for the analysis of cancer regions. Training and deployment of SVM is similar to the approach elaborated in Section 6.4.2.2, and again discriminative thresholds are avoided to prevent complexity due to multiple classes. The parameter selection procedure for SVM predicts the experimental parameters c and γ using grid search on the training samples with k -fold cross validation ($k = 5$). Due to 21 feature configurations used for SVM machine learning (Table 8.2), the parameter selection process is carried out separately for each configuration, yielding various pairs of c, γ . Optimal parameters are selected and applied for different cross validation strategies using the quantitative comparative evaluation scheme for a comprehensive analysis, discussed in Section 8.4.1.1.

7.4.1.2 AdaBoost Ensemble Learning

AdaBoost or adaptive boosting is another supervised machine learning method from the realm of conventional approaches where handcrafted feature extraction is a prerequisite before training the models for classification. Handcrafted low-level and high-level features, either self-designed or state-of-the-art, are used in conjunction with AdaBoost ensemble learners during the training phase. The trained classifiers predict the malignancy levels of unknown samples through a classifier combination and weighted majority voting scheme. Again, for the mathematical interpretation and parameter selection of AdaBoost ensemble learning algorithm, Section 6.4.2.2 can be referred. A set of values of K_{max} after parameter search for each classification configuration with different input feature groups is obtained. These are then applied in several experiments using AdaBoost ensemble learning and classification in the comparative evaluation scheme (Table 8.2).

7.4.1.3 Random Forests

As discussed earlier, random forests are a popular ensemble learning method which constructs many decision trees and the final class is majority voted by the individual trees. It involves the experimental parameters number of trees N_t and number of features for best split m . In order to approximate suitable values N_t and m at which the error stabilizes, the OOB error is computed for each N_t in a predefined range ($N_t \in [1, 1500]$), and each m , where m is considered as \sqrt{M} , $\log_2 M$ and M , M being the total number of features. Example of the estimated parameters has been discussed for cell nuclei classification in Section 6.4.2.2 and similar procedure has been applied for the analysis of cancer regions. Hence, for each configuration involving different kind of image descriptors of the histopathological datasets, the two parameters N_t and m are predicted using the bootstrapping procedure, and OOB error estimates are calculated from the resulting random forest.

7.4.2 Deep Convolutional Neural Networks

Deep learning using convolutional neural networks has gained a lot of positive attention from researchers due to its achievements in various image analysis tasks. In recent years, it has achieved breakthrough performance due to availability of large-scale training data and huge parallelization using GPUs. However, it is still in its early stages of development and needs to be extensively studied in order to be successful in digital histopathology. The author's research in deep learning [Sharma 2017b] is an attempt to contribute to this field by the experiments on gastric cancer WSI for two classification problems, namely, analysis of cancer regions and necrosis detection. This is achieved by developing a self-designed CNN architecture for gastric cancer image analysis, and comparing its performance with previously derived handcrafted features followed by traditional machine methods, AlexNet CNN framework which is successful in general object categorization and an ensemble of the two deep networks. The motivations, developed methodology and experimental setup are discussed in the following sections.

7.4.2.1 Motivation

There are several motivations to explore deep learning in this work. In the knowledge of the author, deep learning methods have not been studied until now for the specified classification problems, namely, the analysis of cancer regions based on immunohistochemistry and necrosis detection in H&E stained histopathological WSI of gastric cancer. One of the most significant promises of deep convolutional neural networks is to replace the requirement of handcrafted feature design and extraction with efficient task-specific learning algorithms. Such algorithms have even shown to outperform many hand-engineered features in several fields [Razavian 2014], including studies in histopathological image analysis. Until the emergence of deep learning methods, the most important and demanding component for image analysis using traditional methods in digital pathology has been the suitable design and selection of handcrafted features describing the complex tissue images, so the author is highly motivated to explore this potential of deep learning. Additionally, as training and classification based on deep learning includes direct processing of image regions, it eliminates the requirement of image segmentation. Thereby, the classification performance is not limited by the quality of results of the cell nuclei segmentation and the optional cell nuclei classification algorithms, an advantage compared to the high-level handcrafted features discussed in Section 7.3.3. Most of the previous deep learning methods in histopathology consider smaller image sizes for training the deep networks, however, in the described experiments, high resolution (512×512 pixels) provides a larger field of view that is preferable for analysis of histological images, as it helps acquiring context information such as neighborhood properties and tissue architecture at higher magnification ($40\times$), but not widely explored due to memory restrictions. The author has used the same CNN architecture for two distinct purposes, namely, cancer classification based on IHC and necrosis detection, in order to explore another quality promised by deep learning *i.e. generalizability*, where the same general purpose learning machine can be applied for a wide range of tasks.

7.4.2.2 AlexNet: A Well-known CNN Architecture for General Image Characterization

The AlexNet CNN framework [Krizhevsky 2012] is a widely known and accepted deep learning algorithm due to its tremendous accomplishment in general image categorization of ILSVRC challenge for the ImageNet database [Deng 2009]. This CNN framework has performed reasonably well for general images containing 1000 categories, such as Airplane, Bird, Cat, Dog and so on. The AlexNet CNN is a complex, deep and wide CNN architecture that contains several layers of computation. Because of the higher computational complexity and memory requirements of AlexNet, the authors of AlexNet have wisely implemented the architecture using two parallel GPUs. It is one of the earliest deep networks with favorable outcomes, so it is considered as baseline or state-of-the-art in the development of deep learning methods for image classification. Hence, for the analysis of malignancy levels according to immunohistochemical response, the proposed CNN method is comparatively evaluated with AlexNet CNN framework for the gastric cancer histological datasets. AlexNet CNN has also been used for quantitatively comparisons in the necrosis detection task of the image pre-analysis stage. The AlexNet CNN architecture implemented for cancer classification in gastric cancer WSI datasets is illustrated in Figure 7.10.

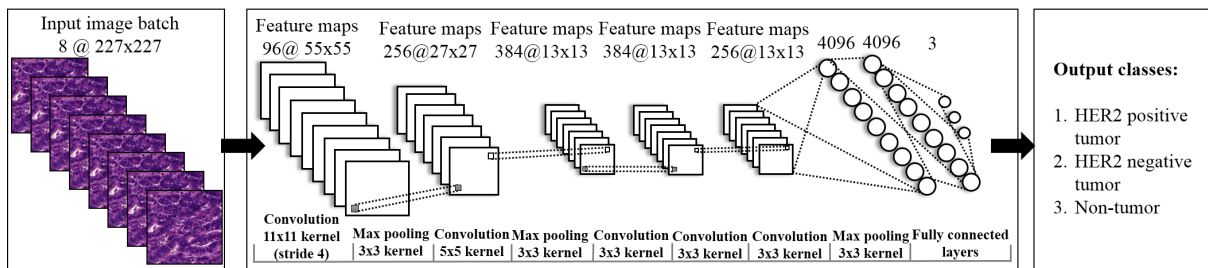


Fig. 7.10 AlexNet CNN architecture implemented for cancer classification in gastric cancer WSI datasets. Adapted from [Krizhevsky 2012]

Regarding the comparative evaluation of classification performance using the AlexNet CNN framework on gastric cancer datasets, the input histological images used for training and classification are cropped to size 227×227 from the center, similar to the original images in AlexNet framework. The learning properties are also aligned with the experiments in the original framework in [Krizhevsky 2012]. However, the number of training rounds and batch sizes are tuned according to the gastric cancer histological datasets, and only one GPU is used for implementation of both CNN architectures in this work.

7.4.2.3 Empirical Analysis of CNN Architectures

Different CNN architectures are empirically studied to observe the behavior of variation in model characteristics (*e.g.* network depth, layer properties, training parameters etc.) by training from scratch on a representative subset of the entire whole slide image data for cancer classification. For this purpose, a smaller dataset from five WSI is first created where image tiles are generated from three WSI for training the empirical CNN configurations, one WSI for validation phase and

one WSI for test phase. This empirical analysis acts as a *pilot study*, as it allows establishment of the feasibility of deep learning methods for the described image analysis problems, considering a high visual complexity of gastric cancer histopathological images. It further facilitates basic understanding of the behavior of deep convolutional neural network architectures and economical utilization of time and space by first reading only a part of the available WSI image data.

The details and results for the empirically evaluated CNN architectures are summarized in Table 7.3. The last shown CNN network is the AlexNet CNN introduced in [Krizhevsky 2012]. There exist possibilities to extend the number of CNN architectures for analysis by using more combinations of connected variables, however, performing an exhaustive evaluation is highly time consuming and restricted by the available hardware, so a few elementary CNN architectures are considered at this stage. A few variants of the reported architectures are also tested without any further improvement in accuracy. On observing the classification performance for both validation and test WSI, learning curves and overall learning behavior with respect to cancer classification, the CNN architecture with nine layers (three convolutional layers, three pooling layers and three fully connected layers) achieves most accurate results for all the three malignancy levels on the representative dataset, and will tentatively improve by using more abundant examples. Based on empirical analysis and insight to accurately model the characteristics of histological images, this self-designed CNN architecture with the best performance on the representative dataset is selected for further evaluation with entire available WSI data. Later, the proposed CNN is also applied for necrosis detection (details in Section 6.2.4.2).

It is observed that for gastric cancer images in general, convolution layers with larger kernels followed by smaller ones lead to a better training procedure with a lower training loss and higher accuracies. Similar relative convolutional kernel sizes are adapted by AlexNet and other popular CNN architectures. Additionally, an architecture with smaller convolutional kernels followed by larger ones was also tested, but it produced reasonable accuracies for only two classes and

Table 7.3 Details of most successful empirically evaluated CNN architectures for cancer classification on gastric cancer representative datasets

Model number	Total number of layers	Number of convolution layers	Kernel sizes: convolution layers	Number of feature maps	Kernel sizes: pooling layers	Number of fully connected layer outputs	Average multi-class accuracy (cancer classification)	
							Validation WSI	Test WSI
1	6	2	3, 3	16, 16	3, 3	128, 3	0.3235	0.3333
2	6	2	7, 3	16, 16	2, 2	256, 4	0.5208	0.5140
3	7	2	9, 9	16, 16	3, 3	128, 128, 3	0.6025	0.5417
4	7	2	7, 5	16, 16	2, 2	256, 128, 3	0.6000	0.5432
5	9	3	7, 5, 3	24, 16, 16	2, 2, 2	256, 128, 3	0.6056	0.5571
6	10	4	9, 7, 5, 3	32, 128, 128, 128	3, 3, 3	2048, 2048, 3	0.5951	0.4376
7	11	5	11, 5, 3, 3, 3	96, 256, 384, 384, 256	3, 3, 3	4096, 4096, 3	0.3948	0.3948

completely failed to classify the HER2 negative tumor class, indicating the importance of larger neighborhoods in histopathological images. Poor performance was also observed in another model with hinge loss function, whereas softmax loss is used in the architectures mentioned in Table 2.1, thus, found more suitable for learning. Reasons for testing moderately deep networks in our experiments are the limited availability of training data (especially for necrosis detection) and system memory. Higher network depths would cause a rise in the number of network parameters leading to non-optimal learning with smaller datasets (overfitting), and also result in higher space and time complexity. However, a more detailed study of the CNN architectural design to determine the relationship between design parameters is a future direction of our research, and can plausibly lead to an improvement of results obtained so far in these experiments.

Another option was to fine-tune a pre-trained CNN, but was intuitively skipped due to the following reasons. As the name suggests, fine-tuning of a pre-trained network is generally performed when the network trained successfully with visually very similar images. For example, AlexNet was originally tested on general object categorization with ILSVRC dataset in [Krizhevsky 2012], and was later fine-tuned for Flickr style images in [Xia 2015]. Such fine-tuning of AlexNet does not appear useful for the entire complex histological gastric cancer data. Furthermore, the classification results even after training AlexNet from scratch during empirical analysis on a smaller representative dataset are not sufficiently adequate. For the selected CNN architecture as well, it is considered more reliable to train from scratch with the entire WSI data. Therefore, for the purpose of completeness, the performance of AlexNet and proposed CNN architecture are comparatively evaluated using the same cross validation strategies for both approaches and training from scratch on the entire WSI data.

7.4.2.4 Proposed Convolutional Neural Network Architecture

The self-designed CNN architecture with the best empirical performance for cancer classification is a purely supervised feed forward network shown in Figure 7.11. This preliminary convolutional neural network consists of nine stages, where the first six stages comprise of convolution, non-linearities and pooling operations, and the last three stages are fully connected layers such as in regular neural networks. A batch size of 8 is used during training phase and 10 during deployment phase to improve efficiency using batch processing. Image mean values are calculated and subtracted from the images during both phases, such as in the AlexNet framework [Krizhevsky 2012]. The detailed descriptions of the various layers are as follows.

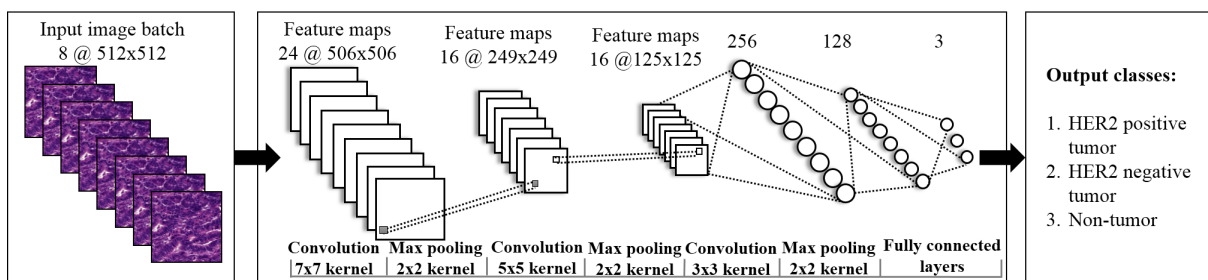


Fig. 7.11 Proposed CNN architecture for cancer classification in gastric cancer WSI

Convolutional Layer

In each convolutional layer, first the output of previous layers is convolved with multiple learned weight matrices called filter masks or learned kernels. Then the result is processed by a non-linear operation to generate the layer output. The linear operation in the k^{th} convolution layer, whose input is denoted by x^{k-1} (output of the $[k-1]^{th}$ layer) comprises of a two-dimensional convolution [Podlozhnyuk 2007] as shown in equation 7.20.

$$o^k[m, n] = x^{k-1}[m, n] * W^k[m, n] + b_k = \sum_{u=-\infty}^{\infty} \sum_{v=-\infty}^{\infty} x^{k-1}[u, v] W^k[m-u, n-v] + b_k \quad (7.20)$$

where, W^k denotes the learned weight matrix and b_k denotes bias or offset of the k^{th} convolution layer. For our experiments, three convolutional layers are used with filter sizes 7×7 , 5×5 and 3×3 and 24, 16 and 16 feature maps respectively. The size of learned filters should be similar to the size of patterns to be detected. This decision is task-dependent and in the described classification problems, subtle details between the cancer regions need to be resolved in contrast to large objects in general image analysis, which explains the choice of filter sizes. These sizes are also restricted by the GPU configuration as convolution is a computationally intensive operation efficiently executed on the GPU. Further, the weight initialization is performed using Gaussian functions with standard deviation 0.01, and bias is kept constant at 0.1.

The outputs o^k are applied to an operation which improves the learning process by increasing non-linear properties of the network, and rectified linear units (ReLU) are selected due to benefits explained in [Krizhevsky 2012]. These are non-saturating activation functions where piece-wise linear tiling can be achieved. The basic ReLU operation is given by equation 7.21. The cascade of operations in the convolutional layer are depicted in Figure 7.12.

$$x^k = \max(0, o^k) \quad (7.21)$$

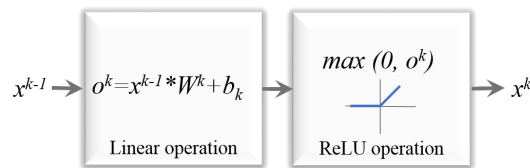


Fig. 7.12 Cascade operations in the k^{th} convolutional layer

Pooling Layer

Subsampling (or down-sampling) refers to reducing the overall size of a signal. In the domain of two-dimensional filter outputs, it can additionally increase the position invariance of filters. For this purpose, a pooling layer performing *max-pooling* [Krizhevsky 2012] is applied after each convolutional layer. Max-pooling involves splitting the filter output matrix into non-overlapping grids and taking the maximum value in each grid as the value in the reduced matrix. Hence, it combines responses at different locations and adds robustness to small spatial variations, thereby

increasing translational invariance, along with reducing spatial resolution. We use a kernel size 2×2 and stride two for maxpooling operation after each convolutional layer.

Fully-connected Layer

After the convolutional and pooling layers, the fully-connected layers implement high-level reasoning in the convolutional neural network. In the fully-connected layer, neurons are connected to the neurons in previous and next layers, such as in conventional neural networks [Kröse 1993]. In the proposed CNN, three fully connected layers are used with 256, 128 and three (or two) outputs respectively, as the number of outputs in last layer depends on the classification problem at hand. Bias is set to 0.01.

Dropout method [Srivastava 2014] assists in reducing overfitting, especially when the available training data is limited such as the WSI data. During each iteration, individual nodes along with incoming and outgoing edges are removed from the network, and are later returned along with their initial weights. In our approach, after each of the first two fully-connected layers, the dropout ratio *i.e.* the probability of dropping any input for both stages is set to 0.25.

Learning Properties

Learning of the convolutional neural network is based on measuring a loss function (also called objective function, error function, cost function) that indicates the error of learned network parameters. The learning objective is to compute the parameters to minimize the loss function. Softmax function [Bishop 2006] (equation 7.22) is the probability of class c_i given input X , where z_i represents the score for i^{th} class among total C classes. The softmax loss E is calculated as negative log likelihood of the softmax function (equation 7.23), where N denotes the length of the class vector.

$$P(c_i|X) = \frac{e^{z_i}}{\sum_{j=1}^C e^{z_j}} \quad (7.22)$$

$$E = -\frac{1}{N} \sum_{n=1}^N \log(P(c_n|X)) \quad (7.23)$$

The method used for optimizing the loss minimization is called *Stochastic Gradient Descent with Momentum* [Sutskever 2013], [Bottou 2012]. For the t^{th} iteration, the update process is denoted by equation 7.24 [Jia 2014], [Bottou 2012], where θ_t denotes the current weight update, θ_{t-1} is the previous weight update, w represents the weights, E is the average loss over the dataset, $\nabla E(w)$ is negative gradient, η is the learning rate and $\mu \in [0, 1]$ represents momentum used for speeding up the convergence of gradient and preventing oscillations.

$$\begin{aligned} \theta_t &= \mu\theta_{t-1} - \eta\nabla E(w_{t-1}) \\ w_t &= w_{t-1} + \theta_t \end{aligned} \quad (7.24)$$

In the learning phase, the two hyperparameters η and μ are required to be experimentally determined. After the empirical analysis stage, η has been set to 0.001 initially, and is decreased

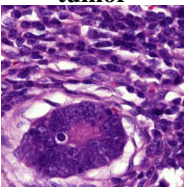
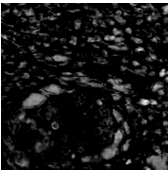
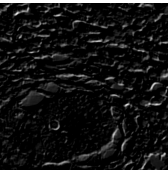
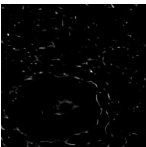



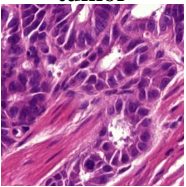
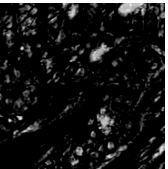
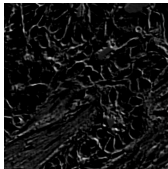

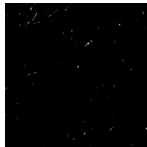
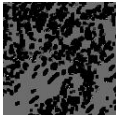

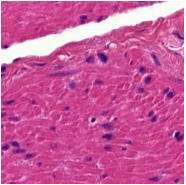
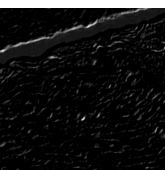
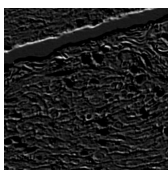
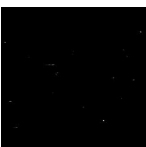
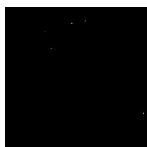
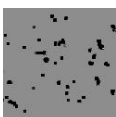

by a factor of 0.1 after every 20,000 iterations, and μ is fixed at 0.9. Around eight training epochs are performed during each training phase after initial experimental analysis of training curves on our datasets.

Visualizing Outputs of CNN Layers

The outputs of the each layer of the CNN can be visualized in order to determine the image characteristics highlighted and learned by the layers. The two-dimensional outputs of the three convolutional layers of the proposed CNN architecture are visually analyzed by forward passing example image tiles of each malignancy level through a pre-trained CNN model, and viewing the outputs of the filters. Similarly, the output of pooling layers can also be visualized. Examples of outputs of the constituent filters in three convolutional layers of the proposed CNN architecture are shown in Table 7.4.

In Table 7.4, the corresponding raw image features captured by each convolutional layer are clearly evident through the example image tiles. Most significantly, it can be observed that the knowledge from the first convolutional layer provides an overall view of the tissue image, including inter-cellular texture along with spatial distributions and architectural properties of tissue components. The filters in the second convolutional layer strongly respond to the darker objects in the image, thereby, highlighting the deeply stained haematoxylin cells which are found more in the tumor classes compared to non-tumor. The third layer generates negative responses to the cell nuclei (haematoxylin) and positive to cytoplasm (eosin), that is similar

Table 7.4 Examples of two-dimensional outputs in the three convolutional layers of the proposed CNN architecture

Tumor type	Convolution layer 1		Convolution layer 2		Convolution layer 3	
HER2 positive tumor 						
HER2 negative tumor 						
Non-tumor 						

to a cell nuclei segmentation result, and helps the deep network to locate the cell nuclei and learn their morphological properties in the tissue image. This visualization indicates the main advantage of deep learning, that the salient characteristics in H&E stained gastric cancer images can be directly learned by deep networks using end-to-end training without the requirement of handcrafting adequate features before the training process.

7.4.2.5 Summary of Experimental CNN Parameters

It can be observed from the description of convolutional neural networks, that the design and development of a suitable convolutional neural network architecture is a complicated process, and requires the selection of a large number of experimental parameters for an optimal performance. This section summarizes the parameter settings for the proposed CNN architecture, selected for classification of gastric cancer histological datasets. Table 7.4.2.5 shows the names, descriptions, allowed values, and selected values for each experimental parameter. The architecture design and parameter definitions refer to the Caffe deep learning framework [Jia 2014].

Table 7.5 Summary of parameter selection for the proposed CNN architecture

Type	Name of parameter	Definition	Allowable values	Selection for proposed CNN
Input parameters and transformations	Batch size (training)	Number of training images to process at one time	Depends on GPU memory	8
	Batch size (testing)	Number of test images to process at one time	Depends on GPU memory	10
	Crop size	Size of the image to consider for cropping from the centre	Positive integer	512
	Scale	A number for mapping of intensities in predefined range	Real number	1
	Type of database	Way in which data is entered into the network	LEVELDB, LMDB, raw images, directly from memory or HDF5 from files on disk	LMDB
Convolutional layer parameters	Number of outputs	Number of convolutional filters	Depends on network architecture	24,16,16
	Kernel size	Specifies height and width of each filter	Depends on network architecture	7,5,3
	Weight filler type	Type of weight initialization used	Constant, Gaussian, Positive Unitball, Uniform, Xavier, Bilinear, MSRA	Gaussian
	Weight filler value	Initialization value	Real number	0.01
	Bias term	Specifies whether to learn and apply a set of additive biases to the filter outputs	True, false	TRUE

Convolutional layer parameters	Bias type	Type of bias used	Constant, Gaussian, Positive Unitball, Uniform, Xavier, Bilinear, MSRA	Constant
	Bias value	Value of bias used	Real number	0.1
	Pad	Number of pixels to (implicitly) add to each side of the input	Positive integer	0
	Stride	Intervals at which to apply the filters to the input	Positive integer	1
	Group	The input and output channels are separated into g groups, such that i^{th} output group channels will be only connected to the i^{th} input group channels.	Positive integer	1
Pooling layer parameters	Pooling method	Method for pooling	Maximum, Average, or Stochastic	Maximum
	Pad	Number of pixels to (implicitly) add to each side of the input	Positive integer	0
	Stride	Intervals at which to apply the filters to the input	Positive integer	2
Activation layer parameters	Type of activation	Helps in selecting from one of the activation functions	ReLU, Sigmoid, TahH, Absolute value, Power, Binomial normal log likelihood	ReLU
Fully connected layer parameters	Number of outputs	Number of filters	Positive integer	256,128,3 (or 2)
	Weight filler type	The type of weight initialization used	Constant, Gaussian, Positive Unitball, Uniform, Xavier, Bilinear, MSRA	Gaussian
	Weight filler value	Initialization value	Real number	0.005,0.005, 0.01
	Bias term	specifies whether to learn and apply a set of additive biases to the filter outputs	True, false	TRUE
	Bias type	Type of bias used	Constant, Gaussian, Positive Unitball, Uniform, Xavier, Bilinear, MSRA	Constant
	Bias value	Value of bias	Real number	0.1
Dropout parameters	Dropout ratio	Ratio of samples to be left out	[0,1]	0.25

Loss layer parameters	Type of loss	Way to compute the loss function	Softmax, Euclidean, Hinge, Information Gain, Sigmoid Cross Entropy	Softmax
Solver parameters	Solver type	Type of back propagation algorithm used to train the network	Stochastic Gradient Descent, AdaDelta, Adaptive Gradient, Adam, RMSprop, Nesterov's Accelerated Gradient	Stochastic Gradient Descent
	Test iterations	Number of test iterations should occur per test interval	Positive integer	168,000 (or 2,000)
	Test interval	Number of times the test phase of the network will be executed	Positive integer	20,000 (or 10,000)
	Base learning rate	Initial learning rate of the network	Real number	0.001
	Learning rate policy	Way in which learning rate should change over time	Step, Multistep, Fixed, Exponential, Polynomial, Sigmoid	Step
	Gamma	Value by which the learning rate should change every time	Real number	0.1
	Step size	Number of iterations between each step	Positive integer	20,000
	Maximum iterations	Number of iterations when the network should stop training	Positive integer	210,000 (or 60,000)
	Momentum	Momentum for gradient descent	[0,1]	0.9
	Weight decay	Factor of (regularization) penalization of large weight	[0,1]	0.001
	Solver mode	Whether training will be performed on CPU or GPU	CPU,GPU	GPU

7.4.2.6 Ensemble of Convolutional Neural Networks

Ensemble learning is a technique where a collection of classifiers or learners can be used to obtain a joint classification decision by combining their predictions. For instance, among the traditional machine learning methods, AdaBoost and random forests have been applied as ensembles of individual classifiers. Similarly, an ensemble of convolutional neural networks [Chen 2016b] is a scheme where the outputs of independently-trained CNNs can be combined to obtain the final prediction. Consider an ensemble E of M networks $\{n_1, n_2, \dots, n_M\} \in E$, where $P_j(c_i|X)$ represents the softmax probability for an instance of class c_i given input X for network n_j .

Hence, the final prediction probability P_E of the ensemble E for that instance will be given by,

$$P_E(c_i|X) = \frac{1}{M} \sum_{j=1}^M P_j(c_i|X) \quad (7.25)$$

An ensemble of CNNs can allow a combination of diverse model properties, thereby, reducing anomalies related to overfitting of the training data. An ensemble of CNNs can also promote a better generalization of classifiers, thereby, leading to a higher accuracy of prediction than the individual CNN classifiers under investigation. Usually, ensembles of multiple trained models of the same CNN architecture are studied in previous literature. However, in this study, after observing the results of the two different CNN architectures, namely, AlexNet framework and the proposed CNN architecture, it will be interesting to consider an ensemble of the two convolutional neural networks as a part of the described experiments.

7.4.3 Classification Strategies

In order to attain optimum classification performance in analysis of cancer regions, two classification strategies are explored for the traditional machine learning methods, *i.e.* *single-stage* classification and *hierarchical* classification. The first strategy performs a classification of tissue regions directly into one of the three malignancy classes. In the second strategy, a two-stage classification is performed. The first stage classifies the regions into two broad classes namely tumor containing HER2 positive tumor and HER2 negative tumor, and non-tumor. In the second stage, regions classified as tumor class are further classified into one of the respective sub-classes. The two classification strategies are shown in Figure 7.13. These have been experimented using the low-level state-of-the-art features, high-level state-of-the-art features and high-level handcrafted features with traditional machine learning methods. Due to higher time complexity involved in training from scratch for hierarchical classification, only the first strategy *i.e.* single-stage classification is applied to the deep learning experiments using convolutional neural networks.

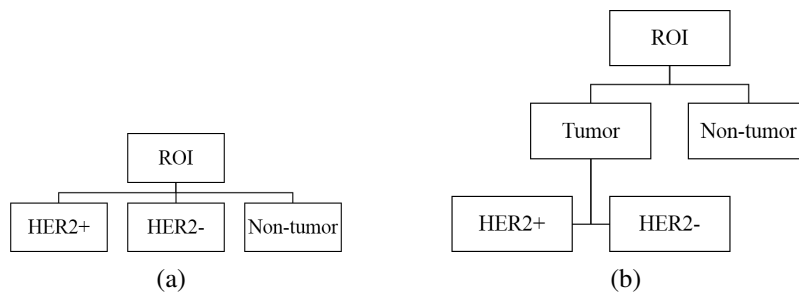


Fig. 7.13 Classification strategies for cancer classification (a) Single-stage classification (b) Hierarchical classification.

7.5 Applications

7.5.1 Computer-aided Diagnosis: Cancer Classification based on Immunohistochemistry

It is already known that H&E staining is a routinely applied technique, but immunohistochemical staining is a specialized procedure in histopathology. So immunohistochemical stains generally involve a higher expenditure and are less common than H&E stain. Especially the HER2 stain has been recently introduced as a biomarker for gastric cancer [Behrens 2015]. HER2 and H&E stained WSI are available but H&E stain is analyzed due to ease of use and lower preparation costs. Pathologists are able to mostly discriminate between malignancy levels by visually observing HER2 stained slides using optical or virtual microscopy methods. However, while observing H&E stained slides, they require a greater time and effort to identify the corresponding tumor areas. Moreover, even in the more distinct HER2 stain, inter-and intra-observer variability were noted for the observations made by ten pathologists in [Behrens 2015] using the same WSI data, prior to being provided for this research. Therefore, a computerized application performing such cancer discrimination can act as the potential solution to these difficulties, more so in the H&E stained tissue regions, and can assist pathologists in computer aided diagnosis of gastric cancer.

The best overall performance for the stage involving analysis of cancer regions is observed with combination of high-level handcrafted and low-level state-of-the-art features. This configuration along with the training samples can be utilized in a computerized application for automatic classification of unknown regions of interest in the H&E stained gastric cancer WSI into one of the three malignancy levels based on their HER2 immunohistochemistry. The information obtained after automatic analysis of gastric cancer can assist pathologists in diagnostic and prognostic tasks. Although the method does not assure perfect accuracy, it can definitely reduce the efforts in specialized staining and manual analysis, as well a inter-and intra-observer variability. As stated before, a limitation of the described methods is the availability of few labeled WSI data that can be overcome by creating annotations on multiple WSI by medical experts covering large-scale biological variations in the population. However, the results are particularly promising when partially labeled data in large-sized single or multiple WSI of the same patient is available for training purpose.

Deep learning methods can also be applied for the described CAD application, due to their reasonable performance compared to traditional machine learning methods. A pre-trained CNN can be deployed to determine the cancer-based categories of unlabeled constituent regions in the WSI. This has been demonstrated by generating a probability map for each category of cancer classification at the WSI-level as follows. During classification of an unknown image tile, the prediction probability estimate of each class can be expressed using a color scale from blue to red representing values in range $[0,1]$. The probability representations of constituent non-overlapping image tiles overlay the original H&E WSI to generate a probability map for

each tumor class. For visualization purpose, the probability map depicts which tissue areas in the WSI have a higher probability of the existence of a particular malignancy level according to its immunohistochemistry, showing the more and the less likely regions belonging to a particular class. An example for cancer classification of one of the H&E WSI is shown in Figure 7.14. Figure 7.14(a) depicts the original WSI with pathologists' annotations showing HER2 positive areas with red and HER2 negative areas with yellow polygons. Figures 7.14(b), 7.14(c) and 7.14(d) are the probability maps for HER2 positive, HER2 negative and non-tumor using the proposed CNN architecture. Because of very high resolution of the WSI, the constituent image tiles appear as small colored pixels at a lower magnification ($0.3\times$). The results obtained are mostly consistent with expert annotations made in the corresponding HER2 WSI, indicating desirable behavior of the proposed CNN method for analysis of H&E stained gastric cancer WSI based on their HER2 immunohistochemistry. Higher confusion and prediction error is observed mainly in the lower south-eastern part of the WSI, especially in the tumor classes. Around 75% of the WSI is correctly predicted, which is consistent with the experimental results.

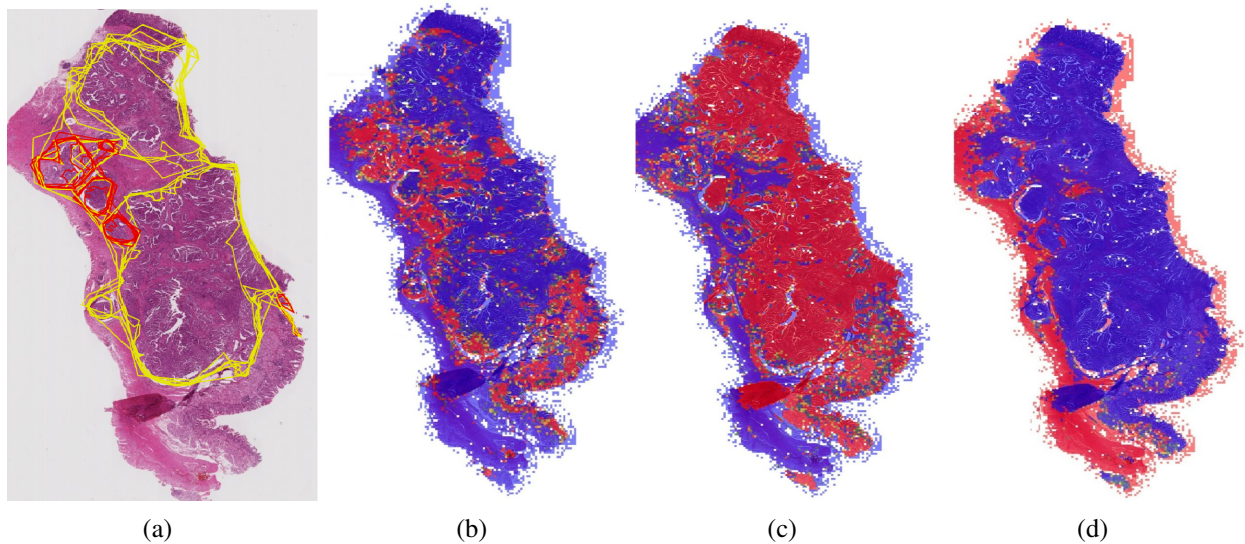


Fig. 7.14 *Illustrative examples of classification result (a) Original H&E WSI with pathologists' annotations for cancer classification based on IHC, and corresponding probability maps using proposed CNN architecture for (a) HER2 positive tumor (b) HER2 negative tumor (c) non-tumor at a low magnification ($0.3\times$).*

7.5.2 Content-based Image Retrieval

Content-based Image Retrieval (CBIR) consists of a group of methods for retrieving relevant digital images by their visual content. The horizon of CBIR systems includes all methods ranging from image similarity functions to highly complex image annotation systems [Datta 2008]. A general CBIR algorithm is composed of two main steps. Firstly, a signature calculation is performed to mathematically describe images based on the characteristics of their visual content,

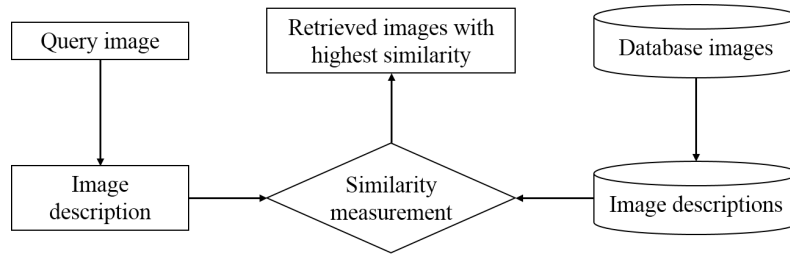


Fig. 7.15 Schematic diagram of a CBIR system

and includes intensity, color, texture, shape, size, location, or their mixtures. The second step assesses the similarity between the query and database images, and generates a similarity score by comparing image descriptors of the query image and those of each of the database images. Finally, the database images with highest similarity score are returned as retrieval results. A typical CBIR system is demonstrated in Figure 7.15.

From the perspective of histological image analysis, manual selection of regions of interest similar to a query region, in collections of large-sized WSI is an uphill task that can be automated using an appropriate CBIR system. Such a system facilitates automatic retrieval of groups of image regions sharing common visual and diagnostic characteristics to a query ROI. It can potentially contribute towards diagnosis, prognosis, teaching and research in the field of histopathology. It can be applied in clinical environments to assist pathologists for inspection and diagnosis of typical regions in a given WSI data, additionally serving as a basis for selecting the relevant treatment method. It can also assist in disease prognostics, allowing pathologists to predict which patients may be susceptible to a specific disease, and also predicting disease outcome, recurrence and survival. Moreover, the system can be useful as a training tool for medical students, researchers and instructors to search large educational repositories of disease-specific histological WSI image data for interesting ROIs. It may help them to visually inspect desirable regions and compare their interpretations with the automatic classification. CBIR system will not only allow the retrieval of cases with patients having similar diagnosis but also cases with visual similarity but different diagnoses, which can be subject to further research. Some examples of previous literature for CBIR in digital pathology is given in [Yang 2007], [Sharma 2012] and [Sridhar 2015].

Specifically in our experiments involving the analysis of cancer regions in H&E stained tissue using HER2 immunohistochemistry, traditional methods can be utilized by extracting suitable handcrafted features to find image regions from the WSI dataset similar to the query region, based on its underlying malignancy level. A simple prototype CBIR application has been demonstrated based on malignancy levels in the gastric cancer WSI datasets, using the features extracted prior to traditional machine learning, namely, RGB histograms as low-level state-of-the-art features, Voronoi diagrams and Delaunay triangulation as high-level state-of-the-art features, $gARG[r_F]$ as high-level handcrafted features and selected subset of the combined features of RGB histogram and $gARG[r_F]$ graphs. The idea involves a similarity calculation of respective feature vectors between query image and all the database images. Different distance metrics are evaluated for

this purpose, namely, Euclidean distance, Manhattan distance, Chebyshev distance, Bray-Curtis distance, Cosine distance and Correlation distance [McCune 2002] and resulting performance measures are relatively observed. Due to the most superior behavior with the Manhattan distance metric, the experimental results depicted in this section are only based on this metric.

The image datasets consisting of non-overlapping image tiles size 512×512 pixels (as described in Section 7.2.1) are used for demonstrating this application. A set of query images are selected randomly from the dataset with 15 images per slide and 55 images per malignancy level, making a total of 165 query images. The remaining 2,988 image are used as the database. Each query image is sequentially fed into the application and similar images are retrieved from the database along with a matching score, which is used to sort the retrieved images in decreasing order of their detected similarity to the input query image. The matching score $s_{q,d}$ between a query image q and a database image d is computed as the reciprocal of the Manhattan distance between their feature vectors \mathbf{f}_a and \mathbf{f}_b each of length K , given by,

$$s_{q,d} = \left(\sum_{k=1}^K |f_{q,k} - f_{d,k}| \right)^{-1} \quad (7.26)$$

For evaluating any CBIR system, a set of performance measures are commonly applied [Makhoul 1999]. The widely used performance measures include precision, recall and F-score. Precision is the fraction of retrieved images that are also relevant, recall is the fraction of the relevant images that are successfully retrieved and F-score is the harmonic mean of precision and recall. Below are the mathematical definitions of precision P , recall R and F-score F , where N_{rr} is the number of relevant retrieved images, N_t is the total number of retrieved images and N_r is the total number of relevant images.

$$P = \frac{N_{rr}}{N_t} \quad (7.27)$$

$$R = \frac{N_{rr}}{N_r} \quad (7.28)$$

$$F = \frac{2PR}{P + R} \quad (7.29)$$

For the described CBIR system, the image is considered relevant if it belongs to a given reference category *i.e.* malignancy level based on the HER2 immunohistochemical response. The numerical results are generated by averaging the calculated performance measures over all the query images. For evaluation purpose, this procedure is repeated for multiple rounds and the overall results are recorded. Illustrative experimental results of the prototype application are depicted as precision-recall (P-R) curves in Figure 7.16 and values at $N_t = 10, 50, 100$ in Table 7.6, for the computed features and Manhattan distance metric.

The experimental results show a reasonable shape of overall P-R curves for most feature extraction methods. As observed for analysis of cancer regions, higher precision is achieved for non-tumor and tumors, but lower among the two tumor subclasses. Familiar trends are seen for the three malignancy levels, where non-tumor queries receive best matches from the WSI database,

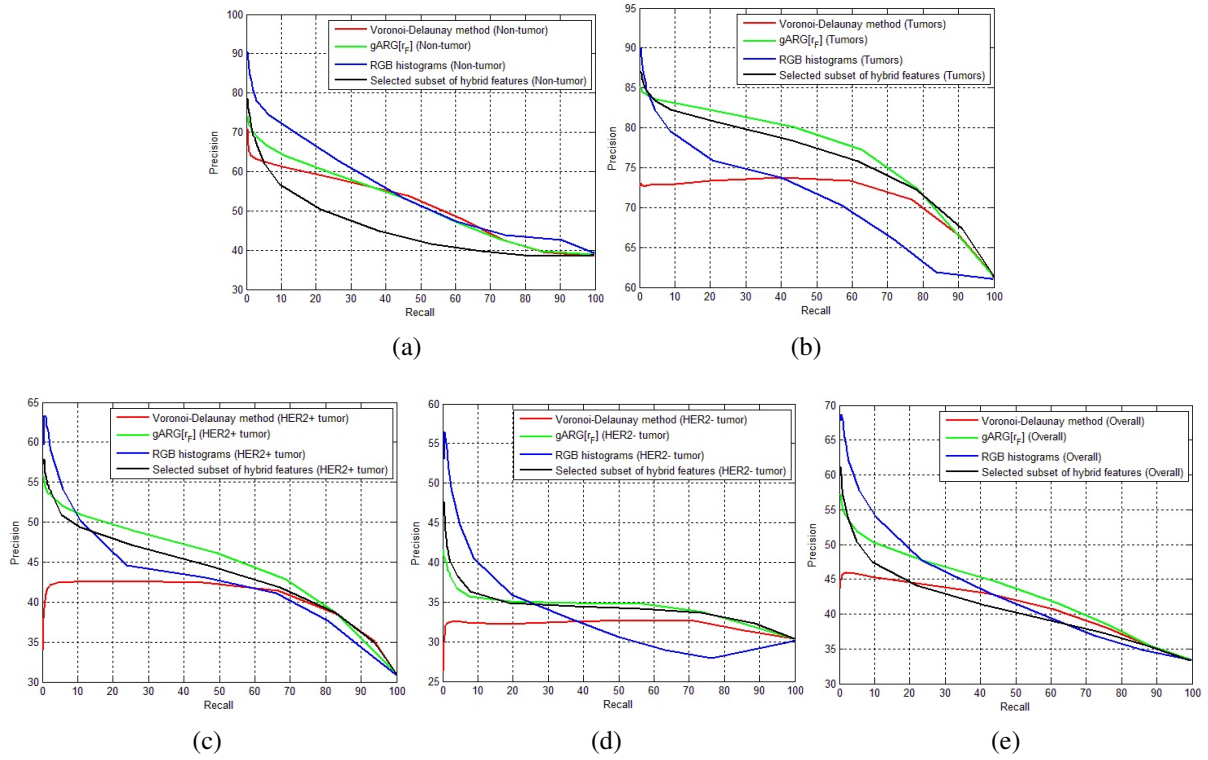


Fig. 7.16 Illustrative experimental results of prototype CBIR application as P-R curves for categories (a) Non-tumor (b) Tumors (c) HER2 positive tumor (d) HER2 negative tumor and (e) overall result using the Manhattan distance metric.

Table 7.6 Per-class experimental results showing performance measures for the prototype CBIR application

Performance measure	Method	Non-tumor	Tumors	HER2+ tumor	HER2-tumor
Precision (10)	Voronoi-Delaunay method	66.70	73.08	37.89	29.87
	$gARG[r_F]$ method	72.64	84.97	55.13	40.85
	RGB histograms	86.16	88.95	63.31	56.44
	Selected subset of hybrid features	75.88	86.85	57.84	45.89
Precision(50)	Voronoi-Delaunay method	63.33	72.71	42.08	32.42
	$gARG[r_F]$ method	69.00	84.07	53.16	38.26
	RGB histograms	77.52	84.56	57.86	48.89
	Selected subset of hybrid features	67.06	84.46	53.35	40.12
Precision(100)	Voronoi-Delaunay method	62.46	72.93	42.39	32.58
	$gARG[r_F]$ method	66.82	83.63	52.10	36.68
	RGB histograms	74.43	82.16	54.14	44.73
	Selected subset of hybrid features	62.11	83.35	50.98	38.23
F-Score($P \leq 0.5$)	Voronoi-Delaunay method	54.03	76.09	44.96	35.23
	$gARG[r_F]$ method	52.42	75.93	56.15	47.20
	RGB histograms	53.01	75.76	54.56	49.24
	Selected subset of hybrid features	36.82	75.93	42.33	35.80

followed by HER2 positive tumor and then HER2 negative tumor. As expected, the Voronoi-Delaunay method is not as effective as the other feature groups in all the cases. Especially for the tumors, the $gARG[r_F]$ handcrafted features are the most effective representation compared to the other feature types showing their unique ability to describe tumors. In the overall result, RGB histograms outperform $gARG[r_F]$ at lower recalls, whereas $gARG[r_F]$ graph-based method outperforms RGB histograms at higher recalls. The selected subset of hybrid features has an intermediate precision at lower recalls, but degrades more steeply than the other two methods, especially because of its unsatisfactory performance for the non-tumor class. Furthermore, the F-scores of the RGB histograms and $gARG[r_F]$ graph-based individual methods are also higher as compared to Voronoi-Delaunay method and hybrid feature subset, proving superiority of the former two feature extraction techniques for the CBIR application.

An interesting observation is lower precision of the CBIR application compared to the supervised traditional machine learning methods using identical feature extraction approaches, which can be attributed to two main reasons. Firstly smaller image size (512×512 pixels) is considered for a larger database, that was also observed to perform less favorably using traditional machine learning methods due to restricted field of view and neighborhood information. Secondly, no intelligent machine learning algorithm is applied and simply distance metrics are calculated in an unsupervised clustering-like process. Nevertheless, the results highlight the capabilities of the state-of-the-art and handcrafted image description methods, highlighting their discriminative power for retrieving images appearing similar to a query image of particular malignancy levels in the WSI database, at smaller image sizes and without any supervised learning.

7.6 Summary

In this chapter, the third stage of the proposed framework is described, which involves the image analysis of gastric cancer in H&E stained WSI, based on the HER2 immunohistochemical response of the tissue. In this stage, firstly, appropriate histological datasets are created for studying a vast range of analysis methods in digital histopathology. Following the dataset generation, two main routes of image analysis are investigated, namely, the *traditional route* with methods for image description and representation followed by traditional machine learning methods, and the *deep learning route* including the study of convolutional neural networks. Among these, novel methodologies are proposed by the author, most significantly, a cell nuclei ARG graph-based method along with eight variants and a self-deigned CNN architecture after extensive empirical analysis. Furthermore, the well known state-of-the-art methods for histopathological image analysis are quantitatively compared with the proposed methods using an elaborate scheme of comparative evaluation. Finally, this stage contributes towards two practical applications, namely, CAD using automatic cancer classification and CBIR in gastric cancer WSI datasets.

Performance Evaluation

Contents

8.1	Introduction	143
8.1.1	Evaluation Methods	144
8.1.2	Performance Metrics	146
8.2	Comparative Evaluation: Necrosis Detection	149
8.3	Comparative Evaluation: Cell Nuclei Classification	151
8.4	Comparative Evaluation: Analysis of Cancer Regions	153
8.4.1	Concept of Comparative Evaluation	153
8.4.2	Low-level State-of-the-art Features	157
8.4.3	High-level State-of-the-art Features	163
8.4.4	High-level Handcrafted Features	164
8.4.5	Combination of Features	173
8.4.6	Deep Learning Methods	175
8.4.7	Overall Observations	177
8.5	Summary	182

8.1 Introduction

Performance evaluation of developed methods is a significant process in any scientific environment. Especially, in digital histopathology, quantitative evaluation of performance of the proposed image analysis methods can facilitate observation, verification and validation to establish their feasibility, efficiency, reproducibility and practical usability in routine and (or) research settings.

A challenge for the performance evaluation of image analysis tasks, is the existence of biological variances among patients in a given population, as every patient has individual

characteristics in their cellular composition which are reflected in the tissue images and are impossible to avoid. So the image datasets should ideally include diverse patients incorporating a broad range of biological variations in any given population. The problem of limited availability of labeled WSI data may be a bottleneck for comprehensive performance evaluation. As stated earlier, the datasets used in the described experiments contains eleven whole slide images of gastric cancer, each from a different patient. As a result, these datasets are heterogeneous in nature, so in order to evaluate the robustness of the proposed system, special cross validation methods have been employed. Heterogeneity also prevails in terms of H&E stain intensity and malignancy levels in the WSI, further testing the reproducibility of the developed methods.

Quantitative evaluation is performed at every stage of the proposed framework, in order to observe the performance of the various feature extraction, automatic detection and classification algorithms in the experimental pipeline. Comparative analysis of the novel approaches provides a relative measure of their performance with respect to the existing and widely known state-of-the-art methods. The chapter first describes the performance evaluation methods and related performance metrics. Then it provides the details of comparative analysis in the various stages of the research framework and also discusses the corresponding observations, causes and implications. The results are further elaborated and can be referred from Appendix B.

8.1.1 Evaluation Methods

The most widely used methods for evaluation of the predictive performance of automatic detection and classification approaches belong to the category of *Cross Validation*. Cross validation is a model validation technique for assessing how the results of a statistical analysis will generalize to an independent data set [Fukunaga 2013]. It is mainly used to estimate how accurately a predictive model will perform in practice. The main goal of cross validation is to evaluate the predictive performance of a classification algorithm while minimizing overfitting by not allowing overlap between training and testing dataset.

Each round of cross validation involves partitioning a dataset into complementary subsets, performing the learning on one subset, called the training set, and validating on the other subset called the validation set or test set. To reduce variability, multiple rounds of cross validation are performed using different partitions and results are averaged over the rounds. Cross validation is a powerful technique recognized as a fair way to properly estimate model prediction performance [Seni 2010]. Three types of cross validation strategies are used to quantitatively evaluate the various parts of the proposed framework.

8.1.1.1 k -fold Stratified Shuffled Split Cross Validation

A k -fold cross validation divides all the samples in k groups of samples, called folds, of nearly equal sizes. The prediction function is learned using $k - 1$ folds, and the fold left out is used for test. Hence, each model is trained using $(k - 1)n/k$ samples and tested using n/k samples. The k -fold stratified shuffled split cross validation is a variation of k -fold cross validation that

performs training and testing of k stratified folds, *i.e.* the folds are made by preserving the percentage of samples for each class, hence, called ‘stratified’. Samples are randomly shuffled before splitting, hence, the word ‘shuffled’ is used. Also, the test folds are non-overlapping with each other. Stratification has been previously established as a better approach than regular cross validation by [Kohavi 1995], hence, selected as an evaluation method.

This type of cross validation has been applied for evaluation of the automatic classification experiments in the second stage of image pre-analysis, namely, necrosis detection and cell nuclei classification, as well as all the classification tasks in the third stage *i.e.* analysis of cancer regions. In most of the described experiments, $k = 3$ or $k = 10$ is used. The choice of k is an important factor to observe the classification performance, as a lower value of k will have lower computational requirements and lower variance between rounds but may lead to higher bias in the trained models, and a higher value of k will have reverse implications, *i.e.* more computationally expensive due to more rounds, higher variance between rounds and lower bias in the trained models. However, in order to reduce variance and not increase bias, cross validations with a lower k can be repeated multiple times as it will lead to different random folds, and their observations can be averaged. Especially for datasets with several categories, using lower k with repeated runs has been proved as a better approach [Kohavi 1995]. Therefore, $k = 3$ has been selected for the multi-class cell nuclei classification with seven categories in the datasets during image pre-analysis stage, and averaged for many runs. Furthermore, for the necrosis detection, $k = 10$ is selected for traditional machine learning using SVM with discriminative thresholds due to smaller datasets and lower computational requirements, whereas $k = 3$ is used for deep learning methods as these are more expensive in terms of computational and memory requirements. Stratified 10-fold cross validations are recommended in general [Kohavi 1995], so $k = 10$ is also applied for the analysis of cancer in the third stage using traditional machine learning methods. For deep learning methods in this stage, $k = 3$ is retained for the same reasons as explained before. The final results for all the described methods are obtained by averaging over multiple experiments.

8.1.1.2 Leave-a-sample-out Cross Validation

Leave-a-sample-out, more popularly called Leave-One-Out (LOO) is a simple cross-validation method where each learning set is created by taking all the samples except one, the test set being the sample left out. Thus, for n total samples, n different training sets and n different test sets are obtained, and each model is trained on $n - 1$ samples leading to n cross validation rounds. Assuming $k < n$, LOO is more computationally expensive than k -fold cross validation. Also, in terms of accuracy, LOO often results in high variance as an estimator for the test error [Breiman 1992]. Hence, in the experiments, though LOO is used to additionally validate the performance of the classification methods in some settings, k -fold stratified shuffled split is preferred over LOO to comparatively analyze and select the most efficient strategy. The LOO cross validation has been applied for the cell nuclei classification experiments in the second stage, however, skipped for the image datasets of necrosis detection and cancer analysis due

to higher computational requirements, especially because deep learning is explored for these problems leading to a high learning phase complexity in addition to large-scale training datasets.

8.1.1.3 Leave-a-patient-out Cross Validation

Cross validations are also performed using the leave-a-patient-out strategy, where in each round, the samples belonging to one patient are excluded from the training process and classifiers are trained by using the samples of remaining patients. The excluded patient's samples are used for validation. This cross validation strategy is used in order to study the robustness and practical usability of the described methods when samples from unknown patients need to be predicted. Intuitively, the resulting prediction rates will be lower than the other two cross validation strategies as in each round, the model has not learned the individual characteristics of the test patient, thus, it is more difficult to predict without any prior visual patient-specific knowledge. This effect is predominantly due to the existing biological variations between individual patients in a population, which is further enhanced by inter-WSI variations in stain intensities and malignancy levels. Hence, this type of cross validation is the most indicative of the performance of the classifiers in practical settings and has been applied in all the experiments in the key stage of this research involving analysis of cancer regions. Due to limited datasets from only four and five WSI during pre-analysis procedures of necrosis detection and cell nuclei classification respectively, leave-a-patient-out is not applied for these experiments.

8.1.2 Performance Metrics

The quantitative metrics used for evaluation of the cross validation performance are described below.

8.1.2.1 Confusion Matrix and Related Quantities

Confusion Matrix

A confusion matrix for multiple classes is computed during each cross validation round. This is defined as a $C \times C$ matrix \mathbf{M} when the actual (or reference) data consists of C number of classes, with element m_{ij} , $i, j \in \{1, 2, \dots, C\}$ as the number of samples of actual class i predicted as class j . A strong classifier results in large values of diagonal elements in the confusion matrix. Also, classifier confusion between two or more classes can be analyzed by observing the entries of this matrix. A confusion matrix is shown in Table 8.1. An accuracy confusion matrix \mathbf{A} can be computed from the number confusion matrix \mathbf{M} by dividing each element by the total number of samples in each class, *i.e.*,

$$a_{ij} = \frac{m_{ij}}{\sum_{k=1}^C m_{ik}} \quad (8.1)$$

Table 8.1 *Elements in the confusion matrix M*

Predicted \ Actual	1	2	. . .	C
1	m_{11}	m_{12}	. . .	m_{1C}
2	m_{21}	m_{22}	. . .	m_{2C}
.
.
.
C	m_{c1}	m_{c2}	. . .	m_{cc}

Per-class Classification Accuracy

This value denotes how well the samples of each class have been classified, and for the i^{th} class, it can be obtained as the i^{th} diagonal element of the accuracy confusion matrix **A**. It can be defined as,

$$PC_i = \frac{m_{ii}}{\sum_{k=1}^C m_{ik}} \quad (8.2)$$

Overall Classification Accuracy

The overall classification accuracy *OCA* is calculated from the confusion matrix **M** as the fraction of the total number of correct predictions as,

$$OCA = \frac{1}{N} \sum_{i=1}^C m_{ii} \quad (8.3)$$

where N is the total number of samples in the test dataset and can be obtained as,

$$N = \sum_{i=1}^C \sum_{j=1}^C m_{ij} \quad (8.4)$$

Balanced Classification Accuracy

The balanced classification accuracy *BCA* is calculated from the confusion matrix as the average of per class accuracy over the number of classes, and defined as,

$$BCA = \frac{1}{C} \sum_{i=1}^C PC_i \quad (8.5)$$

The balanced accuracy avoids bias in case of unbalanced number of samples for each class in the dataset, hence, it provides a better estimation of the classification algorithm compared to overall classification accuracy [Brodersen 2010].

8.1.2.2 Means and Standard Deviations

The overall mean and standard deviation of the mean for a test configuration $\{xyz\}$ over multiple cross validation rounds can be obtained from the above defined classification accuracy metrics for each round. The mean accuracy is calculated by,

$$\mu_{xyz} = \frac{1}{K} \sum_{k=1}^K P_k \quad (8.6)$$

The standard deviation of mean accuracy (also called standard error) is obtained as [Everitt 2006]:

$$\sigma_{xyz} = \sigma / \sqrt{K} \quad (8.7)$$

$$= \sqrt{\frac{1}{K(K-1)} \sum_{k=1}^K (P_k - \mu_{xyz})^2} \quad (8.8)$$

where σ represents sample standard deviation, K is the total number of cross validation rounds and P_k is a given performance metric (per-class accuracy, overall accuracy or balanced accuracy) computed for the k^{th} cross validation round in any given cross validation strategy. Means and standard errors are demonstrated using bar diagrams.

8.1.2.3 Box and Whisker Diagrams

These graphical plots are primarily generated in order to visually observe the performance variation of the classification algorithm between multiple cross validation rounds in each of the cross validation strategies. In the experiments the data series depicted using box and whisker diagrams consists of the per-class accuracy percentage over each of the cross validation rounds. A schematic example of the box and whisker plot is shown in Figure 8.1. In general, these diagrams are a simple and convenient way of representing statistical distributions, in which a rectangle is drawn to represent the first and third quartiles, usually with a line to indicate the median value. A point marker shows the mean value of the statistical distribution. Whiskers denote the minimum and maximum values on each side of the rectangle. The spacing between the different parts of the box indicate the degree of dispersion and skewness in the achieved classification rates over multiple cross validation rounds of the same cross validation strategy.

Box and whisker diagrams have been generated for the individual cross validation strategies such as k-fold stratified shuffled split and leave-a-patient-out in order to indicate inter-round variations in classification performance, however, the process is computationally intensive to accomplish using leave-a-sample-out cross validation due to a vast number of training and deployment rounds same as the number of samples, and only one classification per round. Hence, box and whisker diagrams are not demonstrated for this type of cross validation strategy.

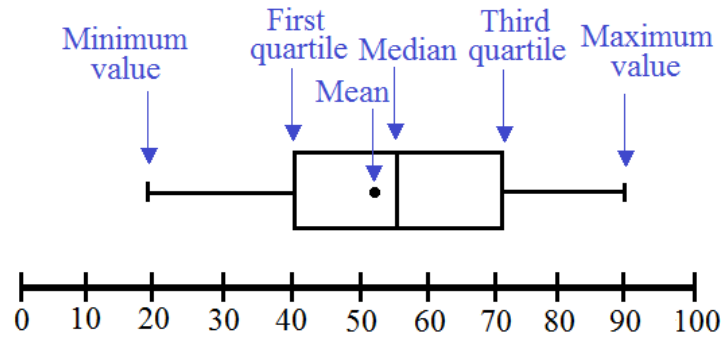
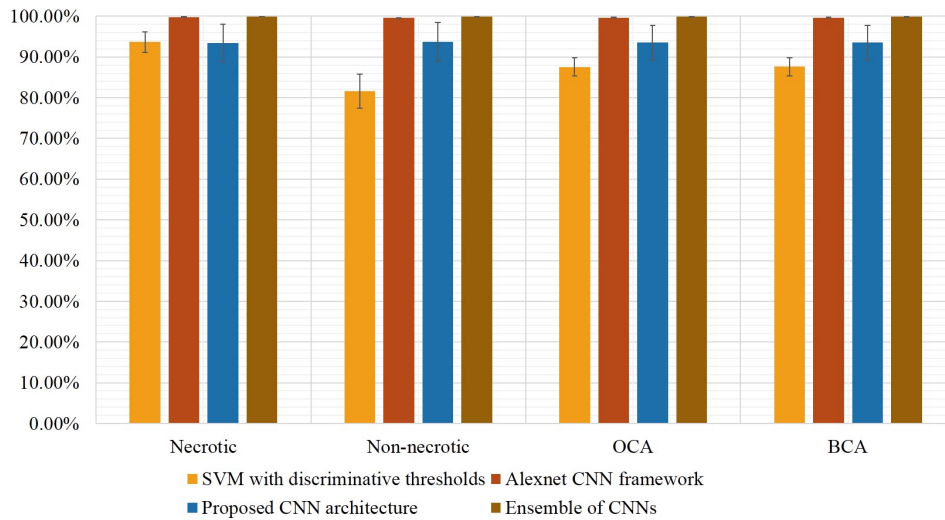


Fig. 8.1 Schematic example of box and whisker plot

8.2 Comparative Evaluation: Necrosis Detection

The overall mean and standard error of classification accuracies, namely, per-class, overall and balanced classification accuracy averaged over multiple experiments in necrosis detection are graphically depicted in Figure 8.2 using k-fold stratified shuffled split cross validation, where $k = 10$ for traditional methods using SVM with discriminative thresholds, and $k = 3$ for deep learning methods. The table for average balanced classification accuracy shows the comparison after averaging over multiple cross validation experiments. The detailed cross validation results including average accuracy confusion matrices for each classification method and comparative box and whisker diagrams for each class are illustrated in Sections B.1.1 and B.2.1.



(a)

Method	SVM with discriminative thresholds	AlexNet CNN framework	Proposed CNN architecture	Ensemble of CNNs
Average BCA	87.66%	99.71%	93.56%	99.89%

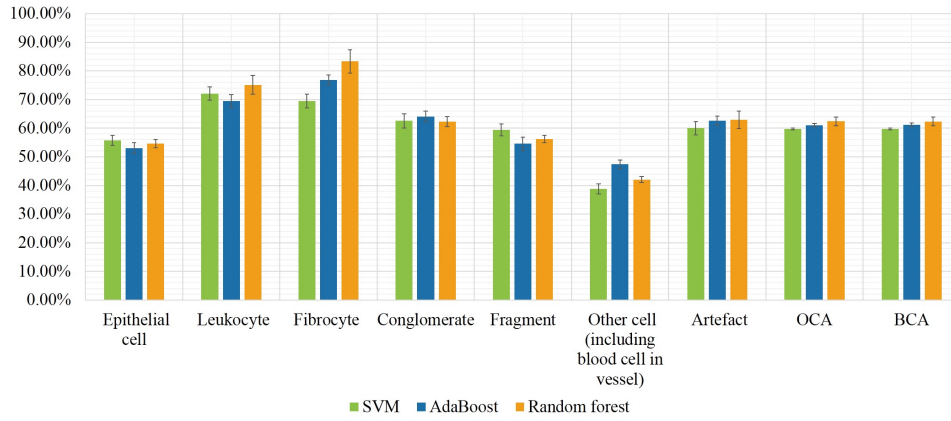
(b)

Fig. 8.2 Overall mean and standard error of classification accuracy for necrosis detection experiments using (a) k-fold stratified shuffled split cross validation (b) Average BCA for all methods over both cross validations.

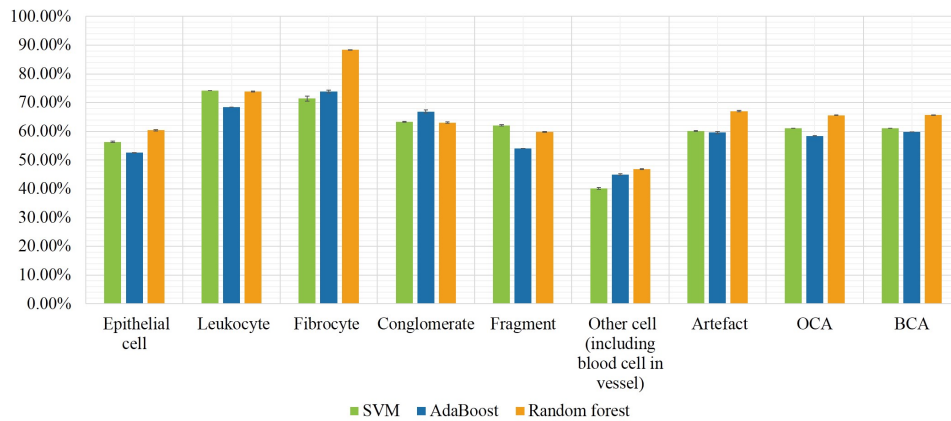
Leave-a-patient-out cross validation is not performed for necrosis detection, as the necrosis datasets have been generated using maximum available labeled areas in four WSI, which are not sufficient to incorporate the biological heterogeneity among all the patients. The author believes that leave-a-patient-out cross validation may not provide adequate evaluation, as labeled data is available for only four whole slide images which may introduce a learning bias during the training phase that adversely affects the classification of incoming samples from the unknown patient. This effect was experimentally observed and the leave-a-patient-out cross validation results are reported in [Sharma 2017b], however, not presented here due to few difficult WSI leading to complexities in individual characteristics of necrotic areas. Furthermore, leave-a-sample-out cross validation is also not applied here due to extremely high computational requirements, owing to very large datasets and long learning times of convolutional neural networks in deep learning methods.

From the mean and standard deviations of classification accuracy using k-fold cross validations, it can be observed that all the three methods achieve good classification rates ($>85\%$) for both categories in the necrosis detection task. The overall and balanced measures yield closely similar values because of nearly balanced number of samples for each category in the binary classification task. Summarizing the experimental results for necrosis detection, it can be observed that ensemble of CNNs has the best classification performance (average $OCA = BCA = 99.89\%$), followed by AlexNet CNN framework (average $OCA = BCA = 99.71\%$), the proposed CNN architecture (average $OCA = BCA = 93.56\%$) and the traditional method involving texture feature extraction followed by SVM classification and discriminative thresholding (average $OCA = 87.58\%$, $BCA = 87.66\%$). On comparing the classification performance of traditional machine learning method versus the deep learning methods, it is clearly observed that for necrosis detection, deep learning methods outperform traditional machine learning method. A similar effect has been observed after addressing the problem with random forest machine learning compared to deep learning in [Sharma 2017b].

Furthermore, the accuracies are slightly higher for necrotic class compared to non-necrotic class, suggesting its distinct appearance in human tissues, which is reinforced by observing the confusion matrices in Appendix B.1.1. A lower detection rate in the SVM-based traditional method is mainly seen for the non-necrotic class, which is not harmful as such false positives can be safely excluded with the correctly classified necrotic data. However, false negative rate should not be high due to possibility of inclusion in the final working datasets, which is fulfilled by all the described methods (average false negative rate = 3.29%). From Figure 8.2 and the box and whisker diagrams in Appendix B.2.1, it is evident that the traditional method shows larger variations between learning rounds followed by proposed CNN architecture and minimum for AlexNet CNN framework. Hence, for the necrosis detection task, traditional textural method using SVM machine learning can be selected for fewer computational requirements during training, and deep learning methods for a more accurate classification performance but are more computationally intensive during the training process. The selected method can be applied for classifying unknown regions as demonstrated in the second stage of the process pipeline,



(a)



(b)

Method	SVM	AdaBoost	Random forest
Average BCA	60.41%	63.89%	63.99%

(c)

Fig. 8.3 Overall mean and standard error of classification accuracy for cell nuclei classification experiments using (a) k -fold stratified shuffled split and (b) leave-a-sample-out cross validations (c) Average BCA for all methods over both cross validations.

allowing exclusion of visual information of necrotic areas before further image analysis in the third stage of research.

8.3 Comparative Evaluation: Cell Nuclei Classification

The overall mean and standard deviation of classification accuracies, namely, per-class, overall and balanced classification accuracy averaged over multiple cross validation experiments, are graphically depicted in Figure 8.3(a) using k -fold stratified shuffled split ($k = 3$) and Figure 8.3(b) using leave-a-sample-out cross validations in cell nuclei classification experiments. The table for average balanced classification accuracy presents overall comparison after averaging over both the cross validation methods. The detailed cross validation results including average accuracy confusion matrices for each classification method and comparative box and whisker diagrams for each class using individual cross validation strategies are shown in Sections B.1.2 and B.2.2.

Leave-a-patient-out cross validation is not performed in these experiments during image pre-analysis stage, as labeled data from the smaller initial working datasets in a few image tiles selected from five out of eleven WSI is available, hence, this cross validation method will not give a correct estimation of the existing inter-patient variance for small subset of the entire whole slide image data. In contrast, leave-a-sample-out cross validation is considered as a more appropriate method for quantitative analysis of cell nuclei classification due to the study of only traditional machine learning algorithms which don't have a high computational requirement in their training procedure compared to deep learning methods.

From the mean and standard error bar diagrams, it is observed that among the seven classes, the recognition rates using all the machine learning methods is highest for fibrocytes, followed by leukocytes, and then conglomerates, artefacts and fragments, validating the ability of the classification methods to strongly distinguish cell nuclei of these classes based on the extracted cell nuclei features. Especially, fibrocytes show the highest classification accuracy due to their distinct 'long and thin' appearance. However, a lower performance is observed for the class other cell, which can be explained as this class contains cell nuclei that are either not visually distinct or not clearly stained in the H&E images, hence, difficult to be distinguished into any specific class, and also include blood cells in vessels, which have a very low occurrence. So this class is not compact due to inconsistent intra-class characteristics. An intermediate accuracy for epithelial cells is because of a varying appearance among the three types of cancer regions, as these cells are mostly affected by cancer proliferation and change appearance with the malignancy levels. The difficulty of classifying fragments is due to the lack of a well-defined visual appearance or structure, and these are available in different shapes and sizes and some may resemble other cell types. From the confusion matrices in Appendix B.1.2, it can be noted that the highest classification confusions are between artefacts, fragments and other cells.

On comparing the classification performance of the three machine learning methods, it can be seen that, for most classes and both cross validations, random forests have highest multi-class classification rates (average $OCA = 64.01\%$, $BCA = 63.99\%$) followed by AdaBoost (average $OCA = 63.38\%$, $BCA = 63.89\%$) and then SVM (average $OCA = 60.39\%$, $BCA = 60.41\%$). However, it has slightly lower recognition rates than the other two classification methods for classes conglomerates and intermediate performance for fragments and other cell classes. The trends of per-class, overall and balanced classification accuracies correspond for both cross validation strategies with minor differences. The overall and balanced classification accuracies are almost equal due to the balanced nature of the cell nuclei dataset over cross validation rounds. Also, standard error for leave-a-sample-out cross validation are low due to lower variance between corresponding trained models in multiple experiments, as inter-round computation is not feasible for each round with only one prediction. Due to a superior classification performance, high predictability and lower inter-round variations as seen from the graphical results, random forests machine learning method has marginally surpassed the other two traditional machine learning methods, hence, selected for the cell nuclei classification of unknown samples. The result of classification using random forests are utilized for subsequent steps, namely, multiresolution

enhancement of segmentation results and determination of tissue composition in the second stage, as well as parts of region analysis based on cancer in the third stage of the research pipeline.

8.4 Comparative Evaluation: Analysis of Cancer Regions

8.4.1 Concept of Comparative Evaluation

8.4.1.1 Scheme for Comparative Evaluation

An extensive quantitative comparison is performed using traditional methods and deep learning methods for the analysis of cancer regions in H&E stained gastric cancer image WSI based on HER2 immunohistochemistry. Among the traditional methods, individual feature groups in the three image description categories are considered, namely, low-level state-of-the-art features including GLCM statistics, Gabor filter-bank responses, local binary patterns, Varma-Zisserman textons, gray histograms, HSV histograms, RGB histograms and other color-based measurements; high-level state-of-the-art features including Voronoi diagram and Delaunay triangulation; and high-level handcrafted features including the eight variants of the cell nuclei ARG. Furthermore, object-level features are considered using only the vertex attributes in the cell nuclei ARG which include global morphological and contour intensity-based features of the cell nuclei segments. Tissue composition is another candidate using the vertex identities in the cell nuclei ARG as a result of cell nuclei classification in each region of interest. These two cases are taken as a part of the high-level handcrafted features as they are derived from the cell nuclei ARG, and examined to establish the significance of neighborhood relationships and spatial arrangements between tissue components. Two combinations of feature groups in the form of hybrid features are also included in the comparative evaluation. Among the deep learning methods, performance is assessed for AlexNet CNN framework, proposed CNN architecture and ensemble of the two networks. The scheme for comparative evaluation is shown in Table 8.2. The detailed experimental results with pictorial representations, observations and discussion for the analysis of cancer regions using the scheme of comparative evaluation can be found in this section, and are elaborated further in Sections B.1.3 and B.2.3 in the form of confusion matrices and box and whisker diagrams respectively. Additionally, the computational requirements of proposed methods using traditional machine learning and deep learning are given in Appendix A.5.

8.4.1.2 Implications from Results of Comparative Evaluation

The overall observations are presented in Section 8.4.7, which can be referred to facilitate further discussion about the implications of results during analysis of cancer regions, including relative advantages and limitations of the described methods.

From the experimental experience of the above described methods, one apparent advantage of using handcrafted features followed by traditional machine learning is their enhanced scope of interpretation in histological images possessing unique characteristics, including large WSI

Table 8.2 *Scheme of comparative evaluation of studied methods for analysis of cancer regions*

Level of feature extraction	Contained feature groups	Machine learning method
Low-level state-of-the-art features	GLCM statistics Gabor filter-bank responses Local binary patterns Varma-Zisserman textons Gray histograms RGB histograms HSV histograms Other color-based measurements	Traditional methods Support vector machines AdaBoost ensemble learning Random forests
High-level state-of-the-art features	Voronoi diagrams and Delaunay triangulation	
High-level handcrafted features	Object-level features: only vertex attributes Tissue composition: only vertex identities $gARG[r_F]$ $gARG[r_A]$ $nsARG[r_F]$ $nsARG[r_A]$ $ncARG[r_F]$ $ncARG[r_A]$ $ncARG_{v+}[r_F]$ $ncARG_{v+}[r_A]$	
Combination of features	Selected subset of hybrid low-level features Selected subset of hybrid low-level and high-level features	
Raw images	-	Deep learning methods AlexNet CNN framework Proposed CNN architecture Ensemble of CNNs

sizes, multiple resolutions and complex appearances. Such interpretations have been established in Table 7.1 based on visual appearance of the tissues in this study, using high-level architectural information including cell structures, arrangements and topology. However, complete control over extracted features is not achieved by the convolutional neural networks as internal abstractions in its layers are not easily understood. This requires an extensive study of the design parameters during the CNN architecture design phase. The requirement has led to the empirical analysis of CNN architectures to determine an optimal framework for the histopathological problems using several permutations and combinations in this work. However, an exhaustive study is a highly time-consuming process and limited by GPU specifications, so currently a few intuitively designed CNN architectures are explored and optimized for the cancer analysis problem. The selected CNN architecture is later applied for detecting necrosis to explore generalizability of CNNs and shows a reasonable performance in both applications. Other possible task-specific deep networks can be studied in future which were limited now due to time and

hardware requirements, including the architectures optimized for necrosis detection. The second bottleneck observed for deep learning methods like all other deep learning applications so far, is the requirement of large-scale datasets containing thousands of images for training unlike the traditional methods. This problem is currently addressed by data augmentation on the available WSI datasets, but it will be a more appropriate direction to expand the available ground truth by using a larger number of patients. It can be emphasized that the extension of current ground truth data with more patients and lower number of images per patient for training, can provide more robust classification algorithms. The third point in favor of traditional machine learning is the currently a lower computational complexity compared to deep learning, that allows traditional methods to be executed in acceptable times on the CPU, however, deep learning using CNNs requires longer training times that are not suitable for CPU execution due to the high algorithm complexity. Lastly, implementing traditional machine learning algorithms require selection of few experimental parameters, in contrast to deep learning where many parameters must be decided before the training phase, as seen in Table 7.4.2.5.

Nevertheless, the traditional methods also exhibit drawbacks compared to deep learning methods. Especially, for the high-level handcrafted feature extraction using graph-based methods, the following challenges are confronted. Firstly, the cell nuclei segmentation algorithm has not achieved best results due to a non-ideal acquisition process. The segmentation has been optimized for gastric cancer images using a multiresolution combination approach and the combined results are enhanced compared to individual magnifications. However, during segmentation evaluation, still some cell nuclei are not correctly detected and this effect can lead to the construction of cell nuclei ARGs with a few missing cell nuclei. Moreover, the cell nuclei contours even in the constituent resolutions are not always accurately estimated, hence, after multiresolution combination, such contours will not be depicting the true cell boundaries that may distort the object-level features of cell nuclei segments, which are important vertex attributes in the cell nuclei ARG. The third disadvantage of the proposed high-level method is that cell nuclei classification shows high discrimination complexity of the cell categories, which introduces additional errors in the $nsARG$, $ncARG$ and $ncARG_{v+}$ variants of the cell nuclei ARGs. All the errors of the cell nuclei segmentation and classification in the image pre-analysis stage prior to graph construction can propagate in the cancer classification process leading to possible misclassifications. In contrast, deep learning uses raw images directly, thereby, eliminating the requirement of the two steps of the image pre-analysis stage, namely cell nuclei segmentation and classification. This further leads to a general advantage of deep learning versus all the traditional methods that is the absence of handcrafted feature design and extraction stage, thereby, saving time and effort for this purpose.

As specified in Section 7.2.1, two image tile sizes *i.e.* 512×512 and 1024×1024 have been tested with the preference to the larger size due to additional contextual information. To verify this assumption, a subset of the experiments using handcrafted high-level features was performed for both sizes and comparatively evaluated using the two datasets described in Section 7.2.1. The calculated metrics show more favorable results for the larger image size as large ROIs

contain more context information with larger fields of view at a high magnification. Specifically, using the $gARG[r_F]$ handcrafted graph-based method, followed by random forest single-stage classification, average BCA for the smaller tile size shows a reduction by $\approx 8\%$. Hence, the result for smaller image size have not been expanded for traditional machine learning methods to reduce space and time complexity and allow a clear visualization of optimal results. However, an apparent observation in favor of deep learning methods over handcrafted features and traditional machine learning methods, is the use of the smaller 512×512 image size due to GPU memory restrictions generating similar classification accuracies. The smaller size provides comparable results for deep learning, to those obtained for larger size and superior than the same size in traditional machine learning. This is a very interesting and encouraging observation in favor of deep convolutional neural networks, and suggests that the raw images contain rich information and are processed by the several layers of CNN which have the potential to extract meaningful discriminative knowledge required for accurate prediction. Moreover, the input image size can be increased to larger ones, *e.g.* 1024×1024 pixels, with the deployment of cascaded or upgraded GPUs, however, it remains a future prospect of the research.

On comparing the computational complexities, it can be observed in Appendix A.5 that among traditional methods, the non-optimized graph construction algorithm for cell nuclei ARG depends on tissue complexity and leads to higher computational requirements compared to state-of-the-art feature extraction methods. Also, the feature combination in hybrid methods causes a high dimensional feature space that increases the time complexity and space requirements for learning and classification for traditional machine learning methods, especially for the AdaBoost ensemble learning algorithm, where a non-optimized code is being used. Among the deep learning methods, the training phase using the proposed CNN, even with a fast (although not optimized) implementation on the GPU, requires around two days for the current datasets. In contrast, AlexNet requires a few hours due to an optimized implementation. Furthermore, an optimized implementation of traditional machine learning methods requires relatively smaller duration to complete training, even for same-sized data [Sharma 2017b]. More details on computational requirement are discussed in Appendix A.5. Improving computational requirements for the non-optimized algorithms is essential and will be addressed in the future.

In the following sections, the overall mean and standard error of classification accuracies, namely, per-class, overall and balanced classification accuracy averaged over multiple cross validation experiments, are graphically shown and observations are discussed for each evaluation configuration using k-fold stratified shuffled split ($k = 10$) and leave-a-patient-out cross validations for analysis of cancer regions. An overall comparison of the balanced classification accuracy averaged over both the cross validation methods is also depicted in tabular form. The detailed cross validation results including average accuracy confusion matrices (using the most successful classification method), and comparative box and whisker diagrams for each class are shown in Sections B.1.3 and B.2.3 for the two individual cross validation strategies.

In general, from the quantitative evaluation of classification performance, it can be observed that mostly, non-tumor and tumors are easily discriminated from each other (in first stage of

hierarchical classification strategy), followed by HER2 positive tumor, and then HER2 negative tumor. Furthermore, from the assessment of leave-a-patient-out cross validation results, it is found that its overall performance is lower than the performance reported using k-fold stratified shuffled split cross validation, along with a higher deviation for all the methods. These two effects are explained in the overall observations in Section 8.4.7.1 showing class-wise and evaluation method-wise comparisons, after the experimental results of individual methods. The leave-a-patient-out performance is a measure of the robustness and practical usability of the system, hence, can be given precedence over k-fold stratified shuffled split to decide the comprehensive accomplishment of a given method.

8.4.2 Low-level State-of-the-art Features

These features include the state-of-the-art pixel-based feature extraction methods, namely, GLCM statistics, Gabor filter-bank responses, local binary patterns, Varma-Zisserman textons, gray histograms, HSV histograms, RGB histograms and other color-based measurements.

8.4.2.1 GLCM Statistics

The k-fold stratified shuffled split ($k = 10$), leave-a-patient-out and overall performance for GLCM based features is shown in Figure 8.4. The SVM hierarchical classification shows most favorable overall results in k-fold stratified shuffled split (average $OCA = 82.04\%$, $BCA = 81.73\%$) and leave-a-patient-out (average $OCA = 52.88\%$, $BCA = 51.32\%$) cross validations. Next best are random forests single-stage, SVM single-stage and random forest hierarchical strategies, which are followed by AdaBoost with a higher difference. In general, GLCM statistics prove to be a powerful feature as a low number of features have been able to provide a reasonable and comparable performance using k-fold stratified shuffled split cross validation. However, there is a large gap in performance using these features in leave-a-patient-out cross validation,

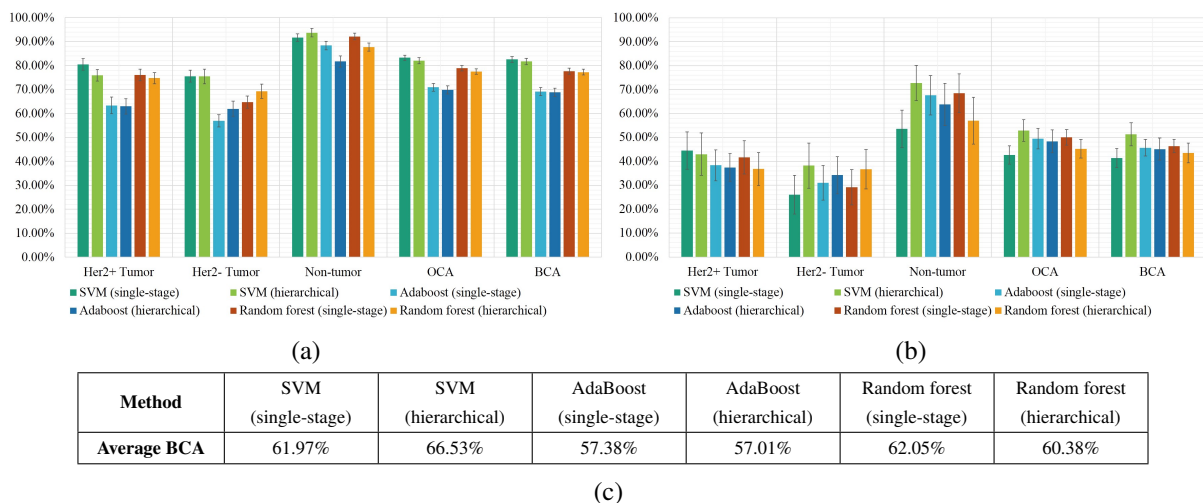


Fig. 8.4 Overall mean and standard error of classification accuracy for cancer classification experiments for GLCM statistics using (a) k-fold stratified shuffled split and (b) leave-a-patient-out cross validation (c) Average BCA for all methods over both cross validations.

i.e. significant when compared to observations of subsequent experiments with handcrafted graph-based feature extraction methods and deep learning methods, indicating that GLCM-based features are less robust to biological variations compared to high-level traditional methods and deep learning methods explored later. From the confusion matrices in Appendix B.1.3, it is interesting to notice higher confusion of HER2 negative class with HER2 positive class, but lower confusion vice-versa, which is also observed for other methods mentioned in the following sections, suggesting that HER2 positive tumor is easier to recognize by trained classifiers compared to HER2 negative tumor.

8.4.2.2 Gabor Filter-bank Responses

The k -fold stratified shuffled split ($k = 10$), leave-a-patient-out and overall classification performance for Gabor filter-bank features is depicted in Figure 8.5. The average BCA obtained using Gabor filters is systematically lower compared to GLCM statistics, showing their limited descriptive power for the described problem. Lowest overall errors are achieved using SVM hierarchical classification in k -fold stratified shuffled split (average $OCA = 79.87\%$, $BCA = 79.71\%$) and leave-a-patient-out (average $OCA = 45.58\%$, $BCA = 42.80\%$), followed by SVM single-stage and random forests hierarchical classification. A noticeable decline is again seen for AdaBoost machine learning method. Furthermore, for the GLCM statistics, large differences are observed between the two types of cross validations. Particularly, the leave-a-patient-out accuracies are extremely low, depicting a lower robustness of Gabor features to heterogeneity among patients leading to less suitability for the cancer classification problem.

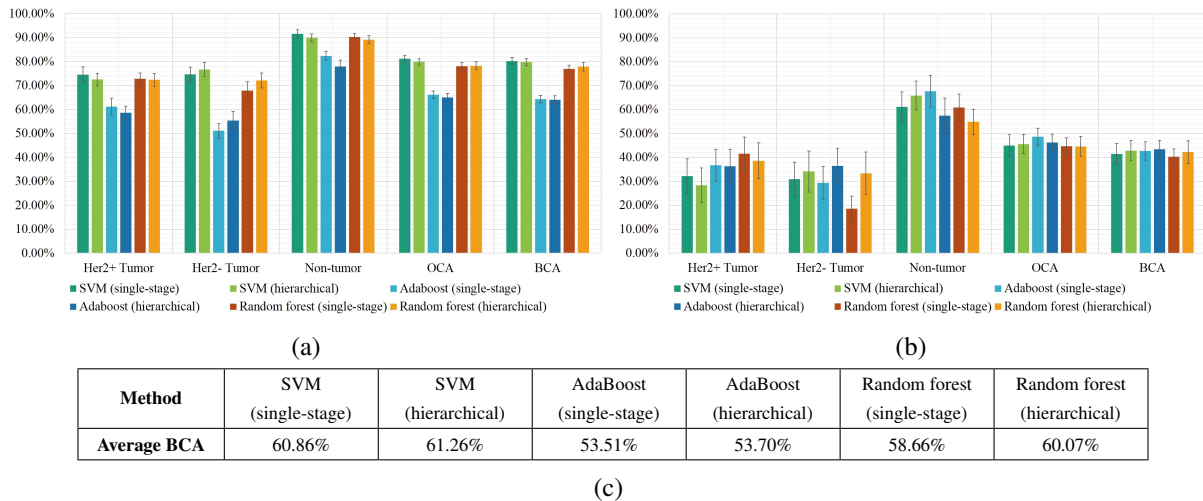


Fig. 8.5 Overall mean and standard error of classification accuracy for cancer classification experiments for Gabor filter-bank responses using (a) k -fold stratified shuffled split and (b) leave-a-patient-out cross validation (c) Average BCA for all methods over both cross validations.

8.4.2.3 Local Binary Patterns

The k -fold stratified shuffled split ($k = 10$), leave-a-patient-out and overall classification performance for local binary patterns is presented in Figure 8.6. At a glance, LBP histograms show

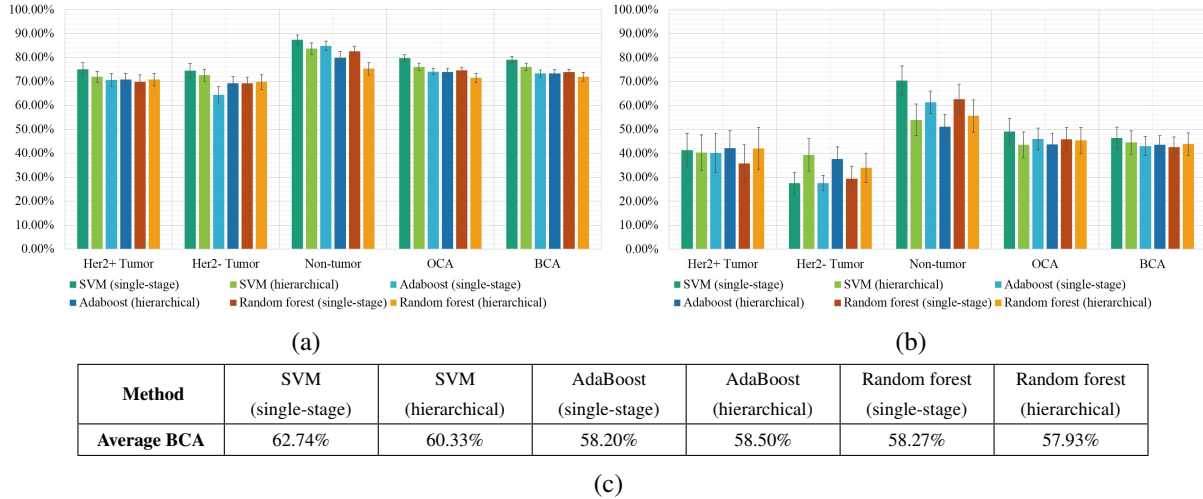


Fig. 8.6 Overall mean and standard error of classification accuracy for cancer classification experiments for LBP histograms using (a) k -fold stratified shuffled split and (b) leave-a-patient-out cross validation (c) Average BCA for all methods over both cross validations.

intermediate classification performance between the previously evaluated features, namely, GLCM statistics and Gabor filter-bank responses for the computer-based analysis of gastric cancer. SVM single-stage classification has best overall results with k -fold stratified shuffled split (average $OCA = 79.70\%$, $BCA = 79.03\%$) and leave-a-patient-out (average $OCA = 49.10\%$, $BCA = 46.45\%$) cross validations. Next best classification is achieved by SVM and AdaBoost in hierarchical strategy with close similarities. Random forests also have a similar performance to AdaBoost ensemble learning. Again, like GLCM and Gabor filters, the leave-a-patient-out accuracies are especially not satisfactory for the two tumor classes, suggesting a low competence of local binary patterns for the specific cancer analysis task.

8.4.2.4 Varma-Zisserman Textons

Figure 8.7 presents the k -fold stratified shuffled split ($k = 10$), leave-a-patient-out and average classification accuracy using Varma-Zisserman texton-based approach. The Varma-Zisserman texton-based approach shows the most promising results compared to the state-of-the-art low-level features discussed so far using k -fold stratified shuffled split, suggesting its higher popularity in histological image analysis. Most superior classification performance is achieved by SVM single-stage method in k -fold stratified shuffled split (average $OCA = 89.56\%$, $BCA = 89.06\%$) and leave-a-patient-out (average $OCA = 48.65\%$, $BCA = 44.30\%$) cross validations, followed by SVM hierarchical method. The other two classification methods show a wider difference to SVM, as random forests and AdaBoost in order. One prominent observation is the huge gap between the two types of validation results especially in the two tumor classes, with a highly

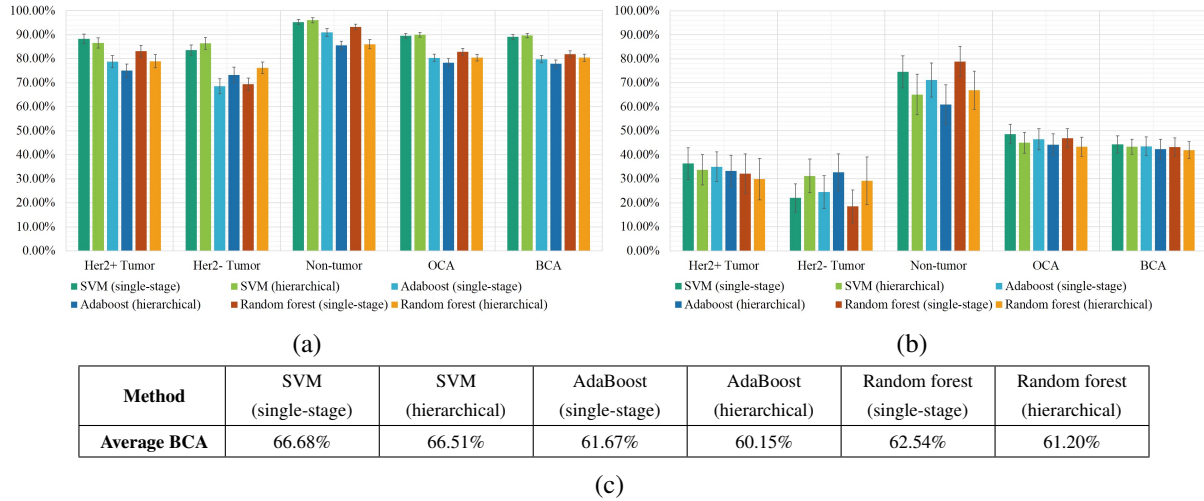


Fig. 8.7 Overall mean and standard error of classification accuracy for cancer classification experiments for Varma-Zisserman textons using (a) k -fold stratified shuffled split and (b) leave-a-patient-out cross validation (c) Average BCA for all methods over both cross validations.

favorable random stratified shuffled split performance but comparatively low leave-a-patient-out response. This effect can be attributed to two main reasons. The unusually high former metric may represent an over-fitting of trained classifiers (especially seen in SVM methods), and lower leave-a-patient-out accuracies depict poor robustness of the texton-based features towards biological variations in population.

8.4.2.5 Gray Histograms

The k -fold stratified shuffled split ($k = 10$), leave-a-patient-out and overall classification performance for gray histograms is shown in Figure 8.8. From the above observations, it can be seen that

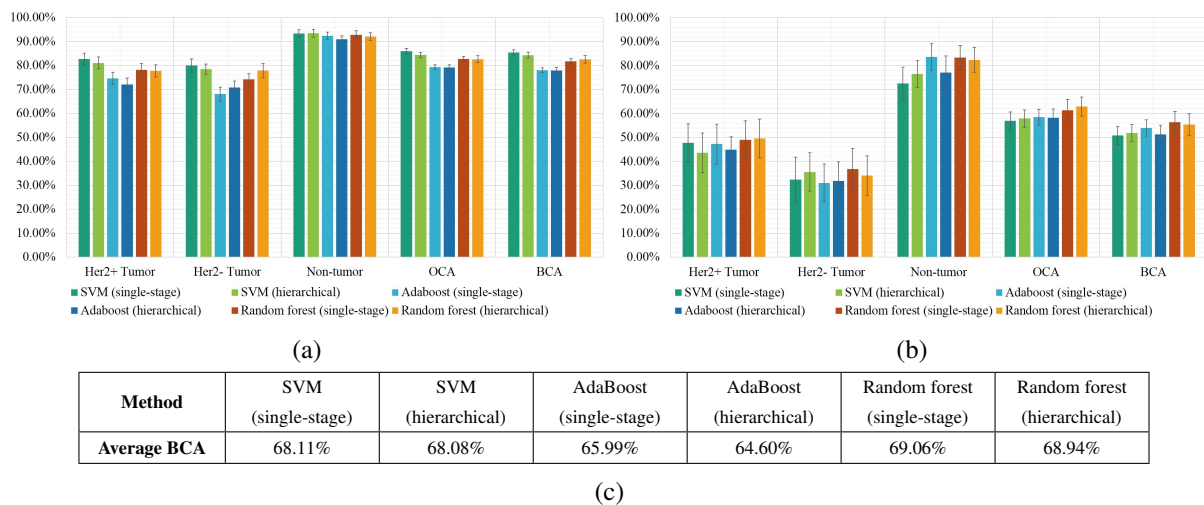


Fig. 8.8 Overall mean and standard error of classification accuracy for cancer classification experiments for gray histograms using (a) k -fold stratified shuffled split and (b) leave-a-patient-out cross validation (c) Average BCA for all methods over both cross validations.

gray histograms have performed superior to all the state-of-the-art low-level methods discussed before. Among the classification methods, random forests single-stage method has highest classification accuracy in k-fold stratified shuffled split (average $OCA = 82.65\%$, $BCA = 81.73\%$) and leave-a-patient-out (average $OCA = 61.28\%$, $BCA = 56.39\%$) cross validations. This is followed by random forest hierarchical classification, SVM and AdaBoost in single-stage and hierarchical strategies respectively. It can be noted that the leave-a-patient-out cross validation shows better performance compared to LBP histograms and Varma-Zisserman textons, hence, the features can be considered more robust to the inter-patient variability in a given population. Therefore, gray histograms prove to be more useful low-level features in the analysis of H&E stained images, and the corresponding intensity distributions obtained due to H&E staining contribute in providing effective visual characteristics to the tissue regions according to the given malignancy levels.

8.4.2.6 HSV Histograms

The k-fold stratified shuffled split ($k = 10$), leave-a-patient-out and overall classification performance for HSV histograms is given in Figure 8.9. For the HSV histograms, it is observed that overall, HSV based features are even more effective than the gray histograms, and show superior performance for the given classification task. We observe that HSV histogram contains a large number of features and interestingly, the AdaBoost ensemble learning method has the highest accuracy in single-stage classification strategy with k-fold stratified shuffled split (average $OCA = 89.33\%$, $BCA = 88.79\%$) and leave-a-patient-out (average $OCA = 65.45\%$, $BCA = 60.98\%$) cross validations, followed by SVM, AdaBoost and random forest in hierarchical counterparts. Random forest single-stage classification show an abruptly low leave-a-patient-out accuracy for HER2 negative tumor class. The success of AdaBoost algorithm for this case is mainly due to a larger number of features, as AdaBoost provides each

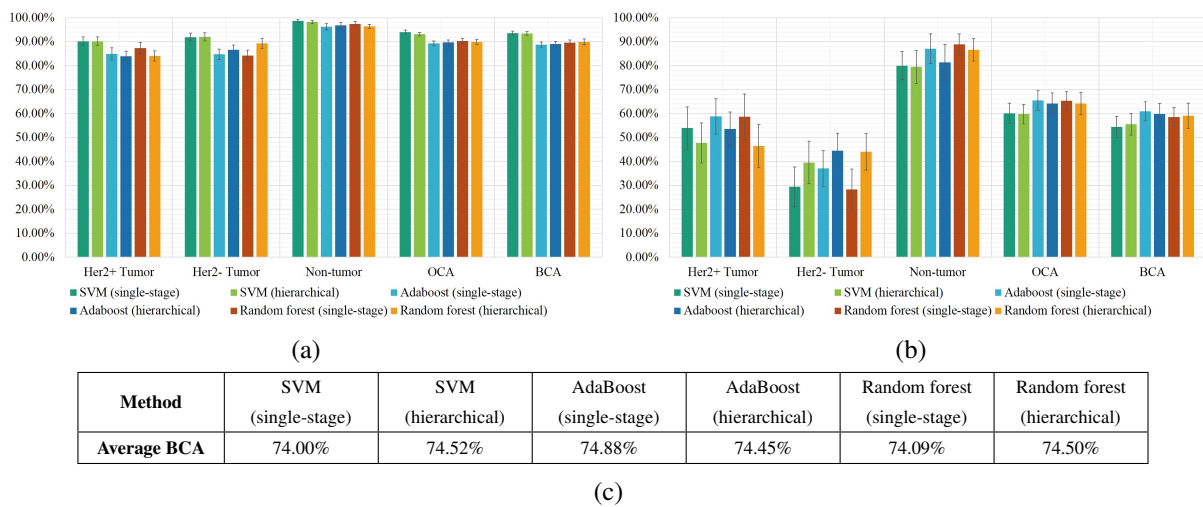


Fig. 8.9 Overall mean and standard error of classification accuracy for cancer classification experiments for HSV histograms using (a) k-fold stratified shuffled split and (b) leave-a-patient-out cross validation (c) Average BCA for all methods over both cross validations.

feature incrementally to the weak learners in order to improve its current decision of the ensemble of classifiers. This effect will also be observed in the subsequent experiments dealing with longer feature vectors. Again, due to higher leave-a-patient-out performance, HSV histograms have proved to be more robust to the heterogeneous nature of our WSI datasets as compared to the previous low-level state-of-the-art feature extraction approaches, thus, constituting a stronger feature set for describing the tissue regions for the cancer classification problem.

8.4.2.7 RGB Histograms

Figure 8.10 presents the k -fold stratified shuffled split ($k = 10$), leave-a-patient-out and average classification accuracy using RGB histograms. It can be observed that the RGB histograms

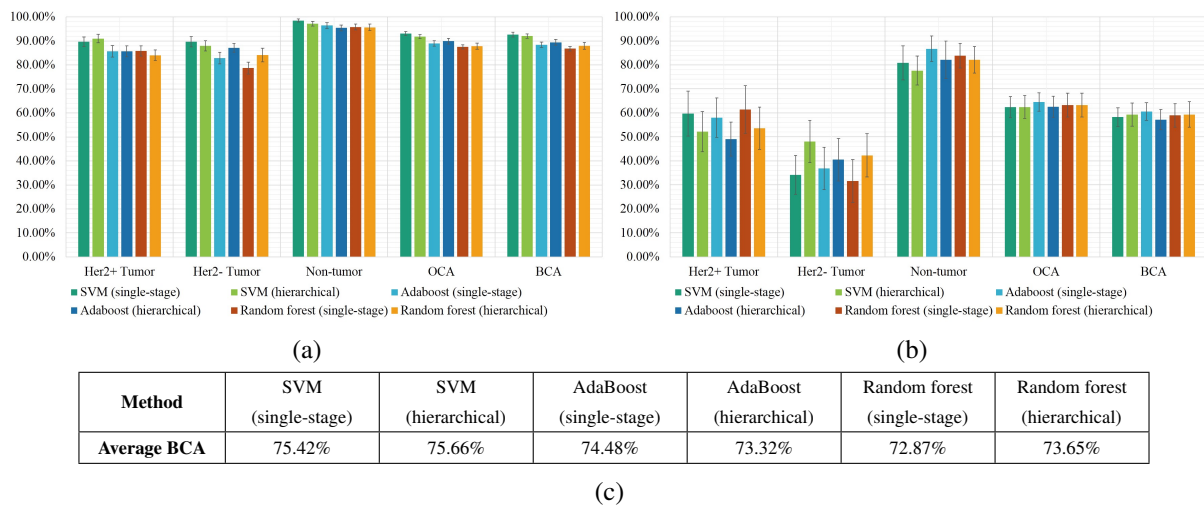


Fig. 8.10 Overall mean and standard error of classification accuracy of cancer classification experiments for RGB histograms using (a) k -fold stratified shuffled split and (b) leave-a-patient-out cross validation (c) Average BCA for all methods over both cross validations.

can provide powerful and descriptive features for discrimination between the three malignancy levels based on HER2 immunohistochemical response. Until now, these features have the most successful classification performance among all the other state-of-the-art low-level feature extraction methods, as they even outperform gray and HSV histograms. SVM hierarchical classification method shows an overall improvement in k -fold stratified shuffled split (average $OCA = 91.77\%$, $BCA = 92.03\%$) and leave-a-patient-out (average $OCA = 62.45\%$, $BCA = 59.28\%$) cross validations, and SVM and AdaBoost single-stage approaches are closest in performance. AdaBoost classification is ranked higher than random forests, which can be attributed to a larger number of extracted features as observed in HSV histograms. The highest leave-a-patient-out accuracy so far suggests that the features have better practical usability for completely unlabeled patient WSI data.

8.4.2.8 Other Color-based Measurements

Figure 8.11 presents the k -fold stratified shuffled split ($k = 10$), leave-a-patient-out and average classification accuracy using the other color-based measurements. From the results, it is evident

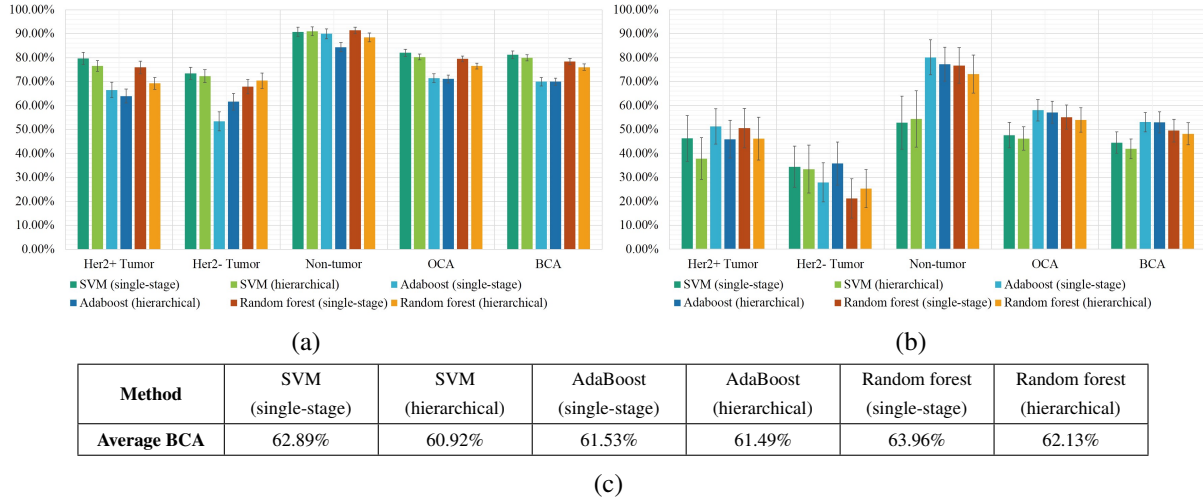


Fig. 8.11 Overall mean and standard error of classification accuracy for cancer classification experiments for other color-based measurements using (a) k -fold stratified shuffled split and (b) leave-a-patient-out cross validation (c) Average BCA for all methods over both cross validations.

that the low-level color-based features, although few in number with six features, show discriminative potential for histological images. These are not as efficient as the histograms explored earlier, however, perform more reasonably compared to other low-level state-of-the-art features such as Gabor filter-bank responses and LBP histograms, especially showing superior leave-a-patient-out accuracies. The most successful classification method is random forest single-stage classification with k -fold stratified shuffled split (average $OCA = 79.53\%$, $BCA = 78.42\%$) and leave-a-patient-out (average $OCA = 55.15\%$, $BCA = 49.51\%$), and SVM single-stage and random forest hierarchical classification follow closely. It can also be observed from the bar diagrams, that the AdaBoost method in both strategies has the most favorable leave-a-patient-out performance. Hence, the combination of the features with AdaBoost can be considered robust to heterogeneity among patients in contrast to some of the other low-level state-of-the-art features.

8.4.3 High-level State-of-the-art Features

In this section, experiments are performed to analyze the graph-based methods involving proximity graphs, specifically Voronoi Diagrams and Delaunay Triangulation. In literature, these graph and their derivatives have shown success in various histopathological problems in digital pathology [Sharma 2015a], hence, can be considered as high-level state-of-the-art methods in histological image analysis. They are comparatively evaluated with the other state-of-the-art low-level and the proposed handcrafted high-level features. The k -fold stratified shuffled split ($k = 10$), leave-a-patient-out and overall classification performance of the Voronoi-Delaunay method is shown in Figure 8.12. It can be observed that for the gastric cancer images, the

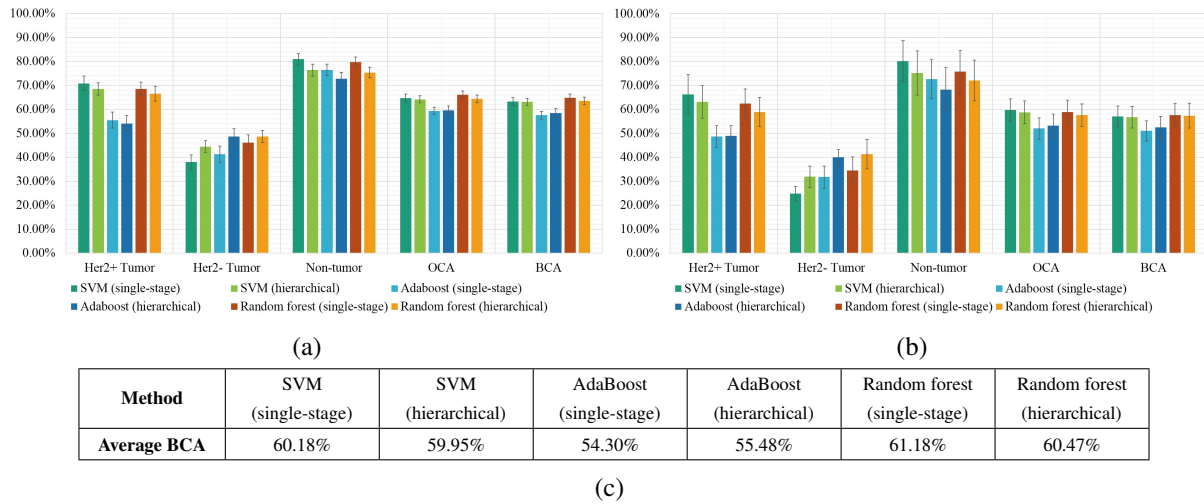


Fig. 8.12 Overall mean and standard error of classification accuracy for malignancy region classification experiments for Voronoi-Delaunay method using (a) *k*-fold stratified shuffled split and (b) leave-a-patient-out cross validation (c) Average BCA for all methods over both cross validations.

Voronoi-Delaunay method has comparable performance to some of the low-level feature extraction approaches such as Gabor filter-bank responses and LBP histograms, however, is inferior to other state-of-the-art methods. Random forest single-stage classification gives most superior results in *k*-fold stratified shuffled split (average $OCA = 66.15\%$, $BCA = 64.80\%$) and leave-a-patient-out (average $OCA = 58.89\%$, $BCA = 57.56\%$). Next best are random forest in single-stage and SVM in hierarchical strategy. These features show more promising results using leave-a-patient-out cross validation, however, the stratified shuffled split performance is not as satisfactory as previously studied low-level methods, and presents potential scope of improvements. In the later experiments using the proposed handcrafted methods, a desirable enhancement of classification performance can be further noticed. Interestingly, a favorable characteristic of this method is a lower gap between corresponding accuracy using the two cross validation approaches, indicating strength of features to handle biological variations.

8.4.4 High-level Handcrafted Features

These include the self-designed graph-based features for analysis of cancer regions including generic cell nuclei ARG, nuclei-specific ARG, nuclei-composite ARG and nuclei composite ARG with additional vertex attributes, each with fixed and adaptive r . Only vertex attributes are also considered with and without the identities of vertices representing tissue composition and object-level features respectively.

8.4.4.1 Object-level Features: Only Vertex Attributes

This section evaluates the classification performance when only the vertex attributes of proposed cell nuclei ARG are considered, and represent the individual cell nuclei object-level properties, without the associated edge attributes and neighborhood information. The overall k-fold stratified shuffled split ($k = 10$), leave-a-patient-out and overall classification performance is shown in Figure 8.13. The best classifier is obtained using random forest hierarchical method with k-fold

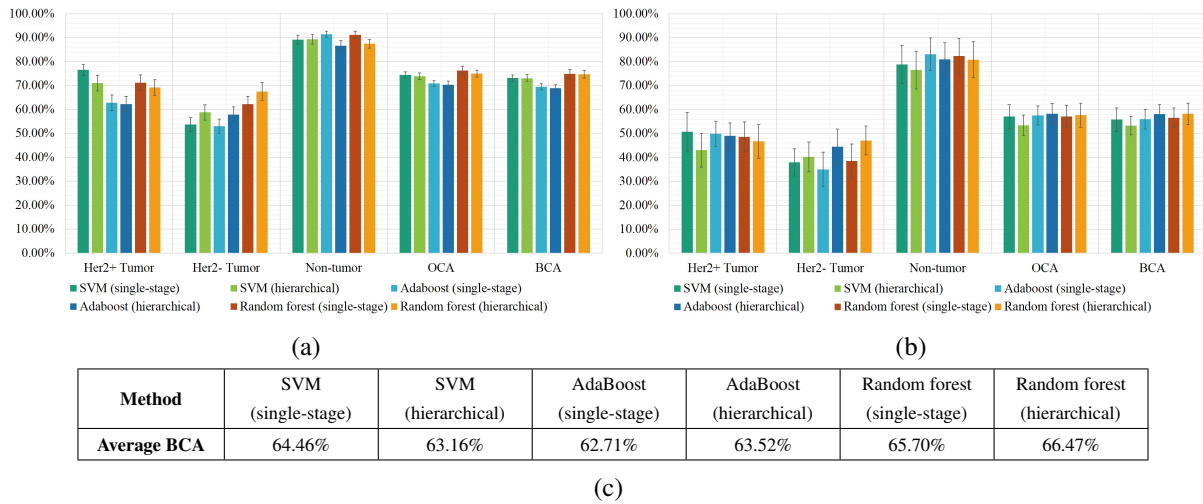


Fig. 8.13 Overall mean and standard error of classification accuracy for region classification experiments for only vertex attributes using (a) k-fold stratified shuffled split and (b) leave-a-patient-out cross validation (c) Average BCA for all methods over both cross validations.

stratified shuffled split (average $OCA = 74.95\%$, $BCA = 74.72\%$) and leave-a-patient-out (average $OCA = 57.61\%$, $BCA = 58.21\%$), closely followed by its single-stage counterpart. SVM and AdaBoost show similar performance with slight precedence of the former. It is interesting to see that this feature extraction method already outperforms a section of the individual low-level and high-level state-of-the-art methods discussed earlier. This strongly suggests the discriminative strength of cell nuclei characteristics based on morphology and contour intensity using H&E stained images due to the distinct staining of the cell nuclei, that would have been difficult to achieve in other stains such as HER2. It is even more exciting to observe later, that most of the proposed cell nuclei ARG methods perform superior to the method comprising only vertex attributes and reflects the importance of neighborhood relationships and spatial arrangements to describe the tissue in addition to individual characteristics of cell nuclei.

8.4.4.2 Tissue Composition: Only Vertex Identities

In this method, cell nuclei are first automatically classified using the cell nuclei classification method, and the result of this step are utilized to obtain the tissue composition as the fraction of each type of cell nuclei segment in each region of interest. The overall k-fold stratified shuffled split ($k = 10$), leave-a-patient-out and average accuracy is shown in Figure 8.14.

Overall superior result are attained by SVM single-stage classification using k-fold stratified

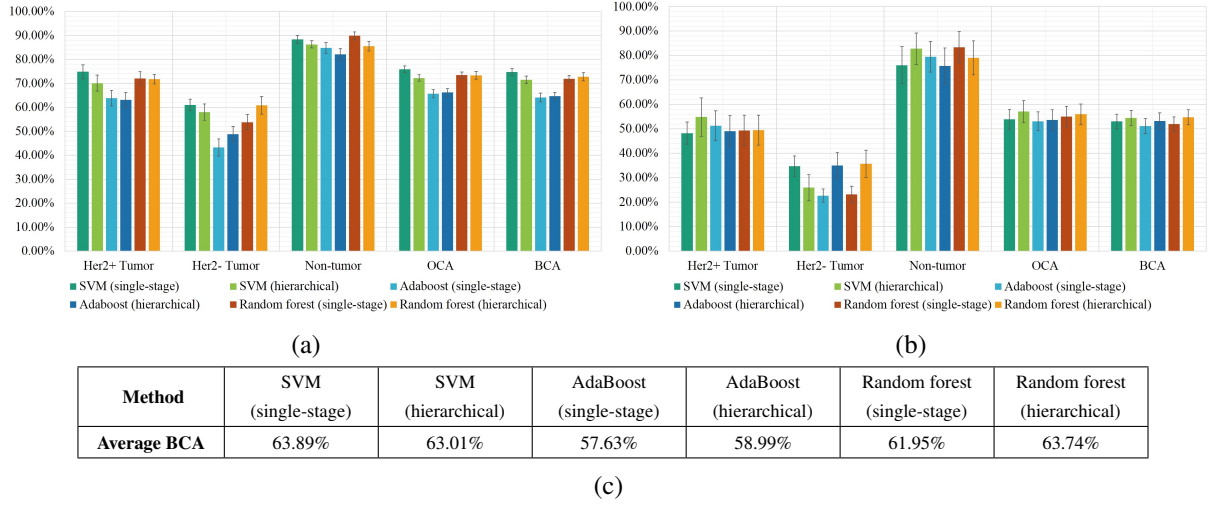


Fig. 8.14 Overall mean and standard error of classification accuracy for region classification experiments for only vertex identities using (a) k-fold stratified shuffled split and (b) leave-a-patient-out cross validation (c) Average BCA for all methods over both cross validations.

shuffled split (average $OCA = 75.87\%$, $BCA = 74.77\%$) and leave-a-patient-out (average $OCA = 53.90\%$, $BCA = 53.02\%$) cross validations followed by random forest hierarchical classification. It is surprising to see that only the seven features related to the tissue composition of each image tile are capable of generating a good classification performance, reflecting a high importance of tissue composition for such discrimination. These features show the potential to be more powerful and improve accuracy when combined with the edge-based and size-based global graph features of a few variants of the cell nuclei ARGs, as it will be observed in subsequent experiments. The performance of this method is lower than raw vertex attributes (especially using leave-a-patient-out), mainly because additional error may be introduced during the cell nuclei classification step.

8.4.4.3 Generic Cell Nuclei ARG and Fixed r : $gARG[r_F]$

The k-fold stratified shuffled split ($k = 10$), leave-a-patient-out and overall classification performance of the $gARG[r_F]$ graph-based method is shown in Figure 8.15. The best classification performance is observed uniformly by SVM single-stage method in both cross validations, *i.e.* k-fold stratified shuffled split (average $OCA = 79.02\%$, $BCA = 77.94\%$) and leave-a-patient-out (average $OCA = 64.48\%$, $BCA = 63.61\%$), followed by SVM and random forests in hierarchical strategy respectively. As observed in experiments of all the earlier image description methods, non-tumor is the most correctly classified class, followed by HER2 positive tumor and then HER2 negative tumor. HER2 negative tumor is least correctly classified due to its intermediate properties, with highest confusion with the other tumor class. Hierarchical classification strategy shows an improvement compared to single-stage classification strategy for AdaBoost

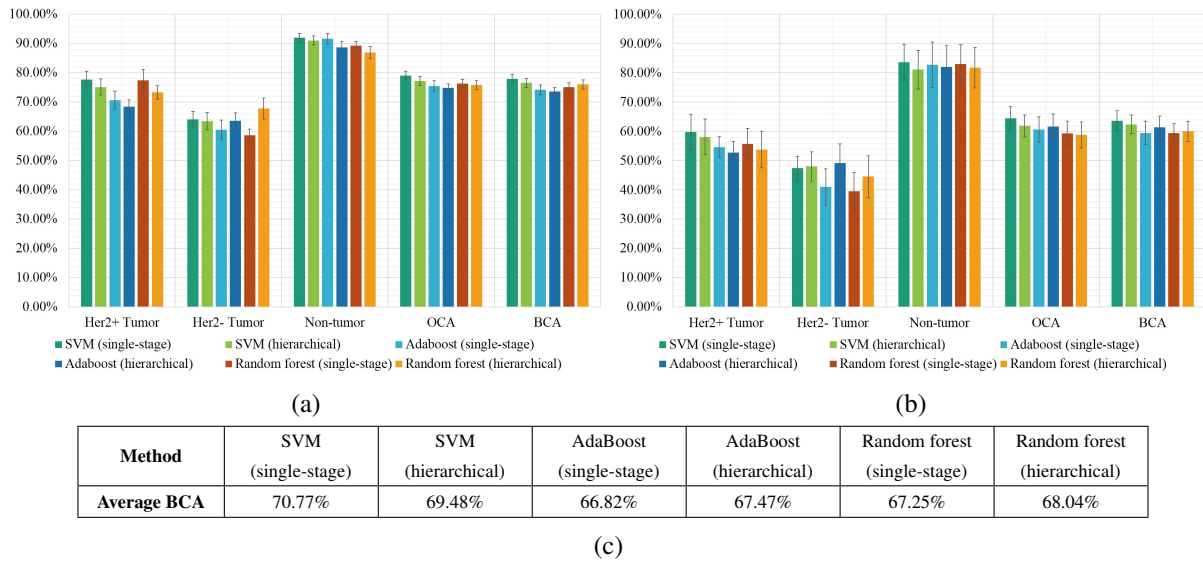


Fig. 8.15 Overall mean and standard error of classification accuracy for malignancy region classification experiments for $gARG[r_F]$ using (a) k -fold stratified shuffled split and (b) leave-a-patient-out cross validation (c) Average BCA for all methods over both cross validations.

and random forests which have the highest classification rate for HER2 negative tumor compared to other classification methods, but these are attained at the cost of higher classification error in non-tumor class. The handcrafted graph-based method has superior performance compared to only vertex attributes, which reinforces the requirement of neighborhood interactions and architectural characteristics of the tissue components.

8.4.4.4 Generic Cell Nuclei ARG and Adaptive r : $gARG[r_A]$

The k -fold stratified shuffled split ($k = 10$), leave-a-patient-out cross validation and overall performance of the $gARG[r_A]$ approach is shown in Figure 8.16. Overall, SVM single-stage and random forest hierarchical classification are most accurate, with SVM outperforming in k -fold stratified shuffled split (average $OCA = 78.90\%$, $BCA = 77.73\%$) and leave-a-patient-out (average $OCA = 60.53\%$, $BCA = 59.46\%$). Next best classification methods include random forest single-stage and both strategies using AdaBoost ensemble learning. As seen for $gARG[r_F]$ method, the random forest hierarchical classification has the highest classification rate for HER2 negative class (average 57.56%) compared to all other classification methods, however, leads to slightly lower accuracy for the non-tumor class. Furthermore, the performance is slightly lower for this approach compared to the $gARG[r_F]$ counterparts, especially for leave-a-patient-out cross validation strategy. The most probable reason for this effect is the automatic calculation of r using a uniform grid assumption, which may not always give optimum values representing tumor tissues accurately.

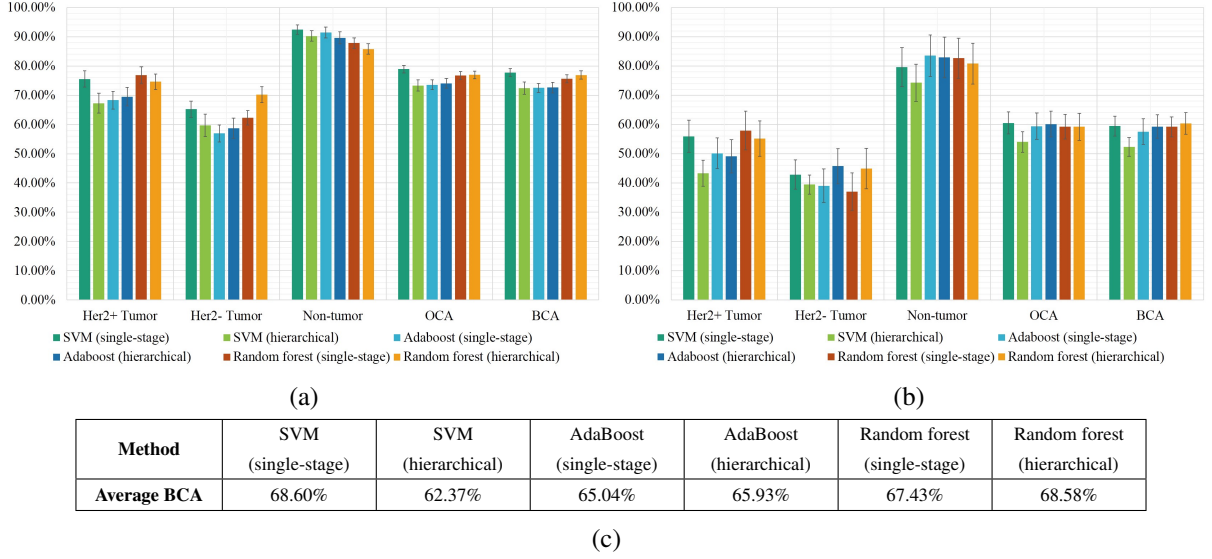


Fig. 8.16 Overall mean and standard error of classification accuracy for malignancy region classification experiments for $gARG[r_A]$ using (a) k -fold stratified shuffled split and (b) leave-a-patient-out cross validation (c) Average BCA for all methods over both cross validations.

8.4.4.5 Nuclei-specific ARG and Fixed r : $nsARG[r_F]$

The overall results for k -fold stratified shuffled split ($k = 10$), leave-a-patient-out cross validation and overall performance using the $nsARG[r_F]$ graph-based handcrafted features is shown in Figure 8.17. Out of the various classification methods, again SVM single-stage classification outperforms others in k -fold stratified shuffled split (average $OCA = 72.43\%$, $BCA = 70.73\%$) and leave-a-patient-out (average $OCA = 57.49\%$, $BCA = 53.89\%$) cross validations, followed

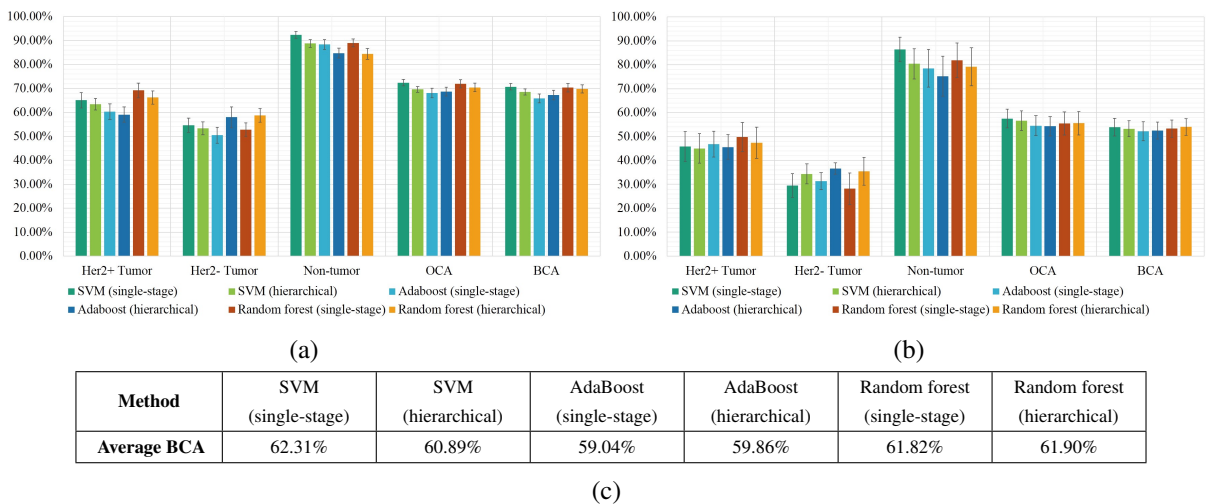


Fig. 8.17 Overall mean and standard error of classification accuracy for malignancy region classification experiments for $nsARG[r_F]$ using (a) k -fold stratified shuffled split and (b) leave-a-patient-out cross validation (c) Average BCA for all methods over both cross validations.

by random forests hierarchical classification. AdaBoost lags behind the other two machine learning methods in both strategies. The classification performance for this case is consistently and significantly lower compared to both the generic cell nuclei ARG approaches and also the nuclei composite ARG approaches observed later, which is most prominently seen in the two tumor classes. This effect can be attributed to three main reasons. Firstly, morphological and contour intensity based vertex attributes of cell nuclei have been omitted due to their prior application in cell nuclei classification step, which depletes the global graph features of useful information about individual vertices (cell nuclei) and affects decision-making process. Secondly, this ARG construction takes into account only selective interactions between the cell nuclei of each type and later combines them, which may not be as meaningful representation of tumors as compared to considering the arrangements of all types of cell nuclei in a neighborhood. Finally, the cell nuclei classification is non-ideal (as seen in previous observations) which may lead to error propagation in construction of nuclei-specific ARGs. However, from the detailed results in Appendix B.1.3, it can be seen that there is an improvement ($> 3\%$) in the classification rate for leave-a-patient-out cross validation in the non-tumor class. This is due to the decision of using imputation (with zero) for a fraction of samples for a uniform feature length, which mostly belong to non-tumor due to absence of certain type of cell nuclei, and this contributes to unique class characteristics resulting in a lower error rate.

8.4.4.6 Nuclei-specific ARG and Adaptive r : $nsARG[r_A]$

The classification performance using k -fold stratified shuffled split ($k = 10$), leave-a-patient-out cross validation and overall performance for the $nsARG[r_A]$ graph-based method is illustrated in Figure 8.18. The SVM hierarchical classification method has an overall best performance using k -

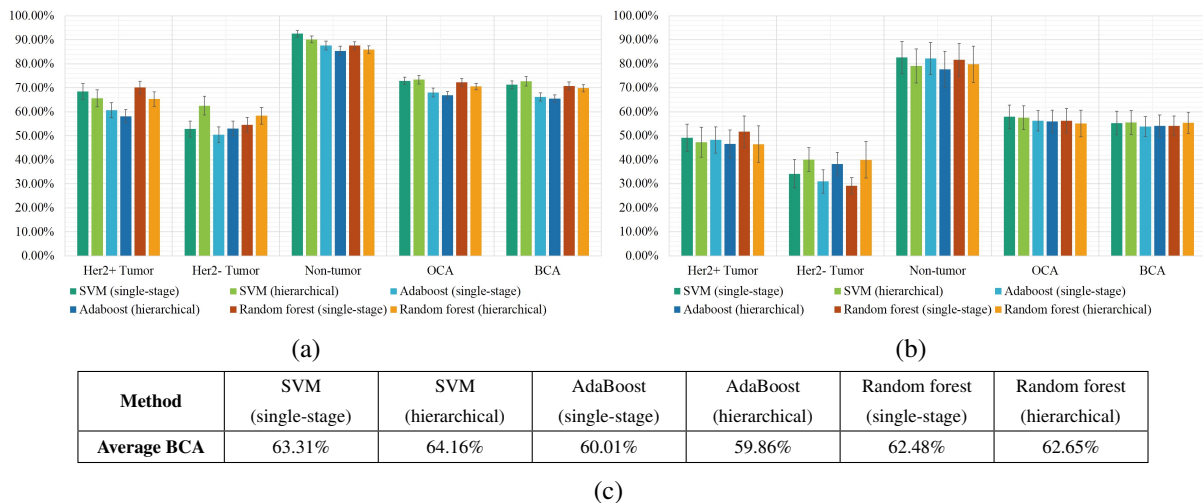


Fig. 8.18 Overall mean and standard error of classification accuracy for malignancy region classification experiments for $nsARG[r_A]$ using (a) k -fold stratified shuffled split and (b) leave-a-patient-out cross validation (c) Average BCA for all methods over both cross validations.

fold stratified shuffled split (average $OCA = 73.43\%$, $BCA = 72.81\%$) and leave-a-patient-out (average $OCA = 57.57\%$, $BCA = 55.51\%$) cross validations, especially showing improvement in classification of HER2 negative tumor class over other strategies. The performance is comparable to $nsARG[r_F]$ approach with similar degradation from generic ARG variants, due to reasons discussed for $nsARG[r_F]$. One diversion from earlier behavior is the slight improvement using adaptive r compared to fixed r for the nuclei-specific ARG construction, however, the difference is not significant for apparent reasoning. From the confusion matrices in Appendix B using hierarchical approach, the confusion between the two tumor classes in the second stage can be clearly seen higher as compared to the classes in the first stage of classification, especially for leave-a-patient-out cross validation.

8.4.4.7 Nuclei-composite ARG and Fixed r : $ncARG[r_F]$

The classification performance using k -fold stratified shuffled split ($k = 10$), leave-a-patient-out cross validation and overall performance for the $ncARG[r_F]$ method is illustrated in Figure 8.19. The SVM single-stage classification outperforms other methods in k -fold stratified shuffled split (average $OCA = 77.29\%$, $BCA = 76.03\%$) and leave-a-patient-out (average $OCA = 64.50\%$, $BCA = 63.20\%$) cross validations, closely followed by random forest hierarchical classification method. This graph-based approach shows superior results compared to the nuclei-specific construction and those comparable to the generic graph $gARG[r_F]$ approach. This is a favorable observation as lower number of global graph features are considered compared to generic cell nuclei ARG construction by excluding the object-level features as vertex attributes and considering tissue composition. The main reason for this behavior is the ability to enhance the useful information by combining individual class identities of each vertex (cell

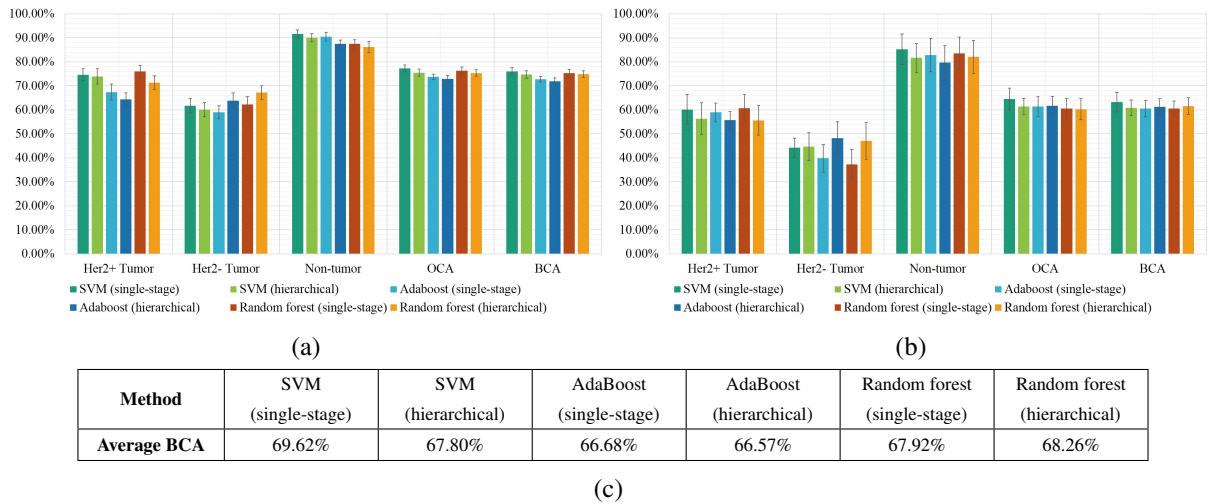


Fig. 8.19 Overall mean and standard error of classification accuracy for region classification experiments for $ncARG[r_F]$ using (a) k -fold stratified shuffled split and (b) leave-a-patient-out cross validation (c) Average BCA for all methods over both cross validations.

nucleus), along with the neighborhood vertex attributes, edge attributes representing spatial relational characteristics as well as size-based graph features. Hence, the $ncARG$ graphs are truly ‘composite’ due to an amalgamation of the underlying tissue composition and spatial arrangements. However, this improvement in result presents a trade-off between the computational complexity and classification performance, as it demands higher computational requirements than the generic cell nuclei ARGs due to the additional step for cell nuclei classification. The $ncARG[r_F]$ graph-based method has a better performance compared to the tissue composition combining only vertex identities without edge linking and edge attributes, that determines the high-level spatial relations between the components can enhance the derived information.

8.4.4.8 Nuclei-composite ARG and Adaptive r : $ncARG[r_A]$

The classification performance using k -fold stratified shuffled split ($k = 10$), leave-a-patient-out cross validation and overall performance for the $ncARG[r_A]$ method is shown in Figure 8.20. Again, the method has close performance using all machine learning approaches, especially it is difficult to decide between SVM single-stage and random forest hierarchical methods, both showing upper hand in classification accuracy compared to the other listed methods. The marginally higher random forest hierarchical accuracies in k -fold stratified shuffled split (average $OCA = 74.80\%$, $BCA = 74.95\%$) and leave-a-patient-out (average $OCA = 59.21\%$, $BCA = 59.82\%$) cross validations depict that there is degradation of performance in the nuclei composite construction as compared to fixed r , which is consistent with the generic ARG results, and strengthens the premise that empirically selected r kept fixed for all types of regions leads to more suitable edge linking compared to a dynamic linking rule.

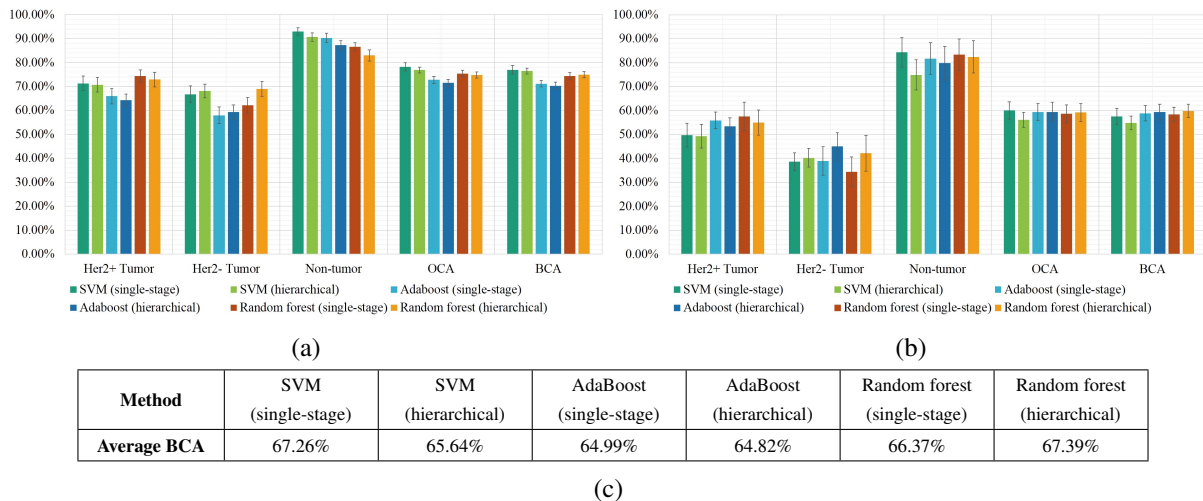


Fig. 8.20 Overall mean and standard error of classification accuracy for region classification experiments for $ncARG[r_A]$ using (a) k -fold stratified shuffled split and (b) leave-a-patient-out cross validation (c) Average BCA for all methods over both cross validations.

8.4.4.9 Nuclei-composite ARG with Additional Vertex Attributes and Fixed r : $ncARG_{v+}[r_F]$

The overall results for k -fold stratified shuffled split ($k = 10$), leave-a-patient-out cross validation and overall performance using the $ncARG_{v+}[r_F]$ handcrafted features is shown in Figure 8.21. Among the classification methods, SVMs in both strategies are superior to random forests and

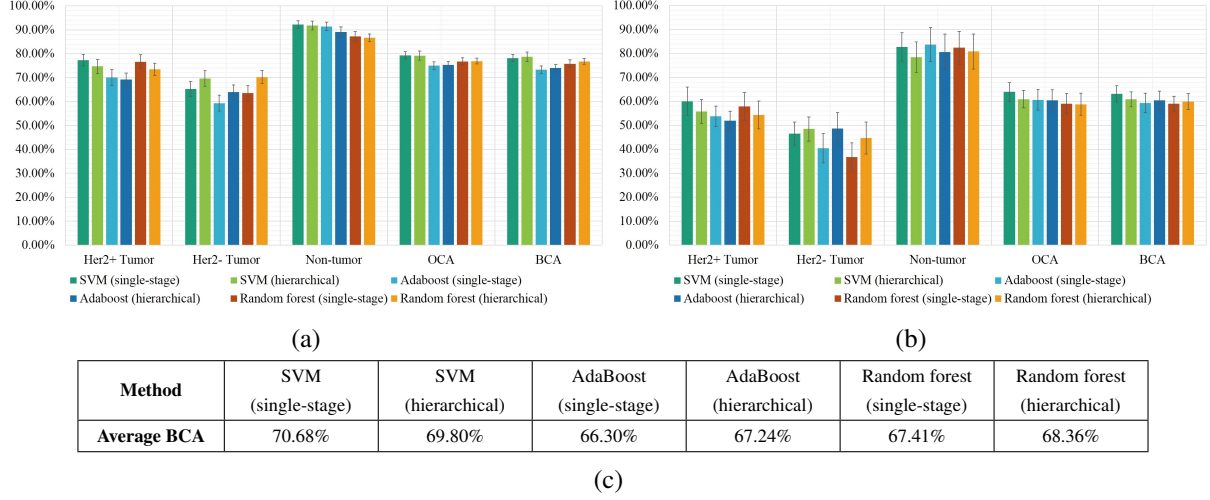


Fig. 8.21 Overall mean and standard error of classification accuracy for region classification experiments for $ncARG_{v+}[r_F]$ using (a) k -fold stratified shuffled split and (b) leave-a-patient-out cross validation (c) Average BCA for all methods over both cross validations.

AdaBoost, with SVM single-stage classification leading in k -fold stratified shuffled split (average $OCA = 79.35\%$, $BCA = 78.23\%$) and leave-a-patient-out (average $OCA = 63.94\%$, $BCA = 63.14\%$) cross validations, followed by random forest hierarchical method. The $ncARG_{v+}[r_F]$ shows an improvement ($\approx 1 - 2\%$) over $ncARG[r_F]$ with highly similar performance trends. On careful observation, it can be concluded that the additional vertex attributes already utilized in cell nuclei classification step when applied for a second time during global graph feature extraction from the cell nuclei ARG lead to only a minor enhancement in the recognition rates, hence, can be avoided without any significant loss.

8.4.4.10 Nuclei-composite ARG with Additional Vertex Attributes and Adaptive r : $ncARG_{v+}[r_A]$

The overall results for k -fold stratified shuffled split ($k = 10$), leave-a-patient-out cross validation and overall performance using the $ncARG_{v+}[r_A]$ handcrafted features is shown in Figure 8.22. SVM single-stage classification has best average results, with k -fold stratified shuffled split (average $OCA = 80.15\%$, $BCA = 79.08\%$) and leave-a-patient-out (average $OCA = 58.99\%$, $BCA = 57.51\%$) cross validations. Random forests and SVM in hierarchical strategy have the next best performance. It can be specifically observed that the leave-a-patient-out accuracies are lower for adaptive r compared to fixed r for this type of ARG

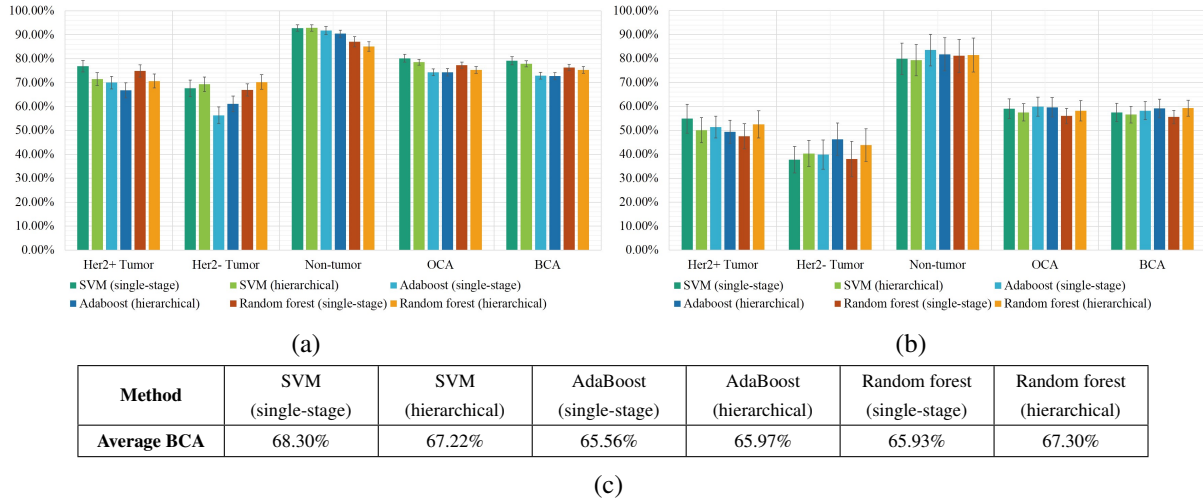


Fig. 8.22 Overall mean and standard error of classification accuracy for region classification experiments for $ncARG_v+[r_A]$ using (a) k -fold stratified shuffled split and (b) leave-a-patient-out cross validation c) Average BCA for all methods over both cross validations.

construction, which indicates that fixed r is consistently a better edge linking approach compared to an automatically determined r . There is also no significant improvement from the $ncARG[r_A]$ counterpart, even a lower leave-a-patient-out performance, again suggesting that the identities of cell nuclei segments along with global information derived from edge attributes with size based features is sufficient for similar classification performance, with the advantage of a shorter length of the feature vectors and lower computational complexity.

8.4.5 Combination of Features

In this section, the classification experiments performed using the selected subset of hybrid low-level features and the selected subset of hybrid low-level and high-level features are comparatively evaluated with respect to individual feature groups, and each other. The combined or hybrid feature sets are reduced using correlation analysis, followed by the traditional machine learning and classification.

8.4.5.1 Selected Subset of Hybrid Low-level Features

Figure 8.23 presents the k -fold stratified shuffled split ($k = 10$), leave-a-patient-out and average classification accuracy using the selected subset of hybrid low-level features. It can be noted here, that SVM hierarchical classification method outperforms other methods with k -fold stratified shuffled split (average $OCA = 93.23\%$, $BCA = 93.49\%$) and leave-a-patient-out (average $OCA = 58.58\%$, $BCA = 57.17\%$), followed by SVM and AdaBoost in single-stage strategies. Random forests are not as successful as the other two methods. As remarked in Section 7.3.4, it was expected that when automatic correlation analysis is applied, correlated but meaningful features may also be removed during feature selection process, and this leads to a slight decline in

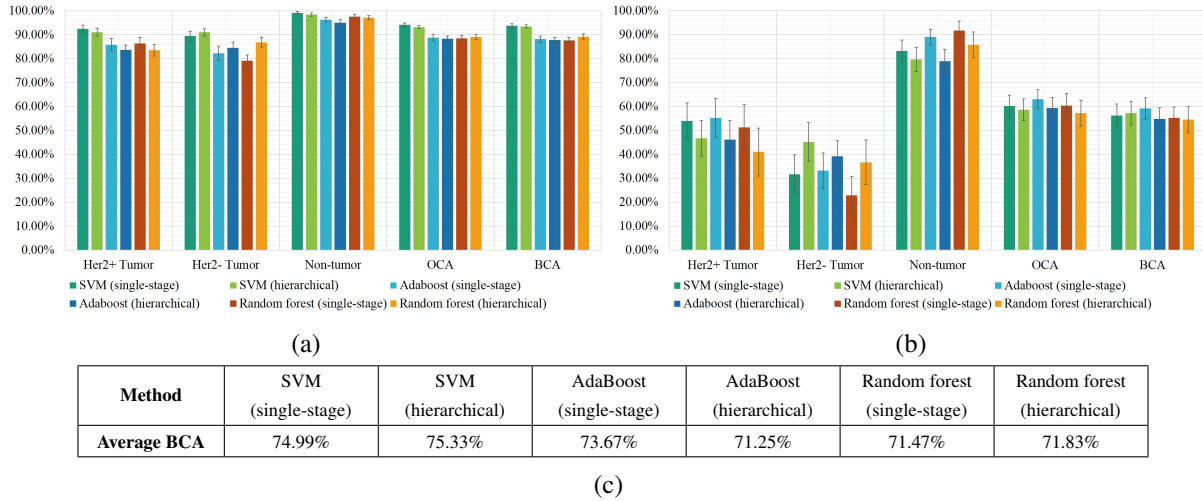


Fig. 8.23 Overall mean and standard error of classification accuracy for cancer classification experiments for selected subset of hybrid low-level features using (a) *k*-fold stratified shuffled split and (b) *leave-a-patient-out* cross validation (c) Average BCA for all methods over both cross validations.

classification accuracy compared to the best performing individual methods. This effect is evident while using the selected subset of hybrid low-level features, which compare favorably to almost all the low-level state-of-the-art features, however, have slightly lower performance compared to the most successful RGB histograms, especially in the *leave-a-patient-out* cross validation. Similar phenomenon was observed in [Sharma 2016], where a comparison between hybrid features before and after correlation analysis has been drawn using AdaBoost classification, revealing that correlation analysis leads to a slightly lower performance due to a reduced feature set. Such a comparison is not made for the configurations described here, as the length of feature vectors will be extremely long without correlation analysis that becomes unmanageable with extensive time requirements. Hence, it can be concluded that, correlation based-feature extraction can be primarily applied to reduce the length of combined feature vectors, without significantly compromising the classification performance.

8.4.5.2 Selected Subset of Hybrid Low-level and High-level Features

The *k*-fold stratified shuffled split ($k = 10$), *leave-a-patient-out* and average classification accuracy using the selected subset of hybrid low-level and high-level features is shown in Figure 8.24. Among the machine learning methods, the SVM method is most successful in both strategies, followed by AdaBoost due to a larger number of image descriptors, and then random forests. The SVM single-stage classification method has the highest rates with *k*-fold stratified shuffled split (average *OCA* = 92.26%, *BCA* = 91.67%) and *leave-a-patient-out* (average *OCA* = 64.10%, *BCA* = 61.09%). The average balanced classification accuracies for the combination of high-level and low-level features are consistently higher for all traditional machine learning methods compared to those for the combination of only low-level features. This observation is strongly evident of the potential of object-based, graph-based, structural and architectural information in histopathological image analysis, and proves that the handcrafted

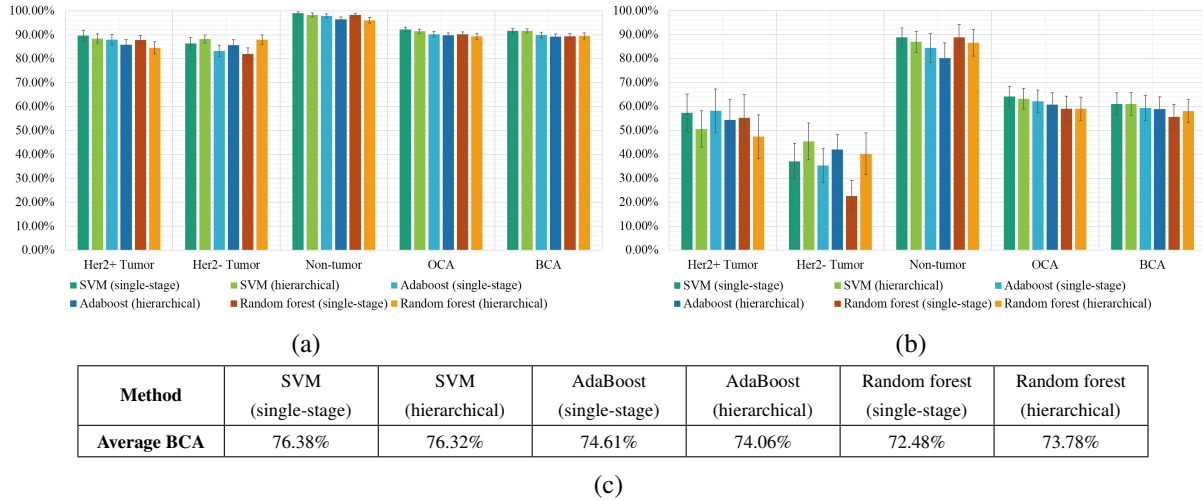


Fig. 8.24 Overall mean and standard error of classification accuracy for cancer classification experiments for selected subset of hybrid low-level and high-level features using (a) k -fold stratified shuffled split and (b) leave-a-patient-out cross validation (c) Average BCA for all methods over both cross validations.

high-level features lead to additional performance enhancement over the state-of-the-art low-level features. Furthermore, the improvement is more significantly shown in the leave-a-patient-out cross validation, suggesting higher robustness of the hybrid subset for high-level and low-level features towards inter-patient variations. Although, a few configurations with individual handcrafted graph-based methods such as $ncARG[r_F]$ and $ncARG_{v+}[r_F]$ show higher leave-a-patient-out accuracies compared to hybrid features, the degradation mainly attributed to information loss during feature selection.

8.4.6 Deep Learning Methods

Using deep learning methods, firstly, several CNN architectures are empirically investigated to determine a suitable design for learning significant characteristics of histological images. A preliminary CNN architecture is proposed and applied for two classification problems, namely, cancer classification based on immunohistochemical response and necrosis detection. The performance of the designed CNN is quantitatively evaluated and compared with the AlexNet CNN framework which is popular in deep learning for general object classification. Ensemble of the two CNNs is also quantitatively analyzed.

The overall results for k -fold stratified shuffled split ($k = 3$) and leave-a-patient-out cross validation using the three deep learning methods, namely, AlexNet CNN framework, proposed CNN architecture and ensemble of CNNs, is comparatively illustrated in Figure 8.25.

It can be noted that using all the three methods, accuracies are higher when trained with the entire available WSI dataset, unlike the initial empirical experiments. This behavior was expected because of subsequent use of augmented datasets built from the entire available WSI data, hence, the networks are trained with large-scale visual information with diverse examples to classify

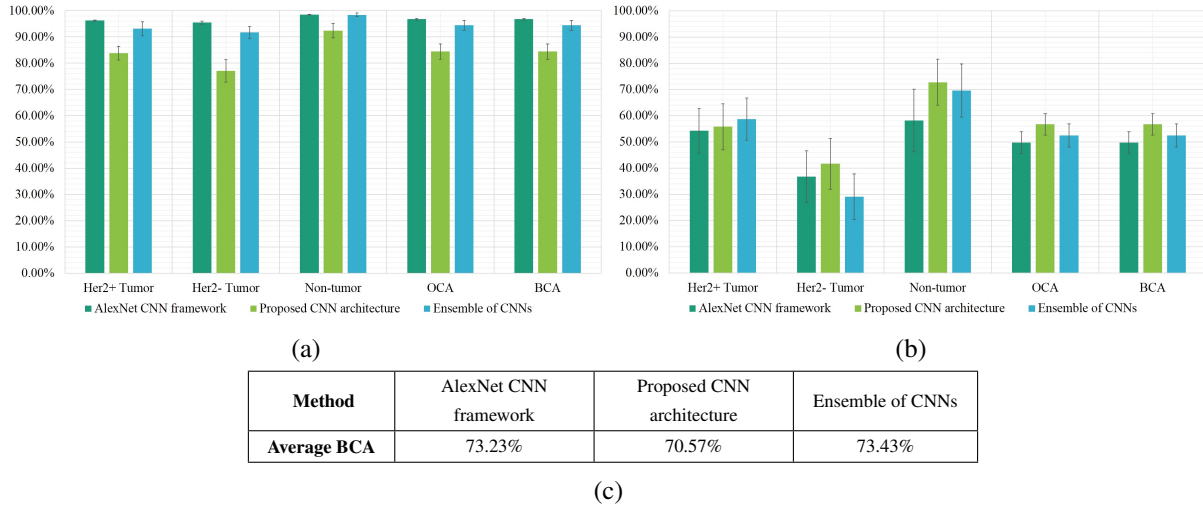


Fig. 8.25 Overall mean and standard error of classification accuracy for region classification experiments for deep learning methods using (a) *k*-fold stratified shuffled split and (b) leave-a-patient-out cross validation (c) Average BCA for all methods over both cross validations.

the unknown images more effectively. Also, overall and balanced classification accuracies are identical due to uniform distribution of samples in the three classes.

From the above graphical representations, on comparing the two CNN architectures, it can be clearly observed that the *k*-fold stratified shuffled split performance of AlexNet (average $OCA = BCA = 96.74\%$) is superior to the proposed CNN (average $OCA = BCA = 84.41\%$) in contrast to the leave-a-patient-out where the observation is reversed with the proposed CNN performing better (average $OCA = BCA = 56.72\%$) than the AlexNet CNN framework (average $OCA = BCA = 49.73\%$). In the *k*-fold stratified shuffled split cross validations, samples in the training and validation dataset may belong to the same patient, hence, superior performance of AlexNet CNN framework is not a clear indicative of a more suitable model for the described problem. On the other hand, a better leave-a-patient-out performance is a more recommendable characteristic for practical application of the method in a clinical setting where unknown samples need to be automatically classified with trained networks of previous cases, and this requirement is better achieved using the proposed CNN architecture. The higher leave-a-patient-out performance of the proposed CNN suggests its higher robustness to biological variations among patients, with lower standard deviations, especially with a desirable improvement in the classification accuracy of HER2 positive and HER2 negative tumor. Its per-class and overall classification accuracies also correspond to the other results obtained using handcrafted and state-of-the-art features in traditional machine learning, thereby, reducing the possibility of any anomaly, which may have occurred during AlexNet training in *k*-fold stratified shuffled split classification due to a high numerical gap in the two types of cross validation accuracies.

When ensemble of CNNs is compared to individual CNNs, it is seen that intermediate results are obtained in both types of cross validations, with *k*-fold stratified shuffled split (average

$OCA = BCA = 94.40\%$) and leave-a-patient-out (average $OCA = BCA = 52.46\%$), leading to an overall best performance by slightly outperforming the individual methods. This is due to the ability to reduce individual errors of each network, by averaging the output of both networks and causing performance enhancement. However, in few leave-a-patient-out experiments, the ensemble shows a decrease in accuracy compared to individual methods due to borderline Softmax probabilities associated with the correctly predicted class, leading to misclassification on averaging. This decline is mainly observed for the HER2 negative tumor class.

From the box and whisker diagrams in Appendix B.2.3, it is evident that for a few difficult WSI during leave-a-patient-out classification, the CNNs completely fail to classify either HER2 positive, HER2 negative or non-tumor class, hence, showing high variations in classification accuracy between cross validation rounds. This effect is more prominent for two to three specific WSI of the eleven WSI. Such low per-class accuracy reflects on a negative aspect of the deep learning methods compared to traditional machine learning methods, where such an effect is seldom observed. Due to complicated abstractions in the CNN layers, the cause is not fully explainable, most probably, the data representations in the CNN do not lead to an appropriate learning process during the leave-a-patient-out configuration for a few patients, due to high inter-patient variability. The most feasible solution is a richer dataset from larger number of patients.

8.4.7 Overall Observations

In summary, on carefully investigating the experimental results, it can be concluded that the proposed methods including the handcrafted graph-based features with traditional machine learning, and the novel self-designed CNN architecture provide promising results that compare favorably with established methods for the computer-based analysis of H&E stained histopathological images of gastric cancer. Overall observations showing comparisons based on different criteria is provided in the following section.

8.4.7.1 Performance Summary: Class-wise and Evaluation Method-wise Comparison

The average performance of all the image analysis methods in the cancer analysis stage, with respect to the different cancer categories in the two types of cross validation evaluation methods is shown in Figure 8.26.

It has been observed in the experimental results earlier, that the per-class accuracies using the six traditional machine learning strategies and deep learning methods show similar trends between the three types of malignancy levels. From Figure 8.26, it is evident that the higher-level classes tumors and non-tumor show the most superior prediction rates compared to the two tumor subclasses. The class-wise results indicate that it is comparatively easier to distinguish non-tumor class (average 88.43%) from both the tumor classes (average 91.60%) by the trained classifiers due to distinct visual appearances, however, more difficulty is experienced to discern between the two tumor types, namely, HER2 positive tumor and HER2 negative tumor. HER2 positive

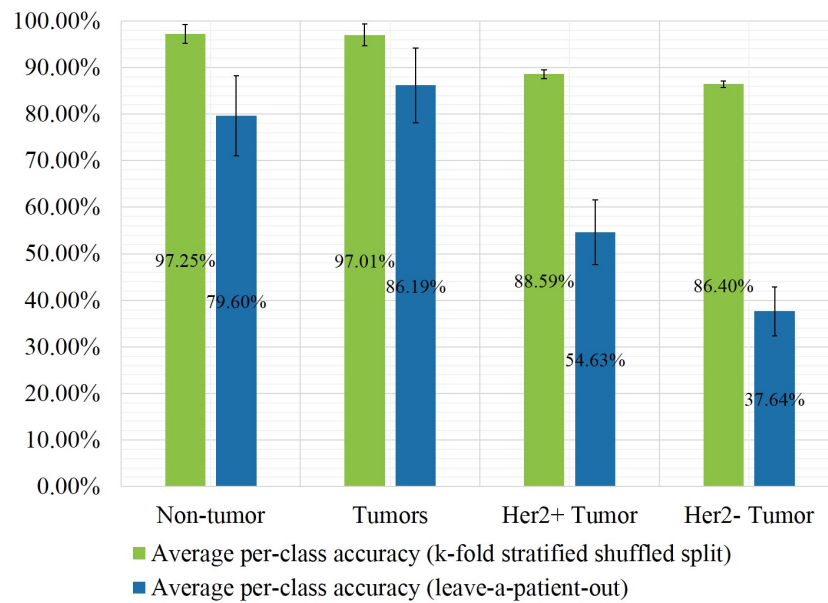


Fig. 8.26 Overall mean and standard error of per-class classification accuracy for all the methods over both cross validations

tumor has intermediate recognition rates (average 71.61%) and discrimination from other two tumor types. HER2 negative tumor has the least accurate classification (average 62.02%) for all methods, that was expected as HER2 negative tumor represents tissue regions where tumor is partially developed, hence there is also less or no immunohistochemical response from cells lying in these regions. This class denotes the beginning of tumor proliferation or an intermediate stage between no malignancy (non-tumor) and high malignancy levels (HER2 positive tumor). Visually, this tumor class has a blurred definition compared to the other two malignancy levels in H&E stain, and shows highest classification confusion with the HER2 positive tumor type, (observed in confusion matrices in Appendix B.1.3). Resolving this confusion was proved as the main challenge to obtain good identification rates for both the tumor classes in H&E stained tissue images.

Moreover, while comparing the two types of evaluation methods, it is observed that the k-fold stratified shuffled split cross validation results are consistently more promising compared to leave-a-patient-out cross validations. In the k-fold stratified shuffled split cross validation, variations in entire sample population have been utilized for a better performance. Hence, these results ensure highly desirable outputs in scenarios where partially annotated data in single or multiple large-sized WSI of same patient is available for training purpose. In contrast, a lower leave-a-patient-out performance, mostly seen for the two tumor subclasses, is predominantly due to the biological variation between the limited number of available cases, which is further enhanced by inter-WSI heterogeneity in stain intensity and malignancy levels. It leads to different appearance-based characteristics for each tissue, thus, is more difficult for the classifiers to learn characteristics of the patient being classified without any prior visual knowledge specific to them. The effect has been more prevalent in a few (2-3) difficult WSI that are observed to show

poor recognition rates, which adversely influences the overall classification measurements of the system using all the described traditional machine learning and deep learning methods. This inter-patient variability also accounts for the higher standard deviations of classification accuracy compared to the other type of cross validation, that can be further noted in the box and whisker diagrams in Appendix B.2.3. In order to improve leave-a-patient-out accuracies, learning needs to be based on more extensive WSI datasets of diverse cases, which are currently limited.

8.4.7.2 Performance Summary: Feature-wise and Strategy-wise Comparison in Traditional Machine Learning

The magnified plots of the overall classification performance of the various feature extraction approaches followed by traditional machine learning methods is provided in Figure 8.27. The resulting line diagram shows the average balanced classification accuracies using both types of cross validations, for 21 different configurations during comparative evaluation, in the order given in Table 8.2. This chart emphasizes on the relative performance of each machine learning strategy with respect to the characteristics of each feature group, and typical trends are observed.

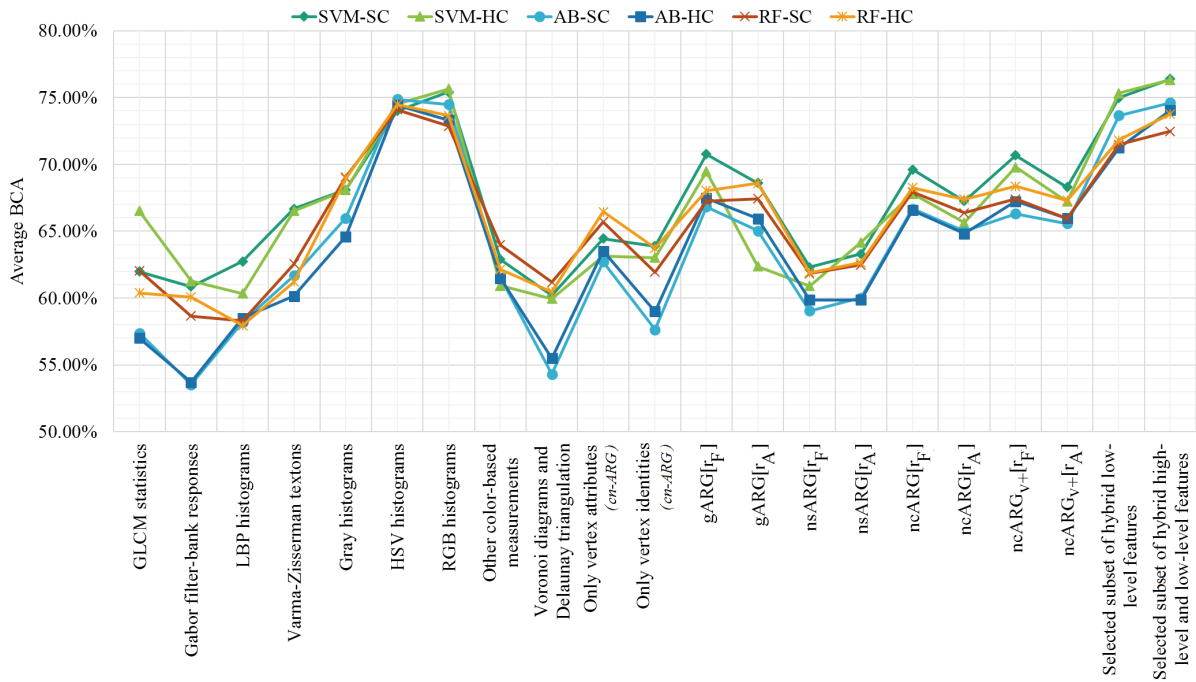


Fig. 8.27 Magnified line diagram showing average BCA for the different feature groups and corresponding traditional machine learning strategies

First and foremost, all the individual curves (each representing a traditional machine learning strategy) have similar trends over the different feature extraction methods. This observation proves that classification performance is more dependent to the suitability of the selected image description method to represent the histological images effectively, as compared to the choice of the traditional machine learning method. Secondly, it can be summarized that, as already

observed in the sections showing experimental results for individual configurations, peaks can be seen in a few feature groups, especially, the hybrid approach for combination of knowledge using low-level state-of-the-art and high-level handcrafted features exhibit the highest prediction performance, followed by hybrid low-level features, state-of-the-art color histograms (RGB and HSV) and high-level handcrafted graph-based features using $gARG[r_F]$, $ncARG_{v+}[r_F]$ and $ncARG[r_F]$ variants of the cell nuclei ARG. Hence, the proposed novel cell nuclei ARG graph-based method compares favorably to most of the state-of-the-art feature extraction approaches. High-level state-of-the-art features using Voronoi diagrams and Delaunay triangulation don't perform satisfactorily, further, $nsARG$ variant of the cell nuclei ARG is observed to be less suitable for the image analysis of gastric cancer regions.

Additionally, from the Figure 8.27, it is difficult to say which of the two classification strategies, namely, single-stage classification or hierarchical classification has led to a better classification. There cannot be certain answer as both the strategies perform comparable, however, hierarchical strategy for two-stage categorization is more successful than the single-stage differentiation mainly of the two tumor classes. Among the traditional machine learning methods, SVM in single-stage strategy and random forest in hierarchical strategy tend to be most superior, and their other counterparts have an intermediate performance. AdaBoost ensemble learning in both strategies is seen inferior to other methods. Specifically, AdaBoost algorithm shows an enhancement compared to other methods with a larger number of extracted features, prominently seen using the hybrid feature sets and the two color histograms, and opposite effect is seen for the smaller feature vectors, for example, GLCM statistics and Gabor filter-bank responses. Such a dependence to the number of derived features is less obvious for SVM and random forest machine learning methods, making them relatively more reliable than AdaBoost.

8.4.7.3 Performance Summary: Method-wise Comparison of Traditional Machine Learning and Deep Learning

As observed in Figure 8.27, the three traditional machine learning methods show similar behavior with few variations for each feature configuration. This observation for traditional methods enforces that designing meaningful features to effectively describe the histological images for estimating appearance-based characteristics of the tissues including texture, color, structure, morphology, architecture and topology, are more significant to accurately discriminate between the cancer categories instead of the selected machine learning method, and careful parameter selection of classifiers can further lead to performance optimization.

An overall comparison of the machine learning methods by averaging the corresponding balanced classification accuracy in the traditional route (with hybrid features) and deep learning route is shown in Figure 8.28 to facilitate a comparison between the two routes.

Among the traditional machine learning methods, SVM shows the most favorable performance over AdaBoost ensemble learning and random forests. A superior performance of the SVM method proves that the derived image descriptors are relevant and good quality, a prerequisite for this machine learning algorithm to perform well. The three deep learning methods

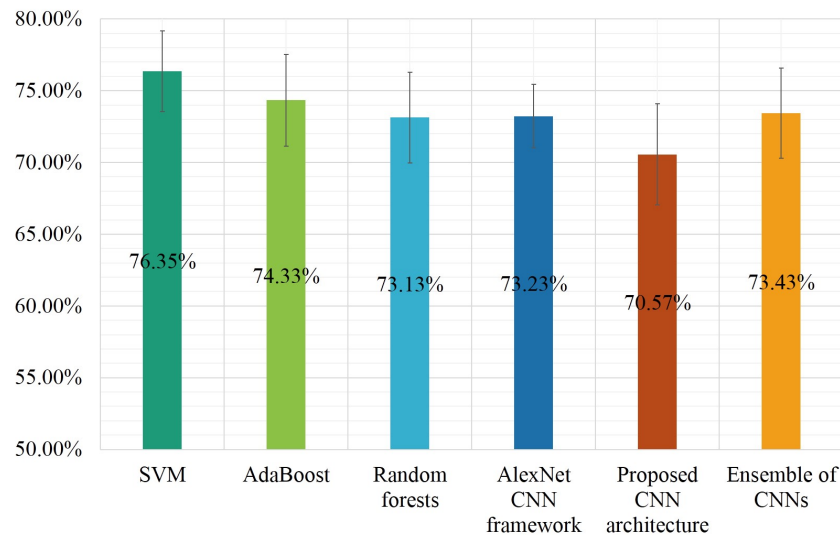


Fig. 8.28 Overall mean and standard error of balanced classification accuracy for traditional machine learning (using hybrid features) and deep learning methods

also have comparable recognition rates, where AlexNet framework has slight better performance than proposed CNN architecture and ensemble learning outperforms both the individual CNN architectures. The proposed CNN architecture shows higher robustness to AlexNet due to a better leave-a-patient-out performance both in empirical analysis (Table 7.4.2.5) and experimental evaluation (details in Section 8.4.6).

On comparing the traditional versus deep learning methods, very small difference is observed between the quantitative performances. SVM has best overall prediction rates followed by AdaBoost, ensemble of CNNs, AlexNet CNN framework, random forests and the proposed CNN architecture. So, it can be concluded that the three deep learning methods have a comparable performance to handcrafted features and traditional machine learning methods, and can adequately represent H&E stained histopathological images of gastric cancer without the requirement of hand-engineered features. The significant overall observation suggests that the choice of suitable image description among the traditional methods, and careful network architectural design among deep learning methods are important requisites for successful realization of the image analysis objectives in digital histopathology.

One interesting observation is that some traditional machine learning methods perform better than deep learning methods for the cancer classification problem. On the other hand, for necrosis detection, deep learning outperforms SVM traditional machine learning. This can be explained by two main reasons. Firstly, as necrosis detection is a textural problem, low-level features using GLCM statistics and Gabor filter-banks have been extracted for the same, however, the results reported for cancer classification in Figure 8.28 are based on combined feature sets using a more comprehensive high-level and low-level information. If the same low-level textural features used in necrosis detection are applied for cancer classification, the performance of traditional machine learning methods will be lower compared to deep learning methods, and

has been already observed in Figure 8.27. Such high-level image description is not possible in necrosis detection as cell nuclei shrink or disappear in necrotic tissues, making segmentation and cell nuclei ARG construction very difficult. Secondly, image tile size in the cancer analysis stage for traditional machine learning is larger (1024×1024 pixels) compared to deep learning (512×512 pixels), but identical (512×512 pixels) for necrosis detection. This larger size for automatic cancer analysis is an advantage to traditional methods due to availability of higher context information, and may have contributed to better performance of SVM and AdaBoost methods. This effect demonstrates the potential ability of deep learning methods and suggests possible performance improvement with future enhancement of GPU capabilities such that deeper network architectures can be investigated with larger image sizes.

8.5 Summary

In this chapter, the results and observations for the conducted experiments in image pre-analysis stage, namely, necrosis detection and cell nuclei classification, and those for analysis of cancer regions using handcrafted features and traditional machine learning methods as well as deep learning methods, are described in detail.

For the pre-analysis stage, results are usually satisfactory for both necrosis detection and cell nuclei classification. The deep learning methods have a favorable performance for automatic detection of necrosis. Similarly, for classifying the cell nuclei segments, handcrafted features with random forests show reasonable detection rates, though lower than necrosis detection due to a more complex multi-class problem.

For the analysis of cancer regions, more favorable performance is achieved using *k-fold stratified shuffled split* compared to *leave-a-patient-out* cross validation. Among the traditional methods, high-level handcrafted graph-based features outperform majority of the state-of-the-art features, and a hybrid feature set after combining low-level and high-level information shows most desirable accuracies. Furthermore, the SVM method has overall superior results than random forests and AdaBoost. The deep learning methods compare favorably to traditional methods, with ensemble of CNNs showing best performance followed by the individual approaches, and self-deigned CNN architecture is found more robust than AlexNet CNN framework.

Statistical Analysis of Stains: Observations and Challenges

Contents

9.1	Introduction	183
9.2	Visualizing Image Statistics: HER2 Stain	184
9.2.1	Visual Inspection	184
9.2.2	Feature Heatmaps	185
9.3	Visualizing Image Statistics: H&E Stain	186
9.3.1	Visual Inspection	186
9.3.2	Feature Heatmaps	187
9.4	Summary	189

9.1 Introduction

In this chapter, image data and feature spaces in the HER2 and H&E stains are comparatively visualized in order to obtain clearer insights to the observations of the conducted experiments. Firstly, a representative image dataset in the HER2 stain is generated in a way similar to the H&E datasets, corresponding to the three malignancy levels based on immunohistochemical response. A set of low-level texture and intensity-based features are extracted from HER2 images and their heatmaps are generated for a better visualization among the three malignancy levels. This process is repeated for H&E images along with additional heatmaps for the high-level state-of-the-art and handcrafted features. The observations of this section assist in further understanding of the image statistics in both HER2 and H&E stains, the related complexity of the described problem and the challenges faced to develop reasonable solutions in the H&E stain.

9.2 Visualizing Image Statistics: HER2 Stain

In order to visualize the nature of images in the three malignancy levels for HER2 stain, a set of five image tiles per class per patient are selected, totaling to 165 image tiles for the WSI dataset. For a fair comparison, the HER2 stained WSI are also tessellated at 1024×1024 pixel size with $40\times$ objective magnification, similar to the H&E images used in traditional machine learning methods. Firstly, a visual inspection of available ground truth WSI data is performed in order to analyze the regions in the defined malignancy levels. Following this, state-of-the-art low-level feature extraction methods, namely, GLCM statistics, LBP histograms, HSV histograms and RGB histograms, are applied to the images, and corresponding feature heatmaps are generated. As cell nuclei segmentation and subsequent graph-based methods are developed specifically for the H&E images, hence, the high-level graph-based feature extraction is not feasible for the HER2 image dataset. Cell nuclei segmentation is much easier in H&E stain compared to HER2 stain, due to the characteristic properties of the former stain, which distinctly marks the cell nuclei along with their boundaries in blue color and the cytoplasm in pink color, so the cell nuclei can be separated from the remaining tissue regions. Such nuclei segregation is not observed using HER2 stain due to different staining properties, thus, developing an automatic cell nuclei segmentation procedure is highly challenging for the HER2 stain and outside the scope of the undertaken image analysis goals focusing on the H&E stain.

9.2.1 Visual Inspection

The HER2 immunohistochemically stained WSI image data is carefully observed using visual inspection of the available ground truth. A few representative examples of image tiles from the three predefined malignancy levels in the HER2 stain are shown in Figure 9.1. On visually exploring the WSI data in the three malignancy classes, distinct characteristics of the tissue regions can be observed as follows. The tissue areas in **non-tumor** class contain smaller and rounder cells of dark gray or bluish color, scattered in a gray or off-white intercellular fluid with a meshy appearance. In **HER2 positive tumor** regions, cells are usually larger, elongated and oval-shaped, are stained light blue-gray with dark blue nuclei and one or more cells are surrounded by brown bubble-like structures. Intercellular fluid is mostly dark brown in color, and intercellular distance is very less due to overlapping cells. Furthermore, cells arrange in dense patterns, and show mitosis as they contain more than one cell nucleus. Intracellular fluid can also be seen inside the cells, that is granular and gray-blue with white spaces in between and cell boundaries are mostly not distinct. The **HER2 negative tumor** class consists of light gray cells with dark gray nucleus in a gray and granular intercellular fluid. The most evident difference of this category with the HER2 positive tumor is the *absence of brown colored regions*. Other properties such as intercellular distance, cell size, mitosis occurrence, cellular patterns and arrangements, and cellular density are intermediate in nature compared to the other two classes.

Therefore, visual inspection of the three tumor types itself suggests that the discrimination between these regions in HER2 stain is a fairly simple task due to mostly unique visual charac-

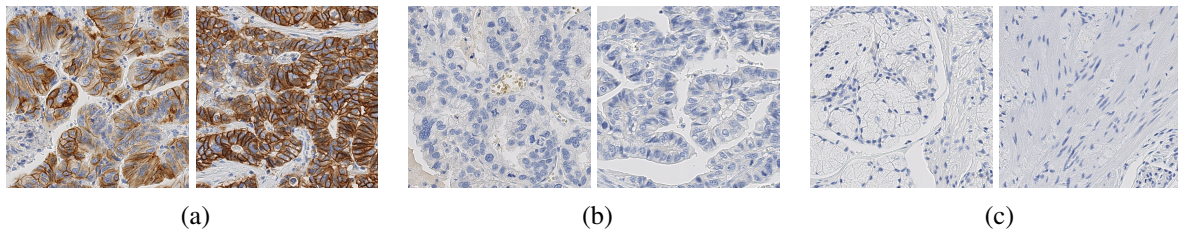


Fig. 9.1 Representative examples of images in three malignancy levels using HER2 stain (a) HER2 positive tumor (b) HER2 negative tumor (c) Non-tumor.

teristics. Also, the computer-aided categorization of malignancy levels in HER2 stain can be achieved with good precision using the established low-level feature extraction methods, further demonstrated in the next section.

9.2.2 Feature Heatmaps

The feature heatmaps generated for HER2 representative dataset corresponding to their GLCM statistics, LBP histograms, HSV histograms and RGB histograms are shown in Figure 9.2. Blue, yellow and pink represent the classes non-tumor, HER2 positive tumor and HER2 negative tumor respectively, and each column represents one feature vector corresponding to each sample. The numeric values of the variables are represented from green to red representing smallest to largest.

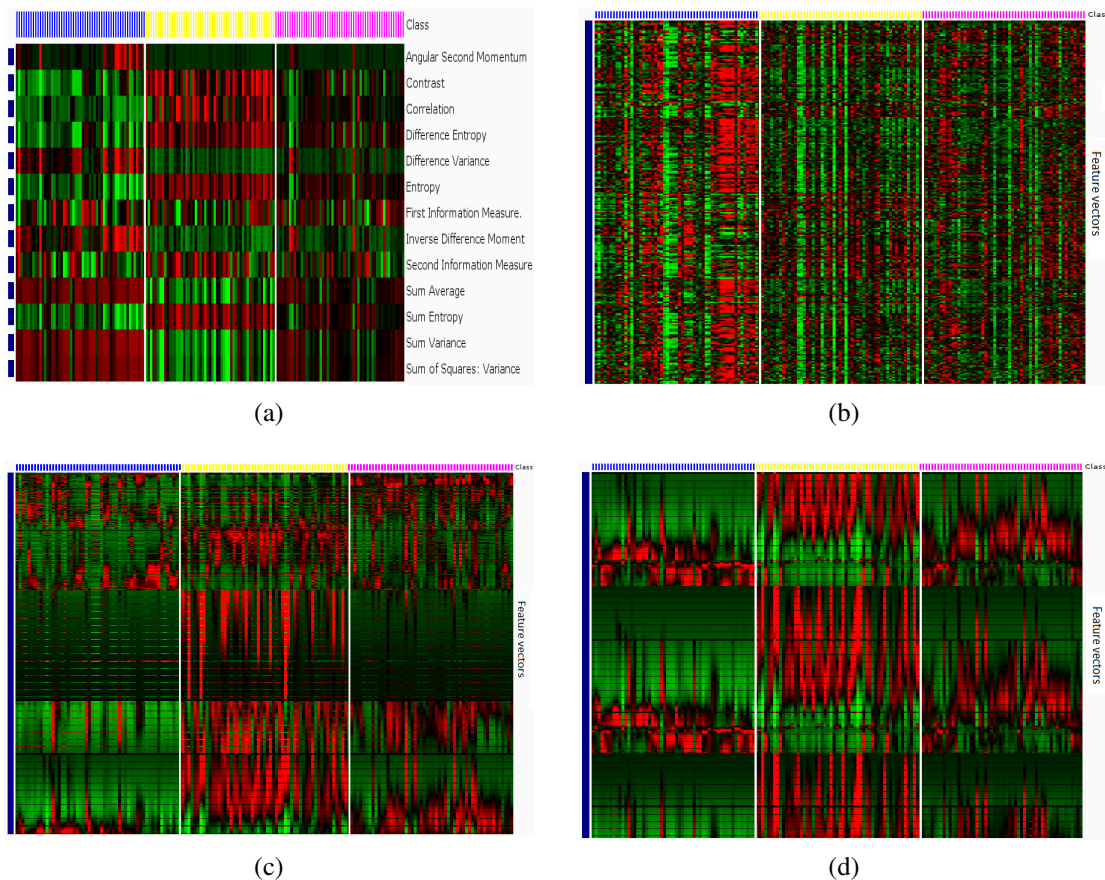


Fig. 9.2 Heatmaps of the extracted feature vectors showing image statistics in the HER2 stain (a) GLCM statistics (b) LBP histograms (c) HSV histograms (d) RGB histograms.

It can be easily observed from the feature heatmaps for the extracted low-level features, that clear distinctions are visible among the image statistics for the three malignancy levels and show clusters of feature vectors of the same malignancy class. A higher similarity can be seen between non-tumor and HER2 negative tumor classes due to absence of the distinct HER2 staining of the tissue, as the HER2 immunohistochemical stain is most prominent in the HER2 positive tumor regions. These feature maps also confirm the analysis during visual inspection. This may lead to the conclusion that the described problem of computer-based classification for analysis of cancer regions is much easier to solve in the original HER2 stain compared to H&E stain, and will result in higher classification performance using the described traditional machine learning and deep learning methods.

9.3 Visualizing Image Statistics: H&E Stain

The final working datasets used for experiments with traditional machine learning methods are also utilized for the visualization purpose, by using visual observation and generation of corresponding feature heatmaps for low-level state-of-the-art features including GLCM statistics, LBP histograms, HSV histograms, RGB histograms and Varma-Zisserman textons, high-level state-of-the-art features *i.e.* Voronoi diagram and Delaunay triangulation, high-level handcrafted cell nuclei ARG-based features and the selected subset of hybrid features.

9.3.1 Visual Inspection

A few representative examples of image tiles from the three predefined malignancy levels in the H&E stain are shown in Figure 9.3. On the visual inspection of image regions in the H&E stain, the following is observed. The **non-tumor** regions consist of smaller cells mostly stained in dark purple color with deep pink intercellular spaces. Cellular density is lower as cells are scattered in the connective tissue with a higher intercellular distance and no distinct arrangements. In contrast, **HER2 positive tumor** regions consist of deeply stained large and elongated cells with a very high cellular density and low intercellular distance, gathered in groups and arranged densely with highly overlapping cell boundaries. Sometimes, only cell nuclei are stained dark purple with a granular light purple cytoplasm, and the intercellular spaces are largely light-to-medium purple. Some cells also show mitosis with multiple nuclei. Most of these observations are similar to those observed in HER2 stain for both classes, with the exception of absence of *distinct staining* of HER2 positive regions. Moreover, **HER2 negative tumor** regions usually show intermediate properties, however, unlike in the HER2 stain, there is no clear distinction with HER2 positive tumor in terms of staining of cells. Also, these areas partly contain elongated larger cells, and partly rounded smaller cells. Cellular arrangements are also beginning to arrange in patterns. The tissue composition of each type of malignancy, along with interactions between the different types of cell nuclei varies. These properties, along with other subtle changes not distinctly noticeable in H&E stain, are attempted to be represented by suitable hand-engineered

features to facilitate an accurate automatic classification using traditional machine learning methods and appropriate deep CNN architectures. An example of such correspondence between tissue malignancy levels and respective cell nuclei ARG has been established in Section 7.3.3.3 (Table 7.1).

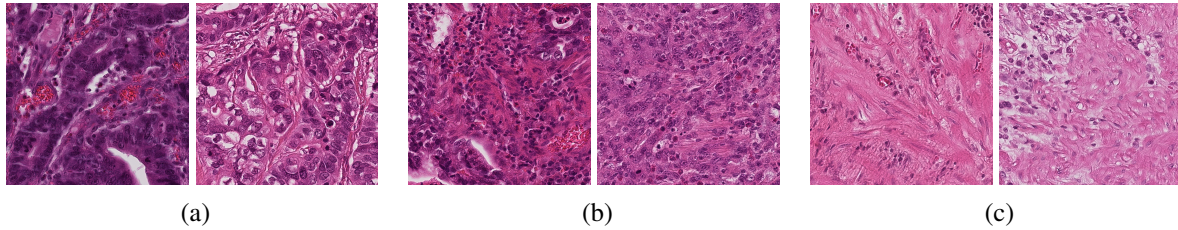


Fig. 9.3 Representative examples of images in three malignancy levels using H&E stain (a) HER2 positive tumor (b) HER2 negative tumor (c) Non-tumor.

9.3.2 Feature Heatmaps

The feature heatmaps generated for the H&E stained image dataset corresponding to their GLCM statistics, LBP histograms, HSV histograms, RGB histograms, Varma-Zisserman textons, Voronoi diagram and Delaunay triangulation based features, $nsARG[r_F]$ handcrafted graph-based features and selected hybrid feature set are shown in Figure 9.4.

From the heatmaps of H&E images using all the mentioned state-of-the-art and handcrafted features, differences in the three types of malignancy levels are visualized mainly for the non-tumor class, with higher similarities in the two tumor classes, namely, HER2 negative tumor and HER2 positive tumor. Such visualization is quite consistent with the visual inspection and observations attained using experiments in traditional machine learning and deep learning methods, showing greater confusion to classify the two tumor classes in general. The image statistics of the two tumor classes in H&E stain show higher closeness with subtler differences not clearly visible in the heatmaps. In some cases such as RGB histograms and graph-based method, the extracted feature vectors are long and their heatmaps require to be magnified to observe the smaller details. On combining low-level and high-level features followed by feature selection, the hybrid feature set shows higher distinction compared to individual features which is also suggested by the classification performance. However, the heatmaps are not always indicative of the final classification accuracy, especially for LBP histograms, and Voronoi diagram and Delaunay Triangulation where higher visual distinction but a lower performance is observed, so, the classification performance finally depends on the overall prediction process after complex learning by classifiers using traditional machine learning approaches.

Therefore, all the feature heatmaps are vividly indicative of the complexity to distinguish between HER2 negative and HER2 positive tumor classes in the H&E stained images resulting in higher prediction confusion of the automatic classification methods. This complexity is also confirmed by the visual inspection of the H&E stained images in the three malignancy levels.

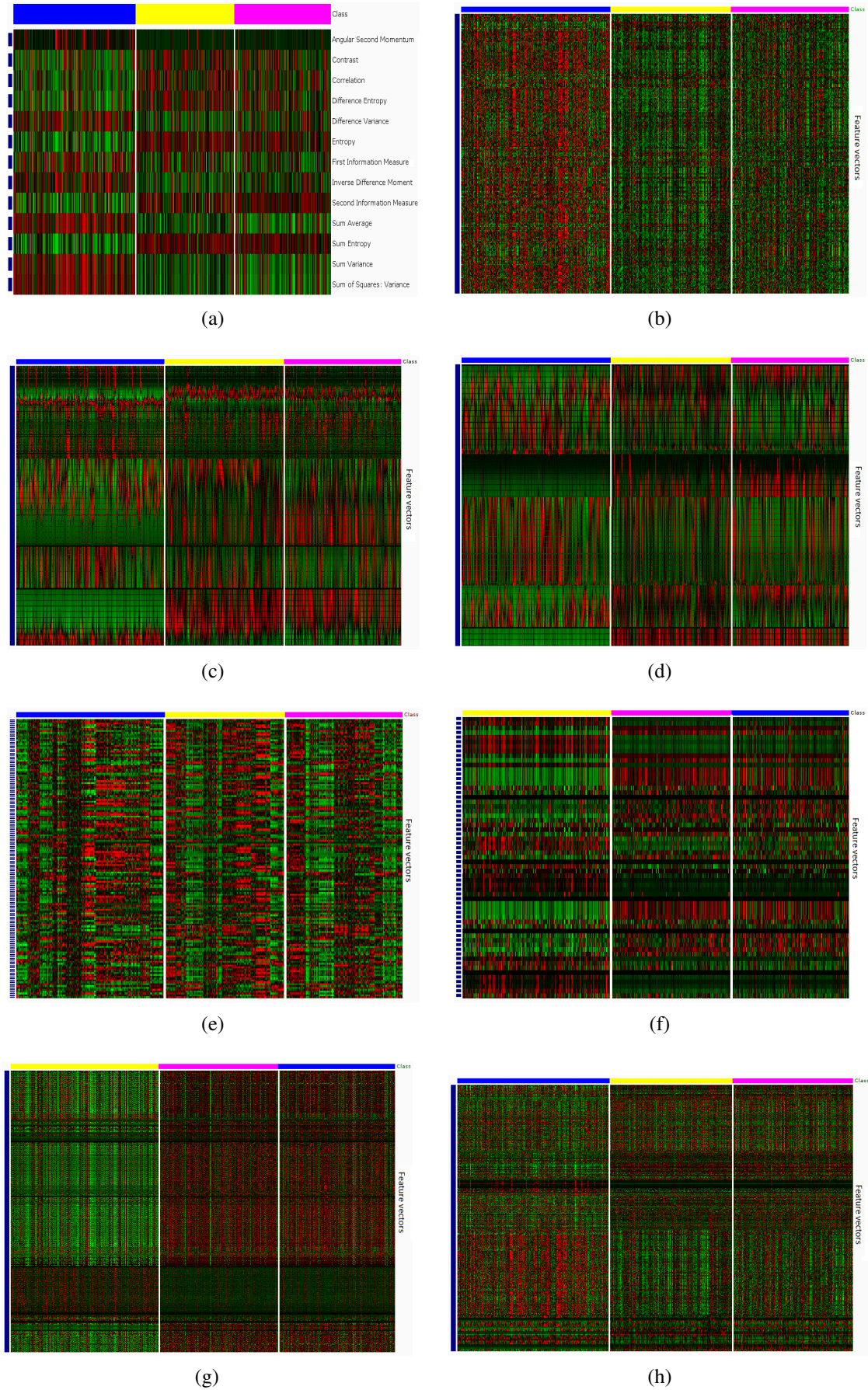


Fig. 9.4 Heatmaps of the extracted feature vectors showing image statistics in the HER2 stain (a) GLCM statistics (b) LBP histograms (c) HSV histograms (d) RGB histograms (e) Varma-Zisserman textons (f) Voronoi diagram and Delaunay triangulations (g) $nsARG[r_F]$ features (h) selected subset of hybrid features.

9.4 Summary

In this chapter, the three defined immunohistochemistry-based malignancy levels are observed with the help of visual inspection, in addition to visualization of generated heatmaps of the respective feature vectors in both HER2 and H&E stains. It is clearly observed that the underlying image statistics of the HER2 stain are quite distinct for the given malignancy regions, however, in the H&E stain, higher similarities are observed especially for the two tumor classes. Such nature of image data presented a strong challenge in this work during the analysis of cancer regions to perform automatic immunohistochemistry-based classification in the H&E stain that led to the development of sophisticated methods for performance optimization. The chapter is also able to highlight the main reasons behind lower classification accuracy for HER2 negative tumor class, which in turn lowers the overall and balanced accuracy of the developed image analysis framework. It can be concluded that due to higher complexity of information in H&E images, the performance (in terms of achieved precision) of the developed methods for the given problem is ultimately limited by the tissue characteristics of the malignancy types in H&E stain. Nevertheless, the proposed handcrafted features with traditional methods and deep learning methods have been specifically designed to capture the finer visual details in the three classes, as discussed in the previous chapters. The results and observations of the developed methods establish that computer-based analysis of gastric cancer tissues is feasible in the more routinely used H&E stain.

Summary

Contents

10.1 Conclusions	191
10.2 Future Work and Recommendations	195
10.2.1 Future Work	195
10.2.2 Recommendations	198

10.1 Conclusions

In the current practice of medicine, visual examination of tissue slides is the routinely used method for diagnosis and prognosis of diseases in histopathology, as the disease characteristics can be identified mainly from tissue appearance. Histological images are inherently complex in nature and contain a wide variety of visual information with different stains, magnifications, tissue types and methods of acquisition. Computer-based analysis of histological images is a primary focus of research in the upcoming domain of digital pathology. Recent literature in this area consists of a wide range of analysis techniques developed for addressing diverse histopathological questions related to specific disease conditions. The main objective of these methods is to assist pathologists in diagnostic and routine tasks, and aid further research in biology and medicine.

Gastric cancer is a major cause of cancer-related deaths in the world. Histopathological observations in gastric cancer tissues is routinely made by pathologists using visual inspection of microscopic slides which requires prolonged manual efforts. Computer-based analysis of gastric cancer whole slide images is an interesting field of research in digital pathology. The research work presented here provides a comprehensive account of the medical image analysis methods studied, modified and developed for effective knowledge representation and description, categorization in multiple scenarios, and applications in histopathological whole slide images of gastric cancer. The study concentrates on the more widely used H&E stain and is largely based on the underlying HER2 immunohistochemical information in the cancer-affected tissue regions.

Each of the three analysis stages, namely, preparation of materials, image pre-analysis and analysis of cancer regions are described in detail with their individual contributions, accomplishments and shortcomings. In the preparatory stage, expert knowledge of pathologists has been utilized to generate the required image datasets for pre-analysis and cancer analysis stages. Semi-automatic registration, annotation transformation between stains and ground truth generation using virtual microscopy tools has been demonstrated for this purpose.

In the image pre-analysis stage, tasks have been undertaken to simplify the process pipeline, specifically the subsequent cancer analysis problem, by exploring techniques mainly for necrosis detection, cell nuclei segmentation, cell nuclei classification and segmentation enhancement. An automatic method using textural analysis with traditional SVM-based learning and deep learning has been proposed for detecting necrosis in histological whole slide images. After quantitative and visual assessment of results, it can be concluded that the necrosis detection method identifies target areas in the selected regions of interest in an effective and timely manner. These smaller necrotic areas have been excluded before analysis of cancer regions. Additionally, in this stage, an advanced fully automatic cell nuclei segmentation method has been applied to isolate cell nuclei followed by multiresolution analysis and segmentation enhancement using a cell nuclei classification approach. The cell nuclei segments are favorably identified as one of the seven categories using object-level features followed by random forest machine learning. Improved cell nuclei segmentation results have been achieved using a multiresolution approach by combining diagnostically useful and eliminating undesirable information from individual magnifications. Such visual information fusion has further assisted computer-aided analysis of cancer regions in the next stage. The tasks in this stage have been adapted and demonstrated as applications, namely, automatic necrosis detection and automatic determination of tissue composition, by deploying the developed methods on unlabeled data that show correspondences to the ground truth generated by expert pathologists. These can be utilized by professionals and computer tools in diagnostic and prognostic tasks.

The cancer analysis stage emphasizes on methods for automatic categorization of the malignancy levels in H&E stained gastric cancer microscopic images based on their HER2 immunohistochemical response. It is difficult to visually differentiate the tumor classes in H&E stain, the reason that H&E is used for primary diagnosis and immunohistochemical staining is subsequently examined. However, there are characteristics not easily discernible by human eyes but can be exploited by automatic analysis systems, and this idea forms the main motivation behind this stage of research. For this purpose, traditional methods and deep learning methods are explored. Among traditional methods, low-level and high-level image descriptors have been studied, followed by traditional machine learning and classification. For the high-level handcrafted feature extraction, a novel approach is developed in the form of cell nuclei attributed relational graphs for knowledge description in gastric cancer, and extensively studied with variations in design and implementation to obtain multiple graph variants. The graph-based methods in general show an overall reasonable performance compared to the state-of-the-art texture, color and intensity-based features, and are especially superior to the morphological features

and identities of individual cell nuclei, reflecting the importance of spatial neighborhoods and topological properties of the tissue. These hand-engineered features are also found to be more informative representations of the cancer tissue compared to the most widely used proximity graphs, leading to better classification accuracy in acceptable computational times. The cell nuclei attributed relational graphs can provide a comprehensive knowledge of tissue architecture by combining properties of individual cell nuclei with their neighborhood interactions and global arrangements. The results based on high-level methods also suggest that perfect segmentation is not a prerequisite for good image classification. It can be stated that the handcrafted graph-based method has proved its ability as potentially powerful tool for describing and analyzing tissue architectures in digital histopathology. Also, combinations of the state-of-the-art and handcrafted features enhance the individual classification rates due to more extensive tissue information, as demonstrated by the hybrid features. The classification accuracy has been improved by increasing image tile size, thereby, enhancing the associated context information with larger fields of view at the highest magnification.

After the feature extraction methods, different traditional machine learning approaches are applied. A comparative analysis between six different classification strategies using three algorithms, namely, support vector machines, AdaBoost ensemble learning and random forests has been made. The classification strategies show similar trends in classification accuracy that vary with the configuration of handcrafted features, reflecting that effective and powerful image representation and description is the most crucial prerequisite before applying any traditional machine learning method. From experimental evidence, the final choice of the machine learning algorithm does not have a major implication, but support vector machines show more suitability in most of the experiments, ensuring that the extracted features are largely relevant. AdaBoost ensemble learning has a superior performance with larger features sets, however, SVM and random forests are quite robust and not particularly affected by the number of features. Hence, the quality of image features contributes more to the overall classification performance of the system compared to the choice of traditional machine learning method. This phenomenon possibly explains the reason why the principle focus of previous and current research related to image analysis in digital histopathology is on the description of tissues and their components in a meaningful manner.

Deep learning methods are also explored in the cancer analysis stage. Firstly, several CNN architectures are empirically investigated to determine a suitable design for learning significant characteristics of histological images. A preliminary self-designed CNN architecture is proposed and applied for two classification problems, namely, cancer classification based on HER2 immunohistochemical response and necrosis detection. The proposed CNN is quantitatively compared with the AlexNet framework which is popular for general object classification, as well as ensemble of the two CNNs. The deep networks are evaluated for necrosis detection to ensure their generalizability. To the satisfaction of the author, deep learning methods are able to keep up the promise of performing well for both the classification problems without the requirement of hand-engineering features. These methods compare favorably and even outperform handcrafted

features followed by traditional machine learning for a subset of experiments, however, these are also superseded by a few traditional methods in other occasions. It can be concluded that, for both the classification problems, the empirically selected and studied deep networks adequately represent H&E stained histopathological images of gastric cancer at high magnification with varying stain intensities and malignancy levels, even for smaller image sizes than traditional methods for one of the problems. Ensembles of these networks further help in combining individual contributions and reducing system variance. Therefore, deep learning is promising and has the potential of understanding and representing histopathological images in a way similar to human visual system and derived handcrafted descriptors.

As described in this study, a part of the undertaken classification tasks are clearly distinguishable by the human visual cortex, namely, necrosis detection and cell nuclei classification. However, the task of cancer classification in H&E stained gastric carcinoma WSI based on HER2 immunohistochemistry is more complex to be visually accomplished, and computerized methods can achieve even superior discrimination rates to human observation showing inter- and intra-observer variability, an important finding which may be fascinating for future researchers. Experimental results in the analysis of cancer regions in H&E stain consistently show that non-tumor can be easily distinguished from tumor, but it is more difficult to distinguish the two tumor classes. Especially HER2 positive tumor is also better distinguishable from the other two types, however, HER2 negative tumor has highest prediction error due to intermediate properties. Nevertheless, the observations based on automatic analysis in H&E stain are highly encouraging, while compared to the manual expert investigation in the more conveniently visualized HER2 stain showing significant subjective variability of observers, already reported in [Behrens 2015].

In the described experiments, the gastric cancer WSI datasets are considered heterogeneous due to the following reasons. Firstly, polygon areas are marked by expert pathologists using HER2 immunohistochemical staining which define varying levels of malignancy in the tissue. Secondly, the WSI datasets also exhibit different degrees of H&E stain. Thirdly, each WSI is acquired from the specimen of one patient, thus, the datasets contain inter-patient biological variability. The encouraging results obtained during the evaluation phase represent the robustness of the proposed framework to heterogeneous WSI data, and its potential ability to be adopted in practical environments. Therefore, the comprehensive goals of this studies to achieve effective computer-based histological image analysis using the routinely examined H&E stained WSI of gastric cancer has been satisfactorily accomplished.

10.2 Future Work and Recommendations

The future directions and outlook of this research, and the useful recommendations are as follows.

10.2.1 Future Work

The future work should chiefly focus on optimizing system performance and computational requirements, and expanding the current database, as explained in the following sections.

10.2.1.1 Performance Optimization

Optimization of the system performance must be pursued in future for both traditional route and deep learning route as follows.

1. **Handcrafted features and traditional machine learning:** The current handcrafted feature extraction and traditional machine learning methods provide satisfactory results, however, there is still scope of improvement, especially in the third stage for analysis of cancer regions. Refining the current image analysis methods to reduce the confusion between tumor types is recognized as the main future direction of this study. This can be attempted by the design of further suitable hand-engineered features followed by exhaustive comparative evaluation with the currently developed and state-of-the-art methods. A refinement of categories in the malignancy levels and a more elaborate classification hierarchy of gastric malignancies at the cellular and tissue levels after reassessment with expert pathologists is also advisable for improved performance.
2. **Deep learning:** The next steps of study will be to enhance the proposed CNN architecture by designing deeper and wider convolutional neural networks for achieving higher classification accuracy for the mentioned classification problems. This can be made possible in the future with an advancement of GPU hardware architectures or using cascades of GPU. This will ensure higher availability of system resources for allowing design and implementation of complex networks and higher resolution images. The proposed CNN architecture will also be applied on other diverse classification problems in histology, to further investigate its generalizability. Other interesting directions include comparative evaluation with successful CNNs in digital histopathology, for example, U-Nets [Ronneberger 2015] have gained recent attention for their accomplishment in histological segmentation that can be extended in H&E WSI based on immunohistochemistry, and ensemble learning can be examined in detail to combine multiple popular networks. Reverse analysis of representations inside the deep networks to identify the particular characteristics that allow the CNNs to make accurate decisions and to understand better why some architectures perform better than others, is also an exciting future research prospect.

Classifiers based on traditional machine learning and deep learning can be integrated in future, in order to harmonize their individual strengths. Traditional methods can be combined with deep

learning by first performing handcrafted feature extraction along with location information to be learned by layers of convolutional neural networks. A reverse approach leading to feature extraction using representations in layers of deep networks and classification by traditional methods can also be investigated in this direction.

10.2.1.2 Computational Optimization

Optimizing computational requirements is one of the most essential measures in order to make a system operable in practical environments, for instance, use in online computer-aided diagnosis and content-based image retrieval software. Attempts towards algorithm optimization in histological image analysis directly at the WSI level can prove pivotal in shaping the future of digital pathology. An approach based on parallel processing frameworks for sharpness adjustment of WSI using distributed computing mechanisms is presented in [Zerbe 2011], and this system can be further explored to process and analyze entire whole slide images. Such parallelization of operations to generate, process and merge the constituent image tiles (or regions) at different magnifications, and yield the resulting classified WSI in real-time may be complex using the present state of technology but can be surely achieved with enhancement of computational power and use of advanced devices for graphics processing such as the GPU in the near future. Therefore, it is advisable that the developed algorithms are modified for real-time analysis of large-sized WSI using advanced parallel processing techniques.

In this work, most of the tasks have low computational requirements, however, can be further optimized using algorithm optimization methods. Among the handcrafted feature extraction approaches, the aim will be to optimize program execution especially for the current graph constructions with corresponding global measures and some of the self-implemented state-of-the-art features. Implementations of the traditional machine learning methods have already been optimized over the past several years through a number of libraries and frameworks developed for data mining in various programming languages. These have been made available as open-source toolkits, for example, OpenCV for C++ [Bradski 2013], SciKit-Learn [Pedregosa 2011] for Python and WEKA [Hall 2009] for JAVA. This work mostly uses optimized programs to implement traditional machine learning, except AdaBoost ensemble learning which has been partly optimized in C# in the Charité working group, but is slower than the other machine learning methods as discussed in Appendix A. Hence, further optimization of AdaBoost is also a prospective research direction.

Among the deep learning methods, training the convolutional neural networks from scratch using the faster implementations on the GPU requires around two days for the proposed CNN architecture and 9 hours for AlexNet CNN framework in the current experimental setup. On the other hand, an optimized implementation of feature extraction followed by traditional machine learning for same sized data requires relatively smaller duration, of the order of a few minutes or hours, to complete the training process even on the CPU. Hence, there is a considerable difference of time requirements for the two distinct learning routes. In this work, performance using deep learning is found comparable to traditional machine learning, and shows potential of

significantly outperforming state-of-the-art technology in future. The trade-off between accuracy versus time efficiency requires further experimental evidence to ensure an upper hand of deep learning algorithms in digital histopathology. Hence, a future consideration will be to optimizing the deep learning algorithms, especially to reduce the present time requirements in the training phase. Space requirements can also be reduced by on-the-fly data generation and augmentation before batch input to the training process.

10.2.1.3 Database Expansion

In the preparatory stage of the process pipeline, ground truth has been generated by expert pathologists by marking manual annotations in the high resolution WSI. Annotating the datasets to generate accurate ground truth is a tedious and exhausting process, therefore, the current experimental setup has been tested on a small number of WSI. Moreover, in the author's knowledge, there are no existing publicly available labeled datasets for the explored research problems. In general, for most of the classification configurations in the cancer analysis stage, more favorable performance is achieved when a few labeled tiles of the same patient are utilized by the system to learn unique biological characteristics of a patient. Additionally, in the presently developed deep learning methods, large number of images are generated using dataset augmentation, however, the size of image datasets for successful training is an experimental question to be further investigated. Hence, extension of the current reference WSI datasets to take into account more biological variations among patients, can facilitate large-scale training to enhance the prediction capabilities and increase the robustness of the proposed system. In order to ensure higher practical operability, the proposed methods should be investigated using more extensive and varied WSI datasets from a larger number of patients.

The proposed histopathological image analysis framework can be tested on other types of tissue images in digital pathology. With data acquisition from different sources including multiple WSI scanners or laboratories, there is a possibility of variations in stain properties between samples. With expanded WSI datasets acquired from several sources in future, a preprocessing step using stain normalization methods [Khan 2014] can be introduced to the current process pipeline. This challenge has not been considered in the present research due to WSI data uniformly acquired from the same WSI scanner and laboratory, and original correlations between the two stain properties (HER2 and H&E) required to be investigated without disturbing their mutual and independent distributions. Other reasons are reduced system complexity, and the initial objective to develop more robust algorithms for heterogeneous datasets with variations to stain distributions.

An option for expanding the scope of different aspects of the framework is to use open-source benchmark datasets made available for specific applications in different challenges, for instance, Mitosis Detection Challenge [Veta 2015] for mitosis detection in breast cancer histopathology images, Camelyon Challenge [ISBI 2016] for automated detection of metastases in H&E stained WSI of lymph node sections, Cell Tracking Challenge [Maška 2014] for tracking moving cells

and Glas Challenge [Sirinukunwattana 2017] to isolate tissue glands in colon cancer images. Such large-scale datasets are currently being developed to facilitate research in deep learning.

10.2.2 Recommendations

A few additional recommendations for image analysis in digital histopathology are as follows. There is a need to consider the information at different magnifications of a whole slide image in order to capture properties at both cellular and histological levels. Navigating a WSI using virtual microscopy software [VMscope 2010b] is made possible at different magnifications using a hierarchical file storage structure. A multi-scale framework using parallel processing techniques is presented in [Huang 2011], where sparse coding and GPU programming are explored and dynamic sampling techniques are used to rapidly identify fields of interest and analyze breast biopsy histopathological whole slide images. The authors report their computational times on WSI comparable to processing times of pathologists. Furthermore, a multiresolution approach to combine visually significant information at different magnifications to improve segmentation results in gastric cancer images has been presented in the pre-analysis stage of this work. For future research, the author recommends analysis of multiresolution or hierarchical data representations, because it has already been observed that different resolutions can provide complimentary information, playing a substantial role to obtain more desirable outcomes than a single resolution.

Deep learning in digital histopathology is still in its early stages of development, and needs to be extensively studied. Presently, it faces a few apparent challenges. For instance, classification problems in histology are hard and diverse with various types of tissues, tumor classes and stains and high similarities in appearance. Further, to train a network for dedicated supervised classification problems, experiments should be performed with vast problem-specific image datasets. This big data requirement can address the challenge of limited availability of ground truth in digital pathology, as the demand from deep learning can positively encourage the scientific community to establish large-scale open-source histopathological repositories for research. A few repositories have already been introduced in academic, industrial and independent areas listed at [DPA 2016] and can be used for this purpose.

Usually, large fields of view are investigated to capture the associated context information, so training deep networks for tissue images can have higher time and space requirements on the GPU. Limited memory configuration of the state-of-the-art graphics cards may restrict GPU-based computation for digital pathology applications. The experimental evidence on the gastric cancer WSI datasets shows comparable description and classification power of deep learning methods versus traditional machine learning methods, so it is still early to predict if deep learning derived abstractions will be able to completely replace handcrafted features in the future of digital pathology. Undoubtedly, in the past years there has been more emphasis on handcrafted features due to certain unique characteristics of tissue images, and deep learning is fairly recently introduced in this field. However, the hardware developments seem promising and may lead to application of wider and deeper networks, the effect of which is still to be examined.

The phenomenon of rapid GPU development has been actively discussed [Needham 2016]. We have also seen that theoretically there are immense possibilities of the number of different deep network architectures that can be designed due to the involvement of several design parameters. Therefore, as the GPU capabilities grow with time, we may soon witness breakthroughs in this area of research.

Other recommendations for future include extending to video or spatio-temporal domain where time is an additional dimension accounting for motion, deformation or changes in the appearance of disease. Three-dimensional data also proves useful for volumetric analysis of cancers, using reconstructions and depth information in thick tissues obtained from surgical resections. Combining information from different modalities, as well as additional knowledge from signals, metadata and other semantic sources may prove beneficial for extracting comprehensive knowledge for more informed diagnostic decisions and treatment planning. This calls for development of sophisticated algorithms to facilitate multimodal analysis of images that have some common characteristics. It is an interesting field where different advanced technologies can be employed to acquire complete information from various sources regarding a particular condition such as different types of microscopic techniques.

Implementation Details

Contents

A.1 Hardware Specifications	201
A.2 Operating Systems	202
A.3 Programming Languages	202
A.4 Supporting Software	203
A.5 Computational Requirements	205
A.6 Examples of Metadata Screenshots	212

A.1 Hardware Specifications

The processes of the proposed image analysis framework have been implemented using two computers and GPU of one of the computers. The first computer has Intel Core i7-3700 processor at 3.40 GHz with 16 GB RAM and Microsoft Windows operating system. It is used for accessing the WSI data, creation of image datasets in the preparatory stage, all the procedures of the pre-analysis stage, state-of-the-art and handcrafted feature extraction followed by traditional machine learning methods during the analysis of cancer regions, and quantitative and comparative evaluation of methods at each stage.

The second computer mostly implements the deep learning modules because of availability of desired GPU. The system specifications for this computer are Intel Core i7-4790 processor at 3.60 GHz with 15 GB RAM and Ubuntu 14.04 operating system. The available GPU hardware is NVIDIA GeForce GTX 660 with 2GB memory. Training and testing of the convolutional neural networks is performed using GPU programming to achieve higher time efficiency over CPU [Strigl 2010].

A.2 Operating Systems

As discussed above, one computer has Microsoft Windows while the other computer (with GPU) has Ubuntu 14.04 as the operating system. The reasons for working with two different operating systems are as follows. Firstly, Microsoft Windows is widely incorporated for routine practice and research in the current medical environments. Hence, the developed methodology and applications are convenient and feasible to use for the medical professionals, scientists, students and persons with limited technical knowledge due to familiarity and user-friendly interface. Moreover, the second computer has been mainly used for implementing deep learning methods, which were (at the time of research) and are still at early stages of development. The documentation of the related software frameworks were observed to be more advanced, flexible and up-to-date for the Linux-based operating systems compared to Windows systems, thus, it was less challenging to organize the deep learning experiments using the second machine. However, with introduction of newer software packages in Windows recently, the deep learning methods can be transferred and adapted to Windows-based machines for practical usability in the medical community.

A.3 Programming Languages

The programming languages considered for implementation of various modules of this research are C#, MATLAB, and Python. However, majority of the modules are written in C#. C# is intended to be a simple, modern, general-purpose, object-oriented programming language [ECMA 2006] and enables developing secure and robust applications, especially for .NET framework. C# and .NET are usually preferred in digital histopathology due to the following reasons. Firstly, several libraries for accessing histological whole slide image data are available in C#, giving ease of access to the large-sized WSI and corresponding annotation data. Also, C# is easy to integrate with Microsoft's technologies which is an important consideration, as Microsoft Windows is the most commonly used operating system for the deployment of medical image processing and analysis applications in practical settings. Additionally, *VMscope GmbH*, the company providing software for WSI image access in this work, is a certified Microsoft partner and has C# modules for reading the WSI data and basic slide level operations, thus, the subsequent analysis methods and applications are also mostly developed in C#. This language is used for most of the process pipeline including the procedures of the preparatory stage, cell nuclei segmentation and feature extraction followed by multiresolution enhancement in the image pre-analysis stage, state-of-the-art and handcrafted feature extraction and AdaBoost ensemble learning method during the analysis of cancer regions, and the generation of all the evaluation results in the research. The integrated development environment (IDE) selected for implementation in C# is Microsoft Visual Studio 2012.

MATLAB (R2010a) is the second programming platform applied in the study, specifically for implementation of Varma-Zisserman texon-based state-of-the-art feature extraction, SVM-based

traditional machine learning method with and without discriminative thresholds, and generation of certain diagrams, *e.g.* ROC curves and P-R curves, for clear visualization. MATLAB is primarily selected because of the availability of corresponding time-efficient open-source modules using the platform. In general, MATLAB offers several advantages such as built-in mathematical operations on arrays which is suitable for two-dimensional data leading to fewer lines of code, and optimized generation of graphical outputs. An additional reason for using MATLAB especially for SVM learning is its low computational requirements of the program, and also facilitates the comparison of computational times with the other two traditional machine learning methods using the other two selected programming languages.

Python programming is used for implementation of a few methods, especially Gabor filter-bank feature extraction and grid-based parameter selection in SVM learning, random forests machine learning, and deep learning methods. The prime reason for selecting this language for developing the deep learning methods is due to the availability of suitable and well-documented open-source software framework for Linux-based system in Python. Random forests are explored using Python libraries again to facilitate comparison in computational times with the other traditional machine learning methods using other programming languages. Python's main advantages are straightforward syntax, fast speed and availability of large number of resources for developers. The IDEs in which the Python modules are written include JetBrains PyCharm and IDLE.

A.4 Supporting Software

The supporting software, libraries and frameworks utilized for different stages of the histological image analysis are described as below, in sequence of their usage.

1. **VM Slide Explorer 2010:** It is a virtual microscopy application that allows viewing and navigating virtual slides efficiently on the monitor [VMscope 2010b]. It is mostly used for viewing the annotation marked by expert pathologists on HER2 immunohistochemically stained WSI, and the corresponding selection of regions of agreement from H&E stained WSI. Also, the reference data for necrosis detection using the splitting approach was generated by expert pathologist from the WSI with the help of this utility. It is proprietary software, and its license is provided by *VMscope GmbH*.
2. **Virtual Slide Access SDK 4.0:** This toolkit allows users to add own functionalities and to use existing image analysis tools in the virtual microscope. The basic functionality of a virtual microscope such as opening of several image formats, navigation, annotations etc., is provided [VMscope 2010a]. The modules of this toolkit are used in C# for accessing WSI data for the succeeding image processing and analysis. It is proprietary software, and its license is provided by *VMscope GmbH*.
3. **CognitionMaster:** It is open-source software that offers object-level image analysis [Wienert 2013]. It allows straightforward development of object-related interactive or fully

automated image analysis solutions. ROIManager is a plugin developed for use with Cognition Master, specifically for ease of annotations in virtual slides, that is also open-source. This software is used in the preparatory stage for generating and reviewing of ground truth annotations for cell nuclei segmentation evaluation, cell nuclei classification and necrosis detection using the merging approach.

4. **ViSPee:** Virtual Slide Processing Environment (ViSPee) provides tools specifically for assisting analysis from remote locations. Image data is stored on a remote server and a web provider is used for accessing it. It is written in Java and being used with Eclipse environment. One of its applications is creating different kinds of annotations for an image, *e.g.* points, lines and polygons. The annotation data is stored in an XML file. It has been extended for creating correspondences between pairs of slides, where one XML file containing registration data is stored for a pair of images. This software is applied during semi-automatic registration between the two stains in the preparatory stage. The utility was developed during Virtual Specimen Scout [VSS 2010] project. It is used by collaborating partners of Charité University Hospital, and not available commercially.
5. **LIBSVM:** LIBSVM [Chang 2011] library is applied for the SVM-based operations. LIBSVM is an open-source library written in C++ and has MATLAB and Python extensions. It provides the options of different SVM kernels supporting classification and regression. Specifically, in this work, its Python tool is used for parameter selection using grid search, and MATLAB modules for SVM training and classification.
6. **Python SciKits:** SciKits (SciPy Toolkits) are open-source add-on packages for SciPy [Jones 2015] consisting of implementation of different scientific tools. A SciKits tool is applied from SciKit-Image [vanderWalt 2014] package for computing state-of-the-art low-level Gabor filter-bank features. Other modules are used from SciKit-Learn [Pedregosa 2011] package for random forest machine learning and classification.
7. **OpenCV (EmguCV):** OpenCV (Open-source Computer Vision) [Bradski 2008] is an open-source library written in C/C++, containing programming functions for computer vision and image analysis. EmguCV wrapper [Shi 2013] is a cross platform .NET wrapper to the OpenCV image processing library. Among state-of-the-art feature extraction methods, HSV histograms are computed and Voronoi-Delaunay graph construction is performed using EmguCV library in C#.
8. **Accord.NET framework:** Accord.NET [Souza 2014] is a framework for scientific computing in .NET, consisting of audio processing, image processing and machine learning libraries written in C#. These libraries are employed for state-of-the-art low-level feature extraction of GLCM measures and LBP histograms.
9. **AForge.NET framework:** AForge.NET [Kirillov 2013] is a .NET framework for computer vision and artificial intelligence written in C#. Its utilities are used to generate a few state-of-the-art low-level features such as RGB histograms and gray histograms in timely manner.
10. **RapidMiner Studio:** RapidMiner is a cross-platform software mainly for data mining, machine learning and analytics [Hofmann 2013]. The basic version of the software is freely

available. It has been utilized for correlation analysis on hybrid feature vectors, followed by removal of correlated features for testing combination of low-level and high-level methods.

11. **Caffe framework:** Caffe [Jia 2014] is an open-source C++ based library with Python and MATLAB bindings for training and deploying CNNs, and also contains reference models such as AlexNet [Krizhevsky 2012] for experimentation. Conversion of augmented image datasets into LMDB, computations of image means and implementation of the proposed CNN architecture are performed using the modules in Caffe framework. The reference implementation of AlexNet is directly utilized from Caffe.
12. **Qlucore Omics Explorer 3.2:** Qlucore is a bioinformatics software [Qlucore 2008], and popular in biotechnology due to its ability to handle large number of samples. It is a proprietary software, and its licensed version has been used. It has been applied for generation of feature heatmaps for visualization purpose, and obtaining statistical inferences for the two stains, HER2 and H&E.

A.5 Computational Requirements

The computational requirements of the crucial program modules in each stage of the image analysis framework are presented in Table A.5 using the hardware mentioned in Section A.1, along with the most important supporting software in their implementation. The modules implemented with assistance of group members in Charité University Hospital have been explicitly cited or mentioned in the remarks (with individual contributions mentioned in the Acknowledgments section). The time requirements are indicative of the average duration and may vary with data characteristics (*e.g.* image complexity, image size, number of cell nuclei etc.). Also, the time indicated does not include the time required for accessing data, for instance, reading and writing WSI or image tiles from or to hard drives, but only the computational time required to perform the specified algorithms.

Table A.1 *Important program modules, supporting software and computational requirements of the proposed framework*

Stage	Program module	Supporting software	Programming platform	Computational requirements	Remarks
Stage 1: Preparation of materials	WSI semi-automatic registration	ViSPEe	JAVA	Time depends on the user to manually generate control points. Triangulation is generated in real-time.	Implemented by group members as a part of the project Virtual Specimen Scout [VSS 2010].

	WSI annotation transformation	Library WsiRegLib as part of Virtual Slide Access SDK 4.0	C#	Average time is <i>ca.</i> 11.5 minutes per annotation transformation.	Implemented with group members.
	WSI tessellations	Virtual Slide Access SDK 4.0	C#	Average time is <i>ca.</i> 0.5 seconds per image tile.	—
Stage 2: Image pre-analysis	Necrosis detection: textural feature extraction	Accord.NET and SciKit-Image	C# and Python	Average time for 11,470 images of smallest size (64×64 pixels) is <i>ca.</i> 5 minutes.	Partly implemented with group members.
	Necrosis detection: SVM machine learning with discriminative thresholds	LibSVM	MATLAB	Average time for training phase is <i>ca.</i> 2.21 seconds for the described datasets and classification is <i>ca.</i> 75 milliseconds per image.	—
	Necrosis detection: AlexNet CNN framework	Caffe	Python	Each training iteration requires 0.15 seconds. Average time for training phase is <i>ca.</i> 2.5 hours for the described datasets and deployment is <i>ca.</i> 0.05 seconds per image.	AlexNet framework is provided in Caffe.
	Necrosis detection: proposed CNN architecture	Caffe	Python	Each training iteration requires 0.65 seconds. Average time for training phase is <i>ca.</i> 11.7 hours for the described datasets and deployment is <i>ca.</i> 0.25 seconds per image.	—
	Necrosis detection: ensemble of CNNs	Caffe	Python and C#	Average classification time is <i>ca.</i> 0.32 milliseconds per image.	Computation of ensemble results is performed in C#.
	Cell nuclei segmentation algorithm	Library CellSegLib of CognitionMaster	C#	Average time for each image of size 1024×1024 pixels is <i>ca.</i> 23 seconds.	Implementation used from [Wienert 2012].

Stage 2: Image pre-analysis	Cell nuclei feature extraction	Most features (27) from libraries of Cognition-Master	C#	Average time for each cell nucleus (sample) is <i>ca.</i> 12.5 milliseconds.	Implementation mostly from [Hufnagl 1984] and [Zerbe 2008].
	Cell nuclei classification: SVM	LibSVM	MATLAB	Average time for training phase is <i>ca.</i> 0.5 seconds for the described datasets and classification is <i>ca.</i> 2 milliseconds per sample.	—
	Cell nuclei classification: AdaBoost	—	C#	Average time for training phase is <i>ca.</i> 51 seconds per sample and classification is <i>ca.</i> 2.7 milliseconds per sample.	AdaBoost implementation by group member.
	Cell nuclei classification: random forests	SciKit-Learn	Python	Average time for training phase is <i>ca.</i> 1.8 seconds for the described datasets and classification is <i>ca.</i> 0.35 seconds per sample.	—
	Multiresolution enhancement	—	C#	Upscaling lower resolution segmentation requires <i>ca.</i> 0.15 seconds per image and multiresolution combination requires <i>ca.</i> 0.2 seconds per image.	—
Stage 3: Analysis of cancer regions	Dataset augmentation	Virtual Slide Access SDK 4.0	C#	Around 80 images are generated each minute.	—
	Low-level state-of-the-art feature extraction: GLCM statistics, LBP histograms, gray histograms, HSV histograms, RGB histograms	Accord.NET and Aforge.NET	C#	Average time is <i>ca.</i> 1.41 seconds per image, and individual times 0.26, 1.58, 0.02, 5.14, 0.03 seconds per image respectively.	—
	Low-level state-of-the-art feature extraction: Gabor filter banks	SciKit-Image	Python	Average time is <i>ca.</i> 0.42 seconds per image.	—

Stage 3: Analysis of cancer regions	Low-level state-of-the-art feature extraction: Varma-Zisserman textons	Open-source RFS filter bank implementation from [Varma 2002]	MATLAB	Average time required for generation of texton dictionary is <i>ca.</i> 4.79 hours, and for feature extraction is <i>ca.</i> 1.7 minutes per image.	RFS implementation is used from available code, generation of texton dictionary and feature extraction are implemented by self.
	Low-level state-of-the-art feature extraction: other color-based measurements	–	C#	Average time is <i>ca.</i> 4.56 seconds per image.	Mean chromaticity concept of group member.
	High-level state-of-the-art feature extraction: Voronoi-Delaunay Graphs graph construction	EmguCV	C#	Average time is 0.25 seconds per image.	Voronoi-Delaunay graph construction is used from library, global feature extraction is implemented by self.
	High-level state-of-the-art feature extraction: Voronoi-Delaunay Graphs global feature extraction	–	C#	Average time is <i>ca.</i> 2.28 seconds per image.	–
	High-level hand-crafted feature extraction: cell nuclei ARG variants graph construction	A part of object-level vertex attributes from libraries of Cognition-Master	C#	Average time is <i>ca.</i> 20 seconds per image but varies with choice of r when fixed.	Additional space requirement for storing graph image and data (optional).
	High-level hand-crafted feature extraction: cell nuclei ARG variants global feature extraction	–	C#	Average time is <i>ca.</i> 8.28 seconds per image for $gARG$ construction.	Global graph feature extraction times for other constructions are similar. Variation in individual times is observed due to graph density and image complexity.

Stage 3: Analysis of cancer regions	Feature correlation analysis	RapidMiner	–	Average time required is <i>ca.</i> 5 seconds for correlation analysis for hybrid set of features.	–
	Region classification: SVM	LibSVM	MATLAB	Average time for training phase is <i>ca.</i> 0.19 seconds for described datasets and classification is <i>ca.</i> 0.43 milliseconds per image.	–
	Region classification: AdaBoost ensemble learning	–	C#	Average time for training phase is <i>ca.</i> 35.23 minutes for described datasets and classification is <i>ca.</i> 2.91 milliseconds per image	AdaBoost implementation by group member. Time requirements depend on number of features (details later).
	Region classification: Random forests	SciKit-Learn	Python	Average time for training phase is <i>ca.</i> 11.61 seconds for described datasets and classification is <i>ca.</i> 3.48 milliseconds per image.	–
	Region classification: AlexNet CNN framework	Caffe	Python	Each training iteration requires 0.15 seconds. Average time for training phase is <i>ca.</i> 8.75 hours for described datasets and deployment is <i>ca.</i> 0.05 seconds per image.	–
	Region classification: proposed CNN architecture	Caffe	Python	Each training iteration requires 0.65 seconds. Average time for training phase is <i>ca.</i> 40.8 hours for described datasets and deployment is <i>ca.</i> 0.25 seconds per image.	–
	Region classification: ensemble of CNNs	Caffe	Python and C#	Average classification time is <i>ca.</i> 0.52 milliseconds per image.	Computation of ensemble results is performed in C#.

	CBIR application	–	C#	One round of retrieval requires <i>ca.</i> 0.4 seconds.	–
Result evaluation	Cell nuclei segmentation evaluation	ROIManager of CognitionMaster	C#	The evaluation metrics for described datasets are calculated in <i>ca.</i> 5.5 seconds. The process depends mostly on the time taken for ground truth data generation.	–
	Evaluation strategies: k-fold stratified shuffled split, leave-a-sample-out and leave one patient out	–	C#	Vary with the number of sample in training and classification phases.	–
	Image statistics: feature heatmaps	Qlucore Omics Explorer 3.2	–	Each feature heatmap is generated in real-time <i>i.e.</i> less than 2 seconds.	–

It may be specified that the *cell nuclei ARG* graph-based methods developed for the two steps, namely, graph construction and global graph feature extraction can be concatenated into a single program as an online procedure, without explicitly saving the graph-based representations of each image and the associated data offline, in order to lower the time and space requirements. These steps have been performed separately in this study to visualize the graphs and organize the computed features from WSI datasets depending on the different types of evaluation methods. Similarly, comprehensive models of trained classifier can be saved offline, and the feature extraction followed by classification of unlabeled input image data can be combined into a single executable application.

The computational requirements for the $gARG[r_F]$ method has been discussed in detail by the author in [Sharma 2016], comparing them with the optimized proximity graph method as follows. The average execution time requirements for constructing the Voronoi-Delaunay graph and cell nuclei ARG for selected image tiles in the WSI histological dataset is shown in Fig. A.1. This time represents the computational time required to identify the graph vertices after segmentation, linking them with edges and also assigning attributes to each vertex and edge for each cell nuclei ARG. It can be noted that the Voronoi-Delaunay algorithm is optimized (EmguCV implementation), thus, the time required is lowest in this case (average 0.25 seconds). Higher computational requirements for the cell nuclei ARG are mainly due to additional attribute calculations for vertex and edges, a geometrical unconstrained nature and a non-optimized algorithm. Further, as the number of vertices increases, time requirements increase for both graph-based representations, with a visibly greater rate for the cell nuclei ARG. This rate also increases with the increase in the assigned values of maximum spanning edge length r . A higher r means more edges and edge attributes, thereby leading to increase in run-time. The time duration does not include global graph feature extraction, which in turn depends on the density

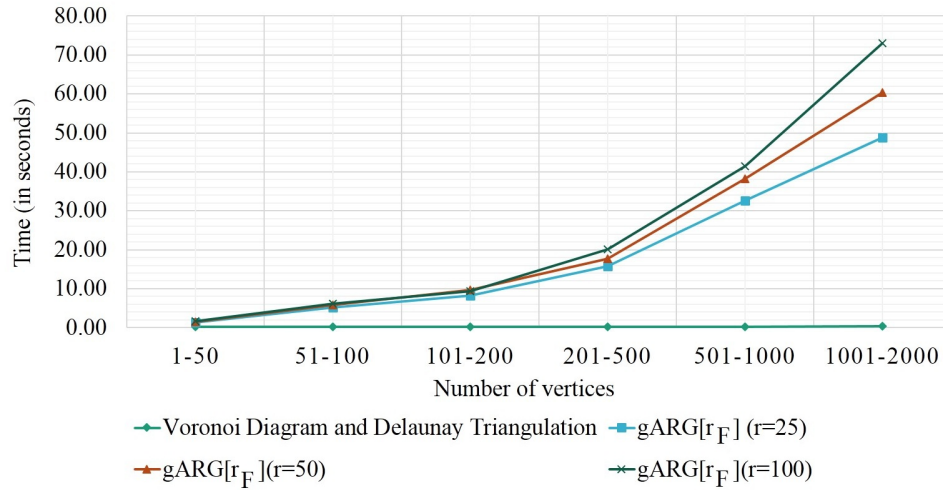


Fig. A.1 Average execution time requirements of graph-based constructions for Voronoi-Delaunay and $gARG[r_F]$ graphs and variations with the number of graph vertices

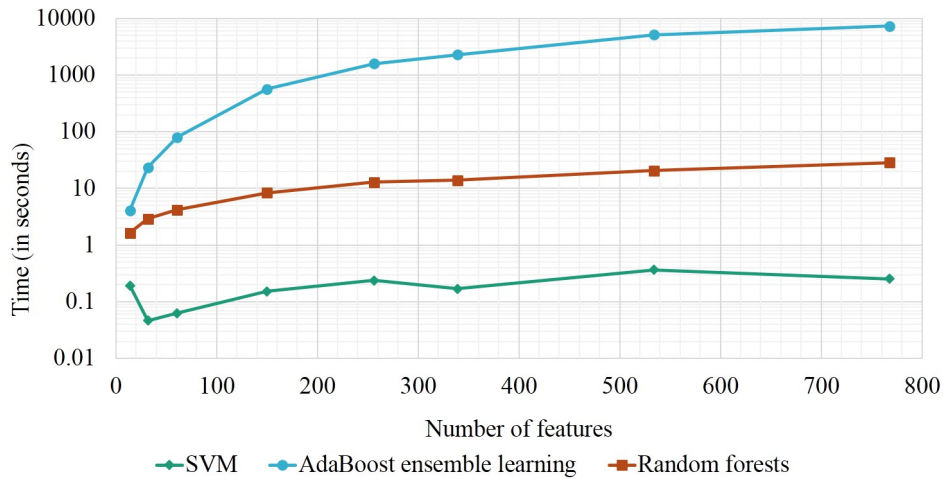


Fig. A.2 Average execution time requirements (log-linear plot) of traditional machine learning methods and variations with number of features during training

of the cell nuclei ARG or Voronoi-Delaunay graph, as the complexity in computation of the graph size-based features increases with increase in the number of vertices and edges.

The computational requirements for different traditional machine learning methods during analysis of cancer regions can be comparatively summarized and demonstrated in Figure A.2. The time is measured in seconds, and is shown in logarithmic scale (base 10) for better comparison. The average times for one training phase (using fixed number of images) of single-stage classification using three traditional machine learning methods, namely, SVM, AdaBoost and random forests, and the variations with the number of features are depicted. As SVM and random forests implementations have been applied using optimized programs from highly developed machine learning libraries, whereas AdaBoost has been programmed by members in Charité, it can be observed that the SVM and random forests are faster compared to AdaBoost ensemble

learning. SVM is the fastest, followed by random forests and then AdaBoost. Also, the dependence of time requirements on the number of features is more prominent in AdaBoost, compared to random forests and SVM, as the current AdaBoost implementation compares the values of each feature incrementally in the classifier, so the time rapidly increases with the number of features. This insists on a need for code optimization to be addressed in the future studies. On the other hand, during the classification stage, AdaBoost has lower average requirements compared to optimized random forest implementation.

A.6 Examples of Metadata Screenshots

Metadata means data about data, in this case, refers to the data that is generated during certain processes on the tissue images, such as corresponding control points generated manually for semi-automatic registration between WSI, point annotations for segmentation evaluation, contour annotations for creation of ground truth of cell nuclei classification and polygon annotations for necrosis detection. In this section, examples of screenshots of the metadata prepared or generated for a few representative processes in the preparatory and pre-analysis stages are demonstrated. These processes include the semi-automatic WSI registration (XML file), annotations for cell nuclei segmentation evaluation (ROI file), and features with class labels for cell nuclei classification (XML learning sample file) for the specified AdaBoost ensemble learning implementation. Such representative examples of metadata screenshots are shown in Figure A.3. XML learning sample files similar to Figure A.3(c) have also been generated for the region-based samples in the third stage of the framework for analysis of cancer regions using the different feature extraction methods with AdaBoost ensemble learning and classification.

```
<?xml version="1.0" encoding="UTF-8"?>
- <WsiRefPntPairs>
  - <metadata>
    <version rev="1.3"/>
    <creator userID="patho-amba"/>
    <imgHandlerUri name="http://http"/>
    <wsiFolder name="vss-stor.charite.de"/>
    <imagePath name="Misc/Reg/Kiel"/>
    <slide1 name="MagenCa05-1-Her2-1" height="121024" width="90531" offsetY="0" offsetX="0" ext="vsf" id=""/>
    <slide2 name="MagenCa05-1-HE-1" height="111553" width="86273" offsetY="0" offsetX="0" ext="vsf" id=""/>
  </metadata>
- <slide1>
  - <refPoints>
    <point name="A0001" id="ce54e0ed-4906-44be-9fe8-bff29fb8c09d" creationRatio="147.9264705882353" y="36982" x="13165"/>
    <point name="A0002" id="3f629a31-43ea-45f5-949e-241b47cbbc05" creationRatio="12.0" y="55121" x="14567"/>
    <point name="A0003" id="0be48f51-a4c0-4924-a332-9c04eca57ace" creationRatio="45.0" y="70470" x="15365"/>
    <point name="A0004" id="ca57dfef-1b91-4d1e-99fa-57f556e570c9" creationRatio="147.9264705882353" y="81803" x="18639"/>
    <point name="A0005" id="87275031-2477-44cf-b765-4fad4e470c81" creationRatio="11.0" y="92667" x="23658"/>
    <point name="A0006" id="a6f7c574-b3e0-41e9-93cc-9c80da080930" creationRatio="146.0" y="103505" x="29198"/>
    <point name="A0007" id="b6f7fd91-15ce-41f8-8075-77e2a3bc854e" creationRatio="7.0" y="28954" x="27483"/>
    <point name="A0008" id="19784fb9-6024-4d24-8a85-8d0f622b238a" creationRatio="7.0" y="24641" x="34631"/>
    <point name="A0009" id="bc24ec68-56ce-482b-af6c-59b0875c90d6" creationRatio="1.0" y="21089" x="44850"/>
    <point name="A0010" id="39fde4a3-8805-46ab-88a1-c1ea1b11cd31" creationRatio="11.0" y="15964" x="52667"/>
  </refPoints>
  - <deLaunay>
    <triangle name="T0000" id="75426426-0720-4292-ad3f-371d656d6fd5" p3="A0044" p2="A0026" p1="A0025"/>
    <triangle name="T0001" id="a34711fb-58db-43da-9434-46d559f30723" p3="A0043" p2="A0025" p1="A0026"/>
    <triangle name="T0002" id="b6d96760-36c3-479c-950f-89c4693bf6ea" p3="A0044" p2="A0002" p1="A0026"/>
    <triangle name="T0003" id="55ec8e4f-8133-4052-844d-3b3d620f89b0" p3="A0044" p2="A0025" p1="A0028"/>
    <triangle name="T0004" id="46384870-e86b-497b-aac2-17219395288f" p3="A0043" p2="A0029" p1="A0025"/>
    <triangle name="T0005" id="3a44ea1c-4ffe-4ca8-ad88-b6a3406ce96e" p3="A0043" p2="A0026" p1="A0023"/>
    <triangle name="T0006" id="e2fae949-168c-40fa-8f3c-31fa55e8c204" p3="A0041" p2="A0026" p1="A0002"/>
    <triangle name="T0007" id="790de0c2-9a43-4ea9-9f28-fb21f5d8a5ba" p3="A0044" p2="A0027" p1="A0002"/>
    <triangle name="T0008" id="33b4284c-9c1e-4984-a381-afbaf6e232d1" p3="A0028" p2="A0025" p1="A0012"/>
    <triangle name="T0009" id="321640a9-714b-43b8-9e64-7843a423e90b" p3="A0044" p2="A0028" p1="A0027"/>
  </deLaunay>
</slide1>
```

(a)

Fig. A.3 Representative examples of metadata screenshots generated during (a) semi-automatic WSI registration (XML file) (b) annotations for cell nuclei segmentation evaluation (ROI file) (c) annotations and features for cell nuclei classification (XML learning sample file)

```

- <LearningSamples Version="1.0">
  <Class Name="Nuclear Fragment" Color="FFFA500"/>
  <Class Name="Conglomerate" Color="FFFFFF00"/>
  <Class Name="Artefact" Color="FFF08080"/>
  <Class Name="Lymphocyte" Color="FF4169E1"/>
  <Class Name="Fibrocyte/border cell of vessels" Color="FFADD8E6"/>
  <Class Name="Epithelial[T]" Color="FF80FF00"/>
  <Class Name="All other cells including blood cells in vessel" Color="FFFF0000"/>
- <Sample Id="906" Class="Nuclear Fragment">
  <Feature Name="MeanChromaticity" Value="0.53482207038304"/>
  <Feature Name="MeanIntensityOnContour" Value="0.484210526315789"/>
  <Feature Name="MeanIntensity" Value="0.484210526315789"/>
  <Feature Name="StdDevIntensityOnContour" Value="0"/>
  <Feature Name="StdDevIntensity" Value="0"/>
  <Feature Name="ContourValue" Value="0.137755184378278"/>
  <Feature Name="GradientFit" Value="0.222222222222222"/>
  <Feature Name="MinimumDistanceToTessellationBorder" Value="0.0349002406379888"/>
  <Feature Name="PixelsAtLayerBorder" Value="0"/>
  <Feature Name="MaximumDistanceToBorder" Value="0.061491457474511"/>
  <Feature Name="ObjectPixels" Value="0.00577528876443822"/>
  <Feature Name="AspectRatioOfBoundingEllipsoid" Value="0.40392661064291"/>
  <Feature Name="MinorAxisOfBoundingEllipsoid" Value="0.0297496859761996"/>
  <Feature Name="MajorAxisOfBoundingEllipsoid" Value="0.0331787038074242"/>
  <Feature Name="AngleOfBoundingEllipsoid" Value="0.460573596770456"/>
  <Feature Name="FormFactorOfContour" Value="0.0324796694856612"/>
  <Feature Name="ConvexityOfContour" Value="0.924353559010614"/>
  <Feature Name="LengthOfContour" Value="0.0216947229676196"/>
  <Feature Name="AreaOfContour" Value="0.0052417532760958"/>
  <Feature Name="FormFactorOfConvexHull" Value="0.0982733497738032"/>
  <Feature Name="LengthOfConvexHull" Value="0.0260731851758618"/>
  <Feature Name="AreaOfConvexHull" Value="0.00446496140457091"/>
  <Feature Name="Feret" Value="0.0318179425987268"/>
  <Feature Name="MinimalRadiusOfEnclosingCenteredCircle" Value="0.0230592437836567"/>
  <Feature Name="MaximalRadiusOfEnclosedCenteredCircle" Value="0.117041147196131"/>
  <Feature Name="Roundness" Value="0.512556213544682"/>
  <Feature Name="FormFactor" Value="0.79166797581584"/>
  <Feature Name="ContrastGLCM" Value="0.101762691146371"/>
  <Feature Name="EntropyGLCM" Value="0.333980861661334"/>
  <Feature Name="EnergyGLCM" Value="0.123160039237633"/>
  <Feature Name="HomogeneityGLCM" Value="0.21522690570261"/>
  <Feature Name="FractalDimension" Value="0.0873574178298831"/>
</Sample>
- <Sample Id="3090" Class="Nuclear Fragment">
  <Feature Name="MeanChromaticity" Value="0.0528500021996748"/>
  <Feature Name="MeanIntensityOnContour" Value="0.0947368421052632"/>
  <Feature Name="MeanIntensity" Value="0.0947368421052632"/>
  <Feature Name="StdDevIntensityOnContour" Value="0"/>
  <Feature Name="StdDevIntensity" Value="0"/>
  <Feature Name="ContourValue" Value="0.143312324929972"/>
  <Feature Name="GradientFit" Value="0.25"/>
  <Feature Name="MinimumDistanceToTessellationBorder" Value="0.0156078620995287"/>
  <Feature Name="PixelsAtLayerBorder" Value="0"/>
  <Feature Name="MaximumDistanceToBorder" Value="0.0849726965747477"/>
  <Feature Name="ObjectPixels" Value="0.00892544627231362"/>
  <Feature Name="AspectRatioOfBoundingEllipsoid" Value="0.664942465878552"/>
  <Feature Name="MinorAxisOfBoundingEllipsoid" Value="0.0609800795296089"/>
  <Feature Name="MajorAxisOfBoundingEllipsoid" Value="0.0297576240003363"/>
  <Feature Name="AngleOfBoundingEllipsoid" Value="0.748628307359114"/>
  <Feature Name="FormFactorOfContour" Value="0.0102473985496022"/>
  <Feature Name="ConvexityOfContour" Value="0.930355075703978"/>
  <Feature Name="LengthOfContour" Value="0.0269994259824338"/>
  <Feature Name="AreaOfContour" Value="0.00840488025305016"/>
  <Feature Name="FormFactorOfConvexHull" Value="0.0488474846805515"/>
  <Feature Name="LengthOfConvexHull" Value="0.0331569613407644"/>
  <Feature Name="AreaOfConvexHull" Value="0.00696231269865294"/>
  <Feature Name="Feret" Value="0.0285379362681978"/>
  <Feature Name="MinimalRadiusOfEnclosingCenteredCircle" Value="0.0304188003117926"/>
  <Feature Name="MaximalRadiusOfEnclosedCenteredCircle" Value="0.149198773361589"/>
  <Feature Name="Roundness" Value="0.764194083729043"/>
  <Feature Name="FormFactor" Value="0.924931460221114"/>
  <Feature Name="ContrastGLCM" Value="0.124710240100589"/>
  <Feature Name="EntropyGLCM" Value="0.394846599444714"/>
  <Feature Name="EnergyGLCM" Value="0.0941418490671467"/>
  <Feature Name="HomogeneityGLCM" Value="0.215704891027608"/>
  <Feature Name="FractalDimension" Value="0.120133426448511"/>
</Sample>
- <LearningSamples Version="1.0">
  <Class Name="Cells" Color="FF00FF40"/>
- <Sample Id="0" Class="Cells">
  <Feature Name="x0" Value="291"/>
  <Feature Name="y0" Value="41"/>
</Sample>
- <Sample Id="1" Class="Cells">
  <Feature Name="x0" Value="322"/>
  <Feature Name="y0" Value="20"/>
</Sample>
- <Sample Id="2" Class="Cells">
  <Feature Name="x0" Value="354"/>
  <Feature Name="y0" Value="12"/>
</Sample>
- <Sample Id="3" Class="Cells">
  <Feature Name="x0" Value="390"/>
  <Feature Name="y0" Value="15"/>
</Sample>
- <Sample Id="4" Class="Cells">
  <Feature Name="x0" Value="261"/>
  <Feature Name="y0" Value="57"/>
</Sample>
- <Sample Id="5" Class="Cells">
  <Feature Name="x0" Value="232"/>
  <Feature Name="y0" Value="71"/>
</Sample>
- <Sample Id="6" Class="Cells">
  <Feature Name="x0" Value="294"/>
  <Feature Name="y0" Value="69"/>
</Sample>
- <Sample Id="7" Class="Cells">
  <Feature Name="x0" Value="297"/>
  <Feature Name="y0" Value="90"/>
</Sample>
- <Sample Id="8" Class="Cells">
  <Feature Name="x0" Value="267"/>
  <Feature Name="y0" Value="133"/>
  <Feature Name="x1" Value="266"/>
  <Feature Name="y1" Value="133"/>
</Sample>

```

(b)

(c)

Detailed Experimental Results

Contents

B.1 Confusion Matrices	215
B.1.1 Comparative Evaluation: Necrosis Detection	215
B.1.2 Comparative Evaluation: Cell Nuclei Classification	216
B.1.3 Comparative Evaluation: Analysis of Cancer Regions	218
B.2 Box and Whisker Diagrams	224
B.2.1 Comparative Evaluation: Necrosis Detection	224
B.2.2 Comparative Evaluation: Cell Nuclei Classification	225
B.2.3 Comparative Evaluation: Analysis of Cancer Regions	226

B.1 Confusion Matrices

The accuracy confusion matrix A averaged over multiple cross validation experiments is demonstrated in the following tables. A color scale is associated with the magnitudes of the matrix starting from lowest values as red and highest as green, in discrete steps of 0.1. The color scale is depicted in Figure B.1. A and P refer to the actual and predicted categories. The last row of each confusion matrix gives the overall classification accuracy (OCA) and balanced classification accuracy (BCA) for the corresponding experimental configuration.

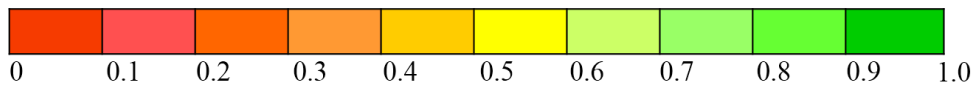


Fig. B.1 Color scale for values in confusion matrices

B.1.1 Comparative Evaluation: Necrosis Detection

Tables B.1 represent the confusion matrices obtained for necrosis detection, where Table B.1(a) shows the results of traditional methods using textural characteristics and SVM classification with

Table B.7 Average accuracy confusion matrix for random forest classification using leave-a-sample-out cross validation

A \ P	Epithelial.	Leukocyte	Fibrocyte	Conglomer.	Fragment	Other cell	Artefact
Epithelial.	0.6042	0.0675	0.1067	0.1292	0.0075	0.0800	0.0050
Leukocyte	0.0350	0.7392	0.0250	0.0308	0.0225	0.1442	0.0033
Fibrocyte	0.0450	0.0000	0.8833	0.0442	0.0042	0.0233	0.0000
Conglomer.	0.1792	0.0292	0.1008	0.6308	0.0150	0.0450	0.0000
Fragment	0.0300	0.0367	0.0400	0.0192	0.5983	0.0800	0.1958
Other cell	0.0988	0.1296	0.0739	0.0548	0.1221	0.4684	0.0523
Artefact	0.0192	0.0201	0.0201	0.0100	0.2107	0.0493	0.6706
OCA=0.6563, BCA=0.6564							

B.1.3 Comparative Evaluation: Analysis of Cancer Regions

The following tables represent the average accuracy confusion matrices for the best performance classification method using both cross validation strategies, namely, k-fold stratified shuffled split ($k = 10$) and leave-a-patient out cross validations. The classification method with highest performance is specified in the respective table captions. A pair of confusion matrices are presented for each configuration in comparative scheme of evaluation (please refer Table 8.2) for analysis of cancer regions. HER2+ refers to HER2 positive tumor, HER2- to HER2 negative tumor and NonT. to non-tumor.

B.1.3.1 Low-level State-of-the-art Features

Table B.8 Average accuracy confusion matrix for GLCM statistics with SVM hierarchical classification (each stage) using (a) k-fold stratified shuffled split (b) leave-a-patient-out cross validation

(a)

Stage 1(AIP)	NonT.	Others	Stage 2(AIP)	HER2+	HER2-
NonT.	0.9373	0.0627	HER2+	0.7940	0.2060
Others	0.0417	0.9583	HER2-	0.2168	0.7832
OCA=0.8204, BCA=0.8173					

(b)

Stage 1(AIP)	NonT.	Others	Stage 2(AIP)	HER2+	HER2-
NonT.	0.7418	0.2582	HER2+	0.5566	0.4434
Others	0.1554	0.8446	HER2-	0.6154	0.3846
OCA=0.5288, BCA=0.5132					

Table B.9 Average accuracy confusion matrix for Gabor filter-bank responses with SVM hierarchical classification (each stage) using (a) *k*-fold stratified shuffled split (b) leave-a-patient-out cross validation

(a)

Stage 1(AIP)	NonT.	Others	Stage 2(AIP)	HER2+	HER2-
NonT.	0.8993	0.1007	HER2+	0.7603	0.2397
Others	0.0491	0.9509	HER2-	0.1920	0.8080
OCA=0.7987, BCA=0.7971					

(b)

Stage 1(AIP)	NonT.	Others	Stage 2(AIP)	HER2+	HER2-
NonT.	0.6634	0.3366	HER2+	0.4503	0.5497
Others	0.2434	0.7566	HER2-	0.5747	0.4253
OCA=0.4558, BCA=0.4280					

Table B.10 Average accuracy confusion matrix for local binary patterns with SVM single-stage classification using (a) *k*-fold stratified shuffled split (b) leave-a-patient-out cross validation

(a)

A \ P	HER2+	HER2-	NonT.
HER2+	0.7514	0.1733	0.0753
HER2-	0.1174	0.7455	0.1372
NonT.	0.0595	0.0660	0.8745
OCA=0.7970, BCA=0.7903			

(b)

A \ P	HER2+	HER2-	NonT.
HER2+	0.4413	0.2794	0.2794
HER2-	0.4008	0.2355	0.3636
NonT.	0.1340	0.1405	0.7255
OCA=0.4910, BCA=0.4645			

Table B.11 Average accuracy confusion matrix for Varma-Zisserman textons with SVM single-stage classification (a) *k*-fold stratified shuffled split (b) leave-a-patient-out cross validation

(a)

A \ P	HER2+	HER2-	NonT.
HER2+	0.8826	0.1020	0.0154
HER2-	0.1091	0.8364	0.0545
NonT.	0.0111	0.0359	0.9529
OCA=0.8956, BCA=0.8906			

(b)

A \ P	HER2+	HER2-	NonT.
HER2+	0.4130	0.4291	0.1579
HER2-	0.4463	0.2479	0.3058
NonT.	0.0425	0.2124	0.7451
OCA=0.4865, BCA=0.4430			

Table B.12 Average accuracy confusion matrix for gray histograms with random forests single-stage classification (a) *k*-fold stratified shuffled split (b) leave-a-patient-out cross validation

(a)

A \ P	HER2+	HER2-	NonT.
HER2+	0.7816	0.1712	0.0472
HER2-	0.1633	0.7425	0.0942
NonT.	0.0400	0.0323	0.9277
OCA=0.8265, BCA=0.8173			

(b)

A \ P	HER2+	HER2-	NonT.
HER2+	0.5101	0.3401	0.1498
HER2-	0.3471	0.4215	0.2314
NonT.	0.0654	0.0784	0.8562
OCA=0.6128, BCA=0.5639			

Table B.13 Average accuracy confusion matrix for HSV histograms with AdaBoost single-stage classification using (a) *k*-fold stratified shuffled split (b) leave-a-patient-out cross validation

(a)				(b)			
A \ P	HER2+	HER2-	NonT.	A \ P	HER2+	HER2-	NonT.
HER2+	0.8493	0.1285	0.0222	HER2+	0.6235	0.2996	0.0769
HER2-	0.0947	0.8474	0.0578	HER2-	0.3388	0.4174	0.2438
NonT.	0.0126	0.0225	0.9649	NonT.	0.0686	0.0523	0.8791
<i>OCA</i> =0.8933, <i>BCA</i> =0.8879				<i>OCA</i> =0.6545, <i>BCA</i> =0.6098			

Table B.14 Average accuracy confusion matrix for RGB histograms with SVM hierarchical classification using (a) *k*-fold stratified shuffled split (b) leave-a-patient-out cross validation

(a)					
Stage 1(AIP)	NonT.	Others	Stage 2(AIP)	HER2+	HER2-
NonT.	0.9719	0.0281	HER2+	0.9210	0.0790
Others	0.0143	0.9857	HER2-	0.1069	0.8931
<i>OCA</i> =0.9177, <i>BCA</i> =0.9203					
(b)					
Stage 1(AIP)	NonT.	Others	Stage 2(AIP)	HER2+	HER2-
NonT.	0.7810	0.2190	HER2+	0.6903	0.3097
Others	0.1493	0.8507	HER2-	0.4378	0.5622
<i>OCA</i> =0.6245, <i>BCA</i> =0.5928					

Table B.15 Average accuracy confusion matrix for other color-based measurements with random forests single-stage classification using (a) *k*-fold stratified shuffled split (b) leave-a-patient-out cross validation

(a)				(b)			
A \ P	HER2+	HER2-	NonT.	A \ P	HER2+	HER2-	NonT.
HER2+	0.7592	0.2024	0.0384	HER2+	0.5749	0.3320	0.0931
HER2-	0.1967	0.6792	0.1242	HER2-	0.4380	0.2727	0.2893
NonT.	0.0303	0.0555	0.9142	NonT.	0.0458	0.1830	0.7712
<i>OCA</i> =0.7953, <i>BCA</i> =0.7842				<i>OCA</i> =0.5515, <i>BCA</i> =0.4951			

B.1.3.2 High-level State-of-the-art Features

Table B.16 Average accuracy confusion matrix for Voronoi-Delaunay method with random forest single-stage classification using (a) *k*-fold stratified shuffled split (b) leave-a-patient-out cross validation

(a)				(b)			
A \ P	HER2+	HER2-	NonT.	A \ P	HER2+	HER2-	NonT.
HER2+	0.6856	0.2384	0.0760	HER2+	0.6194	0.2794	0.1012
HER2-	0.3500	0.4617	0.1883	HER2-	0.3967	0.3678	0.2355
NonT.	0.1000	0.1032	0.7968	NonT.	0.1144	0.1275	0.7582
<i>OCA</i> =0.6615, <i>BCA</i> =0.6480				<i>OCA</i> =0.5889, <i>BCA</i> =0.5756			

B.1.3.3 High-level Handcrafted Features

Table B.17 Average accuracy confusion matrix for only vertex attributes (object-level features) of cell nuclei ARG with random forest hierarchical classification (each stage) using (a) k -fold stratified shuffled split (b) leave-a-patient-out cross validation

(a)					
Stage 1(AIP)	NonT.	Others	Stage 2(AIP)	HER2+	HER2-
NonT.	0.8742	0.1258	HER2+	0.7262	0.2738
Others	0.0531	0.9469	HER2-	0.2892	0.7108
$OCA=0.7495, BCA=0.7472$					
(b)					
Stage 1(AIP)	NonT.	Others	Stage 2(AIP)	HER2+	HER2-
NonT.	0.8039	0.1961	HER2+	0.5149	0.4851
Others	0.0798	0.9202	HER2-	0.5550	0.4450
$OCA=0.5761, BCA=0.5821$					

Table B.18 Average accuracy confusion matrix for only vertex identities (tissue composition) of the cell nuclei ARG with SVM single-stage classification using (a) k -fold stratified shuffled split (b) leave-a-patient-out cross validation

(a)				(b)			
A \ P	HER2+	HER2-	NonT.	A \ P	HER2+	HER2-	NonT.
HER2+	0.7490	0.2186	0.0324	HER2+	0.4818	0.4453	0.0729
HER2-	0.2397	0.6107	0.1496	HER2-	0.4628	0.3182	0.2190
NonT.	0.0261	0.0902	0.8837	NonT.	0.0458	0.1993	0.7549
$OCA=0.7587, BCA=0.7477$				$OCA=0.5390, BCA=0.5302$			

Table B.19 Average accuracy confusion matrix for handcrafted $gARG[r_F]$ method with SVM single-stage classification using (a) k -fold stratified shuffled split (b) leave-a-patient-out cross validation

(a)				(b)			
A \ P	HER2+	HER2-	NonT.	A \ P	HER2+	HER2-	NonT.
HER2+	0.7765	0.1895	0.0340	HER2+	0.6154	0.3320	0.0526
HER2-	0.2669	0.6405	0.0926	HER2-	0.4132	0.4421	0.1446
NonT.	0.0693	0.0111	0.9196	NonT.	0.0294	0.1405	0.8301
$OCA=0.7902, BCA=0.7794$				$OCA=0.6448, BCA=0.6361$			

Table B.20 Average accuracy confusion matrix for handcrafted $gARG[r_A]$ method with SVM single-stage classification using (a) k -fold stratified shuffled split (b) leave-a-patient-out cross validation

(a)				(b)			
A \ P	HER2+	HER2-	NonT.	A \ P	HER2+	HER2-	NonT.
HER2+	0.7555	0.2073	0.0372	HER2+	0.5709	0.3522	0.0769
HER2-	0.2587	0.6521	0.0893	HER2-	0.4256	0.4132	0.1612
NonT.	0.0131	0.0627	0.9242	NonT.	0.0490	0.1634	0.7876
$OCA=0.7890, BCA=0.7773$				$OCA=0.6053, BCA=0.5946$			

Table B.21 Average accuracy confusion matrix for handcrafted $nsARG[r_F]$ method with SVM single-stage classification using (a) k -fold stratified shuffled split (b) leave-a-patient-out cross validation

(a)				(b)			
A \ P	HER2+	HER2-	NonT.	A \ P	HER2+	HER2-	NonT.
HER2+	0.6510	0.3077	0.0413	HER2+	0.4858	0.4534	0.0607
HER2-	0.3149	0.5463	0.1388	HER2-	0.5083	0.3058	0.1860
NonT.	0.0170	0.0588	0.9242	NonT.	0.0229	0.1111	0.8660
$OCA=0.7243, BCA=0.7073$				$OCA=0.5749, BCA=0.5389$			

Table B.22 Average accuracy confusion matrix for handcrafted $nsARG[r_A]$ method with SVM hierarchical classification (each stage) using (a) k -fold stratified shuffled split (b) leave-a-patient-out cross validation

(a)					
Stage 1(A P)	NonT.	Others	Stage 2(A P)	HER2+	HER2-
NonT.	0.9062	0.0938	HER2+	0.7176	0.2824
Others	0.0711	0.9289	HER2-	0.3159	0.6841
$OCA=0.7343, BCA=0.7281$					
(b)					
Stage 1(A P)	NonT.	Others	Stage 2(A P)	HER2+	HER2-
NonT.	0.7908	0.2092	HER2+	0.5622	0.4378
Others	0.1125	0.8875	HER2-	0.5408	0.4592
$OCA=0.5757, BCA=0.5551$					

Table B.23 Average accuracy confusion matrix for handcrafted $ncARG[r_F]$ method with SVM single-stage classification using (a) k -fold stratified shuffled split (b) leave-a-patient-out cross validation

(a)				(b)			
A \ P	HER2+	HER2-	NonT.	A \ P	HER2+	HER2-	NonT.
HER2+	0.7466	0.2065	0.0470	HER2+	0.6113	0.3279	0.0607
HER2-	0.2711	0.6174	0.1116	HER2-	0.4298	0.4215	0.1488
NonT.	0.0209	0.0621	0.9170	NonT.	0.0458	0.1078	0.8464
$OCA=0.7729, BCA=0.7603$				$OCA=0.6450, BCA=0.6320$			

Table B.24 Average accuracy confusion matrix for handcrafted $ncARG[r_A]$ method with random forest hierarchical classification using (a) k -fold stratified shuffled split (b) leave-a-patient-out cross validation

(a)					
Stage 1(A P)	NonT.	Others	Stage 2(A P)	HER2+	HER2-
NonT.	0.8297	0.1703	HER2+	0.7643	0.2357
Others	0.0510	0.9490	HER2-	0.2752	0.7248
$OCA=0.7480, BCA=0.7495$					
(b)					
Stage 1(A P)	NonT.	Others	Stage 2(A P)	HER2+	HER2-
NonT.	0.8197	0.1803	HER2+	0.5872	0.4128
Others	0.0757	0.9243	HER2-	0.6095	0.3905
$OCA=0.5921, BCA=0.5982$					

Table B.25 Average accuracy confusion matrix for handcrafted $ncARG_{v+}[r_F]$ method with SVM single-stage classification using (a) k -fold stratified shuffled split (b) leave-a-patient-out cross validation

(a)				(b)			
A \ P	HER2+	HER2-	NonT.	A \ P	HER2+	HER2-	NonT.
HER2+	0.7733	0.1951	0.0316	HER2+	0.6194	0.3279	0.0526
HER2-	0.2579	0.6521	0.0901	HER2-	0.4298	0.4256	0.1446
NonT.	0.0105	0.0680	0.9216	NonT.	0.0327	0.1438	0.8235
$OCA=0.7935, BCA=0.7823$				$OCA=0.6394, BCA=0.6314$			

Table B.26 Average accuracy confusion matrix for handcrafted $ncARG_{v+}[r_A]$ method with SVM single-stage classification using (a) k -fold stratified shuffled split (b) leave-a-patient-out cross validation

(a)				(b)			
A \ P	HER2+	HER2-	NonT.	A \ P	HER2+	HER2-	NonT.
HER2+	0.7684	0.1927	0.0389	HER2+	0.5668	0.3482	0.0850
HER2-	0.2405	0.6760	0.0835	HER2-	0.4711	0.3595	0.1694
NonT.	0.0111	0.0610	0.9279	NonT.	0.0492	0.1574	0.7934
$OCA=0.8015, BCA=0.7908$				$OCA=0.5899, BCA=0.5751$			

B.1.3.4 Combination of Features

Table B.27 Average accuracy confusion matrix for selected subset of hybrid low-level features with SVM hierarchical classification using (a) k -fold stratified shuffled split (b) leave-a-patient-out cross validation

(a)						(b)					
Stage 1(A P)	NonT.	Others	Stage 2(A P)	HER2+	HER2-	Stage 1(A P)	NonT.	Others	Stage 2(A P)	HER2+	HER2-
NonT.	0.9843	0.0157	HER2+	0.9224	0.0776	NonT.	0.8105	0.1895	HER2+	0.6228	0.3772
Others	0.0135	0.9865	HER2-	0.0776	0.9224	Others	0.1575	0.8425	HER2-	0.5754	0.4246
$OCA=0.9177, BCA=0.9203$						$OCA=0.5858, BCA=0.5717$					

Table B.28 Average accuracy confusion matrix for selected subset of hybrid low-level and high-level features with SVM single-stage classification using (a) k -fold stratified shuffled split (b) leave-a-patient-out cross validation

(a)				(b)			
A \ P	HER2+	HER2-	NonT.	A \ P	HER2+	HER2-	NonT.
HER2+	0.8964	0.0931	0.0105	HER2+	0.6397	0.3401	0.0202
HER2-	0.0992	0.8636	0.0372	HER2-	0.4091	0.3471	0.2438
NonT.	0.0000	0.0098	0.9902	NonT.	0.0163	0.0948	0.8889
$OCA=0.9226, BCA=0.9167$				$OCA=0.6410, BCA=0.6109$			

B.1.3.5 Deep Learning Methods

Table B.29 Average accuracy confusion matrix for AlexNet CNN framework in single-stage classification using (a) *k*-fold stratified shuffled split (b) leave-a-patient-out cross validation

(a)

A \ P		P		
		HER2+	HER2-	NonT.
HER2+	0.9625	0.0349	0.0026	
HER2-	0.0355	0.9541	0.0104	
NonT.	0.0034	0.0116	0.9850	
OCA = BCA=0.9674				

(b)

A \ P		P		
		HER2+	HER2-	NonT.
HER2+	0.5422	0.3581	0.0997	
HER2-	0.3183	0.3679	0.3138	
NonT.	0.1184	0.2999	0.5817	
OCA = BCA=0.4973				

Table B.30 Average accuracy confusion matrix for proposed CNN architecture in single-stage classification using (a) *k*-fold stratified shuffled split (b) leave-a-patient-out cross validation

(a)

		P		
		HER2+	HER2-	NonT.
A				
HER2+		0.8381	0.1427	0.0191
HER2-		0.1779	0.7701	0.0519
NonT.		0.0178	0.0585	0.9237
		OCA = BCA=0.8441		

(b)

		P		
		HER2+	HER2-	NonT.
A				
HER2+		0.5577	0.3798	0.0624
HER2-		0.3615	0.4163	0.2223
NonT.		0.0486	0.2237	0.7277
		OCA = BCA=0.5672		

Table B.31 Average accuracy confusion matrix for ensemble of CNNs in single-stage classification using (a) *k*-fold stratified shuffled split (b) leave-a-patient-out cross validation

(a)

		P		
		HER2+	HER2-	NonT.
A \	HER2+	0.9281	0.0360	0.0358
	HER2-	0.0457	0.9127	0.0416
	NonT.	0.0050	0.0115	0.9835
	OCA = BCA=0.9440			

(b)

		P		
		HER2+	HER2-	NonT.
A \	HER2+	0.5870	0.3253	0.0877
	HER2-	0.4063	0.2908	0.3029
	NonT.	0.0731	0.2308	0.6961
	OCA = BCA=0.5246			

B.2 Box and Whisker Diagrams

The box and whisker diagrams of average per-class classification accuracy presenting a comparison of classification performance of the methods explored in the various stages of the research pipeline, are illustrated in the following sections.

B.2.1 Comparative Evaluation: Necrosis Detection

The box and whisker diagrams of average per-class classification accuracy, comparatively depicting the classification performance of the three methods, namely, the traditional method using textural feature extraction followed by SVM classification with discriminative thresholds, and deep learning methods *i.e.* Alexnet CNN framework, proposed CNN architecture and ensemble of CNNs are summarized in Figure B.2. The performance is quantitatively evaluated using *k*-fold stratified shuffled split ($k = 10$ for SVM-based method and $k = 3$ for deep learning) cross validation.

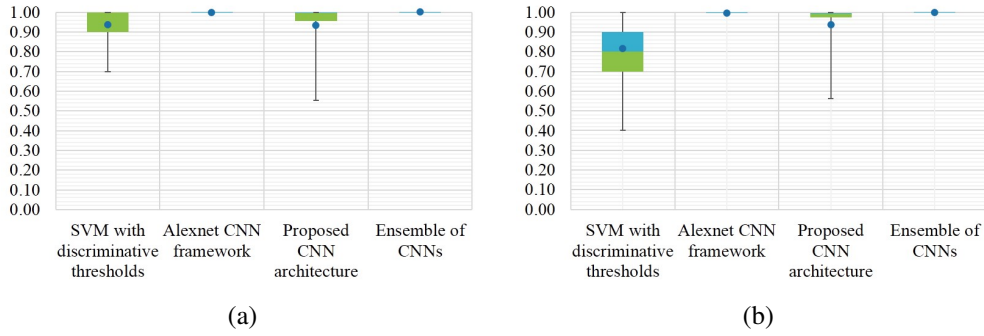


Fig. B.2 Box and whisker diagrams of average per-class classification accuracy for necrosis detection experiments using k -fold stratified shuffled split cross validation for (a) necrosis (b) non-necrosis.

B.2.2 Comparative Evaluation: Cell Nuclei Classification

The box and whisker diagrams of average per-class classification accuracy showing a comparison of classification performance of the three machine learning methods *i.e.* SVM classification method, AdaBoost method and random forests, evaluated using k -fold stratified shuffled split cross validation ($k = 3$) are given in Figure B.3.

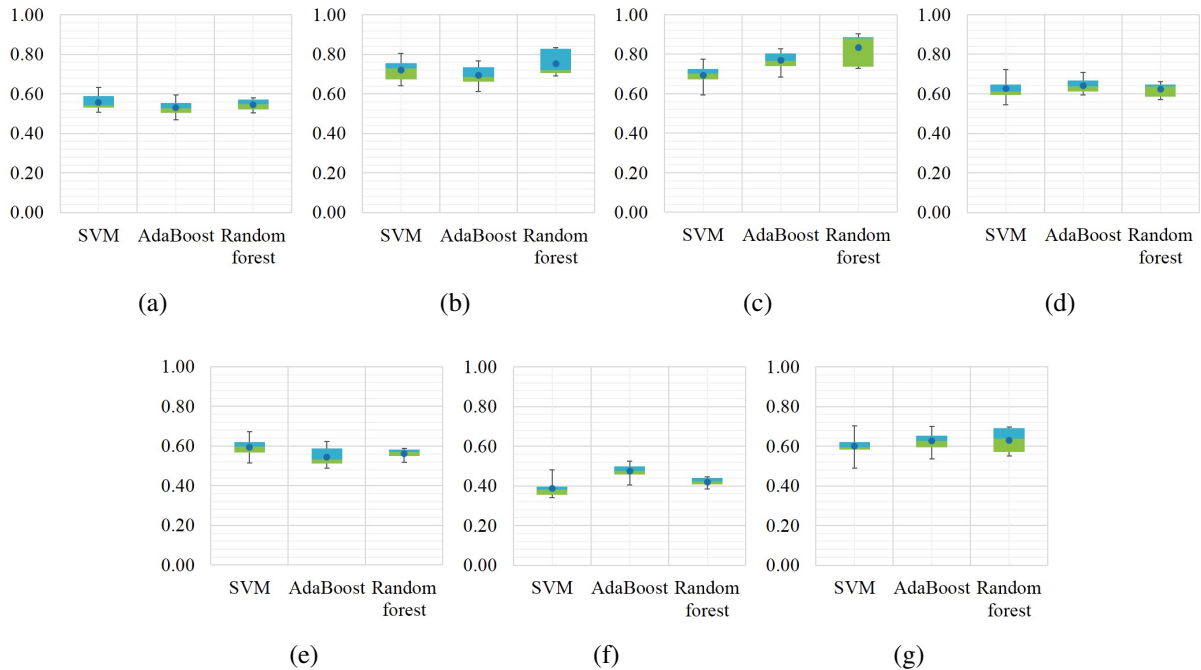


Fig. B.3 Box and whisker diagrams of average per-class classification accuracy for cell nuclei classification experiments using k -fold stratified shuffled split cross validation for a) epithelial cell (b) leukocyte (c) fibrocyte (or border cell) (d) conglomerate (e) fragment (f) other cell (including blood cell in vessel) (g) artefact.

B.2.3 Comparative Evaluation: Analysis of Cancer Regions

The box and whisker diagrams of average per-class classification accuracy for each configuration in the scheme of evaluation (please refer Table 8.2), showing a comparison of the classification performance of the traditional machine learning methods *i.e.* SVM, AdaBoost and random forests, evaluated using k-fold stratified shuffled split ($k = 10$) and leave-a-patient-out cross validations are shown below. SC, HC refer to single-stage classification and hierarchical classification, and AB, RF are AdaBoost and random forests respectively.

B.2.3.1 Low-level State-of-the-art Features

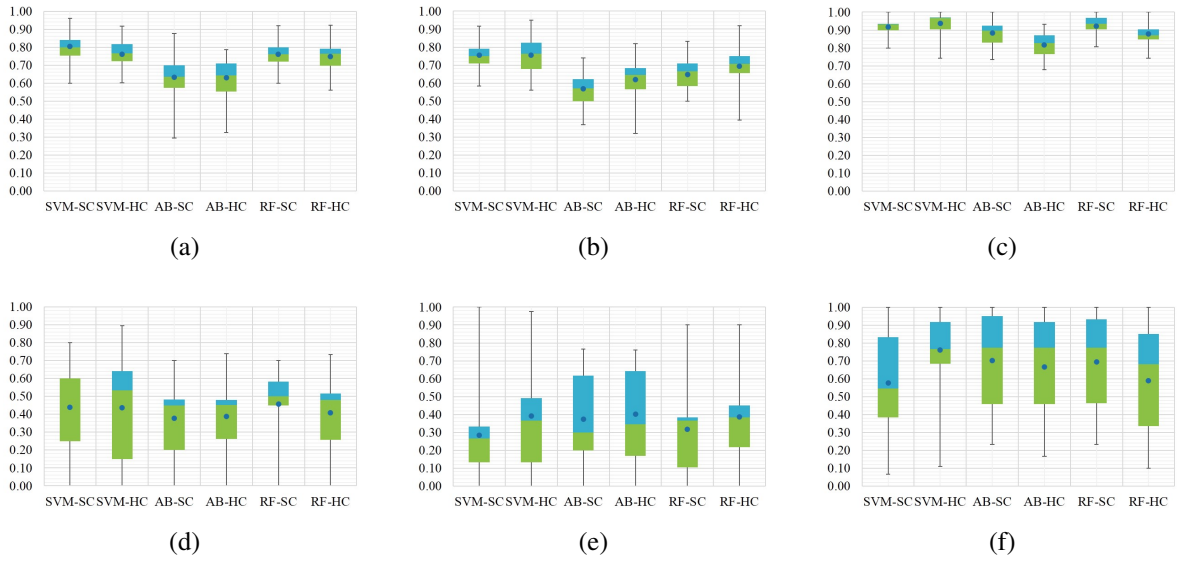


Fig. B.4 Box and whisker diagrams of average per-class classification accuracy of GLCM statistics for (a) HER2 positive tumor (b) HER2 negative tumor (c) Non-tumor using k-fold stratified shuffled split and (d),(e),(f) using leave-a-patient-out cross validations.

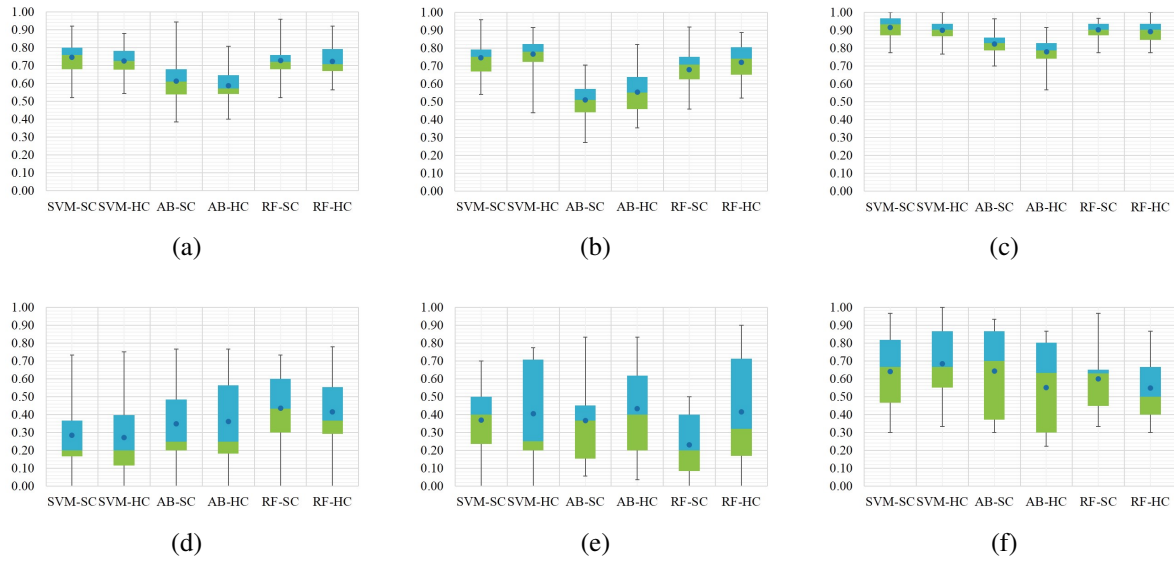


Fig. B.5 Box and whisker diagrams of average per-class classification accuracy of Gabor filter-bank responses for (a) HER2 positive tumor (b) HER2 negative tumor (c) Non-tumor using k -fold stratified shuffled split and (d),(e),(f) using leave-a-patient-out cross validations.

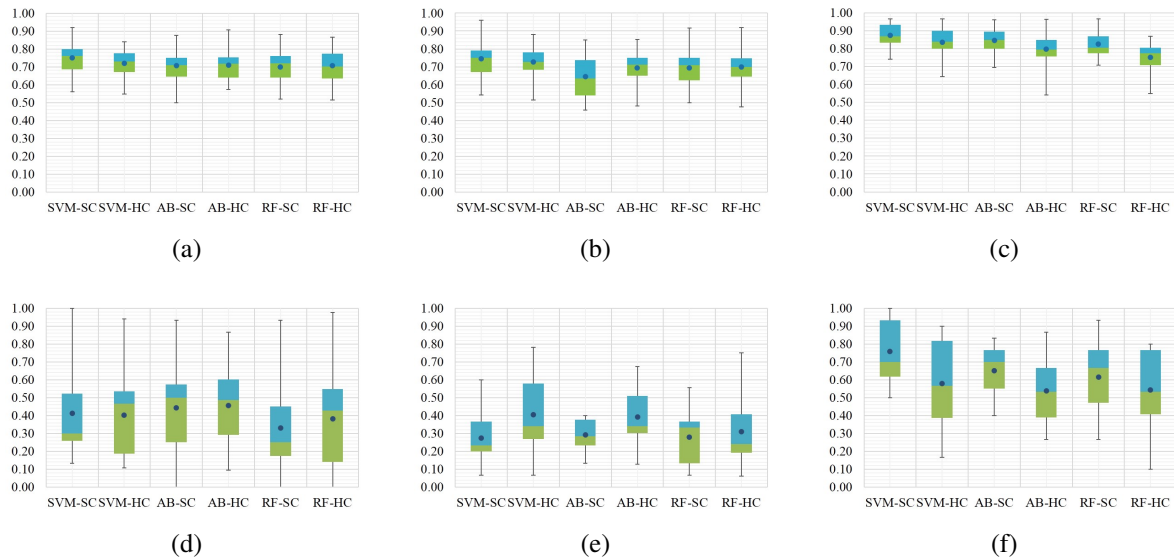


Fig. B.6 Box and whisker diagrams of average per-class classification accuracy of local binary patterns for (a) HER2 positive tumor (b) HER2 negative tumor (c) Non-tumor using k -fold stratified shuffled split and (d),(e),(f) using leave-a-patient-out cross validations.

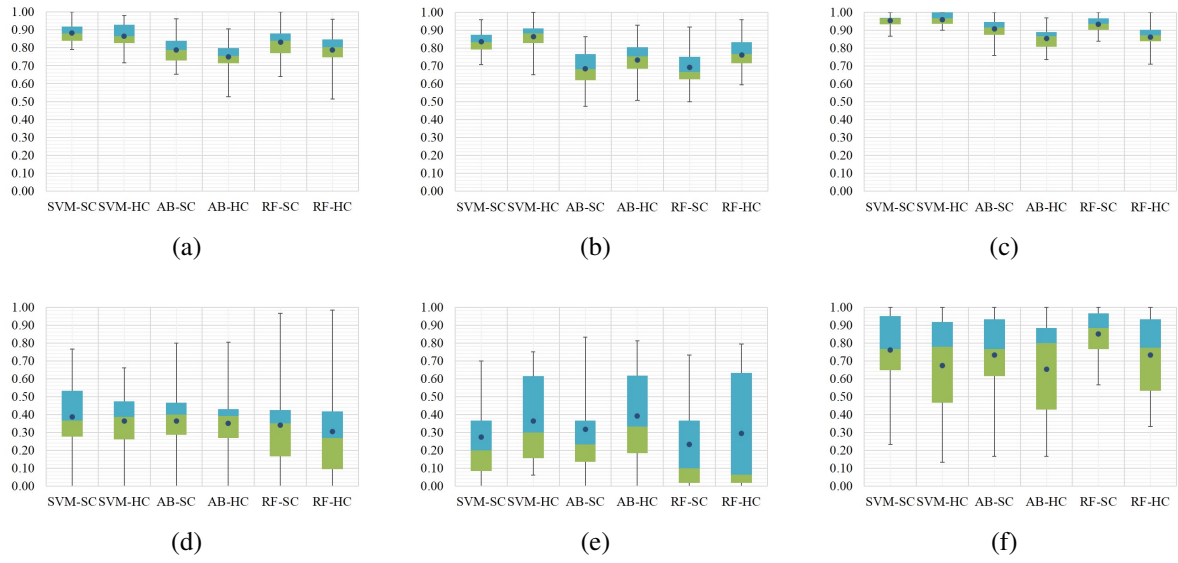


Fig. B.7 Box and whisker diagrams of average per-class classification accuracy of Varma-Zissserman textons for (a) HER2 positive tumor (b) HER2 negative tumor (c) Non-tumor using k -fold stratified shuffled split and (d),(e),(f) using leave-a-patient-out cross validations.

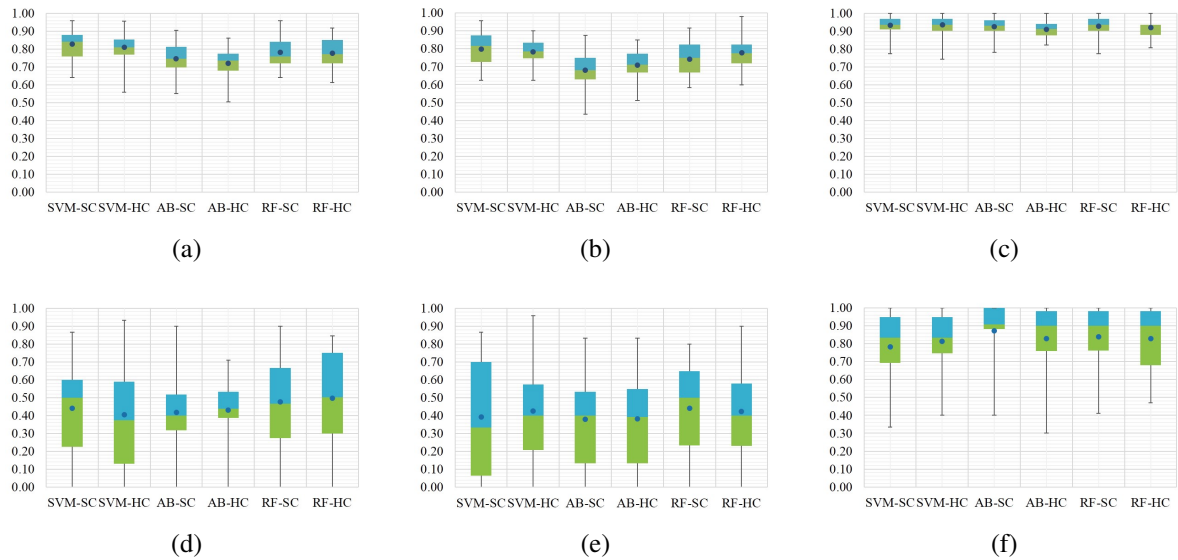


Fig. B.8 Box and whisker diagrams of average per-class classification accuracy of gray histograms for (a) HER2 positive tumor (b) HER2 negative tumor (c) Non-tumor using k -fold stratified shuffled split and (d),(e),(f) using leave-a-patient-out cross validations.

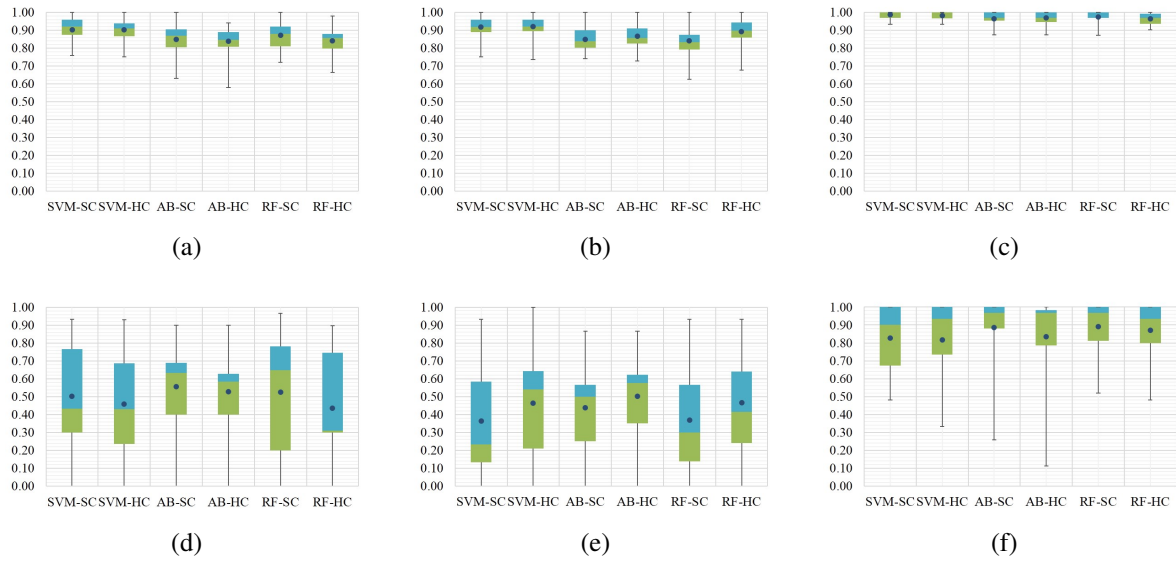


Fig. B.9 Box and whisker diagrams of average per-class classification accuracy of HSV histograms for (a) HER2 positive tumor (b) HER2 negative tumor (c) Non-tumor using k -fold stratified shuffled split and (d),(e),(f) using leave-a-patient-out cross validations.

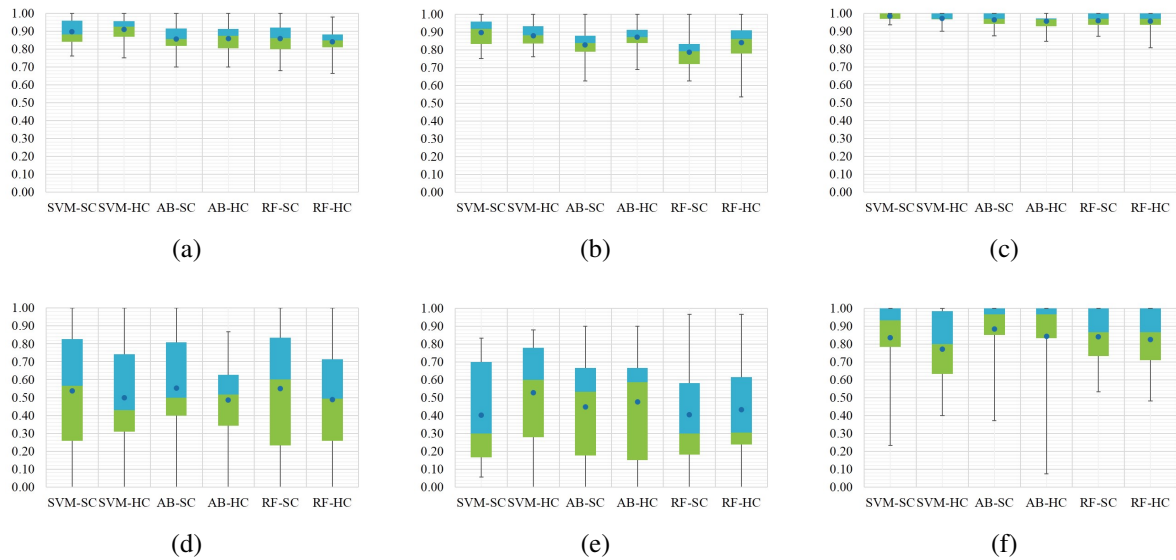


Fig. B.10 Box and whisker diagrams of average per-class classification accuracy of RGB histograms for (a) HER2 positive tumor (b) HER2 negative tumor (c) Non-tumor using k -fold stratified shuffled split and (d),(e),(f) using leave-a-patient-out cross validations.

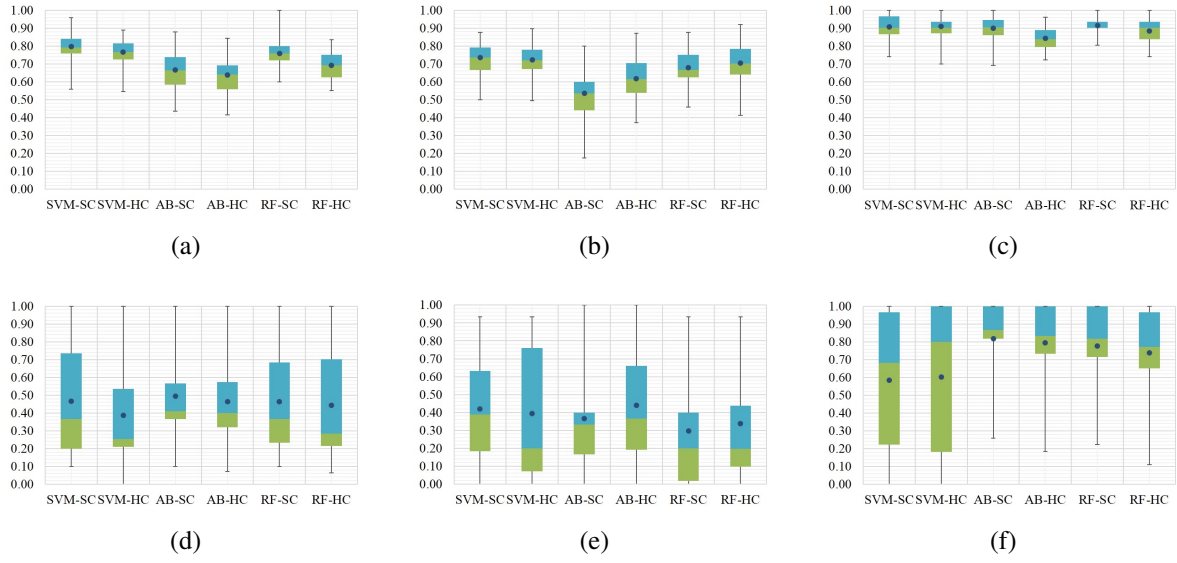


Fig. B.11 Box and whisker diagrams of average per-class classification accuracy of other color-based measurements for (a) HER2 positive tumor (b) HER2 negative tumor (c) Non-tumor using k -fold stratified shuffled split and (d),(e),(f) using leave-a-patient-out cross validations.

B.2.3.2 High-level State-of-the-art Features

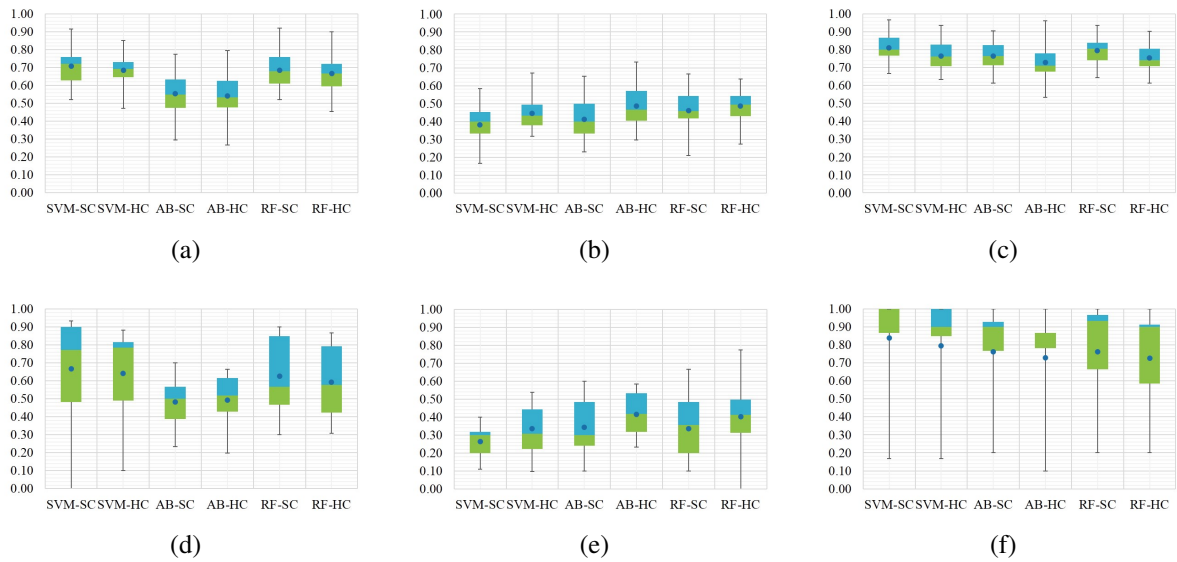


Fig. B.12 Box and whisker diagrams of average per-class classification accuracy for the Voronoi-Delaunay method for (a) HER2 positive tumor (b) HER2 negative tumor (c) Non-tumor using k -fold stratified shuffled split and (d),(e),(f) using leave-a-patient-out cross validations.

B.2.3.3 High-level Handcrafted Features

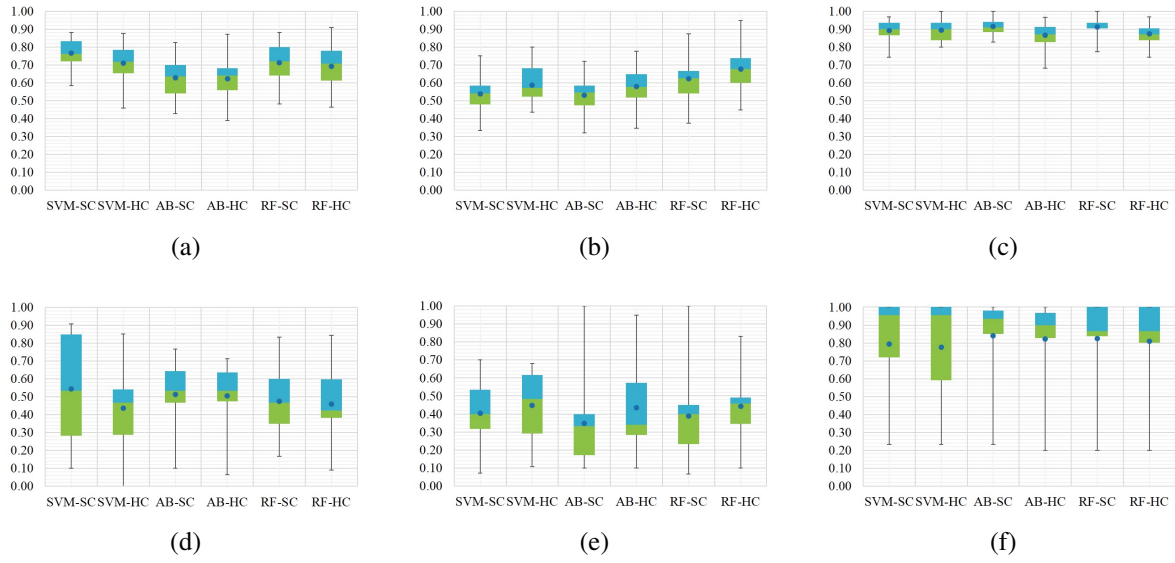


Fig. B.13 Box and whisker diagrams of average per-class classification accuracy for only vertex attributes of cell nuclei ARG for (a) HER2 positive tumor (b) HER2 negative tumor (c) Non-tumor using k -fold stratified shuffled split and (d),(e),(f) using leave-a-patient-out cross validations.

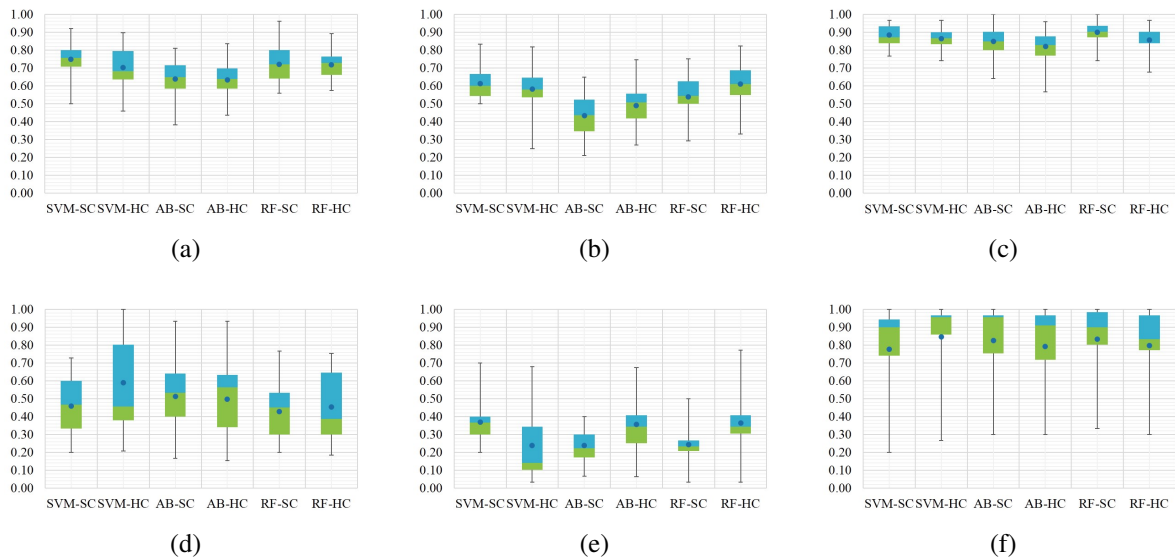


Fig. B.14 Box and whisker diagrams of average per-class classification accuracy for only vertex identities (or tissue composition) of cell nuclei ARG for (a) HER2 positive tumor (b) HER2 negative tumor (c) Non-tumor using k -fold stratified shuffled split and (d),(e),(f) using leave-a-patient-out cross validations.

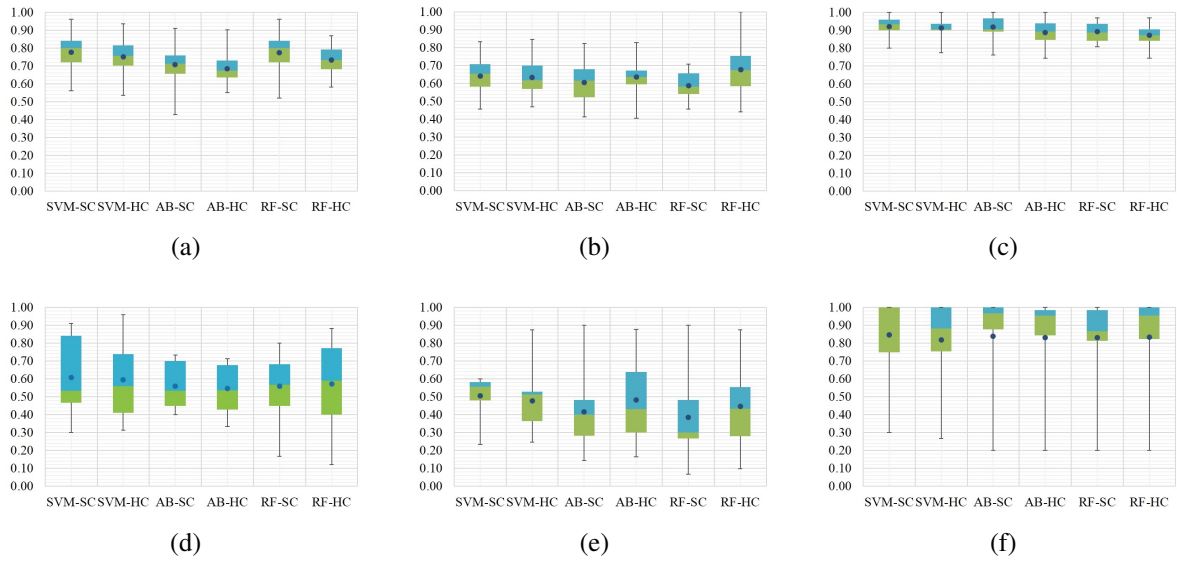


Fig. B.15 Box and whisker diagrams of average per-class classification accuracy for handcrafted $gARG[r_F]$ method for (a) HER2 positive tumor (b) HER2 negative tumor (c) Non-tumor using k -fold stratified shuffled split and (d),(e),(f) using leave-a-patient-out cross validations.

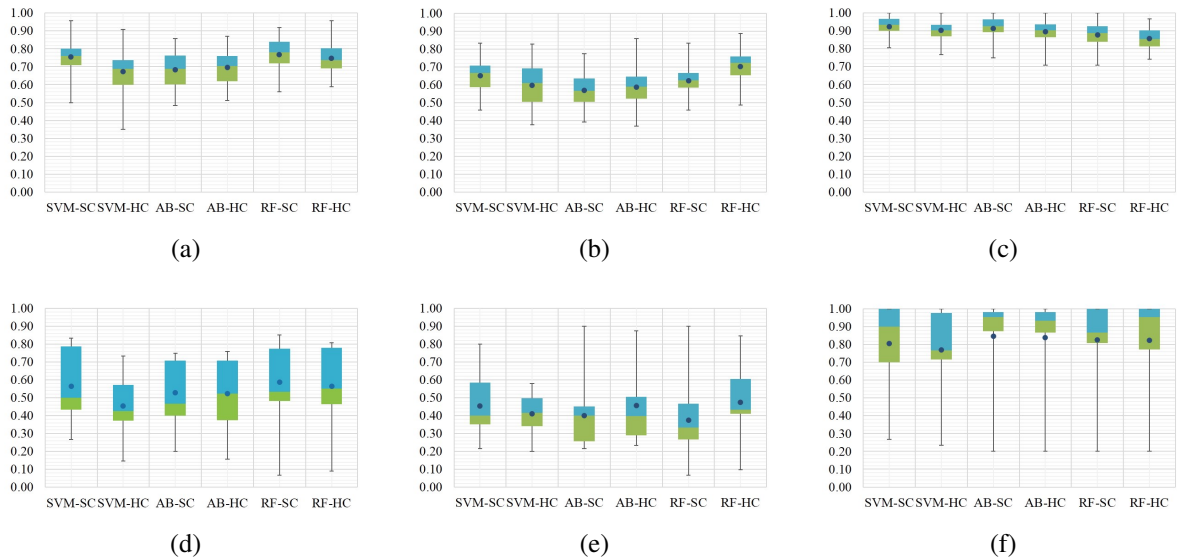


Fig. B.16 Box and whisker diagrams of average per-class classification accuracy for handcrafted $gARG[r_A]$ method for (a) HER2 positive tumor (b) HER2 negative tumor (c) Non-tumor using k -fold stratified shuffled split and (d),(e),(f) using leave-a-patient-out cross validations.

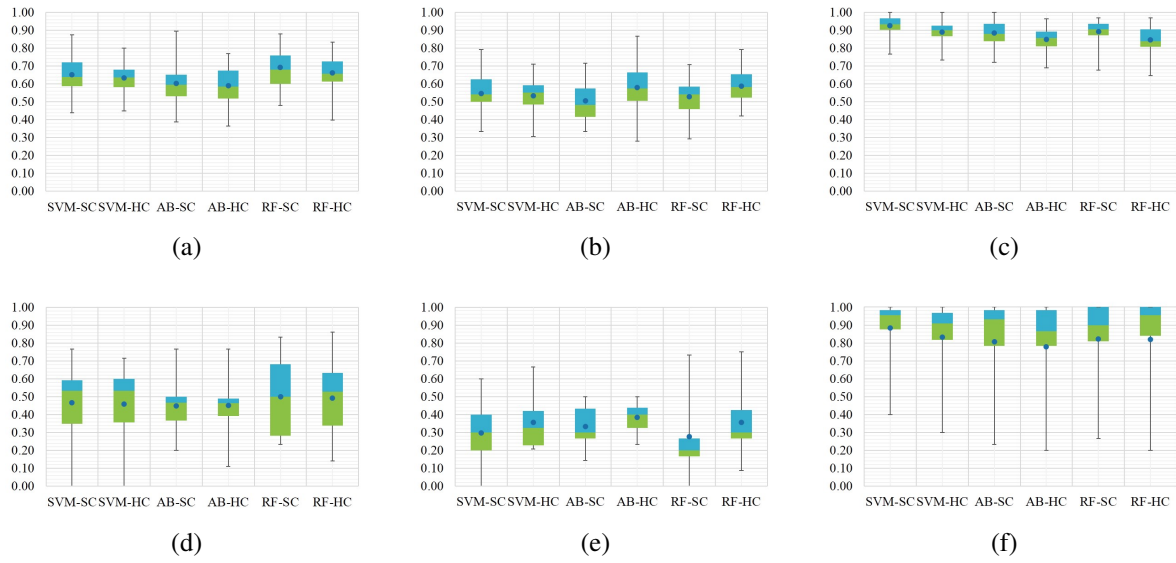


Fig. B.17 Box and whisker diagrams of average per-class classification accuracy for handcrafted $nsARG[r_F]$ method for (a) HER2 positive tumor (b) HER2 negative tumor (c) Non-tumor using k -fold stratified shuffled split and (d),(e),(f) using leave-a-patient-out cross validations.

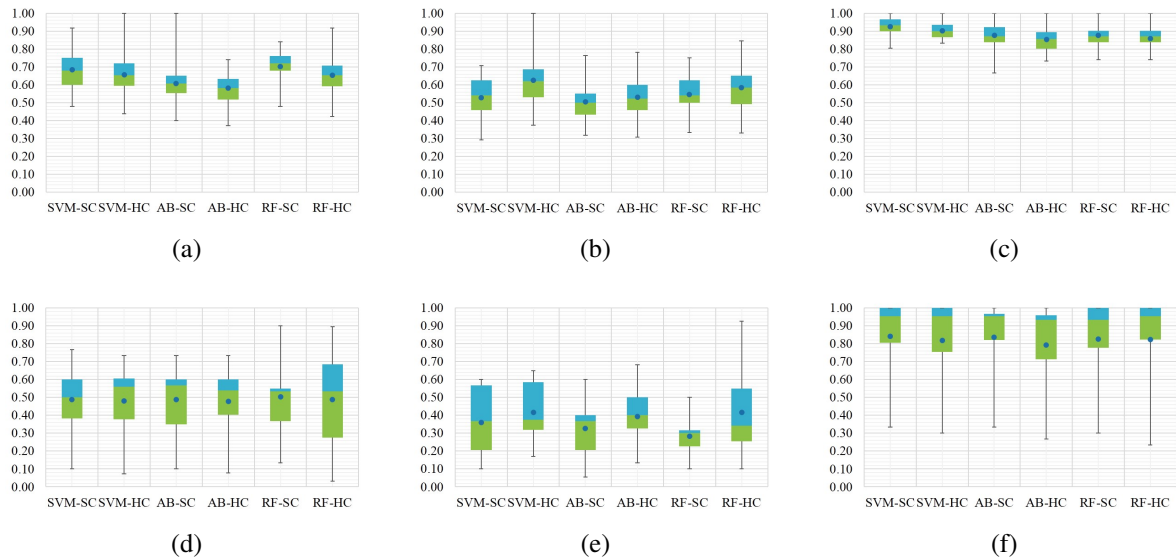


Fig. B.18 Box and whisker diagrams of average per-class classification accuracy for handcrafted $nsARG[r_A]$ method for (a) HER2 positive tumor (b) HER2 negative tumor (c) Non-tumor using k -fold stratified shuffled split and (d),(e),(f) using leave-a-patient-out cross validations.

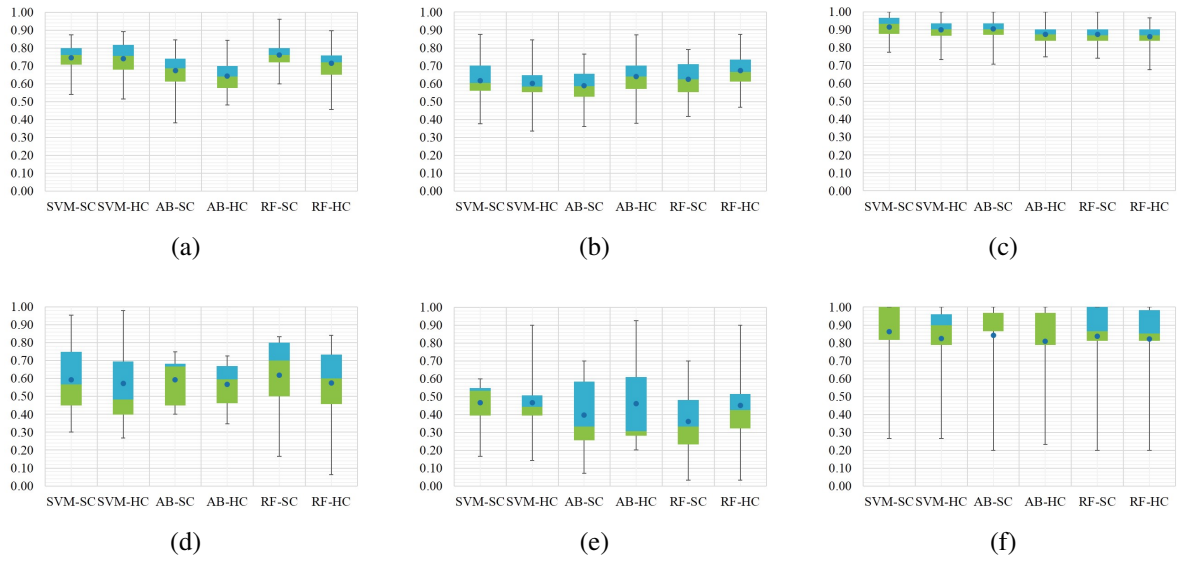


Fig. B.19 Box and whisker diagrams of average per-class classification accuracy for handcrafted $ncARG[r_F]$ method for (a) HER2 positive tumor (b) HER2 negative tumor (c) Non-tumor using k -fold stratified shuffled split and (d),(e),(f) using leave-a-patient-out cross validations.

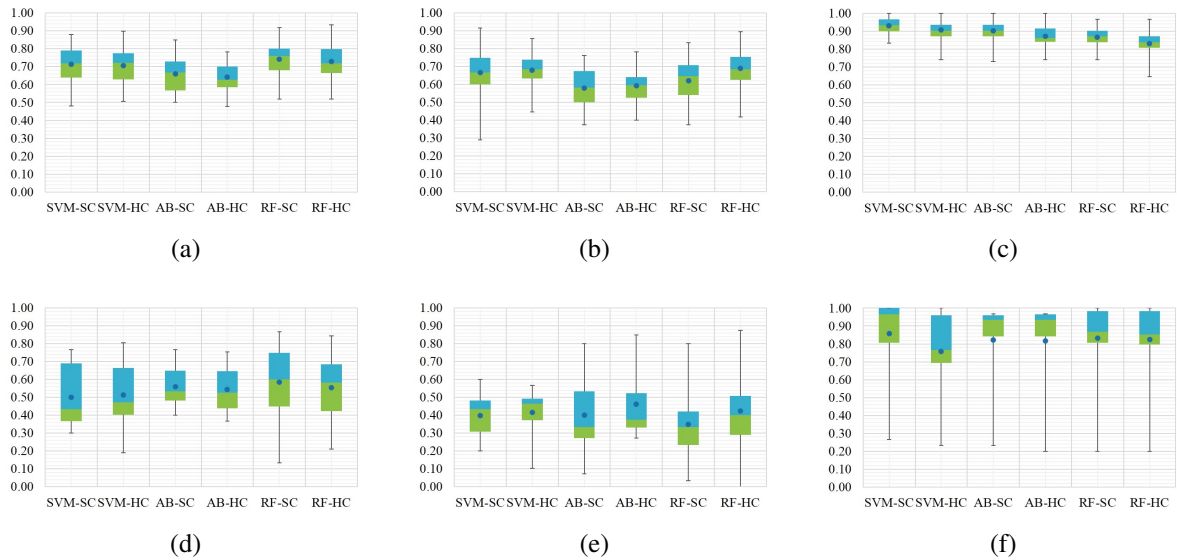


Fig. B.20 Box and whisker diagrams of average per-class classification accuracy for handcrafted $ncARG[r_A]$ method for (a) HER2 positive tumor (b) HER2 negative tumor (c) Non-tumor using k -fold stratified shuffled split and (d),(e),(f) using leave-a-patient-out cross validations.

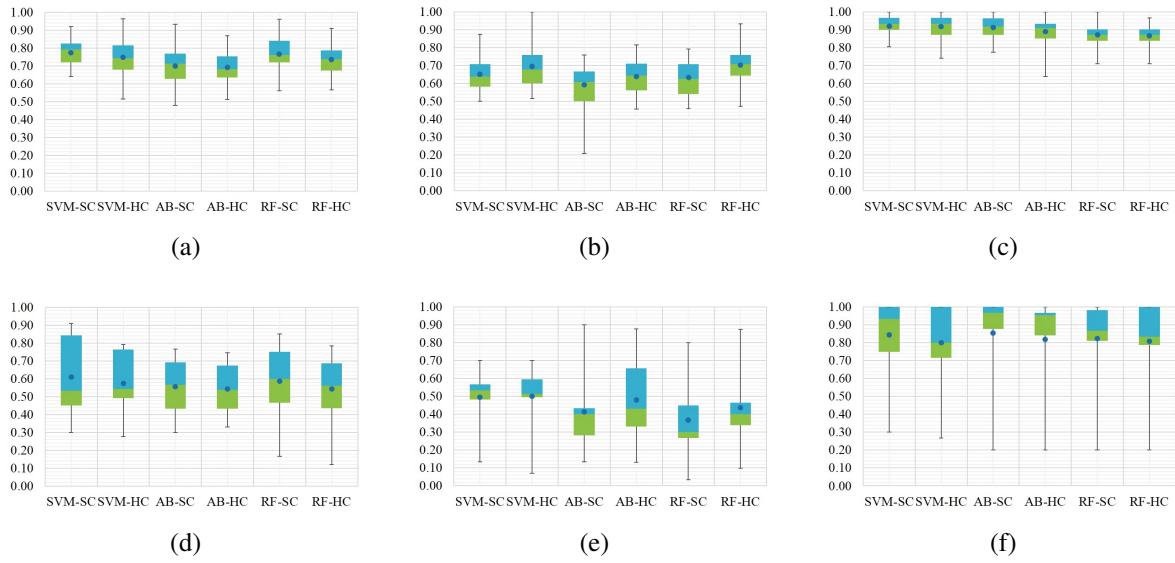


Fig. B.21 Box and whisker diagrams of average per-class classification accuracy for handcrafted $ncARG_{v+}[r_F]$ method for (a) HER2 positive tumor (b) HER2 negative tumor (c) Non-tumor using k -fold stratified shuffled split and (d),(e),(f) using leave-a-patient-out cross validations.

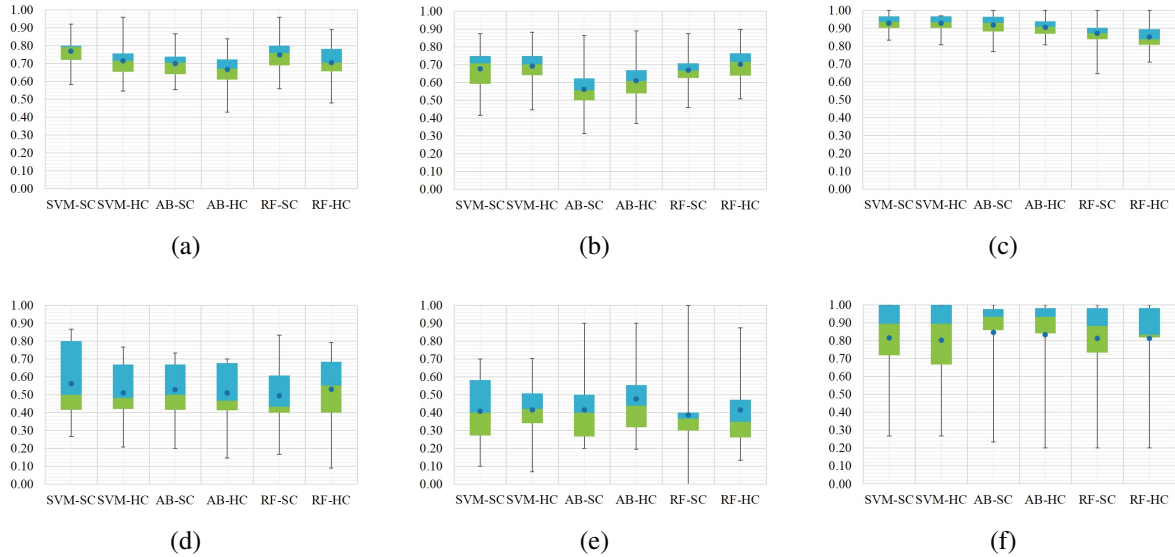


Fig. B.22 Box and whisker diagrams of average per-class classification accuracy for handcrafted $ncARG_{v+}[r_A]$ method for (a) HER2 positive tumor (b) HER2 negative tumor (c) Non-tumor using k -fold stratified shuffled split and (d),(e),(f) using leave-a-patient-out cross validations.

B.2.3.4 Combination of Features

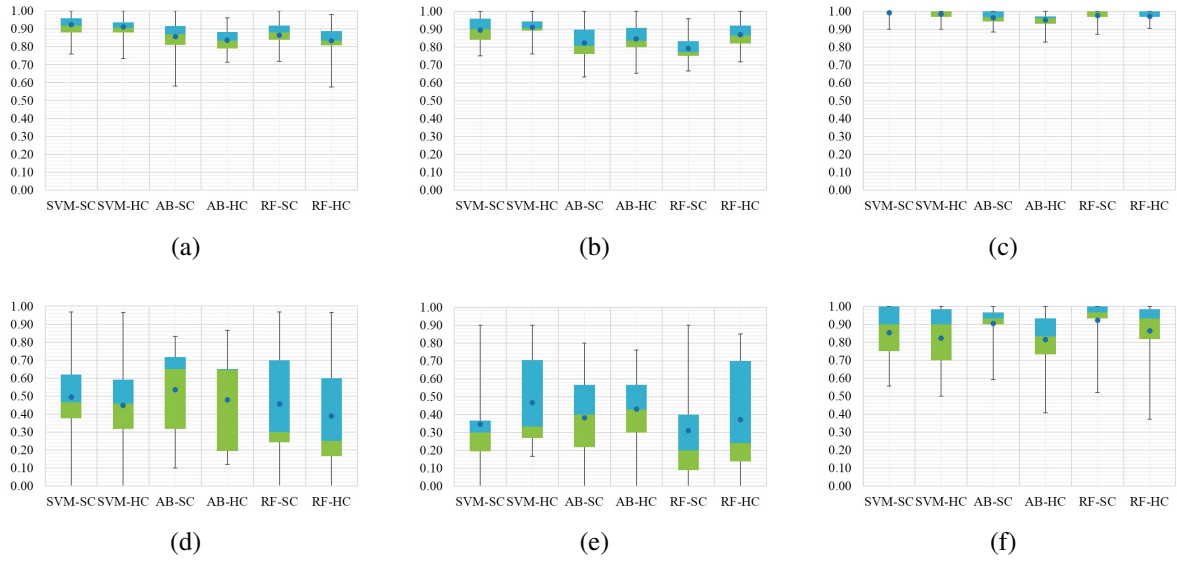


Fig. B.23 Box and whisker diagrams of average per-class classification accuracy of selected subset of hybrid low-level features for (a) HER2 positive tumor (b) HER2 negative tumor (c) Non-tumor using k -fold stratified shuffled split and (d),(e),(f) using leave-a-patient-out cross validations.

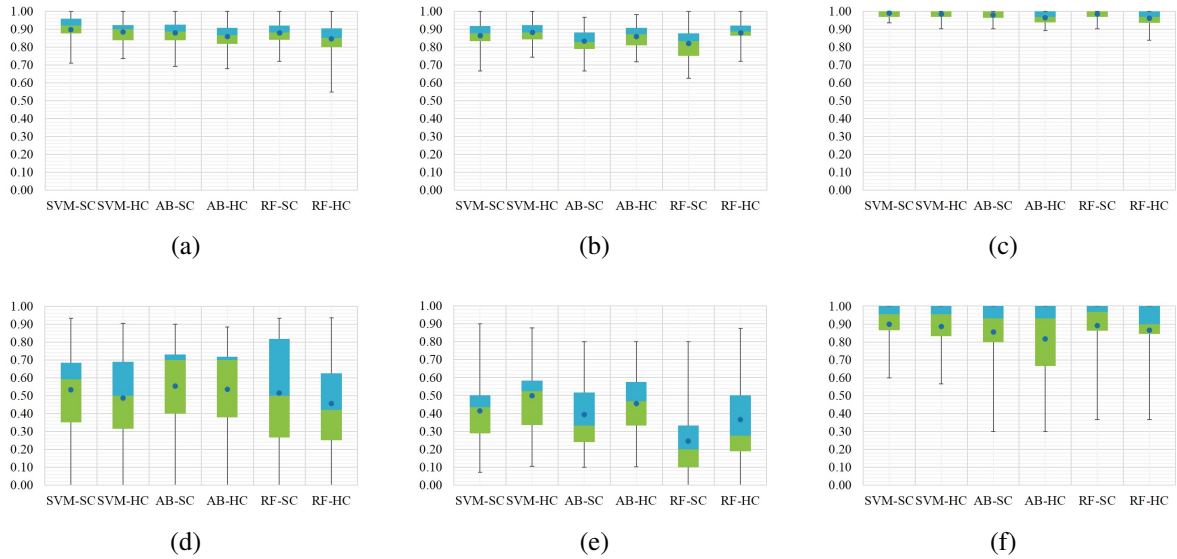


Fig. B.24 Box and whisker diagrams of average per-class classification accuracy of selected subset of hybrid low-level and high-level features for (a) HER2 positive tumor (b) HER2 negative tumor (c) Non-tumor using k -fold stratified shuffled split and (d),(e),(f) using leave-a-patient-out cross validations.

B.2.3.5 Deep Learning Methods

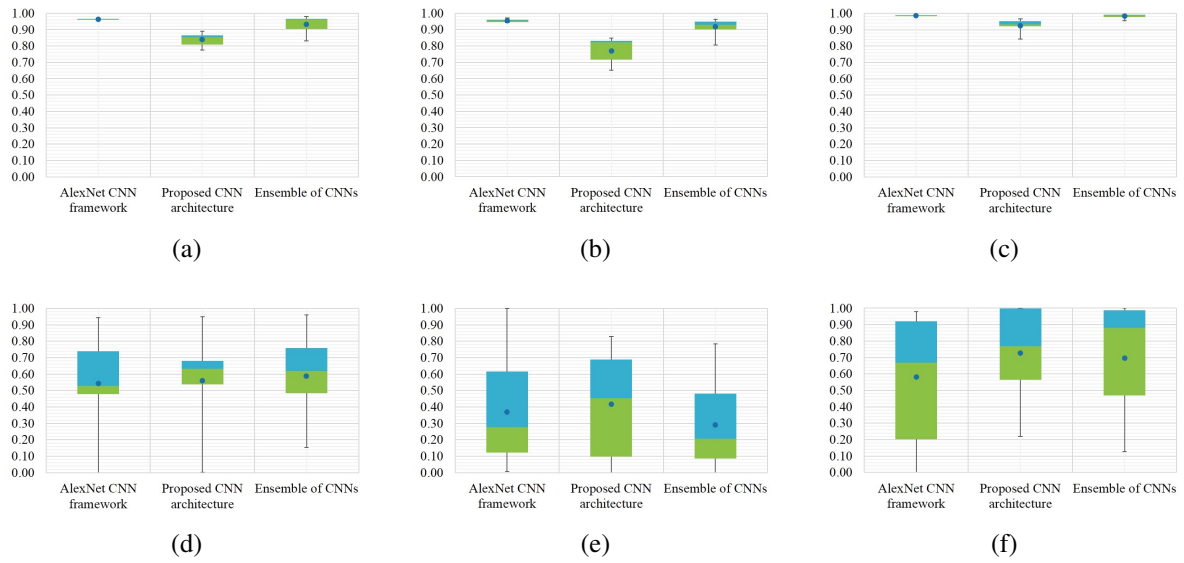


Fig. B.25 Box and whisker diagrams of average per-class classification accuracy for deep learning methods for (a) HER2 positive tumor (b) HER2 negative tumor (c) Non-tumor using k -fold stratified shuffled split and (d),(e),(f) using leave-a-patient-out cross validations.

References

- [Ahonen 2006] T. Ahonen, A. Hadid and M. Pietikainen. *Face description with local binary patterns: Application to face recognition*. Pattern Analysis and Machine Intelligence, IEEE Transactions on, volume 28(12):pages 2037–2041 (2006).
- [Akakin 2012] H. C. Akakin and M. N. Gurcan. *Content-based microscopic image retrieval system for multi-image queries*. Information Technology in Biomedicine, IEEE Transactions on, volume 16(4):pages 758–769 (2012).
- [Albertson 1997] M. O. Albertson. *The irregularity of a graph*. Ars Combinatoria, volume 46:pages 219–225 (1997).
- [Ali 2013] S. Ali, R. Veltri, J. A. Epstein, C. Christudass and A. Madabhushi. *Cell cluster graph for prediction of biochemical recurrence in prostate cancer patients from tissue microarrays*. In *SPIE Medical Imaging*, pages 86760H–86760H. International Society for Optics and Photonics (2013).
- [Alomari 2012] R. S. Alomari, S. Ghosh, V. Chaudhary and O. Al-Kadi. *Local binary patterns for stromal area removal in histology images*. In *SPIE Medical Imaging*, volume 8315, pages 831524–831524–7 (2012). URL <http://dx.doi.org/10.1117/12.911007>.
- [Alon 1997] N. Alon, R. Yuster and U. Zwick. *Finding and counting given length cycles*. Algorithmica, volume 17(3):pages 209–223 (1997).
- [Altunbay 2010] D. Altunbay, C. Cigir, C. Sokmensuer and C. Gunduz-Demir. *Color graphs for automated cancer diagnosis and grading*. Biomedical Engineering, IEEE Transactions on, volume 57(3):pages 665–674 (2010).
- [Amidror 2002] I. Amidror. *Scattered data interpolation methods for electronic imaging systems: a survey*. J. Electronic Imaging, volume 11(2):pages 157–176 (2002). URL <http://dblp.uni-trier.de/db/journals/jei/jei11.html#Amidror02>.
- [Anderson 1997] 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*. The Journal of pathology, volume 181(4):pages 374–380 (1997).
- [Anton 1998] F. Anton, D. Mioc and C. M. Gold. *Dynamic additively weighted voronoi diagrams made easy*. In *Proceedings of the 10th Canadian Conference on Computational Geometry (CCCG'98)* (1998).
- [Anton 2009] F. Anton, D. Mioc and C. Gold. *The Voronoi diagram of circles and its application to the visualization of the growth of particles*. In *Transactions on Computational Science III*, pages 20–54. Springer (2009).
- [Antonini 1992] M. Antonini, M. Barlaud, P. Mathieu and I. Daubechies. *Image coding using wavelet transform*. IEEE Transactions on image processing, volume 1(2):pages 205–220 (1992).
- [Arslan 2013] S. Arslan, T. Ersahin, R. Cetin-Atalay and C. Gunduz-Demir. *Attributed Relational Graphs for Cell Nucleus Segmentation in Fluorescence Microscopy Images*. IEEE Transactions on Medical Imaging, volume 32(6):pages 1121–1131 (2013). ISSN 0278-0062.
- [Baish 2000] J. W. Baish and R. K. Jain. *Fractals and cancer*. Cancer research, volume 60(14):pages 3683–3688 (2000).
- [Bancroft 2008] J. D. Bancroft and M. Gamble. *Theory and practice of histological techniques*. Elsevier Health Sciences (2008).
- [Basavanhally 2010] A. N. Basavanhally, S. Ganesan, S. Agner, J. P. Monaco, M. D. Feldman, J. E. Tomaszewski, G. Bhanot and A. Madabhushi. *Computerized image-based detection and grading of lymphocytic infiltration in HER2+ breast cancer histopathology*. Biomedical Engineering, IEEE Transactions on, volume 57(3):pages 642–653 (2010).

- [Behrens 2015] H.-M. Behrens, V. S. Warneke, C. Böger, N. Garbrecht, E. Jüttner, W. Klapper, M. Mathiak, I. Oschlies, U. Rudolph, C. Stuhlmann-Laeisz *et al.* *Reproducibility of Her2/neu scoring in gastric cancer and assessment of the 10% cut-off rule*. *Cancer medicine*, volume 4(2):pages 235–244 (2015).
- [Bilgin 2007] C. Bilgin, C. Demir, C. Nagi and B. Yener. *Cell-graph mining for breast tissue modeling and classification*. In *Engineering in Medicine and Biology Society, 2007. EMBS 2007. 29th Annual International Conference of the IEEE*, pages 5311–5314. IEEE (2007).
- [Bilgin 2010] C. C. Bilgin, P. Bullough, G. E. Plopper and B. Yener. *ECM-aware cell-graph mining for bone tissue modeling and classification*. *Data mining and knowledge discovery*, volume 20(3):pages 416–438 (2010).
- [Bishop 2006] C. M. Bishop. *Pattern Recognition and Machine Learning (Information Science and Statistics)*. Springer-Verlag New York, Inc., Secaucus, NJ, USA (2006). ISBN 0387310738.
- [Boser 1992] B. E. Boser, I. M. Guyon and V. N. Vapnik. *A training algorithm for optimal margin classifiers*. In *Proceedings of the fifth annual workshop on Computational learning theory*, pages 144–152. ACM (1992).
- [Bosman 2010] F. T. Bosman, F. Carneiro, R. H. Hruban, N. D. Theise *et al.* *WHO classification of tumours of the digestive system*. Ed. 4. World Health Organization (2010).
- [Bottou 2012] L. Bottou. *Stochastic gradient descent tricks*. In *Neural Networks: Tricks of the Trade*, pages 421–436. Springer (2012).
- [Bovik 1990] A. C. Bovik, M. Clark and W. S. Geisler. *Multichannel texture analysis using localized spatial filters*. *IEEE transactions on pattern analysis and machine intelligence*, volume 12(1):pages 55–73 (1990).
- [Bowyer 1981] A. Bowyer. *Computing dirichlet tessellations*. *The Computer Journal*, volume 24(2):pages 162–166 (1981).
- [Bradley 2005] A. P. Bradley, M. Wildermoth and P. Mills. *Virtual microscopy with extended depth of field*. In *Digital Image Computing: Techniques and Applications, 2005. DICTA'05. Proceedings 2005*, pages 35–35. IEEE (2005).
- [Bradski 2008] G. Bradski and A. Kaehler. *Learning OpenCV: Computer vision with the OpenCV library*. "O'Reilly Media, Inc." (2008).
- [Bradski 2013] G. Bradski and A. Kaehler. *Learning OpenCV: Computer Vision in C++ with the OpenCV Library*. O'Reilly Media, Inc., 2nd edition (2013). ISBN 1449314651, 9781449314651.
- [Breiman 1992] L. Breiman and P. Spector. *Submodel selection and evaluation in regression. the x-random case*. *International statistical review/revue internationale de Statistique*, pages 291–319 (1992).
- [Breiman 2001] L. Breiman. *Random forests*. *Machine learning*, volume 45(1):pages 5–32 (2001).
- [Brodersen 2010] K. H. Brodersen, C. S. Ong, K. E. Stephan and J. M. Buhmann. *The balanced accuracy and its posterior distribution*. In *Pattern recognition (ICPR), 2010 20th international conference on*, pages 3121–3124. IEEE (2010).
- [Brouwer 2011] A. E. Brouwer and W. H. Haemers. *Spectra of graphs*. Springer Science & Business Media (2011).
- [Byun 2002] H. Byun and S.-W. Lee. *Applications of support vector machines for pattern recognition: A survey*. In *Pattern recognition with support vector machines*, pages 213–236. Springer (2002).

- [Caicedo 2011] J. C. Caicedo, F. A. González and E. Romero. *Content-based histopathology image retrieval using a kernel-based semantic annotation framework*. Journal of biomedical informatics, volume 44(4):pages 519–528 (2011).
- [Candes 2000] E. J. Candes and D. L. Donoho. *Curvelets: A surprisingly effective nonadaptive representation for objects with edges*. Technical report, DTIC Document (2000).
- [Chang 2011] C.-C. Chang and C.-J. Lin. *LIBSVM: a library for support vector machines*. ACM Transactions on Intelligent Systems and Technology (TIST), volume 2(3):page 27 (2011).
- [Chapelle 2002] O. Chapelle, V. Vapnik, O. Bousquet and S. Mukherjee. *Choosing multiple parameters for support vector machines*. Machine learning, volume 46(1-3):pages 131–159 (2002).
- [Chekkoury 2012] A. Chekkoury, P. Khurd, J. Ni, C. Bahlmann, A. Kamen, A. Patel, L. Grady, M. Singh, M. Groher, N. Navab *et al.* *Automated malignancy detection in breast histopathological images*. In *SPIE Medical Imaging*, pages 831515–831515. International Society for Optics and Photonics (2012).
- [Chen 2006] Y. W. Chen and C. J. Lin. *Combining SVMs with various feature selection strategies*. In *Feature extraction*, pages 315–324. Springer (2006).
- [Chen 2016a] H. Chen, X. Qi, L. Yu and P.-A. Heng. *Dcan: Deep contour-aware networks for accurate gland segmentation*. In *Proceedings of the IEEE conference on Computer Vision and Pattern Recognition*, pages 2487–2496 (2016).
- [Chen 2016b] L. Chen. *Learning Ensembles of Convolutional Neural Networks* (2016).
- [Chu 2011] H. Chu. *Mdb: A memory-mapped database and backend for openldap*. LDAP’11 (2011).
- [Chua 2012] T. Chua and N. Merrett. *Clinicopathologic factors associated with HER2-positive gastric cancer and its impact on survival outcomes—a systematic review*. Int J Cancer, volume 130(12):pages 2845–56 (2012).
- [Chung 2001] S. M. Chung and Q. Wang. *Content-based retrieval and data mining of a skin cancer image database*. In *Information Technology: Coding and Computing, 2001. Proceedings. International Conference on*, pages 611–615. IEEE (2001).
- [Ciregan 2012] D. Ciregan, U. Meier and J. Schmidhuber. *Multi-column deep neural networks for image classification*. In *Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on*, pages 3642–3649 (2012). ISSN 1063-6919.
- [Cireşan 2013] D. C. Cireşan, A. Giusti, L. M. Gambardella and J. Schmidhuber. *Mitosis detection in breast cancer histology images with deep neural networks*. In *Medical Image Computing and Computer-Assisted Intervention—MICCAI 2013*, pages 411–418. Springer (2013).
- [Coleman 1983] T. F. Coleman and J. J. Moré. *Estimation of Sparse Jacobian Matrices and Graph Coloring Problems*. SIAM Journal on Numerical Analysis, volume 20(1):pages 187–209 (1983).
- [Cortes 1995] C. Cortes and V. Vapnik. *Support-vector networks*. Machine learning, volume 20(3):pages 273–297 (1995).
- [Cosatto 2013] E. Cosatto, P.-F. Laquerre, C. Malon, H.-P. Graf, A. Saito, T. Kiyuna, A. Marugame and K. Kamijo. *Automated gastric cancer diagnosis on H&E-stained sections; ltraining a classifier on a large scale with multiple instance machine learning*. In *SPIE Medical Imaging*, pages 867605–867605. International Society for Optics and Photonics (2013).
- [Cruz-Roa 2014] A. Cruz-Roa, A. Basavanhally, F. González, H. Gilmore, M. Feldman, S. Ganesan, N. Shih, J. Tomaszewski and A. Madabhushi. *Automatic detection of invasive ductal carcinoma in whole slide images with convolutional neural networks*. In *SPIE Medical Imaging*, pages 904103–904103. International Society for Optics and Photonics (2014).

- [Czumaj 2008] A. Czumaj and C. Sohler. *Testing Euclidean minimum spanning trees in the plane*. ACM Transactions on Algorithms (TALG), volume 4(3):page 31 (2008).
- [Darro 1993] F. Darro, A. Kruczynski, C. Etievant, J. Martinez, J.-L. Pasteels and R. Kiss. *Characterization of the differentiation of human colorectal cancer cell lines by means of voronoi diagrams*. Cytometry, volume 14(7):pages 783–792 (1993).
- [Datta 2008] R. Datta, D. Joshi, J. Li and J. Z. Wang. *Image retrieval: Ideas, influences, and trends of the new age*. ACM Computing Surveys (CSUR), volume 40(2):page 5 (2008).
- [Daugman 1985] J. G. Daugman. *Uncertainty relation for resolution in space, spatial frequency, and orientation optimized by two-dimensional visual cortical filters*. JOSA A, volume 2(7):pages 1160–1169 (1985).
- [De Wouwer 2000] V. De Wouwer, V. Marck, V. Dyck *et al.* *Wavelets as chromatin texture descriptors for the automated identification of neoplastic nuclei*. Journal of microscopy, volume 197(1):pages 25–35 (2000).
- [Delaunay 1934] B. Delaunay. *Sur la sphère vide. A la mémoire de Georges Voronoï*. Bulletin of Academy of Sciences of the USSR, (6):pages 793–800 (1934). URL http://www.mathnet.ru/php/archive.phtml?wshow=paper&jrnid=im&paperid=4937&option_lang=eng.
- [Deligdisch 1993] L. Deligdisch, H. Kerner, C. Cohen, D. Dargent and J. Gil. *Morphometric differentiation between responsive tumor cells and mesothelial hyperplasia in second-look operations for ovarian cancer*. Human pathology, volume 24(2):pages 143–147 (1993).
- [Demir 2004] C. Demir, S. H. Gultekin and B. Yener. *Spectral analysis of cell-graphs of cancer*. Department of Computer Science, Rensselaer Polytechnic Institute, Troy, New York (2004).
- [Demir 2005a] C. Demir, S. H. Gultekin and B. Yener. *Augmented cell-graphs for automated cancer diagnosis*. In *ECCB/JBI*, page 12 (2005). URL <http://dblp.uni-trier.de/db/conf/eccb/eccb2005.html#DemirGY05>.
- [Demir 2005b] C. Demir, S. H. Gultekin and B. Yener. *Learning the topological properties of brain tumors*. IEEE/ACM Transactions on Computational Biology and Bioinformatics (TCBB), volume 2(3):pages 262–270 (2005).
- [Deng 2009] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li and L. Fei-Fei. *ImageNet: A Large-Scale Hierarchical Image Database*. In *CVPR09* (2009).
- [Deng 2012] L. Deng. *Three classes of deep learning architectures and their applications: a tutorial survey*. APSIPA transactions on signal and information processing (2012).
- [Deng 2014] L. Deng and D. Yu. *Deep learning: Methods and applications*. Foundations and Trends in Signal Processing, volume 7(3–4):pages 197–387 (2014).
- [DeVita 2010] V. T. DeVita, T. S. Lawrence and S. A. Rosenberg. *Cancer: principles and practice of oncology-advances in oncology*, volume 1. Lippincott Williams & Wilkins (2010).
- [Diamond 2004] J. Diamond, N. H. Anderson, P. H. Bartels, R. Montironi and P. W. Hamilton. *The use of morphological characteristics and texture analysis in the identification of tissue composition in prostatic neoplasia*. Human Pathology, volume 35(9):pages 1121–1131 (2004).
- [Díaz 2010] G. Díaz and E. Romero. *Histopathological image classification using stain component features on a PLSA model*. In *Progress in Pattern Recognition, Image Analysis, Computer Vision, and Applications*, pages 55–62. Springer (2010).
- [DiFranco 2011] M. D. DiFranco, G. O’Hurley, E. W. Kay, R. W. G. Watson and P. Cunningham. *Ensemble based system for whole-slide prostate cancer probability mapping using color texture features*. Computerized medical imaging and graphics, volume 35(7):pages 629–645 (2011).

- [Dirichlet 1850] L. G. Dirichlet. *Über die Reduction der positiven quadratischen Formen mit drei unbestimmten ganzen Zahlen*. Journal für die reine und angewandte Mathematik, volume 40:pages 209–227 (1850). URL <http://eudml.org/doc/147457>.
- [Doyle 2006] S. Doyle, C. Rodriguez, A. Madabhushi, J. Tomaszewski and M. Feldman. *Detecting prostatic adenocarcinoma from digitized histology using a multi-scale hierarchical classification approach*. In *Engineering in Medicine and Biology Society, 2006. EMBS'06. 28th Annual International Conference of the IEEE*, pages 4759–4762. IEEE (2006).
- [Doyle 2008] S. Doyle, S. Agner, A. Madabhushi, M. Feldman and J. Tomaszewski. *Automated grading of breast cancer histopathology using spectral clustering with textural and architectural image features*. In *Biomedical Imaging: From Nano to Macro, 2008. ISBI 2008. 5th IEEE International Symposium on*, pages 496–499. IEEE (2008).
- [Doyle 2012] S. Doyle, M. Feldman, J. Tomaszewski and A. Madabhushi. *A Boosted Bayesian Multiresolution Classifier for Prostate Cancer Detection From Digitized Needle Biopsies*. IEEE Transactions on Biomedical Engineering, volume 59(5):pages 1205–1218 (2012). ISSN 0018-9294.
- [DPA 2016] DPA. *Whole Slide Imaging Repository*, Digital Pathology Association. <https://digitalpathologyassociation.org/whole-slide-imaging-repository> (2016).
- [ECMA 2006] ECMA. *C# language specification (4th ed.)*. <http://www.ecma-international.org/publications/standards/Ecma-334.htm> (2006). URL <http://www.ecma-international.org/publications/standards/Ecma-334.htm>.
- [Edelsbrunner 1987] H. Edelsbrunner. *Algorithms in combinatorial geometry*, volume 10. Springer Science & Business Media (1987).
- [Eppstein 1999] D. Eppstein. *Spanning trees and spanners*. Handbook of computational geometry, pages 425–461 (1999).
- [Esgiar 1998] A. N. Esgiar, R. N. Naguib, B. S. Sharif, M. K. Bennett and A. Murray. *Microscopic image analysis for quantitative measurement and feature identification of normal and cancerous colonic mucosa*. Information Technology in Biomedicine, IEEE Transactions on, volume 2(3):pages 197–203 (1998).
- [Everitt 2006] B. S. Everitt. *The Cambridge dictionary of statistics*. Cambridge University Press (2006).
- [Falconer 2004] K. Falconer. *Fractal geometry: mathematical foundations and applications*. John Wiley & Sons (2004).
- [Farjam 2007] R. Farjam, H. Soltanian-Zadeh, K. Jafari-Khouzani and R. A. Zoroofi. *An image analysis approach for automatic malignancy determination of prostate pathological images*. Cytometry Part B: Clinical Cytometry, volume 72(4):pages 227–240 (2007).
- [Ficsor 2006] L. Ficsor, V. Varga, L. Berczi, P. Miheller, A. Tagscherer, M. L.-c. Wu, Z. Tulassay and B. Molnar. *Automated virtual microscopy of gastric biopsies*. Cytometry Part B: Clinical Cytometry, volume 70(6):pages 423–431 (2006).
- [Ficsor 2008] L. Ficsor, V. S. Varga, A. Tagscherer, Z. Tulassay and B. Molnar. *Automated classification of inflammation in colon histological sections based on digital microscopy and advanced image analysis*. Cytometry Part A, volume 73(3):pages 230–237 (2008).
- [Fiddy 1985] M. Fiddy. *The radon transform and some of its applications*. Journal of Modern Optics, volume 32(1):pages 3–4 (1985).
- [Fisher 1921] R. A. Fisher. *On the probable error of a coefficient of correlation deduced from a small sample*. Metron, volume 1:pages 3–32 (1921).

- [Fogel 1989] I. Fogel and D. Sagi. *Gabor filters as texture discriminator*. Biological Cybernetics, volume 61(2):pages 103–113 (1989). ISSN 0340-1200. URL <http://dx.doi.org/10.1007/BF00204594>.
- [Fortune 1987] S. Fortune. *A sweepline algorithm for voronoi diagrams*. Algorithmica, volume 2(1-4):pages 153–174 (1987).
- [Freund 1995] Y. Freund and R. E. Schapire. *A decision-theoretic generalization of on-line learning and an application to boosting*. In *European conference on computational learning theory*, pages 23–37. Springer (1995).
- [Friedman 2001] J. Friedman, T. Hastie and R. Tibshirani. *The elements of statistical learning*, volume 1. Springer series in statistics Springer, Berlin (2001).
- [Fu 1974] K. S. Fu. *Syntactic methods in pattern recognition*. Elsevier (1974).
- [Fukunaga 2013] K. Fukunaga. *Introduction to statistical pattern recognition*. Academic press (2013).
- [Gabriel 1969] K. R. Gabriel and R. R. Sokal. *A new statistical approach to geographic variation analysis*. Systematic Biology, volume 18(3):pages 259–278 (1969).
- [Gonzales 2009] R. C. Gonzales and R. E. Woods. *Digital Image Processing*. New Jersey: Prentice Hall, 3 edition (2009).
- [Goode 2007] A. Goode, M. Chen, A. Tarachandani, L. B. Mummert, R. Sukthankar, C. Helfrich, A. Stefanni, L. Fix, J. Saltzman and M. Satyanarayanan. *Interactive Search of Adipocytes in Large Collections of Digital Cellular Images*. In *ICME*, pages 695–698. IEEE (2007). URL <http://dblp.uni-trier.de/db/conf/icmcs/icme2007.html#GoodeCTMSHSFSS07>.
- [Gorelick 2013] L. Gorelick, O. Veksler, M. Gaed, J. A. Gómez, M. Moussa, G. Bauman, A. Fenster and A. D. Ward. *Prostate Histopathology: Learning Tissue Component Histograms for Cancer Detection and Classification*. IEEE Transactions on Medical Imaging, volume 32(10):pages 1804–1818 (2013). ISSN 0278-0062.
- [Graham 1985] R. L. Graham and P. Hell. *On the history of the minimum spanning tree problem*. Annals of the History of Computing, volume 7(1):pages 43–57 (1985).
- [Green 1978] P. J. Green and R. Sibson. *Computing Dirichlet tessellations in the plane*. The Computer Journal, volume 21(2):pages 168–173 (1978).
- [Gunduz-Demir 2007] C. Gunduz-Demir. *Mathematical modeling of the malignancy of cancer using graph evolution*. Mathematical biosciences, volume 209(2):pages 514–527 (2007).
- [Gunduz 2004] C. Gunduz, B. Yener and S. Gultekin. *The cell graphs of cancer*. In *ISMB/ECCB (Supplement of Bioinformatics)*, pages 145–151 (2004). URL <http://dblp.uni-trier.de/db/conf/ismb/ismb2004.html#GunduzYG04>.
- [Gurcan 2009] M. N. Gurcan, L. E. Boucheron, A. Can, A. Madabhushi, N. M. Rajpoot and B. Yener. *Histopathological image analysis: A review*. Biomedical Engineering, IEEE Reviews in, volume 2:pages 147–171 (2009).
- [Guyon 2003] I. Guyon and A. Elisseeff. *An Introduction to Variable and Feature Selection*. J. Mach. Learn. Res., volume 3:pages 1157–1182 (2003). ISSN 1532-4435. URL <http://dl.acm.org/citation.cfm?id=944919.944968>.
- [Hall 2009] M. Hall, E. Frank, G. Holmes, B. Pfahringer, P. Reutemann and I. H. Witten. *The WEKA Data Mining Software: An Update*. SIGKDD Explor. Newsl., volume 11(1):pages 10–18 (2009). ISSN 1931-0145. URL <http://doi.acm.org/10.1145/1656274.1656278>.
- [Hamilton 1987] P. Hamilton, D. Allen, P. Watt, C. Patterson and J. Biggart. *Classification of normal colorectal mucosa and adenocarcinoma by morphometry*. Histopathology, volume 11(9):pages 901–911 (1987).

- [Hamilton 1994] P. Hamilton, N. Anderson, P. Bartels and D. Thompson. *Expert system support using Bayesian belief networks in the diagnosis of fine needle aspiration biopsy specimens of the breast*. J Clin Pathol, volume 47(4):pages 329–36 (1994). ISSN 0021-9746. URL <http://www.biomedsearch.com/nih/Expert-system-support-using-Bayesian/8027370.html>.
- [Hamilton 1997] P. W. Hamilton, P. H. Bartels, D. Thompson, N. H. Anderson, R. Montironi and J. M. Sloan. *Automated location of dysplastic fields in colorectal histology using image texture analysis*. The Journal of pathology, volume 182(1):pages 68–75 (1997).
- [Hamon 2013] J. Hamon. *Optimisation combinatoire pour la sélection de variables en régression en grande dimension: Application en génétique animale*. Ph.D. thesis, Université des Sciences et Technologie de Lille-Lille I (2013).
- [Haralick 1973] R. M. Haralick, K. S. Shanmugam and I. Dinstein. *Textural features for image classification*. IEEE Transactions on Systems, Man and Cybernetics, volume 3(6):pages 610–621 (1973).
- [Hartigan 1979] J. A. Hartigan and M. A. Wong. *Algorithm as 136: A k-means clustering algorithm*. Journal of the Royal Statistical Society. Series C (Applied Statistics), volume 28(1):pages 100–108 (1979).
- [Heikkilä 2006] M. Heikkilä and M. Pietikäinen. *A texture-based method for modeling the background and detecting moving objects*. Pattern Analysis and Machine Intelligence, IEEE Transactions on, volume 28(4):pages 657–662 (2006).
- [Hinton 2005] G. E. Hinton. *What kind of graphical model is the brain?* In IJCAI, volume 5, pages 1765–1775 (2005).
- [Hofmann 2013] M. Hofmann and R. Klinkenberg. *RapidMiner: Data mining use cases and business analytics applications*. CRC Press (2013).
- [Hopcroft 1973] J. Hopcroft and R. Tarjan. *Algorithm 447: Efficient algorithms for graph manipulation*. Communications of the ACM, volume 16(6):pages 372–378 (1973).
- [Hou 2015] L. Hou, D. Samaras, T. M. Kurç, Y. Gao, J. E. Davis and J. H. Saltz. *Efficient Multiple Instance Convolutional Neural Networks for Gigapixel Resolution Image Classification*. CoRR, volume abs/1504.07947 (2015). URL <http://arxiv.org/abs/1504.07947>.
- [Huang 2011] C.-H. Huang, A. Veillard, L. Roux, N. Loménie and D. Racocanu. *Time-efficient sparse analysis of histopathological whole slide images*. Computerized medical imaging and graphics, volume 35(7):pages 579–591 (2011).
- [Hufnagl 1983] P. Hufnagl and K. Voss. *Ein zeitoptimaler Konturfolgealgorithmus (A time-optimal contour search algorithm)*. Digitale Bildverarbeitung, Wiss. Beitr. d. TU Dresden, pages 18–26 (1983).
- [Hufnagl 1984] P. Hufnagl, A. Schlosser and K. Voss. *Merkmale der Form, Größe und Lage digitaler Objekte*. Bild und Ton., volume 37:pages 293–298 (1984).
- [Hughes 1968] G. Hughes. *On the mean accuracy of statistical pattern recognizers*. IEEE Transactions on Information Theory, volume 14(1):pages 55–63 (1968). ISSN 0018-9448.
- [Hunter 1972] I. Hunter. *Practical Section Cutting and Staining*. Journal of clinical pathology, volume 25(5):page 460 (1972).
- [Intriligator 1978] M. D. Intriligator. *Econometric models, techniques, and applications*. Technical report, Prentice-Hall Englewood Cliffs, NJ (1978).
- [ISBI 2016] ISBI. *ISBI challenge on cancer metastasis detection in lymph node*. <https://camelyon16.grand-challenge.org/> (2016).
- [Jafari-Khouzani 2003] K. Jafari-Khouzani and H. Soltanian-Zadeh. *Multiwavelet grading of pathological images of prostate*. Biomedical Engineering, IEEE Transactions on, volume 50(6):pages 697–704 (2003).

- [Jaromczyk 1992] J. W. Jaromczyk and G. T. Toussaint. *Relative neighborhood graphs and their relatives*. Proceedings of the IEEE, volume 80(9):pages 1502–1517 (1992).
- [Jaulent 2000] M. Jaulent, C. Le Bozec, Y. Cao and P. Degoulet. *A property concept frame representation for flexible image-content retrieval in histopathology databases*. Proc AMIA Symp., pages 379–83 (2000).
- [Jia 2014] Y. Jia, E. Shelhamer, J. Donahue, S. Karayev, J. Long, R. Girshick, S. Guadarrama and T. Darrell. *Caffe: Convolutional architecture for fast feature embedding*. In *Proceedings of the ACM International Conference on Multimedia*, pages 675–678. ACM (2014).
- [Johnson 1939] W. Johnson and R. Mehl. *Reaction Kinetics in processes of nucleation and growth*. Transactions of American Institute of Mining and Metallurgical Engineers, volume 135:pages 416–458 (1939).
- [Jones 2015] E. Jones, T. Oliphant, P. Peterson *et al.* *SciPy: Open source scientific tools for Python, 2001*. URL <http://www.scipy.org>, volume 73:page 86 (2015).
- [Takeji 2006] Y. Takeji, S. Yamaguchi, D. Yoshida, K. Tanoue, M. Ueda, A. Masunari, T. Utsunomiya, M. Imamura, H. Honda, Y. Maehara *et al.* *Development and assessment of morphologic criteria for diagnosing gastric cancer using confocal endomicroscopy: an ex vivo and in vivo study*. Endoscopy, volume 38(09):pages 886–890 (2006).
- [Kayser 1986] K. Kayser, M. Shaver, F. Modlinger, K. Postl and J. Moyers. *Neighborhood Analysis of Low Magnification Structures (Glands) in Healthy, Adenomatous, and Carcinomatous Colon Mucosa*. Pathology-Research and Practice, volume 181(2):pages 153–158 (1986).
- [Kayser 1987] K. Kayser. *Neighborhood condition and application of syntactic structure analysis in histopathology*. Acta Stereol, volume 6:pages 373–384 (1987).
- [Kayser 1988] K. Kayser. *Application of structural pattern recognition in histopathology*. In *Syntactic and structural pattern recognition*, pages 115–135. Springer (1988).
- [Kayser 1989] K. Kayser and H. Stute. *Minimum Spanning Tree, Voronoi's Tessellation, Johnson-Mehl Diagrams in Human Lung Carcinoma*. Pathology-Research and Practice, volume 185(5):pages 729–734 (1989).
- [Kayser 1990] K. Kayser, H. Stute, J. Bubenzer and J. Paul. *Combined morphometrical and syntactic structure analysis as tools for histomorphological insight into human lung carcinoma growth*. Analytical cellular pathology: the journal of the European Society for Analytical Cellular Pathology, volume 2(3):pages 167–178 (1990).
- [Kayser 1991] K. Kayser, K. Sandau, G. Böhm, K. D. Kunze and J. Paul. *Analysis of soft tissue tumors by an attributed minimum spanning tree*. Analytical and quantitative cytology and histology/the International Academy of Cytology [and] American Society of Cytology, volume 13(5):pages 329–334 (1991).
- [Keenan 2000] S. J. Keenan, J. Diamond, W. Glenn McCluggage, H. Bharucha, D. Thompson, P. H. Bartels and P. W. Hamilton. *An automated machine vision system for the histological grading of cervical intraepithelial neoplasia (CIN)*. The Journal of pathology, volume 192(3):pages 351–362 (2000).
- [Khan 2014] A. M. Khan, N. Rajpoot, D. Treanor and D. Magee. *A Nonlinear Mapping Approach to Stain Normalization in Digital Histopathology Images Using Image-Specific Color Deconvolution*. IEEE Transactions on Biomedical Engineering, volume 61(6):pages 1729–1738 (2014). ISSN 0018-9294.
- [Khurd 2010] P. Khurd, C. Bahlmann, P. Maday, A. Kamen, S. Gibbs-Strauss, E. M. Genega and J. V. Frangioni. *Computer-aided Gleason grading of prostate cancer histopathological images using texton forests*. In *Biomedical Imaging: From Nano to Macro, 2010 IEEE International Symposium on*, pages 636–639. IEEE (2010).

- [Kimeldorf 1970] G. S. Kimeldorf and G. Wahba. *A correspondence between bayesian estimation on stochastic processes and smoothing by splines*. The Annals of Mathematical Statistics, volume 41(2):pages 495–502 (1970).
- [Kirillov 2013] A. Kirillov. *Aforge.NET framework*. <http://www.aforgenet.com> (2013).
- [Kirkpatrick 1985] D. G. Kirkpatrick and J. Radke. *A framework for computational morphology*. In G. T. Toussaint, editor, *Computational Geometry, Machine Intelligence and Pattern Recognition*, volume 2, pages 217–248. Elsevier, North-Holland (1985).
- [Knerr 1990] S. Knerr, L. Personnaz and G. Dreyfus. *Single-layer learning revisited: a stepwise procedure for building and training a neural network*. In *Neurocomputing*, pages 41–50. Springer (1990).
- [Kohavi 1995] R. Kohavi *et al.* *A study of cross-validation and bootstrap for accuracy estimation and model selection*. In *Ijcai*, volume 14, pages 1137–1145. Stanford, CA (1995).
- [Kong 2009] J. Kong, O. Sertel, H. Shimada, K. Boyer, J. Saltz and M. Gurcan. *Computer-aided Evaluation of Neuroblastoma on Whole-slide Histology Images: Classifying Grade of Neuroblastic Differentiation*. *Pattern Recogn.*, volume 42(6):pages 1080–1092 (2009). ISSN 0031-3203. URL <http://dx.doi.org/10.1016/j.patcog.2008.10.035>.
- [Kothari 2011] S. Kothari, J. H. Phan, R. A. Moffitt, T. H. Stokes, S. E. Hassberger, Q. Chaudry, A. N. Young and M. D. Wang. *Automatic batch-invariant color segmentation of histological cancer images*. In *Biomedical Imaging: From Nano to Macro, 2011 IEEE International Symposium on*, pages 657–660. IEEE (2011).
- [Kotsiantis 2007] S. B. Kotsiantis, I. Zaharakis and P. Pintelas. *Supervised machine learning: A review of classification techniques*. In I. G. Maglogiannis, editor, *Emerging Artificial Intelligence Applications in Computer Engineering: Real Word AI Systems with Applications in EHealth, HCI, Information Retrieval and Pervasive Technologies*, pages 3–24. IOS Press (2007).
- [Krizhevsky 2012] A. Krizhevsky, I. Sutskever and G. E. Hinton. *Imagenet classification with deep convolutional neural networks*. In *Advances in neural information processing systems*, pages 1097–1105 (2012).
- [Kroemer 2009] G. Kroemer, L. Galluzzi, P. Vandenabeele, J. Abrams, E. Alnemri, E. Baehrecke, M. Blagosklonny, W. El-Deiry, P. Golstein, D. Green *et al.* *Classification of cell death: recommendations of the Nomenclature Committee on Cell Death 2009*. *Cell Death & Differentiation*, volume 16(1):pages 3–11 (2009).
- [Kröse 1993] B. Kröse and P. van der Smagt. *An introduction to neural networks* (1993).
- [Krysko 2008] D. V. Krysko, T. V. Berghe, K. D’Herde and P. Vandenabeele. *Apoptosis and necrosis: detection, discrimination and phagocytosis*. *Methods*, volume 44(3):pages 205–221 (2008).
- [Langner 2006] C. Langner, G. Hutterer, T. Chromecki, S. Leibl, P. Rehak and R. Zigeuner. *Tumor necrosis as prognostic indicator in transitional cell carcinoma of the upper urinary tract*. The Journal of urology, volume 176(3):pages 910–914 (2006).
- [Lauren 1965] P. Lauren. *The two histological main types of gastric carcinoma, an attempt at a histoclinical classification*. *Acta Pathol. Microbiol. Scand.*, volume 64:pages 31–49 (1965).
- [Lawson 1977] C. L. Lawson. *Software for C^1 Surface Interpolation*. In *Mathematical Software III (J. Rice, editor)*, pages 161–194 (1977).
- [Le 2012] Q. V. Le, J. Han, J. W. Gray, P. T. Spellman, A. Borowsky and B. Parvin. *Learning invariant features of tumor signatures*. In *Biomedical Imaging (ISBI), 2012 9th IEEE International Symposium on*, pages 302–305. IEEE (2012).

- [LeCun 1989] Y. LeCun, B. Boser, J. S. Denker, D. Henderson, R. E. Howard, W. Hubbard and L. D. Jackel. *Backpropagation applied to handwritten zip code recognition*. Neural computation, volume 1(4):pages 541–551 (1989).
- [Lecun 1998] Y. Lecun, L. Bottou, Y. Bengio and P. Haffner. *Gradient-based learning applied to document recognition*. Proceedings of the IEEE, volume 86(11):pages 2278–2324 (1998). ISSN 0018-9219.
- [LeCun 2015] Y. LeCun, Y. Bengio and G. Hinton. *Deep learning*. Nature, volume 521(7553):pages 436–444 (2015).
- [Leek 1999] R. Leek, R. Landers, A. Harris and C. Lewis. *Necrosis correlates with high vascular density and focal macrophage infiltration in invasive carcinoma of the breast*. British journal of cancer, volume 79(5-6):page 991 (1999).
- [Linder 2012] N. Linder, J. Konsti, R. Turkki, E. Rahtu, M. Lundin, S. Nordling, C. Haglund, T. Ahonen, M. Pietikäinen and J. Lundin. *Identification of tumor epithelium and stroma in tissue microarrays using texture analysis*. Diagn Pathol, volume 7(22):pages 1596–7 (2012).
- [Lingas 1994] A. Lingas. *A linear-time construction of the relative neighborhood graph from the Delaunay triangulation*. Computational Geometry, volume 4(4):pages 199–208 (1994).
- [Litjens 2016] G. Litjens, C. I. Sánchez, N. Timofeeva, M. Hermesen, I. Nagtegaal, I. Kovacs, C. Hulsbergen-Van De Kaa, P. Bult, B. Van Ginneken and J. Van Der Laak. *Deep learning as a tool for increased accuracy and efficiency of histopathological diagnosis*. Scientific reports, volume 6 (2016).
- [Lloyd 1982] S. Lloyd. *Least squares quantization in PCM*. Information Theory, IEEE Transactions on, volume 28(2):pages 129–137 (1982).
- [Luo 2003] B. Luo, R. C. Wilson and E. R. Hancock. *Spectral embedding of graphs*. Pattern recognition, volume 36(10):pages 2213–2230 (2003).
- [Mäenpää 2003] T. Mäenpää. *The local binary pattern approach to texture analysis - extensions and applications*. Ph.D. thesis, University of Oulu (2003).
- [Majno 1995] G. Majno and I. Joris. *Apoptosis, oncosis, and necrosis. an overview of cell death*. The American journal of pathology, volume 146(1):page 3 (1995).
- [Makhoul 1999] J. Makhoul, F. Kubala, R. Schwartz and R. Weischedel. *Performance Measures For Information Extraction*. In *In Proceedings of DARPA Broadcast News Workshop*, pages 249–252 (1999).
- [Mandelbrot 1967] B. B. Mandelbrot. *How long is the coast of britain*. Science, volume 156(3775):pages 636–638 (1967).
- [Mandelbrot 1983] B. B. Mandelbrot. *The fractal geometry of nature*, volume 173. Macmillan (1983).
- [Marčelja 1980] S. Marčelja. *Mathematical description of the responses of simple cortical cells**. JOSA, volume 70(11):pages 1297–1300 (1980).
- [Maška 2014] M. Maška, V. Ulman, D. Svoboda, P. Matula, P. Matula, C. Ederra, A. Urbiola, T. España, S. Venkatesan, D. M. Balak *et al.* *A benchmark for comparison of cell tracking algorithms*. Bioinformatics, volume 30(11):pages 1609–1617 (2014).
- [Matula 1980] D. W. Matula and R. R. Sokal. *Properties of Gabriel graphs relevant to geographic variation research and the clustering of points in the plane*. Geographical analysis, volume 12(3):pages 205–222 (1980).
- [McCune 2002] B. McCune, J. Grace, B. McCune and J. Grace. *Distance measures*. Analysis of Ecological Communities. MjM Software Design, Gleneden Beach (2002).

- [Meijer 1992] G. Meijer, P. Van Diest, J. Fleege and J. Baak. *Syntactic structure analysis of the arrangement of nuclei in dysplastic epithelium of colorectal adenomatous polyps*. Analytical and quantitative cytology and histology/the International Academy of Cytology [and] American Society of Cytology, volume 14(6):pages 491–498 (1992).
- [Moore 2011] K. L. Moore, A. M. Agur and A. F. Dalley. *Essential clinical anatomy*. Lippincott Williams & Wilkins (2011).
- [Morabia 2013] A. Morabia. *A history of epidemiologic methods and concepts*. Birkhäuser (2013).
- [Movellan 2002] J. R. Movellan. *Tutorial on Gabor filters*. Open Source Document (2002).
- [Mukundan 2000] R. Mukundan, S. Ong and P. Lee. *Discrete orthogonal moment features using chebyshev polynomials*. Tpq, volume 1:page 0 (2000).
- [Muro-Cacho 2000] C. A. Muro-Cacho, A. B. Cantor and M. Morgan. *Prognostic factors in malignant gastrointestinal stromal tumors*. Annals of Clinical & Laboratory Science, volume 30(3):pages 239–247 (2000).
- [Najafabadi 2015] M. M. Najafabadi, F. Villanustre, T. M. Khoshgoftaar, N. Seliya, R. Wald and E. Muharemagic. *Deep learning applications and challenges in big data analytics*. Journal of Big Data, volume 2(1):pages 1–21 (2015).
- [Nedzved 2007] A. Nedzved, A. Belotserkovsky, T. Lehmann and S. Ablameyko. *Morphometrical feature extraction on color histological images for oncological diagnostics*. In *5th International Conference on Biomedical Engineering*, pages 14–16 (2007).
- [Needham 2016] P. Needham, A. W. Götz and R. C. Walker. *Why Graphics Processing Units*. Electronic Structure Calculations on Graphics Processing Units: From Quantum Chemistry to Condensed Matter Physics, pages 1–22 (2016).
- [O’Callaghan 1975] J. F. O’Callaghan. *An alternative definition for neighborhood of a point*. IEEE Transactions on Computers, volume 24(11):pages 1121–1125 (1975).
- [Ojala 1996] T. Ojala, M. Pietikäinen and D. Harwood. *A comparative study of texture measures with classification based on featured distributions*. Pattern recognition, volume 29(1):pages 51–59 (1996).
- [Ojala 2002] T. Ojala, M. Pietikäinen and T. Mäenpää. *Multiresolution gray-scale and rotation invariant texture classification with local binary patterns*. Pattern Analysis and Machine Intelligence, IEEE Transactions on, volume 24(7):pages 971–987 (2002).
- [Okabe 2009] A. Okabe, B. Boots, K. Sugihara and S. N. Chiu. *Spatial tessellations: concepts and applications of Voronoi diagrams*, volume 501. John Wiley & Sons (2009).
- [Orlov 2008] N. Orlov, L. Shamir, T. Macura, J. Johnston, D. M. Eckley and I. G. Goldberg. *WND-CHARM: Multi-purpose image classification using compound image transforms*. Pattern recognition letters, volume 29(11):pages 1684–1693 (2008).
- [Otsu 1979] N. Otsu. *A Threshold Selection Method from Gray-Level Histograms*. IEEE Transactions on Systems, Man, and Cybernetics, volume 9(1):pages 62–66 (1979). ISSN 0018-9472.
- [Pawlina 2006] W. Pawlina and M. Ross. *Histology: A Text and Atlas*. Lippincott Williams & Wilkins, Baltimore, MD (2006).
- [Pearson 1895] K. Pearson. *Note on regression and inheritance in the case of two parents*. Proceedings of the Royal Society of London, volume 58(347-352):pages 240–242 (1895).
- [Pedregosa 2011] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot and E. Duchesnay. *Scikit-learn: Machine learning in Python*. Journal of Machine Learning Research, volume 12:pages 2825–2830 (2011).

- [Podlozhnyuk 2007] V. Podlozhnyuk. *Image convolution with CUDA*. NVIDIA Corporation white paper, June, volume 2097(3) (2007).
- [Pollheimer 2010] M. J. Pollheimer, P. Kornprat, R. A. Lindtner, L. Harbaum, A. Schlemmer, P. Rehak and C. Langner. *Tumor necrosis is a new promising prognostic factor in colorectal cancer*. *Human pathology*, volume 41(12):pages 1749–1757 (2010).
- [Preparata 1985] F. P. Preparata and M. I. Shamos. *Introduction*. In *Computational Geometry*, pages 1–35. Springer (1985).
- [Prewitt 1977] J. M. Prewitt. *Interactive decision-making for picture processing*. In *Decision and Control including the 16th Symposium on Adaptive Processes and A Special Symposium on Fuzzy Set Theory and Applications, 1977 IEEE Conference on*, pages 373–379. IEEE (1977).
- [Prewitt 1978] J. Prewitt and S. Wu. *An application of pattern recognition to epithelial tissues*. In *Proceedings of the Annual Symposium on Computer Application in Medical Care*, page 15. American Medical Informatics Association (1978).
- [Prewitt 1979] J. M. Prewitt. *Graphs and grammars for histology: An introduction*. In *Proceedings of the Annual Symposium on Computer Application in Medical Care*, page 18. American Medical Informatics Association (1979).
- [Qlucore 2008] Qlucore. *Qlucore - The D.I.Y Bioinformatics Software*. <http://www.qlucore.com> (2008).
- [Qureshi 2008] H. Qureshi, O. Sertel, N. Rajpoot, R. Wilson and M. Gurcan. *Adaptive discriminant wavelet packet transform and local binary patterns for meningioma subtype classification*. In *Medical Image Computing and Computer-Assisted Intervention—MICCAI 2008*, pages 196–204. Springer (2008).
- [Rajpoot 2003] K. Rajpoot and N. M. Rajpoot. *Hyperspectral colon tissue cell classification*. MSc DSIP Dissertation, Faculty of Computing Sciences & Engineering, De Montfort University, UK (2003).
- [Rajpoot 2004] K. Rajpoot and N. Rajpoot. *SVM optimization for hyperspectral colon tissue cell classification*. In *Medical Image Computing and Computer-Assisted Intervention—MICCAI 2004*, pages 829–837. Springer (2004).
- [Ramesh 2012] N. Ramesh, B. Dangott, M. E. Salama and T. Tasdizen. *Isolation and two-step classification of normal white blood cells in peripheral blood smears*. *Journal of pathology informatics*, volume 3 (2012).
- [Ramos-Vara 2014] J. Ramos-Vara and M. Miller. *When tissue antigens and antibodies get along: revisiting the technical aspects of immunohistochemistry—the red, brown, and blue technique*. *Veterinary pathology*, volume 51(1):pages 42–87 (2014).
- [Razavian 2014] A. Razavian, H. Azizpour, J. Sullivan and S. Carlsson. *CNN features off-the-shelf: an astounding baseline for recognition*. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, pages 806–813 (2014).
- [Rochow 1994] T. G. Rochow and P. A. Tucker. *Introduction to microscopy by means of light, electrons, X-rays, or acoustics*. Springer Science & Business Media (1994).
- [Rokach 2010] L. Rokach. *Ensemble-based classifiers*. *Artificial Intelligence Review*, volume 33(1-2):pages 1–39 (2010).
- [Rokach 2014] L. Rokach and O. Maimon. *Data mining with decision trees: theory and applications*. World scientific (2014).
- [Rolls 2011] G. Rolls. *An Introduction to Specimen Preparation*. <http://www.leicabiosystems.com/pathologyleaders/an-introduction-to-specimen-preparation/> (2011).

- [Romo 2011] D. Romo, E. Romero and F. González. *Learning regions of interest from low level maps in virtual microscopy*. Diagnostic Pathology, volume 6(Suppl 1):page S22 (2011).
- [Ronneberger 2015] O. Ronneberger, P. Fischer and T. Brox. *U-Net: Convolutional Networks for Biomedical Image Segmentation*. In *Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, volume 9351 of *LNCS*, pages 234–241. Springer (2015). URL <http://lmb.informatik.uni-freiburg.de/Publications/2015/RFB15a>. (available on arXiv:1505.04597 [cs.CV]).
- [Roula 2002] M. Roula, J. Diamond, A. Bouridane, P. Miller and A. Amira. *A multispectral computer vision system for automatic grading of prostatic neoplasia*. In *Proceedings IEEE International Symposium on Biomedical Imaging*, pages 193–196 (2002).
- [Rozenberg 1993] G. Rozenberg and A. Salomaa. *Current trends in theoretical computer science: essays and tutorials*, volume 40. World Scientific (1993).
- [Rugge 2015] M. Rugge, M. Fassan and D. Y. Graham. *Epidemiology of Gastric Cancer*. In V. E. Strong, editor, *Gastric Cancer*, pages 23–34. Springer International Publishing (2015). ISBN 978-3-319-15825-9. URL http://dx.doi.org/10.1007/978-3-319-15826-6_2.
- [Ruifrok 2001] A. C. Ruifrok, D. A. Johnston *et al.* *Quantification of histochemical staining by color deconvolution*. Analytical and quantitative cytology and histology, volume 23(4):pages 291–299 (2001).
- [Rüschoff 2010] J. Rüschoff, M. Dietel, G. Baretton, S. Arbogast, A. Walch, G. Monges, M.-P. Chenard, F. Penault-Llorca, I. Nagelmeier, W. Schlake *et al.* *HER2 diagnostics in gastric cancer—guideline validation and development of standardized immunohistochemical testing*. Virchows Archiv, volume 457(3):pages 299–307 (2010).
- [Saeger 2009] K. Saeger. *VMscope GmbH-Virtuelle Mikroskopie*. <http://www.vmscope.de/virtuelle-mikroskopie> (2009).
- [Sanfeliu 1981] A. Sanfeliu. *An application of a distance measure between graphs to the analysis of muscle tissue patterns*. School of Electrical Engineering, Purdue University (1981).
- [Scholkopf 2001] B. Scholkopf and A. J. Smola. *Learning with kernels: support vector machines, regularization, optimization, and beyond*. MIT press (2001).
- [Sengupta 2005] S. Sengupta, C. M. Lohse, B. C. Leibovich, I. Frank, R. H. Thompson, W. S. Webster, H. Zincke, M. L. Blute, J. C. Cheville and E. D. Kwon. *Histologic coagulative tumor necrosis as a prognostic indicator of renal cell carcinoma aggressiveness*. Cancer, volume 104(3):pages 511–520 (2005).
- [Seni 2010] G. Seni and J. F. Elder. *Ensemble methods in data mining: improving accuracy through combining predictions*. Synthesis Lectures on Data Mining and Knowledge Discovery, volume 2(1):pages 1–126 (2010).
- [Sertel 2009] O. Sertel, J. Kong, H. Shimada, U. Catalyurek, J. H. Saltz and M. N. Gurcan. *Computer-aided prognosis of neuroblastoma on whole-slide images: Classification of stromal development*. Pattern recognition, volume 42(6):pages 1093–1103 (2009).
- [Shamos 1975] M. I. Shamos and D. Hoey. *Closest-point problems*. In *Foundations of Computer Science, 1975., 16th Annual Symposium on*, pages 151–162. IEEE (1975).
- [Sharma 2012] H. Sharma, A. Alekseychuk, P. Leskovsky, O. Hellwich, R. Anand, N. Zerbe and P. Hufnagl. *Determining similarity in histological images using graph-theoretic description and matching methods for content-based image retrieval in medical diagnostics*. Diagnostic pathology, volume 7(1):page 134 (2012).
- [Sharma 2015a] H. Sharma, N. Zerbe, S. Lohmann, K. Kayser, O. Hellwich and P. Hufnagl. *A review of graph-based methods for image analysis in digital histopathology*. The Diagnostic Pathology Journal, volume 1(61) (2015).

- [Sharma 2015b] H. Sharma, N. Zerbe, D. Heim, S. Wienert, H.-M. Behrens, O. Hellwich and P. Hufnagl. *A Multi-resolution Approach for Combining Visual Information using Nuclei Segmentation and Classification in Histopathological Images*. In *Proceedings of the 10th International Conference on Computer Vision Theory and Applications (VISAPP 2015)*. Scitepress, pages 37–46 (2015).
- [Sharma 2015c] H. Sharma, N. Zerbe, I. Klempert, S. Lohmann, B. Lindequist, O. Hellwich and P. Hufnagl. *Appearance-based Necrosis Detection Using Textural Features and SVM with Discriminative Thresholding in Histopathological Whole Slide Images*. In *Bioinformatics and Bioengineering (BIBE), 2015 IEEE International Conference on*, pages 1–6. IEEE (2015).
- [Sharma 2016] H. Sharma, N. Zerbe, D. Heim, S. Wienert, S. Lohmann, O. Hellwich and P. Hufnagl. *Cell nuclei attributed relational graphs for efficient representation and classification of gastric cancer in digital histopathology*. volume 9791, pages 97910X–97910X–19 (2016). URL <http://dx.doi.org/10.1117/12.2216843>.
- [Sharma 2017a] H. Sharma, N. Zerbe, C. Böger, S. Wienert, O. Hellwich and P. Hufnagl. *A comparative study of cell nuclei attributed relational graphs for knowledge description and categorization in histopathological gastric cancer whole slide images*. In *In press: 30th IEEE International Symposium on Computer-Based Medical Systems - IEEE CBMS 2017* (2017).
- [Sharma 2017b] H. Sharma, N. Zerbe, I. Klempert, O. Hellwich and P. Hufnagl. *Deep convolutional neural networks for automatic classification of gastric carcinoma using whole slide images in digital histopathology*. In *Press: Computerized Medical Imaging and Graphics* (2017).
- [Shi 2013] S. Shi. *Emgu CV Essentials*. Packt Publishing Ltd (2013).
- [Shuttleworth 2002a] J. Shuttleworth, A. Todman, R. Naguib, B. Newman and M. Bennett. *Colour texture analysis using co-occurrence matrices for classification of colon cancer images*. In *IEEE Canadian Conference on Electrical and Computer Engineering*, volume 2, pages 1134–1139 (2002). ISSN 0840-7789.
- [Shuttleworth 2002b] J. Shuttleworth, A. Todman, R. Naguib, B. Newman and M. Bennett. *Multiresolution colour texture analysis for classifying colon cancer images*. In *Medicine and Biology, 24th Annual Conference and the Annual Fall Meeting of the Biomedical Engineering Society EMBS/BMES, Proceedings of the Second Joint*, volume 2, pages 1118,1119 (2002). ISSN 0840-7789.
- [Simeonov 2006] R. Simeonov and G. Simeonova. *Fractal dimension of canine mammary gland epithelial tumors on cytologic smears*. *Veterinary Clinical Pathology*, volume 35(4):pages 446–448 (2006).
- [Sirinukunwattana 2016] K. Sirinukunwattana, S. E. A. Raza, Y.-W. Tsang, D. R. Snead, I. A. Cree and N. M. Rajpoot. *Locality sensitive deep learning for detection and classification of nuclei in routine colon cancer histology images*. *IEEE transactions on medical imaging*, volume 35(5):pages 1196–1206 (2016).
- [Sirinukunwattana 2017] K. Sirinukunwattana, J. P. Pluim, H. Chen, X. Qi, P.-A. Heng, Y. B. Guo, L. Y. Wang, B. J. Matuszewski, E. Bruni, U. Sanchez *et al.* *Gland segmentation in colon histology images: The glas challenge contest*. *Medical image analysis*, volume 35:pages 489–502 (2017).
- [Sommer 2012] C. Sommer, L. Fiaschi, F. A. Hamprecht and D. W. Gerlich. *Learning-based mitotic cell detection in histopathological images*. In *Pattern Recognition (ICPR), 2012 21st International Conference on*, pages 2306–2309. IEEE (2012).
- [Sorenson 2008] R. L. Sorenson and T. C. Brelje. *Atlas of Human Histology-A Guide to Microscopic Structure of Cells, Tissues and Organs*, chapter Gastrointestinal Tract, pages 223–246. University of Minnesota Bookstores, 2nd edition (2008).
- [Sorkine 2009] O. Sorkine. *Least-squares rigid motion using SVD*. Technical notes, volume 120(3) (2009).
- [Souza 2014] C. R. Souza. *The Accord.NET Framework*. <http://www.accord-framework.net> (2014).

- [Spanhol 2016] F. A. Spanhol, L. S. Oliveira, C. Petitjean and L. Heutte. *Breast cancer histopathological image classification using convolutional neural networks*. In *Neural Networks (IJCNN), 2016 International Joint Conference on*, pages 2560–2567. IEEE (2016).
- [Spyridonos 2001] P. Spyridonos, D. Ravazoula, K. Cavouras, G. Berberidis and P. Nikiforidis. *Computer-based grading of haematoxylin-eosin stained tissue sections of urinary bladder carcinomas*. *Informatics for Health and Social Care*, volume 26(3):pages 179–190 (2001).
- [Sridhar 2015] A. Sridhar, S. Doyle and A. Madabhushi. *Content-based image retrieval of digitized histopathology in boosted spectrally embedded spaces*. *Journal of Pathology Informatics*, volume 6(1):page 41 (2015). URL <http://www.jpathinformatics.org/article.asp?issn=2153-3539;year=2015;volume=6;issue=1;spage=41;epage=41;aulast=Sridhar;t=6>.
- [Srivastava 2014] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever and R. Salakhutdinov. *Dropout: A simple way to prevent neural networks from overfitting*. *The Journal of Machine Learning Research*, volume 15(1):pages 1929–1958 (2014).
- [Stenkvis 1987] B. Stenkvis, R. Bergström, U. Brinne, I. Hesselius, A. Kiviranta, H. Nordgren, L. Schnürer, U. Stendahl, S. Sténson, J. Söderström *et al.* *Automatic analysis of Papanicolaou smears by digital image processing*. *Gynecologic oncology*, volume 27(1):pages 1–14 (1987).
- [Strigl 2010] D. Strigl, K. Kofler and S. Podlipnig. *Performance and scalability of GPU-based convolutional neural networks*. In *2010 18th Euromicro Conference on Parallel, Distributed and Network-based Processing*, pages 317–324. IEEE (2010).
- [Sucaet 2014] Y. Sucaet and W. Waelput. *Digital Pathology*. SpringerBriefs in Computer Science. Springer (2014).
- [Sutskever 2013] I. Sutskever, J. Martens, G. Dahl and G. Hinton. *On the importance of initialization and momentum in deep learning*. In *Proceedings of the 30th international conference on machine learning (ICML-13)*, pages 1139–1147 (2013).
- [Swinson 2002] D. E. Swinson, J. L. Jones, D. Richardson, G. Cox, J. G. Edwards and K. J. O’Byrne. *Tumour necrosis is an independent prognostic marker in non-small cell lung cancer: correlation with biological variables*. *Lung Cancer*, volume 37(3):pages 235–240 (2002).
- [Tabesh 2007] A. Tabesh, M. Teverovskiy, H.-Y. Pang, V. P. Kumar, D. Verbel, A. Kotsianti and O. Saidi. *Multifeature prostate cancer diagnosis and Gleason grading of histological images*. *Medical Imaging, IEEE Transactions on*, volume 26(10):pages 1366–1378 (2007).
- [Thiran 1996] J. P. Thiran and B. Macq. *Morphological feature extraction for the classification of digital images of cancerous tissues*. *Biomedical Engineering, IEEE Transactions on*, volume 43(10):pages 1011–1020 (1996).
- [Thomas 2008] G. K. Thomas, K. C. Cheng, J. Z. Wang, Y. Liu *et al.* *Towards efficient automated characterization of irregular histology images via transformation to frieze-like patterns*. In *Proceedings of the 2008 international conference on Content-based image and video retrieval*, pages 581–590. ACM (2008).
- [Tjandra 2006] J. J. Tjandra, G. J. Clunie, A. H. Kaye and J. Smith. *Textbook of surgery*. Wiley Online Library (2006).
- [Tkalčič 2003] M. Tkalčič and J. F. Tasič. *Colour spaces: perceptual, historical and applicational background*. In *EUROCON 2003. Computer as a Tool. The IEEE Region 8*, volume 1, pages 304–308. IEEE (2003).
- [Tomassini 2010] M. Tomassini. *Introduction to graphs and networks*. Information Systems Department, HEC, University of Lausanne, Switzerland (2010).
- [Toussaint 1980] G. T. Toussaint. *The relative neighbourhood graph of a finite planar set*. *Pattern recognition*, volume 12(4):pages 261–268 (1980).

- [Toussaint 1991] G. T. Toussaint. *Some unsolved problems on proximity graphs*. In D. Dearholt and F. Harary, editors, *Proceedings of the First Workshop on Proximity Graphs. Memoranda in Computer and Cognitive Science M CCS*, pages 91–224. Citeseer (1991).
- [Ulam 1966] S. Ulam. *Patterns of growth of figures: Mathematical aspects*. In G. Kepes, editor, *Module, Proportion, Symmetry, Rhythm*, pages 64–74. Braziller (1966).
- [Vakkila 2004] J. Vakkila and M. T. Lotze. *Inflammation and necrosis promote tumour growth*. *Nature Reviews Immunology*, volume 4(8):pages 641–648 (2004).
- [Van Diest 1995] P. Van Diest, K. Kayser, G. Meijer and J. Baak. *Syntactic structure analysis*. *Pathologica*, volume 87(3):pages 255–262 (1995).
- [vanderWalt 2014] S. van der Walt, J. L. Schönberger, J. Nunez-Iglesias, F. Boulogne, J. D. Warner, N. Yager, E. Gouillart, T. Yu and the scikit-image contributors. *Scikit-image: Image processing in Python*. *PeerJ*, volume 2:page e453 (2014). ISSN 2167-8359. URL <http://dx.doi.org/10.7717/peerj.453>.
- [vanDiest 1992] P. J. van Diest, J. C. Fleege and J. Baak. *Syntactic structure analysis in invasive breast cancer: analysis of reproducibility, biologic background, and prognostic value*. *Human pathology*, volume 23(8):pages 876–883 (1992).
- [Varma 2002] M. Varma and A. Zisserman. *Classifying images of materials: Achieving viewpoint and illumination independence*. In *Computer Vision—ECCV 2002*, pages 255–271. Springer (2002).
- [Varma 2005] M. Varma and A. Zisserman. *A statistical approach to texture classification from single images*. *International Journal of Computer Vision*, volume 62(1-2):pages 61–81 (2005).
- [Vert 2004] J.-P. Vert, K. Tsuda and B. Schölkopf. *A primer on kernel methods*. *Kernel Methods in Computational Biology*, pages 35–70 (2004).
- [Veta 2015] M. Veta, P. J. Van Diest, S. M. Willems, H. Wang, A. Madabhushi, A. Cruz-Roa, F. Gonzalez, A. B. Larsen, J. S. Vestergaard, A. B. Dahl *et al.* *Assessment of algorithms for mitosis detection in breast cancer histopathology images*. *Medical image analysis*, volume 20(1):pages 237–248 (2015).
- [VMscope 2010a] VMscope. *VMscope Plugin Store*. <http://vmscope.com/doc/VM-PluginStore-eng.pdf> (2010).
- [VMscope 2010b] VMscope. *VMscope Products*. <http://vmscope.com/produkte.html> (2010).
- [Volkman 1996] L. Volkman. *Estimations for the number of cycles in a graph*. *Periodica Mathematica Hungarica*, volume 33(2):pages 153–161 (1996).
- [Voronoi 1907] G. Voronoi. *Nouvelles applications des paramètres continus à la théorie des formes quadratiques*. *Journal für die reine und angewandte Mathematik (Crelle's Journal)*, volume 133:pages 198–287 (1907).
- [VSS 2010] VSS. *Virtual Specimen Scout, Charité- Universitätsmedizin Berlin*. <http://www.specimen-scout.de/home/> (2010). URL <http://www.specimen-scout.de/home/>.
- [Walker 1994] R. F. Walker, P. Jackway, B. Lovell and I. Longstaff. *Classification of cervical cell nuclei using morphological segmentation and textural feature extraction*. In *Intelligent Information Systems, 1994. Proceedings of the 1994 Second Australian and New Zealand Conference on*, pages 297–301. IEEE (1994).
- [Wallis 2010] W. D. Wallis. *A beginner's guide to graph theory*. Springer Science & Business Media (2010).
- [Wang 2014] J. Wang, J. D. MacKenzie, R. Ramachandran and D. Z. Chen. *Identifying neutrophils in H&E staining histology tissue images*. In *Medical Image Computing and Computer-Assisted Intervention—MICCAI 2014*, pages 73–80. Springer (2014).

- [Warneke 2011] V. S. Warneke, H.-M. Behrens, J. T. Hartmann, H. Held, T. Becker, N. T. Schwarz and C. Röcken. *Cohort study based on the seventh edition of the TNM classification for gastric cancer: proposal of a new staging system*. Journal of Clinical Oncology, pages JCO–2010 (2011).
- [Warneke 2013] V. S. Warneke, H. Behrens, C. Böger, T. Becker, F. Lordick, M. Ebert and C. Röcken. *Her2/neu testing in gastric cancer: evaluating the risk of sampling errors*. Annals of Oncology, volume 24(3):pages 725–733 (2013).
- [Watson 1981] D. F. Watson. *Computing the n-dimensional delaunay tessellation with application to voronoi polytopes*. The computer journal, volume 24(2):pages 167–172 (1981).
- [Weidner 2009] N. Weidner, R. J. Cote, S. Suster and L. M. Weiss. *Modern Surgical Pathology: 2-Volume Set, Expert Consult-Online & Print*. Elsevier Health Sciences (2009).
- [Weinstein 2009] R. S. Weinstein, A. R. Graham, L. C. Richter, G. P. Barker, E. A. Krupinski, A. M. Lopez, K. A. Erps, A. K. Bhattacharyya, Y. Yagi and J. R. Gilbertson. *Overview of telepathology, virtual microscopy, and whole slide imaging: prospects for the future*. Human pathology, volume 40(8):pages 1057–1069 (2009).
- [West 2000] D. B. West. *Introduction to Graph Theory*. Prentice Hall, 2 edition (2000). ISBN 0130144002.
- [Weyn 1999] 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*. The Journal of pathology, volume 189(4):pages 581–589 (1999).
- [Wienert 2012] S. Wienert, D. Heim, K. Saeger, A. Stenzinger, M. Beil, P. Hufnagl, M. Dietel, C. Denkert and F. Klauschen. *Detection and Segmentation of Cell Nuclei in Virtual Microscopy Images: A Minimum-Model Approach*. Scientific Reports, volume 2 (2012). URL <http://dx.doi.org/10.1038/srep00503>.
- [Wienert 2013] S. Wienert, D. Heim, M. Kotani, B. Lindequist, A. Stenzinger, M. Ishii, P. Hufnagl, M. Beil, M. Dietel, C. Denkert and F. Klauschen. *CognitionMaster: an object-based image analysis framework*. Diagnostic Pathology, volume 8(1):pages 1–8 (2013). URL <http://dx.doi.org/10.1186/1746-1596-8-34>.
- [Wikipedia 2017] Wikipedia. *Stomach — wikipedia, the free encyclopedia*. <https://en.wikipedia.org/w/index.php?title=Stomach> (2017). [Online; accessed 7-February-2017].
- [Wilson 2005] R. C. Wilson, E. R. Hancock and B. Luo. *Pattern vectors from algebraic graph theory*. Pattern Analysis and Machine Intelligence, IEEE Transactions on, volume 27(7):pages 1112–1124 (2005).
- [Wiltgen 2003] M. Wiltgen, A. Gerger and J. Smolle. *Tissue counter analysis of benign common nevi and malignant melanoma*. International journal of medical informatics, volume 69(1):pages 17–28 (2003).
- [Wong 1983] E. Wong and K. Fu. *A Parallel Algorithm for Muscle Tissue Images Classification*. In *Computer Applications in Medical Care, 1983. Proceedings. The Seventh Annual Symposium on*, pages 751–754. American Medical Informatics Association (1983).
- [Wu 2004] T.-F. Wu, C.-J. Lin and R. C. Weng. *Probability estimates for multi-class classification by pairwise coupling*. Journal of Machine Learning Research, volume 5(Aug):pages 975–1005 (2004).
- [Xia 2015] Y. Xia. *Fine-tuning for image style recognition* (2015).
- [Xu 2016] J. Xu, X. Luo, G. Wang, H. Gilmore and A. Madabhushi. *A Deep Convolutional Neural Network for segmenting and classifying epithelial and stromal regions in histopathological images*. Neurocomputing, volume 191:pages 214–223 (2016). ISSN 0925-2312. URL <http://www.sciencedirect.com/science/article/pii/S0925231216001004>.

- [Xue 2007] Z. Xue, S. Antani, L. R. Long, J. Jeronimo and G. R. Thoma. *Investigating CBIR techniques for cervicographic images*. In *AMIA Annual Symposium Proceedings*, volume 2007, page 826. American Medical Informatics Association (2007).
- [Yang 2007] L. Yang, R. Jin, R. Sukthankar, B. Zheng, L. Mummert, M. Satyanarayanan, M. Chen and D. Jukic. *Learning distance metrics for interactive search-assisted diagnosis of mammograms*. In *Medical Imaging*, pages 65141H–65141H. International Society for Optics and Photonics (2007).
- [Yang 2009] L. Yang, W. Chen, P. Meer, G. Salaru, L. A. Goodell, V. Berstis and D. J. Foran. *Virtual microscopy and grid-enabled decision support for large-scale analysis of imaged pathology specimens*. *Information Technology in Biomedicine, IEEE Transactions on*, volume 13(4):pages 636–644 (2009).
- [Yu 2008] F. Yu and H. H. Ip. *Semantic content analysis and annotation of histological images*. *Computers in biology and medicine*, volume 38(6):pages 635–649 (2008).
- [Zaitoun 1998] A. Zaitoun, H. Al Mardini and C. Record. *Quantitative assessment of gastric atrophy using the syntactic structure analysis*. *Journal of clinical pathology*, volume 51(12):pages 895–900 (1998).
- [Zerbe 2008] N. Zerbe. *Analyse serieller histologischer Schnitte im Hinblick auf die automatische Bestimmung gleichartiger Partikel benachbarter Schnittstufen*. Diplomarbeit, Fachhochschule für Technik und Wirtschaft Berlin (2008).
- [Zerbe 2011] N. Zerbe, P. Hufnagl and K. Schlüns. *Distributed computing in image analysis using open source frameworks and application to image sharpness assessment of histological whole slide images*. *Diagn Pathol*, volume 6(Suppl 1):page S16 (2011).
- [Zerbino 1993] D. Zerbino. *[biopsy: its history, current and future outlook]*. *Likars' ka sprava/Ministerstvo okhorony zdorov'ia Ukrainy*, (3-4):pages 1–9 (1993).
- [Zhao 2005] D. Zhao, Y. Chen and N. Correa. *Statistical categorization of human histological images*. In *Image Processing, 2005. ICIP 2005. IEEE International Conference on*, volume 3, pages III–628. IEEE (2005).
- [Zhou 2002] Z.-H. Zhou, Y. Jiang, Y.-B. Yang and S.-F. Chen. *Lung cancer cell identification based on artificial neural network ensembles*. *Artificial Intelligence in Medicine*, volume 24(1):pages 25–36 (2002).
- [Zimmer 2005] H. Zimmer. *Voronoi and delaunay techniques*. *Proceedings of Lecture Notes, Computer Sciences*, volume 8 (2005).
- [Zitova 2003] B. Zitova and J. Flusser. *Image registration methods: a survey*. *Image and vision computing*, volume 21(11):pages 977–1000 (2003).