



HAL
open science

Automated Mitosis Detection in Color and Multi-spectral High-Content Images in Histopathology : Application to Breast Cancer Grading in Digital Pathology

Humayun Irshad

► **To cite this version:**

Humayun Irshad. Automated Mitosis Detection in Color and Multi-spectral High-Content Images in Histopathology : Application to Breast Cancer Grading in Digital Pathology. Human health and pathology. Université de Grenoble, 2014. English. NNT : 2014GRENM007 . tel-01685237

HAL Id: tel-01685237

<https://theses.hal.science/tel-01685237>

Submitted on 16 Jan 2018

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

THÈSE

Pour obtenir le grade de

DOCTEUR DE L'UNIVERSITÉ DE GRENOBLE

Spécialité : **Computer Science**

Arrêté ministériel :

Présentée par

Humayun IRSHAD

Thèse dirigée par **Prof. Daniel RACOCEANU**
et codirigée par **Dr. Ludovic ROUX**

préparée au sein **Image Pervasive Access Laboratory (IPAL) UMI CNRS**
et de **École Doctorale Mathématiques, Sciences et Technologies de l'Information, Informatique**

Automated Mitosis Detection in Color and Multi-spectral High-Content Images in Histopathology: Application to Breast Cancer Grading in Digital Pathology

Thèse soutenue publiquement le **20th January 2014**,
devant le jury composé de :

Prof., William PUECH

Université Montpellier 2, France, Rapporteur

A/Prof., Nasir RAJPOT

Warwick University, UK, Rapporteur

Prof., Jean-Marc CHASSERY

DR CNRS, Grenoble, France, Examineur, Président

Prof., Frédérique CAPRON

GHU-PS, Paris, France, Examineur

Dr., Alexandre GOUAILLARD

Temasys Communications, Singapore, Examineur

Dr., Jacques KLOSSA

TRIBVN, Chatillon, France, Examineur

Prof., Daniel RACOCEANU

UPMC-CNRS, Paris, France, IPAL UMI CNRS, Singapore, Directeur de thèse

Dr., Ludovic ROUX

UJF Grenoble 1, France, IPAL UMI CNRS, Singapore, Co-Directeur de thèse



*To my beloved wife, Saba HAYAT,
for all she sacrificed with love*

Acknowledgment

To complete a PhD thesis project is very difficult by working alone. This is a fact that I experienced during all these years since I started and all the way through the end. Therefore, I desire to thank all the people who were involved or contributed in one way or another to this challenging pursuit.

I would like to express my gratitude to my supervisors Prof. Daniel RACOCEANU from Université Pierre et Marie Curie (UPMC), Paris and director of Image and Pervasive Access Lab (IPAL), UMI CNRS 2955 Singapore, and Dr. Ludovic ROUX from Université Joseph Fourier, Grenoble for their jointly supervision, experience and professionalism generously displayed throughout my PhD duration. This thesis would have not come to life without their vision and collaborative efforts. They offered me the possibility of three years research on MICO (MICroscopy COgnitive) project at IPAL. I would like to thank my advisor Dr. Alexandre GOUAILLARD from Temasys Communications, Singapore for providing me his vision and expertise to finalize my PhD Thesis. He has been very supportive and encouraging at all times. Thanks also to my jury members for their valuable inputs to and viewpoints on my research.

I would also like to thank to our collaborators in MICO project, Prof. Frédérique CAPRON from Pitié-Salpêtrière Hosiptal (GHU-PS), Dr. Jacques KLOSSA from TRIBVN, Prof. Patrick BRÉZILLON from LIP6/UPMC, Mr. Youssouf MHOMA from THALES-TC and Dr. Dirk COLAERT from AGFA Healthcare for their valuable inputs. Special acknowledgement goes to Prof. Frédérique CAPRON and Dr. Gilles Le NAOUR from GHU-PS who provided me all the knowledge about the breast cancer grading and all the slides.

I would like to thank to all co-authors of the papers published at conferences and journals. I would like to thank Dr. Jacques KLOSSA for his willingness to use his platform in evaluation of my research work. He also provided us Multispectral Dataset for mitosis detection.

I am also thankful to all the members' of IPAL team and other researchers; I just recall a few, Prof. Mounir MOKHTARI, Prof. Nicolas LOMENIE, Dr. Maria KULIKOVA, Dr. Antoine VEILLARD, Dr. Sepehr JALALI, Dr. Hamdi ALOULOU, Dr. Thibaut TIBERGHIE, Antoine FAGETTE, Stephane RIGAUD, Olivier MORERE, Michal MUKAWA, who encouraged me and help me to enjoy this "slice of life". I remember our lunch times when sharing ideas - mixed with humor - opened a new perspective for me on better understanding.

Special thanks to beloved wife, Saba HAYAT, for her love and patience. She has sacrificed a lot, and supported me with her profound love along the way. Without her, I could not have imagined to obtain my PhD degree. In fact, this is not my accomplishment alone, but our joint accomplishment as a family.

Of course, great thanks to my family for their unconditional love and support through all these years, even if I didn't listen to them and come back home. My parents, Muham-

mad IRSHAD and Nargis PARVEEN, encouraged me to go and finish my PhD degree. I hope I have served as an example for my younger brother, Umar IRSHAD, who just started his undergraduate study.

Last, but not the least, thanks to my friends in Singapore and Malaysia who supported me over the years. Our life away from our homes would not have been so much fun without them. Also, thanks to all of my friends bringing us together and making us feel at home. Special thanks to Dr. Ehsan YOUNIS for his delicious and versatile foods. And finally, many thanks to all of our friends, near of far, for all their encouragement and support.

Resume

IRSHAD HUMAYUN

Intellectual Property

"Mitosis Detector - Mitosis detector for histopathology", Humayun Irshad, Ludovic Roux, Daniel Racoceanu, copyright CNRS (Statement Software) No DL 05963-01 for the IPAL UMI 2955, 2013.

Education

Jan-2014	Doctor of Philosophy	UJF Grenoble 1, France
Jun-2009	Master of Science	NUCES Islamabad, Pakistan
Sep-2005	Bachelor of Science	UCP Lahore, Pakistan

Trainings and Summer Schools

Oct-2013	Business English	British Council Singapore
Oct-2013	Leading a Scientific Project	instn CEA CNRS France
Oct-2013	Occupational First Aid	St. John Ambulance Singapore
Sep-2011	CIMST Multimodal BioMedical Imaging	ETH Zurich, Switzerland
Jun-2011	Machine Learning Summer School	NUS, Singapore
May-2011	RadioPath 2011 Course	NUHS, Singapore

Distinctions & Honors

1. Ranked second of 17 participants during **ICPR 2012 Contest** on Mitosis Detection in Breast Cancer Histological Images
2. Winner of the **World Summit Youth Award 2009** out of 612 participant projects
3. Runner up for the **P@sha ICT Award 2009** in Research and Development Category
4. Best undergraduate student at University of Central Punjab Lahore Pakistan

Experience

Nov-2010 - present	Research Engineer	IPAL-CNRS Singapore
jan-2010 - Oct-2010	Lecturer	CIIT Islamabad, Pakistan
Feb-2009 - Jan-2010	Research Engineer	nexGIN RC Islamabad, Pakistan
Mar-2008 - Jan-2009	Software Engineer	Cogilent Solutions Pakistan
Jan-2007 - Dec-2008	Teaching Assistant	NUCES Islamabad, Pakistan
Jan-2006 - Aug-2006	Teaching Assistant	GIKI Swabi, Pakistan
Jan-2005 - Aug-2006	Software Engineer	RIDOS Solutions Pakistan

Publications

Scientific Journals

H. Irshad, A. Gouaillard, L. Roux, D. Racoceanu, "Multispectral Band Selection and Spatial Characterization: Application to Mitosis Detection in Breast Cancer Histopathology", in *Computerized Medical Imaging and Graphics (CMIG)*, (Submitted).

H. Irshad, A. Veillard, L. Roux, D. Racoceanu, "Methods for Nuclei Detection, Segmentation and Classification in Digital Histopathology: A Review. Current Status and Future Potential", in *IEEE Reviews on Biomedical Engineering (RBME)*, 2013, issue 99, pp. 1.

H. Irshad, I. Hassan, J. Iqbal, A. R. Aghdam, M. Kamalpour, "m-Health System Support For LHWs Working in Rural Areas", in *Journal Science Intenational-Lahore*, July-Sept., 2013, Vol. 25, issue 3, pp. 653-655.

H. Irshad, "Automated Mitosis Detection in Histopathology using Morphological and Multi-channel Statistics Features", in *Journal of Pathology Informatics (JPI)*, May, 2013, vol. 4, issue 1, pp. 10.

L. Roux, D. Racoceanu, N. Loménie, M. Kulikova, **H. Irshad**, J. Klossa, F. Capron, C. Genestie, G. L. Naour, M. N. Gurcan, "Mitosis detection in breast cancer histological images An ICPR 2012 contest", in *Journal of Pathology Informatics (JPI)*, May, 2013, vol. 4, issue 1, pp. 8.

H. Irshad, S. Jalali, L. Roux, D. Racoceanu, L. J. Hwee, G. L. Naour, F. Capron, "Automated Mitosis Detection using Texture, SIFT Features and HMAX Biologically Inspired Approach", in *Journal of Pathology Informatics (JPI)*, March, 2013, vol. 4, issue 2, pp. 12.

Technical White Paper (pubmed Indexed)

H. Irshad, S. Rigaud, A. Gouaillard, "Primal / Dual Mesh with Application to Triangular / Simplex Mesh and Delaunay / Voronoi", in *Insight Journal*, January-December, 2012.

International Conferences (with published proceedings)

H. Irshad, A. Gouaillard, L. Roux, D. Racoceanu, "Spectral Band Selection for Mitosis Detection in Histopathology", in *IEEE 11th International Symposium on Biomedical Imaging (ISBI)*, Beijing China, Apr-May, 2014.

H. Irshad, L. Roux, D. Racoceanu, "Multi-channels Statistical and Morphological Features based Mitosis Detection in Breast Cancer Histopathology", in *Proc. of 35th International Conference of IEEE Engineering in Medicine and Biology Society (EMBC)*, Osaka, Japan, July, 2013, pp. 6091-6094.

H. Irshad, L. Roux, O. Morère, D. Racoceanu, G. L. Naour, and F. Capron, "Détection automatique et calcul du compte de mitoses sur lames H&E", in *Recherche en Imagerie et Technologies pour la Santé (RITS)*, Bordeaux, France, April, 2013.

S. Naz, **H. Irshad**, H. Majeed, "Image Segmentation using Fuzzy Clustering: A Survey", in *Proc. of IEEE International Conference on Emerging Technologies (ICET)*, Islamabad, Pakistan, October, 2010, pp. 181-186.

H. Irshad, M. Kamran, A. B. Siddiqui, A. Hussain, "Image Fusion using Computational Intelligence: A Survey", in *Proc. of IEEE Second International Conference on Environment and Computer Science (ICECS)*, Dubai, UAE, December, 2009, pp. 128-132.

International Conferences & Workshops (with no proceedings)

H. Irshad, L. Roux, D. Racoceanu,, "Multi-channel Statistics Features based Mitosis Detection in Histopathology", in *International Workshop on Pattern Recognition and Healthcare Analytics (IWPRHA), 21st International Conference on Pattern Recognition (ICPR)*, Tsukuba, Japan, November, 2012.

H. Irshad, S. Jalali, L. Roux, D. Racoceanu, L. J. Hwee, G. L. Naour, F. Capron, "Automated Mitosis Detection Using Texture, SIFT Features and HMAX Biologically Inspired Approach", in *Workshop on Histopathology Image Analysis (HIMA), 15th International Conference on Medical Image Computing and Computer Assisted Intervention (MICCAI)*, Nice, France, October, 2012.

H. Irshad, S. Athar, F. Shahzad, M. Farooq, "M-Health System with focus on Antenatal Care for Rural Areas", in *First International eHealth Conference (e-HAP)*, Karachi, Pakistan, January, 2010.

H. Irshad, S. Athar, F. Shahzad, A. Bashir, F. Jehan, "On The Move Ultrasound Diagnosis on Mobile", in *First International eHealth Conference (e-HAP)*, Karachi, Pakistan, January, 2010.

J. Afridi, M. Kamran, **H. Irshad**, S. Khan, M. Farooq, "Use of CDSS on the Personal Digital Assistant of the Medical Expert", in *First International eHealth Conference (e-HAP)*, Karachi, Pakistan, January, 2010.

Contents

Abstract	1
Acronyms	5
Notations	7
List of Figures	9
List of Tables	13
1 Role of Image Analysis in Histopathology	15
1.1 Introduction	15
1.2 Histopathology	15
1.3 Histopathology Imaging	16
1.4 Computer Aided Diagnosis Systems in Histopathology	18
1.5 Cancer and Grading System	20
1.6 Motivation of Our Study	25
1.7 Thesis Strucure	27
1.8 Conclusion	27
2 Review of Quantitative Image Analysis Methods in Histopathology	29
2.1 Introduction	29
2.2 Image-Processing Methods	30
2.3 Preprocessing	37
2.4 Nuclei Detection, Segmentation and Classification Methods	39
2.5 Spectral and Spatial Characterization	51
2.6 Performance Metrics	52
2.7 Evaluation Methods	54
2.8 Inspection and Editing Software	55
2.9 Limitations and Challenges in Previous Frameworks	55
2.10 Overview of Proposed Framework and Scientific Contributions	57
2.11 Conclusion	59
3 Automated Mitosis Detection in Color (RGB) Images	61
3.1 Introduction	61
3.2 Challenges in Mitosis Count	62
3.3 Color Dataset	63
3.4 Textural based Mitosis detection in Color images (TMC) Framework	63
3.5 Intensity, Textural and Morphology based Mitosis detection in Color images (ITM ² C) Framework	81

3.6	MICO Platform Prototype	96
3.7	Conclusion	100
4	Automated Mitosis Detection in Multispectral Images	105
4.1	Introduction	106
4.2	Multispectral Dataset	106
4.3	Multispectral Intensity, Textural & Morphology-based Mitosis detection in Multispectral images (MITM ³) Framework	106
4.4	Experiments and Results	114
4.5	Discussion	121
4.6	Conclusion	127
5	Orientable 2 - Manifold Meshes and Dynamic Sampling	129
5.1	Introduction	129
5.2	Surfaces and Meshes	130
5.3	Duality	133
5.4	Implement Duality in ITK	134
5.5	Validation	138
5.6	Dynamic Sampling for Cyto-Nuclear Atypia Score in MICO Platform . . .	140
5.7	Conclusion	141
6	Overall Conclusion and Future Perspectives	147
6.1	Conclusion	149
6.2	Future Perspectives	151
A	Glossary	153
	Bibliography	155
	Index	165

Abstract

Automated Mitosis Detection in Color and Multi-spectral High-Content Images in Histopathology: Application to Breast Cancer Grading in Digital Pathology

Digital pathology represents one of the major and challenging evolutions in modern medicine. Pathological exams constitute not only the gold standard in most of medical protocols, but also play a critical and legal role in the diagnosis process. Diagnosing a disease after manually analyzing numerous biopsy slides represents a labor-intensive work for pathologists. Thanks to the recent advances in digital histopathology, the recognition of histological tissue patterns in a high-content Whole Slide Image (WSI) has the potential to provide valuable assistance to the pathologist in his daily practice. Histopathological classification and grading of biopsy samples provide valuable prognostic information that could be used for diagnosis and treatment support. Nottingham grading system is the standard for breast cancer grading. It combines three criteria, namely tubule formation (also referenced as glandular architecture), nuclear atypia and mitosis count. Manual detection and counting of mitosis is tedious and subject to considerable inter- and intra-reader variations. The main goal of this dissertation is the development of a framework able to provide detection of mitosis on different types of scanners and multispectral microscope.

The main contributions of this work are eight fold. First, we present a comprehensive review on state-of-the-art methodologies in nuclei detection, segmentation and classification restricted to two widely available types of image modalities: H&E (Hematoxylin Eosin) and IHC (Immunohistochemical). Second, we analyse the statistical and morphological information concerning mitotic cells on different color channels of various color models that improve the mitosis detection in color datasets (Aperio and Hamamatsu scanners). Third, we study oversampling methods to increase the number of instances of the minority class (mitosis) by interpolating between several minority class examples that lie together, which make classification more robust. Fourth, we propose three different methods for spectral bands selection including relative spectral absorption of different tissue components, spectral absorption of H&E stains and mRMR (minimum Redundancy Maximum Relevance) technique. Fifth, we compute multispectral spatial features containing pixel, texture and morphological information on selected spectral bands, which leverage discriminant information for mitosis classification on multispectral dataset. Sixth, we perform a comprehensive study on region and patch based features for mitosis classification. Seven, we perform an extensive investigation of classifiers and inference of the best one for mitosis classification. Eight, we propose an efficient and generic strategy to explore large images like WSI by combining computational geometry tools with a local signal measure of relevance in a dynamic sampling framework.

The evaluation of these frameworks is done in MICO (COgnitive MIcroscopy, ANR TecSan project) platform prototyping initiative. We thus tested our proposed frameworks

on MITOS international contest dataset initiated by this project. For the color framework, we manage to rank second during the contest. Furthermore, our multispectral framework outperforms significantly the top methods presented during the contest. Finally, our frameworks allow us reaching the same level of accuracy in mitosis detection on brightlight as multispectral datasets, a promising result on the way to clinical evaluation and routine.

Keywords

Histopathology, Digital Patholgy, Breast Cancer, Multispectral imaging, Mitotic Count, Nuclei Detection, Nuclei Segmentation, Nuclei Classification,

Détection Automatique de Mitoses dans des Images Histopathologiques Haut-Contenu, Couleur et Multispectrales : Application à la Gradation du Cancer du Sein en Pathologie Numérique

La pathologie numérique constitue l'une des évolutions majeures de la médecine moderne. Les examens pathologiques représentent la référence médicale et légale de la plupart des protocoles médicaux, occupant ainsi une place essentielle dans le processus de diagnostic. Le diagnostic histopathologique d'une maladie par analyse manuelle - au microscope - de nombreuses lames de biopsie, représente un travail intensif, laborieux, pour les pathologistes. Grâce aux progrès récents dans l'histopathologie numérique, la reconnaissance de tissus histologiques dans une image de lame entière à haut contenu a le potentiel de fournir une aide précieuse au médecin, dans sa pratique quotidienne. La gradation de lames de biopsie fournit des informations pronostiques précieuses qui pourraient être utilisées pour le diagnostic et le traitement. Le système de gradation Nottingham représente le standard actuel pour la gradation du cancer du sein. Il combine trois critères, à savoir les architectures glandulaires, les atypies nucléaires et le compte de mitoses. La détection et le comptage manuel des mitoses est un travail fastidieux, sujet à des variations inter-et intra- observateur considérables. L'objectif principal de cette thèse de doctorat est le développement d'un système capable de fournir une détection des mitoses sur des images provenant de différents types de scanners rapides automatiques ainsi que d'un microscope multispectral.

Les principales contributions de ce travail portent sur huit aspects. Tout d'abord, nous présentons un examen complet de l'état de l'art des méthodes de détection, segmentation et classification de noyaux limitée à deux types de modalités d'images largement répandues : H&E (hématoxyline et éosine) et IHC (immunohistochimie). Deuxièmement, nous analysons les informations statistiques et morphologiques concernant les mitoses dans différents canaux de couleurs pour différents modèles de couleurs qui améliorent la détection des mitoses dans les images couleurs (scanners Aperio et Hamamatsu). Troisièmement, nous étudions des méthodes de sur-échantillonnage pour augmenter le nombre d'instances

de la classe minoritaire (mitose) par interpolation entre plusieurs exemples de classes minoritaires qui sont proches les uns des autres, ce qui rend la classification plus robuste. Quatrièmement, nous proposons trois méthodes différentes pour la sélection des bandes spectrales comprenant l'absorption spectrale relative des différents composants des tissus, l'absorption spectrale des colorants H&E et la technique mRMR (minimum Redundancy Maximum Relevance). Cinquièmement, nous calculons les caractéristiques spatiales multispectrales au niveau des pixels, des textures et des informations morphologiques sur les bandes spectrales sélectionnées, qui exploitent l'information discriminante pour la classification des mitoses sur les données multispectrales. Sixièmement, nous procédons à une étude approfondie du calcul de signatures au niveau de la région d'une mitose ou d'un carré englobant une mitose pour le classement de celles-ci. Septièmement, nous effectuons une étude approfondie des classifieurs, afin d'identifier le plus approprié et efficace pour la classification des mitoses. Huitièmement, nous proposons une stratégie performante et générique pour explorer les lames virtuelles, stratégie combinant des outils de géométrie algorithmique avec une mesure locale de pertinence dans le cadre d'un échantillonnage dynamique.

L'évaluation des différents systèmes proposés est effectuée dans le cadre du projet MICO (MICROSCOPIE COGNITIVE, projet ANR TecSan piloté par notre équipe). Dans ce contexte, les systèmes proposés ont été testés sur les données du benchmark international MITOS. En ce qui concerne les images couleur, notre système s'est ainsi classé en deuxième position du concours selon la valeur du critère F-mesure. Par ailleurs, notre système de détection de mitoses sur images multispectrales surpasse largement les meilleurs résultats obtenus durant le concours. Sur les images multispectrales, nous obtenons ainsi le même niveau de précision dans la détection de mitoses que sur les images couleur, avec une précision accrue des bandes spectrales pertinentes, permettant ainsi la mise au point d'un processus ciblé, un résultat prometteur en vue de la validation et de l'utilisation de la détection automatique de mitoses dans un environnement clinique.

Mots-clefs

Histopathologie, Pathologie Numérique, Cancer du Sein, Images Multispectrales, Segmentation et Classification de Noyaux, Détection de Mitoses

Acronyms

Acronyms	Description
Acc	Accuracy
ACM	Active contour model
BR	Blue-ratio
CAD	Computer aided diagnosis
CD	Centroid distance
CNA	Cyto-nuclear atypia
CNN	convolutional neural networks
CV	Cross validation
EM	Expectation maximization
ER	Error rate
DoG	Difference of Gaussian
DT	Decision tree
FCM	Fuzzy c-means clustering
FL	Follicular lymphoma
FNAR	False negative area ratio
FP	False Positives
FPAR	False positive area ratio
FPR	False positive rate
Gcut	Graph cut
GLCM	Grey level co-occurrence matrix
GLRLM	Grey level run-length matrix
GMM	Gaussian mixture model
GT	Ground truth
GVF	Gradient vector flow
HC	Haralick Co-occurrence
HD	Hausdorff distance
H&E	Hematoxylin & Eosin
HES	Hematoxylin Eosin Saffron
HPF	High power field
IDM	Inverse difference moment
IHC	Immunohistochemical
ITM ² C	Intensity, Textural & Morphology based Mitosis detection in Color images
JI	Jaccard index
LDA	Linear discriminant analysis
LoG	Laplacian of Gaussian
L-SVM	Linear support vector machine
MAD	Mean average distance
MAP	Maximum a posteriori

MI	Mutual Information
MITM ³	Multispectral Intensity, Textural & Morphology based Mitosis detection in Multispectral images
MLP	Multilayer Perceptron
MMSF	Morphological & Multispectral Statistical Features
MRF	Markov random field
mRMR	minimum redundancy maximum relevance
MSI	Multispectral Imaging
Ncut	Normalized cut
NGS	Nottingham grading system
NL-SVM	Non-linear support vector machine
OR	Overlap ratio
PPV	Positive predictive value
RL	Run-length
ROI	Region of interest
RST	Radial symmetry transform
SB	Spectral band (multi or hyper)
SDE	Segmentation distortion evaluation
SIFT	Scale-invariant feature transform
SVM	Support vector machine
TMA	Tissue micro array
TMC	Textural based Mitosis detection in Color images
TNR	True negative rate
TP	True positives
TPR	True positive rate
WSI	Whole slide image

Table 1: Description of Acronyms

Notations

Notations	Description
\mathbb{E}	Energy or Force
$H(\cdot)$	Heaviside function
\mathcal{N}	Gaussian distribution
I	Image of size (m, n)
$I(x, y)$	Image pixel value at position x and y
$I(i)$	i^{th} pixel value of image I
$d(i, j)$	Distance between pixel i and j
I_i	Subset (region) of image
$N(i)$	Neighbors of pixel ' i '
U	Total number of pixels in image
P	Probability
$G(V, E)$	Graph with V vertices and E edges
$P_G(u, v)$	Set of paths connecting 2 vertices (u, v)
T	Threshold value
w	Weight
μ	Mean (average)
σ^2	Variance
σ	Standard deviation
ϕ or Φ	Level set
ψ or Ψ	Shape
ϖ	Contour
$\delta(\cdot)$	Dirac delta function

Table 2: Description of Notations

List of Figures

1.1	Examples of H&E, HES and IHC stained breast cancer images.	17
1.2	Different types of breast cancer [7] (A: Ducts, B: Lobules, C: Dilated section of duct to hold milk, D: Nipple, E: Fat, F: Pectoralis major muscle, G: Chest wall/rib wall, H: Normal Duct Nuclei, I: Ductal cancer nuclei breaking the basement membrane, J: Basement membrane, K: Ductal cancer nuclei, L: Lumen /center of duct, M: Normal lobular nuclei, N: Lobular cancer nuclei breaking the basement membrane and O: Lobular cancer nuclei).	21
1.3	Range of breast ductal cancer ([7]).	22
1.4	Pathobiologic events associated with breast ductal cancer ([5]).	23
1.5	Different phases of mitosis.	24
1.6	Different types of nuclei.	26
1.7	Examples of challenges in nuclei detection and classification.	26
2.1	Results of segmentation and separation using different methods on same area of an image.	47
2.2	Segmentation results using ACM methods on probability and Hematoxylin stained image [173].	47
2.3	The count of performance metrics used in nuclei detection, segmentation and classification.	54
2.4	MITOS Dataset generated by Aperio scanner, Hamamatsu scanner and multispectral microscopy. These HPF are selected and annotated by senior pathologist.	59
3.1	Example of ground truth mitosis nuclei for Aperio (first row) and Hamamatsu (second row) scanners.	62
3.2	Some example of non mitosis nuclei for Aperio (first row) and Hamamatsu (second row) scanners. The non mitosis nuclei are located in the centre of each image.	62
3.3	Example of apoptosis and dust particle that looks similar to mitosis nuclei for Aperio (first row) and Hamamatsu (second row) scanners.	63
3.4	TMC Framework.	64
3.5	RGB and BR Image.	65
3.6	Histogram analysis of different channels on Aperio dataset.	66
3.7	Histogram analysis of different channels on Hamamatsu dataset.	66
3.8	Different steps of candidate detection on Aperio image.	68
3.9	Different steps of candidate detection on Hamamatsu image.	69
3.10	The four directions of adjacency that are defined for calculating texture features.	70

3.11	The used architecture of MLP contains one input layer with nodes equal to number of features, one hidden layer and one output layer with two classes.	75
3.12	Candidate detection results (FM and PPV metrics) on four color channels.	78
3.13	TMC classification results using single channel texture features with four classifiers.	79
3.14	ITM ² C Framework.	82
3.15	Histogram analysis of selected channels on Aperio dataset.	83
3.16	Histogram analysis of selected channels on Hamamatsu dataset.	83
3.17	Candidate detection results on selected eight color channels.	86
3.18	ITM ² C classification results using single channel features with four classifiers.	87
3.19	Visual results of mitosis detection on Aperio images (green circles represent TP, blue circles represent FN and yellow circles represent FP).	89
3.20	Visual results of mitosis detection on Hamamatsu images (green circles represent TP, blue circles represent FN and yellow circles represent FP).	90
3.21	Classification results of region and patches based features with L-SVM classifier.	91
3.22	Aperio (first row) and Hamamatsu (second row) patches of mitosis nuclei on which texture features are computed.	93
3.23	The ROC curves of classification result using patch based features with L-SVM classifier.	94
3.24	The margin curve illustrating the prediction margin between mitosis and non mitosis class.	95
3.25	Comparison of ITM ² C framework results with MITOS contest result on Aperio Dataset. IDSIA: Dalle Molle Institute for Artificial Intelligence Research [35], SUTECH: Shiraz university of technology, NEC: NEC Corporation [113], Warwick [87].	96
3.26	Comparison of ITM ² C framework results with MITOS contest result on Hamamatsu Dataset. NEC: NEC Corporation [113].	97
3.27	Some examples of FNs. The missed mitotic nuclei are located in the center of each image. First row images (a-d) from Aperio and second row images (e-f) are from Hamamatsu Dataset.	97
3.28	Some examples of FPs. The false mitotic nuclei are located in the center of each image. First row images (a-f) from Aperio and second row images (g-l) are from Hamamatsu Dataset.	98
3.29	Architecture of MICO 2.0.	98
3.30	MICO ANR TecSan project partners.	99
3.31	Stereology flow used for mitosis score over a ROI.	100
3.32	Territories and frames.	101
3.33	Frames analyzed by ITM ² C Framework are displayed on TRIBVN Calopix platform. The color code is based on the number of mitosis detected in the frame (from blue for zero mitosis to red for 10 or more mitosis).	102
3.34	Snapshot of Mitosis Detection video in MICO 2.0 platform [6].	102
4.1	SBs of the multispectral microscope and examples for each SB.	107
4.2	MITM ³ Framework.	108
4.3	Example of different components of breast tissue in H&E stained histopathological image. Left image sample is a taken from spectral band 8, focal plane 6 of multispectral microscope; right image is taken from Aperio Slide Scanner.	109

4.4	Normalized absorption spectra of four tissue components in 10 spectral bands (SBs). Note that SB 1 (white band), in nature, is different from other SBs and may serve as reference as it covers the whole visible spectrum and contains all the information that other bands are containing, although at a lower resolution. It is separated from other SBs by a dotted line.	110
4.5	Normalized plot of the hematoxylin (blue line) and eosin (red line) dye absorption spectra in MSI and the difference of hematoxylin and eosin (green line).	110
4.6	Relevant contribution of each SB in accumulated MI.	112
4.7	Different steps in candidate detection on breast cancer MSI histopathology.	113
4.8	Top three ranked focal planes using candidate detection results in all SBs.	114
4.9	Candidate detection results on selected SBs.	116
4.10	Classification results using different patch size based MMSF (L-SVM classifier).	117
4.11	Classification results using single SB MMSF with L-SVM classifier.	118
4.12	Plot of FM using SBs selection. Result from using all SBs from left to the current, e.g. SB 2 result uses SB 8, 9, 7, 6, 2. This order is taken from the mRMR ranking. First vertical dotted line shows that selecting first two SBs features matches the previous best result. Second vertical dotted line highlights the overall best result by selecting features up to SBs 0 which shows 25% increased in FM.	119
4.13	Plot of TPR, PPV and FM with L-SVM classifier using the order of SBs selection. Result from using all SBs from left to the current, e.g. SB 2 result uses SB 8, 9, 7, 6, 2. This order is taken from the mRMR ranking.	120
4.14	Classification results on different subset of MMSF using 5-Fold CV.	122
4.15	Comparison of (MITM ³) framework results with MITOS contest result. Vertical dotted line is used to separate the result of contestant's method and proposed method.	123
4.16	The ROC curve of classification result using selected SBs MMSF with L-SVM classifier.	124
4.17	The margin curve illustrating the prediction margin between mitosis and non mitosis class.	124
4.18	The proposed framework results on MITOS Aperio, Hamamatsu and Multispectral Datasets.	126
4.19	Multispectral (first), Aperio (second row) and Hamamatsu (third row) patches of mitosis on which texture features are computed.	126
5.1	Switching from HPF to WSI Analysis.)	130
5.2	Examples of Meshes	131
5.3	Triangulation and Simplex Mesh	132
5.4	Examples of Dual Meshes. (We also sampled the border points.)	133
5.5	a) A 1-Simplex mesh and its dual; b) A 2-Simplex mesh and its dual triangulation; c) same as (b). The dual of the triangulation boundary is considered to extract the simplex mesh.	134
5.6	Existing data structures	135
5.7	QuadEdgeMeshWithDual data structure	136
5.8	Dual borders management options	138
5.9	Primal to dual mesh	138
5.10	Primal to dual mesh with inside hole	139

5.11	Delaunay to Voronoi Mesh	139
5.12	Non-Planer Mesh containing (Triangulation and Simplex Mesh)	140
5.13	Dynamic sampling method applied over a ROI for CNA score.	142
5.14	Dynamic sampling method applied over WSI. The incrementally construction of Voronoi Diagram (VD) (first row) and its Intensity Map (IM) (second row)of score are shown. Each cell contains a single frame at its center. The bright color in IM represents higher CNA score (mean higher degree of malignancy). After 200 iteration, the whole WSI area is being explored. No area seems favored. After 300 iteration, the algorithm converges towards a high CNA. After 500 iteration, the sampling is very dense around this area and remains sparse in others.	143
5.15	Dynamic sampling method applied over WSI. The incrementally construction of Voronoi Diagram (VD) (first row) and its Intensity Map (IM) (second row)of score are shown. Each cell contains a single frame at its center. The bright color in IM represents higher CNA score (mean higher degree of malignancy). After 200 iteration, the whole WSI area is being explored. No area seems favored. After 300 iteration, the algorithm converges towards a high CNA. After 400 iteration, the sampling is very dense around this area and remains sparse in others.	144

List of Tables

1	Description of Acronyms	6
2	Description of Notations	7
1.1	A selection of the most wellknown digital slide scanners in pathology ([147]).	19
1.2	Nottingham Grading System	24
2.1	A summary of state-of-the-art nuclei detection and segmentation frame- works in histopathology	36
2.2	Summary of Nuclei Features used in Histopathology	50
2.3	List of Evaluation Techniques Used in Previous Studies	55
3.1	Number of HPFs and mitosis nuclei in training and evaluation data sets. . .	63
3.2	Resolution of the Aperio and Hamamatsu scanners.	64
3.3	Notation for HC Features	70
3.4	TMC Classification Results on MITOS Aperio Evaluation Dataset (GT = 100).	77
3.5	TMC Classification Results on MITOS Hamamatsu Evaluation Dataset (GT = 100).	80
3.6	Classification Results on MITOS Aperio Full Dataset using 5-Fold CV (GT = 326).	80
3.7	Classification Results on MITOS Hamamatsu Full Dataset using 5-Fold CV (GT = 326).	81
3.8	ITM ² C Classification Results on MITOS Aperio Evaluation Dataset (GT = 100).	88
3.9	ITM ² C Classification Results on MITOS Hamamatsu Evaluation Dataset (GT = 100).	92
3.10	Patch sizes in pixels and μm on the Aperio and Hamamatsu dataset.	92
4.1	SBs Mutual Information (MI) Measure.	112
4.2	Different Rankings of SBs. The upper dotted line shows that SBs 7,8 and 9 are at top three positions in these ranking. The lower dotted line shows that SBs 4 and 5 are at bottom three positions.	115
4.3	Patch sizes in pixels and μm on the MSI Dataset.	116
4.4	Region vs Patch based MITM ³ Classification Results on MSI Evaluation Dataset (GT = 98).	118
4.5	Classification Results on MSI Dataset using 5-Fold CV (GT = 322).	120
4.6	Classification Result on MSI Dataset with SB 1 vs other SBs using 5-Fold CV (GT = 322).	121
4.7	Classification Result on MSI Dataset with blue, green and red SBs (GT = 322) using 5-Fold CV.	122

5.1	QuadEdgeMesh Data Structure	135
5.2	Summaries of changes in new data structure	136

Chapter 1

Role of Image Analysis in Histopathology

Résumé du chapitre

L'objectif principal de cette recherche est l'étude des techniques d'analyse quantitative d'images pour la détection, la segmentation et la classification de noyaux en histopathologie. Bien que ce soit un problème pour le diagnostic et le pronostic de nombreux types de cancers, nous concentrerons notre recherche au développement d'un système automatisé pour la quantification des noyaux, et plus spécifiquement les noyaux de mitoses dans le cancer du sein en histopathologie. Ce chapitre présente les motivations pour l'aide au diagnostic en histopathologie, et plus particulièrement l'importance de l'analyse quantitative d'images pour aider à la gradation du cancer du sein. Nous présentons brièvement la préparation des lames de biopsie et la typologie de l'imagerie en histopathologie. Nous expliquons aussi brièvement le cancer du sein et le système de gradation de Nottingham. Enfin, nous énumérons quelques types de noyaux présents dans les images histologiques de cancer du sein et nous mettons en évidence les défis liés à leur détection.

1.1 Introduction

The main goal of this research is the study of quantitative image analysis techniques for nuclei detection, segmentation and classification in histopathology. Although this is a problem of interest in diagnosis and prognosis of many types of cancer, we focus our research to develop an automated framework for quantification of nuclei, specifically mitosis nuclei, in breast cancer histopathology. This chapter presents the motivation behind computer aided diagnosis (CAD) systems in histopathology, more specifically the importance of quantitative image analysis in breast cancer grading. We shortly present the specimen preparation and the typology of imaging in histopathology. We also briefly explain the breast cancer and the corresponding grading system. At last, we list the different types of nuclei in breast cancer histological images and highlight the challenges in their detection.

1.2 Histopathology

Histopathology is the microscopic examination of nuclei¹ morphology and tissue distribution supplemented with in situ molecular information for a purpose of studying

1. Thereafter, we use the word nuclei to refer both a nucleus and cell(s) in that context.

the manifestation of disease. Its understanding is a key link between anatomy and areas of physiology, pharmacology, molecular biology and pathology. In clinical practice, histopathologists examine histology slides under microscope and identify morphological and structural characteristics at various scales (e.g., nuclei, tubule, follicles, etc.), as an indicator for cancer presence. These characteristics have similar resemblance in benign and malignant tissues.

Histopathologists play a key role both in diagnosing disease entities and determining biomarkers related to the prognosis and response to specific therapy of malignant tumors [120]. The integrative role of histopathologists, particularly in diagnosing malignant tumors and screening for biomarkers related to patients' response to molecular targeted therapy, upgrades their responsibility in therapy decision [108]. Besides, the diagnosis from histopathology image remains the "gold standard" in diagnosing considerable number of diseases including almost all types of cancers [152].

1.3 Histopathology Imaging

1.3.1 Specimen Preparation

In clinical pathology, a biopsy sample is extracted from suspicious lesion and histological sections are placed onto glass slides for a microscopic examination. The slide preparation normally consists of a sequence of steps. First, biopsy samples are extracted from suspicious lesion. Second, fixation is used to stop the metabolic activities in the tissue and preserve the nuclei and their morphology and architectural structure. Third, tissue is processed using dehydration, clearing, infiltration and embedding [30, 127, 182]. Dehydration is applied to remove water, the main constituent of the tissue. Clearing is used to clean the dehydrant (e.g., alcohols) and make the tissue transparent. Infiltration refers to the saturation of the tissue constituents (e.g., nuclei and vacuities). Embedding removes the clearing agent and becomes solidified in molds to provide sufficient external support for sectioning. Next step is sectioning in which the embedded tissue sample is cut into thin sections (e.g., $5\mu m$ thick for light microscopy) to be placed on a slide. Finally, these thin sections are mounted to a glass slide and stained with one or more pigments. The objective of staining is to enhance the contrast and highlight specific intra- or extra cellular structures under the microscope.

Hematoxylin and Eosin (H&E), and Immunohistochemical (IHC) stainings are two widespread staining protocols in histopathology. H&E staining has been used by pathologists for over a hundred years [53] and is still indispensable for recognizing various tissue types and the morphological changes that form the basis of contemporary cancer diagnosis. Hematoxylin stains DNA rich nuclei in blue, while eosin stains cytoplasm in dark pink shade, muscle in medium pink shade, stroma and connective tissue in light pink shade (see Fig. 1.1(a)). Nuclei show varying cell-type and cancer-type specific patterns of concentrated heterochromatin and prominent nucleoli that are diagnostically very important. Due to incorrect concentration and/or pH of eosin, it ends up being two shades of eosin, instead of three shades of eosin, with the muscle and stroma/connective tissue having the same shade. Adding saffron in Hematoxylin Eosin Saffron (HES) gives the stroma/connective tissue a more yellowish shade (see Fig. 1.1(b)). With HES, it is easier to differentiate muscle from stroma/connective tissue. These dyes provide useful visual clues for the segmentation of nuclei.

A limitation of hematoxylin staining is that it is incompatible with immunofluorescence. It is useful, however, to stain one serial paraffin section from a tissue in which

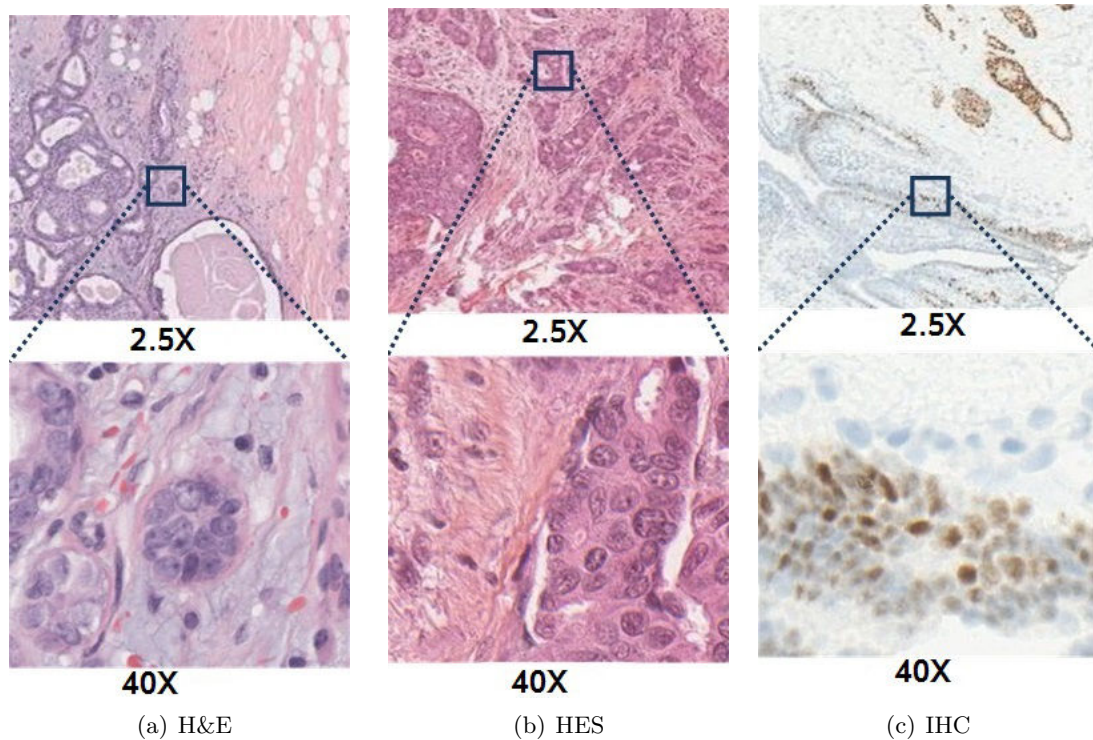


Figure 1.1: Examples of H&E, HES and IHC stained breast cancer images.

immunofluorescence will be performed. Hematoxylin, generally without eosin, is useful as a counter stain for many immunohistochemical or hybridization procedures that use colorimetric substrates (such as alkaline phosphatase or peroxidase).

IHC aims at identifying tissue components by the interaction of target antigens with specific antibodies tagged with a visible label. Samples can be viewed by either light or fluorescence microscopy. Fig. 1.1(c) shows an example of IHC under light microscopy. Using specific tumor markers, physicians use IHC to diagnose a cancer as benign or malignant, determine the stage and grade of a tumor, and identify the cell type and origin of a metastasis to find the site of the primary tumor [38].

1.3.2 Fast Slide Scanners

After slide preparation and staining, next step is digital image acquisition using a fast slide scanner. Slide scanners capture digital images that contain relevant information about specimen at a microscopic level. They are capable of digitizing complete slides at high magnifications. The two main groups of acquisition modalities for microscopic images (WSI devices) can be distinguished: modalities using both a motorized microscope and a camera and slide scanners [169, 147, 126]. For illumination, uniform light spectrum is used to highlight the tissue slide. The imaging system uses one or multiple lenses to magnify the sample and captures digital image with a charge coupled device (CCD) camera.

All manufacturers of slide scanners were invited to participate in the 2nd International Scanner Contest (ISC) 2012 in Berlin, under the auspices of the European Society of Pathology, German Society of Pathology, and the Berufsverband Deutscher Pathologen e.V. (German Professional Organization of Pathologists) [132, 41]. This ISC 2012

contained five domains to evaluate different capabilities of participating systems [2]. During the contest, seven scanning systems (Metafer-VSlide SFx80, NanoZoomer HT 2.0, Panoramic Desk, Panoramic 250, TISSUEScope 4000, UltraFastScanner UFS, VS120 S5) from 6 participants (3DHitech, Hamamatsu, Huron, Metasystems, Olympus, Philips) were evaluated using 32 tests. A list of slide scanners with corresponding specification are shown in Table 1.1.

Color (RGB) Microscopy

Color microscopy is the standard technique in pathology to study the morphology of tissue. In color microscopy, the tissue sample is exposed under visible light from illumination source. A condenser lens is used for focusing the light onto the sample. The dense areas of sample absorb some component of light and produces contrast in the sample. The magnification is limited by the resolving power possible with the wavelength of visible light.

Multispectral Microscopy

In the recent years, scientists have begun to exploit multispectral imaging technology [181, 99, 52, 101, 98, 21, 100, 89] on microscopy image analysis with the objective of determining the makeup of nuclei and other tissue constituents. The principle of this technology is to employ both the visible light and beyond (e.g., ultraviolet and infrared) in the electromagnetic spectrum to collect and process specimens more comprehensively. With data richer than regular brightfield imaging technologies using only visible light, tissue constituents can be more easily identified by using their unique spectral signatures. The main disadvantage of this technique is high cost and complexity for data processing and storage. Its advantage resides into the possibility to control the bandwidth, according to the correlation with the requested patterns. This method could also be used in combination with a fast low magnification color acquisition analysis, by focusing on a precise high-magnified region of interest and a precise (set of) bandwidth(s).

1.4 Computer Aided Diagnosis Systems in Histopathology

After specimen preparation and image production, the resulting digital histology images are analyzed by histopathologists. Manually analyzing numerous slides represents a labour-intensive work for histopathologists, leading to an important inter- and intra-observer variability. Thanks to recent advances in digital histopathology, the recognition of histological tissue patterns in a high-content whole slide images (WSI) within a CAD framework, constitute an important environment for quantitative second opinion and provides diagnostic numerical support. CAD is a blooming interdisciplinary field, combining elements of artificial intelligence and digital image processing with medical knowledge.

Researchers in histopathology have been familiar with the importance of quantitative analysis of histopathological images. These analyses are used to confirm the presence or the absence of disease and also to help in disease progression evaluation. Being important in diagnostic pathology, this quantitative assessment is also used to understand the ground realities for specific diagnostic being rendered like specific chromatin texture in the cancerous nuclei, which may indicate certain genetic abnormalities. In addition, quantitative characterization of pathology imagery is important not only for clinical applications (e.g., to reduce/eliminate inter- and intra-observer variations in diagnosis) but also for research applications (e.g., to recognize the abnormalities for drug discovery [188] and to understand the biological mechanisms of the disease process) [66]. As a consequence, the use of CAD

System	Vendor	Objective/NA	# Slides	Light path	File format
ACIS [®] III	Dako www.dako.com/	Nikon 4x, 10x, 20x, 40x, 60x	100	Halogen lamp	JPEG, TIFF, BMP
Applied Imaging ArioI [®]	www.genetix.com/	Motorized microscope. 1.25x, 5x, 10x, 20x (0.36 μ m/pixel), 40x (0.18 μ m/pixel)	50	Fluorescence	JPEG
dotSlide TM	www.olympus.co.uk/	Motorized microscope. 2x plapon, 10x, 20x, 40x up-Isapo	1 (50)	Halogen lamp, Fluorescence, polarisation	Olympus VSI , JPEG, TIFF , JPEG2000
GoodSpeed [®]	www.alphelys.com/	Motorized microscope	4 or 8	Dark field and fluorescence	JPEG
iSan 2.0 TM	www.bioimagine.com/	Olympus 20x/0.50 (0.46 μ m/pixel) Plan Fluor Olympus 40x/0.75 (0.23 μ m/pixel) Plan Fluor	160	Bright-field, integrated LED	JPEG2000, BIF, TIFF
iScan Coreo Au TM	www.bioimagine.com/	Olympus 4x (0.1 NA), 10x (0.3 NA), 20x (0.5 NA), and 40x (0.75 NA)	160	Bright-field, integrated LED	JPEG2000, BIF, TIFF
iScan Concerto TM	www.bioimagine.com/	Olympus 4x (0.13 NA), 10x (0.3 NA), 20x (0.5 NA), and 40x (0.75 NA) or 60x (0.9 NA, 0.15 μ m/pixel)	80	Integrated LED, Fluorescence	JPEG2000, BIF, TIFF
Pannoramic Desk [®]	www.3dhitech.com/	20x/0.80 (0.23 μ m) Plan Apo. 40x/0.95 Plan Apo	1	Bright-field, halogen lamp	Mirax format, JPEG, TIFF
Pannoramic MidI [®]	www.3dhitech.com/	20x/0.80 (0.23 μ m) Plan Apo. 40x/0.95 Plan Apo	12	Fluorescence	Mirax format, JPEG, TIFF
Pannoramic Scan [®]	www.3dhitech.com/	20x/0.80 (0.23 μ m) Plan Apo. 40x/0.95 Plan Apo	150	Fluorescence	Mirax format, JPEG, TIFF
Pannoramic 250 [®]	www.3dhitech.com/	20x/0.80 (0.23 μ m) Plan Apo. 40x/0.95 Plan Apo	250	Fluorescence	Mirax format, JPEG, TIFF
NanoZoomer [®] HT	www.hamamatsu.com/	20x/0.7. Modes: x20 (0.46 μ m/pixel) x40 (0.23 μ m/pixel)	210	Fluorescence	JPEG, TIFF
NanoZoomer [®] RS	www.hamamatsu.com/	20x/0.75. Modes: x10 (0.92 μ m/pixel) x20 (0.46 μ m/pixel) x40 (0.23 μ m/pixel)	6	Fluorescence	JPEG, TIFF
Nuance FX	www.cri-inc.com/	pectral range 420 \AA 720 nm	1	Bright field and fluorescence multispectral imaging	.im3 for multispectral; 24-bit TIFF for RGB
ScanScope [®] CS	www.aperio.com/	20x/0.75 Plan Apo	5	Bright-field, halogen lamp	TIFF (Aperio SVS), CWS
ScanScope [®] GL	www.aperio.com/	20x/0.75 Plan Apo	1	Bright-field, halogen lamp	TIFF (Aperio SVS), CWS
ScanScope [®] GL-E	www.aperio.com/	20x/0.75 Plan Apo	1	Bright-field, halogen lamp	Composite WebSlide (CWS)
ScanScope [®] OS	www.aperio.com/	60x/1.35 Plan Apo	1	Oil immersion	TIFF (Aperio SVS), CWS
ScanScope [®] XT	www.aperio.com/	20x/0.75 Plan Apo	120	Bright-field, halogen lamp	TIFF (Aperio SVS), CWS
Toco TM	www.claro-inc.jp/	Carl Zeiss Aplan 20x, 40x	1	Bright-field, flat LED	Claro format, JPEG
Vassalo TM	www.claro-inc.jp/	Carl Zeiss Aplan 20x, 40x	80 or 20	Bright-field, flat LED	Claro format, JPEG

Table 1.1: A selection of the most wellknown digital slide scanners in pathology ([147]).

in histopathology can substantially enhance the efficiency and accuracy of histopathologists and overall benefit of the healthcare service.

The research on quantitative image analysis in histopathology imagery can be traced back to the works of Bartels et al. [16], Thiran et al. [168] and Mouroutis et al. [125] but has largely been overlooked, perhaps due to the lack of computational resources and the relatively high cost of digital imaging equipment for pathology. The last few years witnessed a remarkable increase in research studies on digital pathology applications, thanks to recent advances in high-throughput whole slide tissue scanning technology. Quantitative analysis has recently become a vital part in most CAD systems.

Quantitative analysis in histopathology has been conducted for numerous cancer detection and grading applications, including brain [67, 8, 91, 31], breast [164, 60, 65, 133, 39, 79, 80, 17, 45, 73, 176], cervix [85, 71], liver [74], lung [27, 28] and prostate [177, 128, 12] cancer grading. These systems consist of conventional image processing and analysis tools including preprocessing, object detection and segmentation, feature computation, feature selection and classification.

1.5 Cancer and Grading System

The nuclei normally replace themselves through nuclei division (i.e., mitosis), healthy new nuclei take over as old nuclei die out (i.e., apoptosis). Mutation is a phenomenon during nuclei division that results in several different types of changes in gene sequences. Consequently, mutation can turn on respectively off certain genes in nuclei. These nuclei changes induce ability to divide without control or order, producing more similar nuclei. The normal proliferation of nuclei may result in the gross enlargement of an organ, called hyperplasia. Abnormal proliferation of nuclei may result in tumor (neoplasm). Tumor may be benign, pre-malignant (carcinoma in situ) or malignant (invasive cancer). In benign tumors, nuclei proliferate gradually, being close to "normal" (i.e. small, circular, uniform shape, with homogeneous texture) in appearance. They do not invade surrounding tissues and body parts. In malignant tumors, nuclei proliferate rapidly and eventually can spread beyond the original tumor to other parts of the body.

1.5.1 Breast Cancer

Breast cancer refers to a malignant tumor that has developed from nuclei in the breast. Normally, breast cancer either begins in the lobules (glands producing the milk) nuclei, or in the ducts (passages that drain milk from the lobules to the nipple). Breast cancer can often begin in the stromal tissues, which include the fatty and fibrous connective tissues of the breast. Depending on proliferation of nuclei, breast cancer can be classified as carcinoma in situ and infiltrating cancer. In carcinoma in situ, epithelial proliferation is located in the ducts and lobules, without infiltration of the neighboring parts of the organ (see in Figure 1.2(a), 1.2(c)). Infiltration cancer is a cancer invading the mammary tissue, evolving locally then making metastasis (see in Figure 1.2(b), 1.2(d)). A range of cell proliferation in ducts of the breast is described in Figure 1.3.

The molecular, cellular and pathological processes that occur in the transformation from normal tissue to carcinoma in situ tissue and then to breast cancer tissue are shown in Figure 1.4. The main changes that cause breast cancer, including the accumulation of genetic changes, oncogene expression (e.g., HER2/neu), and the loss of normal cell-cycle regulation, appear to have occurred by the time ductal carcinoma in situ (DCIS) is present. At this stage, main clinical features of a subsequent invasive breast cancer are already determined although additional events, including tissue invasion and changes in the surrounding stroma, characterize the invasive tumor [5].

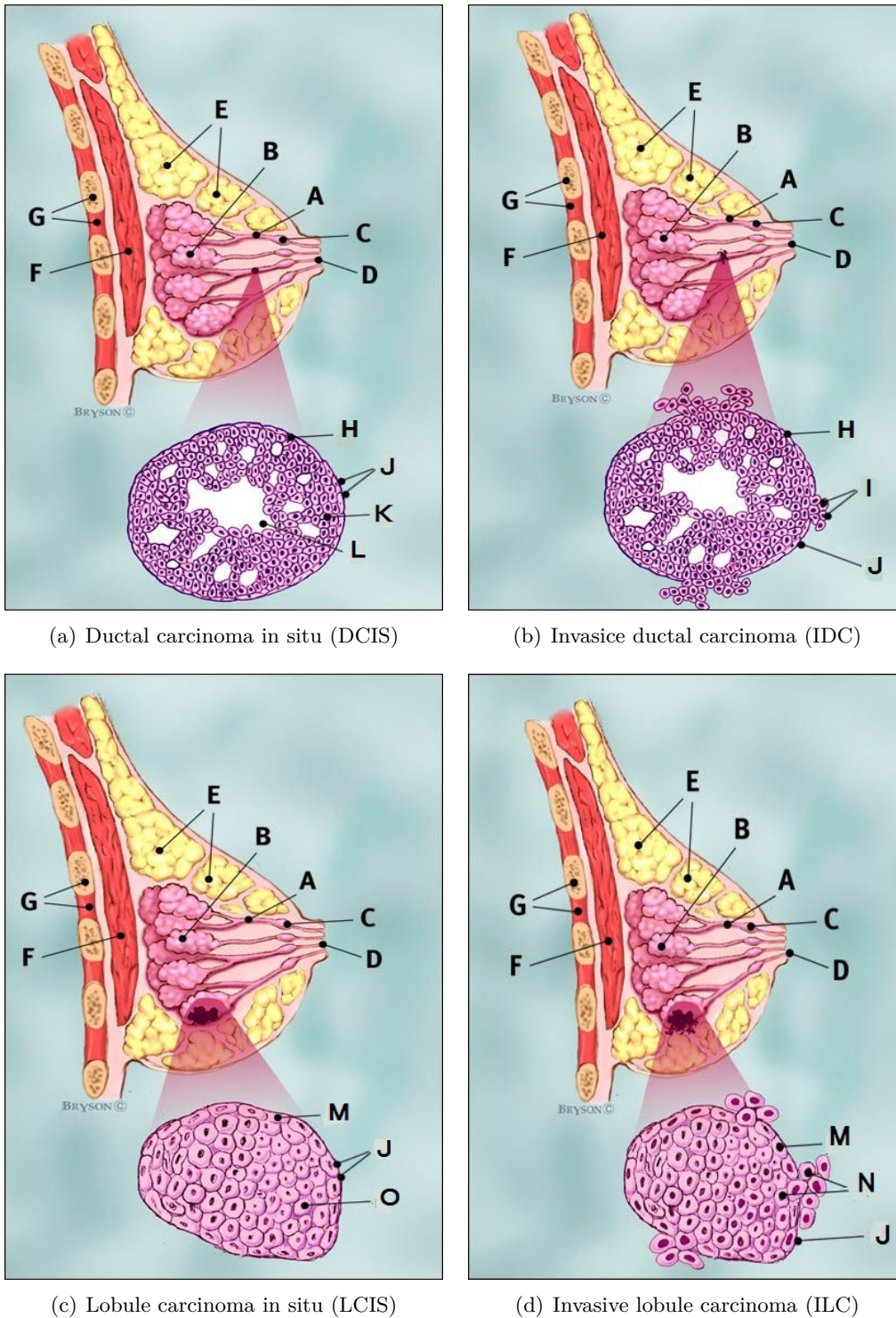


Figure 1.2: Different types of breast cancer [7] (A: Ducts, B: Lobules, C: Dilated section of duct to hold milk, D: Nipple, E: Fat, F: Pectoralis major muscle, G: Chest wall/rib wall, H: Normal Duct Nuclei, I: Ductal cancer nuclei breaking the basement membrane, J: Basement membrane, K: Ductal cancer nuclei, L: Lumen /center of duct, M: Normal lobular nuclei, N: Lobular cancer nuclei breaking the basement membrane and O: Lobular cancer nuclei).

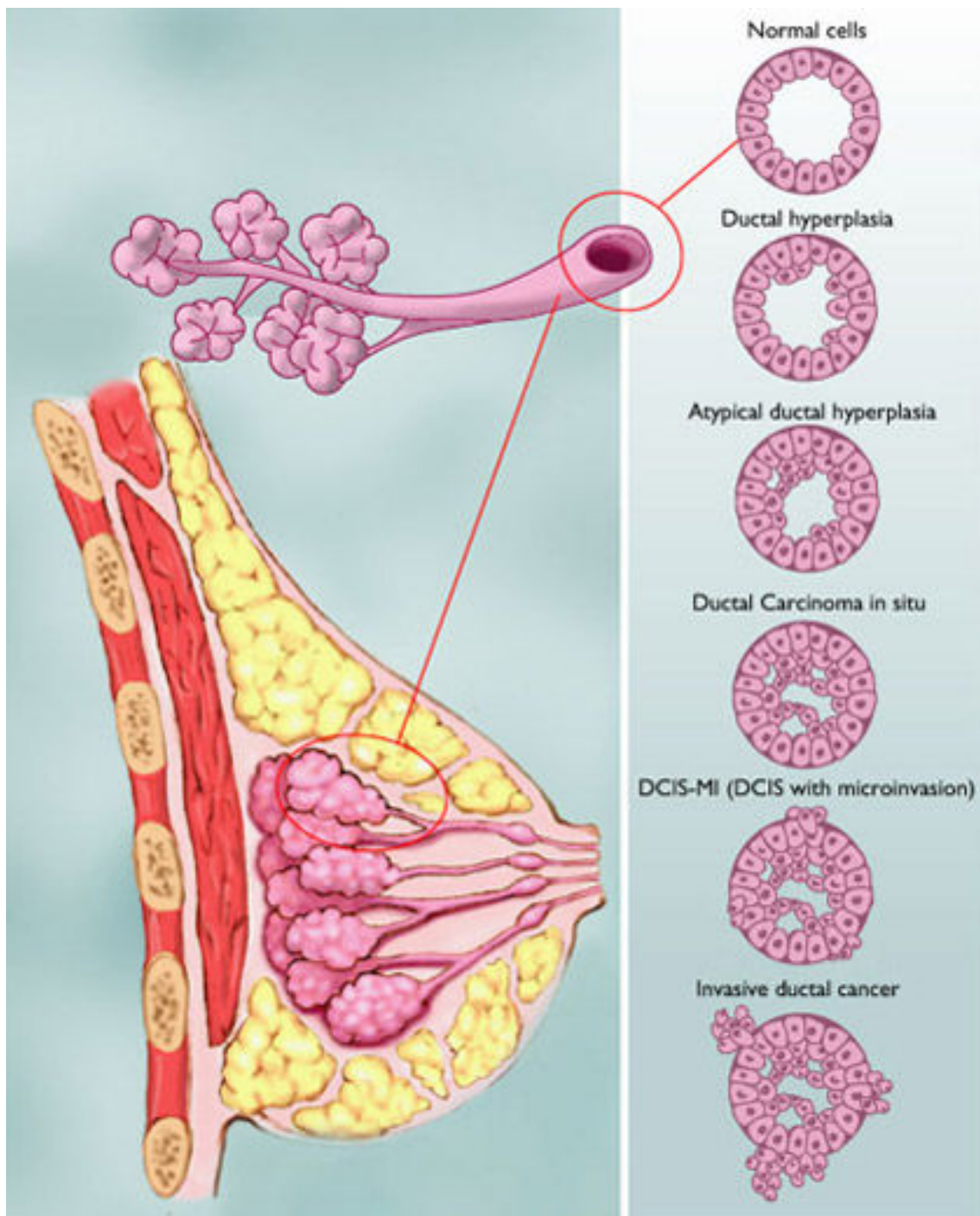


Figure 1.3: Range of breast ductal cancer ([7]).

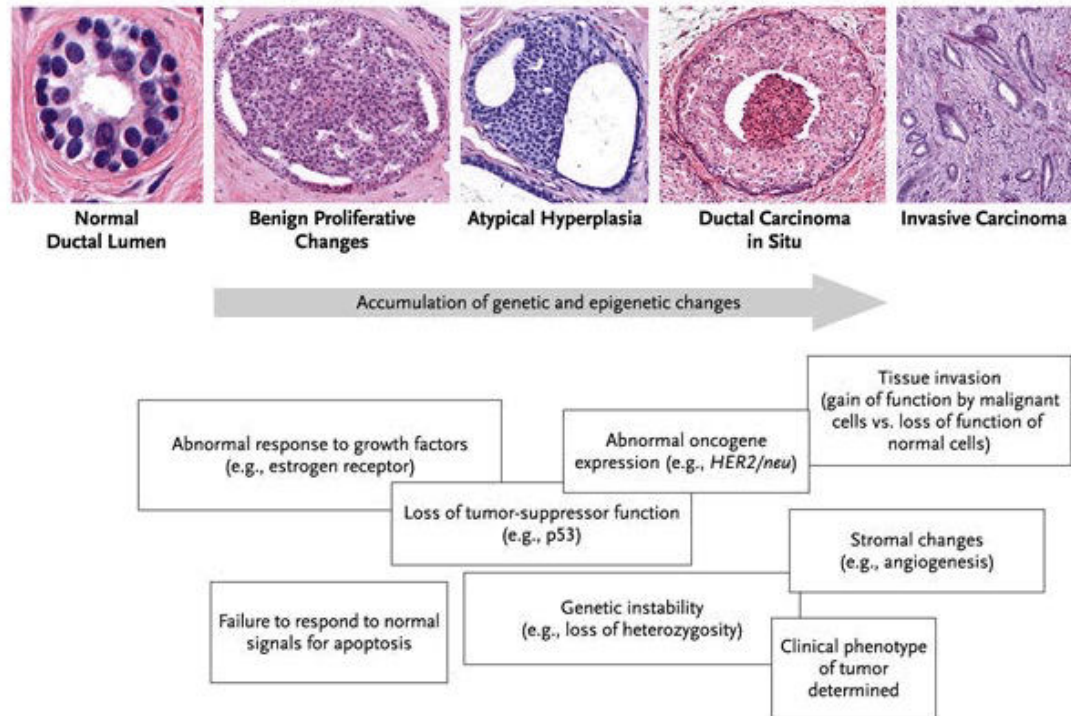


Figure 1.4: Pathobiologic events associated with breast ductal cancer ([5]).

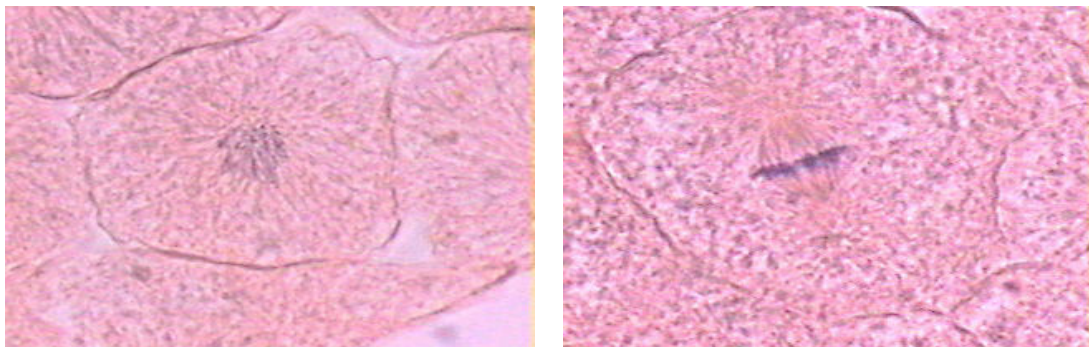
1.5.2 Grading System

Cancer grading refers to how the cancer nuclei look under the microscope compared with normal tissue nuclei. It is different from staging that measures the size of neoplasm and its invasion and metastasis. Infiltrative breast cancers are graded by Nottingham grading system (NGS), an international grading system recommended by World Health Organization. NGS has been proposed by Elston and Ellis [48]. It compares the appearances of the breast cancer tissue to the appearances of normal breast tissue. It influences the prognosis and can affect treatment response. It has three grades I, II and III obtained from the addition of three criteria: gland formation, nuclear atypia and mitotic count. Each of the three criteria are rated 1, 2 or 3 (see Table 1.2). The assessment of the criteria of the grade is semi-quantitative. The addition of the three criteria gives the final grade of breast cancer. The minimum possible score is 3 (1+1+1) and the maximum possible score is 9 (3+3+3). Grade III is assigned to any patient with a score of 8 or 9. Grade II refers to scores of 6 or 7 while grade I refers to scores of 3, 4 or 5.

The mitosis count, alone or within the grade, is currently the best labeling of cell proliferation, and furthermore it is crucial and independent prognosis factor. In NGS, it is stated that Mitotic activity is best assessed at the periphery of the tumor where active growth is most likely. A minimum of 10 high power fields (HPFs) at 40X magnification is assessed [48] by identify truly mitotic nuclei according to Van Diest and Baak criteria (no nucleus membrane, basophilic cytoplasm, hairy extensions clearly recognizable, either as a ball, or on a plane, or as two balls, see Figure 1.5).

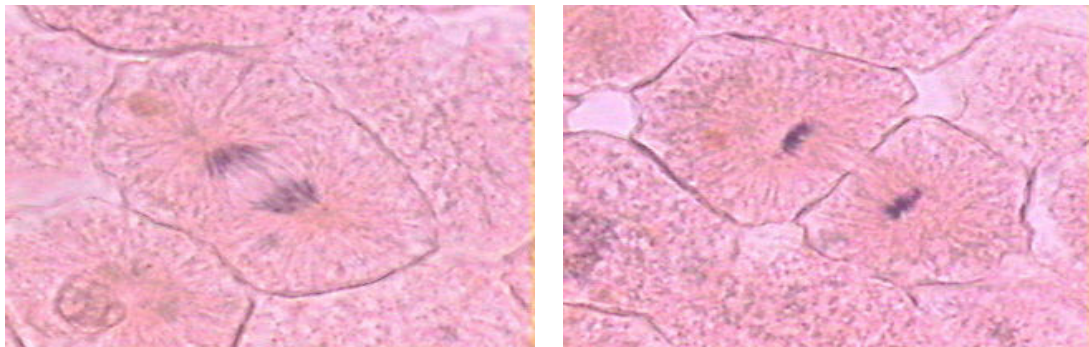
Criteria	Score	Description
Gland formation	1	more than 75% of the tumor forms gland
	2	10 – 75% of the tumor forms gland
	3	less than 10% of the tumor forms gland
Nuclear atypia	1	small, regular uniform nuclei
	2	moderate increase in size and variability
	3	marked variation
Mitosis counts	1	less than 11 mitosis in 10 HPF
	2	between 11 and 20 mitosis in 10 HPF
	3	greater than 20 mitosis in 10 HPF

Table 1.2: Nottingham Grading System



(a) **Prophase** is a stage of mitosis in which the chromatin condenses (i.e. becomes shorter and fatter) into a highly ordered structure called chromosomes, carrying genetic information, align in the middle of the cell before being separated into two daughter nuclei. **Mitosis in prophase stage is not considered for mitosis count.**

(b) **Metaphase** is a stage of mitosis in the eukaryotic cell cycle in which condensed & highly coiled chromosomes, carrying genetic information, align in the middle of the cell before being separated into each of the two daughter nuclei.



(c) **Anaphase** is the stage of mitosis when chromosomes separate in a eukaryotic cell. Each chromatid moves to opposite poles of the cell, the opposite ends of the mitotic spindle, near the microtubule organizing centers.

(d) **Telophase** is a stage of mitosis in an eukaryotic cell in which the effects of prophase and prometaphase events are reversed. Two daughter nuclei form in the cell. The nuclear envelopes of the daughter nuclei are formed from the fragments of the nuclear envelope of the parent cell. As the nuclear envelope forms around each pair of chromatids, the nucleoli reappear.

Figure 1.5: Different phases of mitosis.

1.6 Motivation of Our Study

Worldwide, breast cancer accounts for 22.9% of all cancer in women [23]. It is estimated that about 1 in 8 US women will develop invasive breast cancer over the course of her lifetime [7]. In 2011, an estimated 230,480 new cases of invasive breast cancer were diagnosed in women in the United States, along with 57,650 new cases of non-invasive (in situ) breast cancer [7]. About 39,520 women in the United States were expected to die in 2011 from breast cancer, though death rates have been decreasing since 1990, especially in women under 50. These decreases are thought to be the result of treatment advances, earlier detection through screening and increased awareness [7]. Prognostic assessments and successful treatments for breast cancer vary highly depending on the cancer type, stage, treatment and geographical location of the patient.

Histopathological examination is based on the visual observation of chromatin texture, shapes and sizes of nuclei, size of nucleoli, thickness of nuclear membrane, and regularity of nuclear contour of the population of tumor nuclei that can also be analyzed using quantitative image analysis techniques. Mostly, these image analysis techniques provide more objective prognostic clues, which may be insufficiently observed and quantified by human visual examination. Thus, a computer assisted quantitative image analysis in histopathology is likely to improve the diagnostic and prognostic capabilities and to boost the efficiency of histopathologists by giving a reliable second opinion. These quantitative tools for tissue characterization are also important for understanding the biological mechanism of disease progression.

The most difficult challenge in quantitative image analysis is represented by the spatial analysis, more specifically by the automated nuclei detection, segmentation and classification [54]. The objective of nuclei classification is to assign different labels to different types of nuclei as normal, cancer, mitotic, apoptosis, lymphocytes etc. Quantitative image analysis in cytology has been studied for years and numerous solutions [180, 162, 34, 137, 61] have thus been proposed in the literature. The application of these solutions to histopathology is rather complicated due to the radical differences between the two imaging modalities and to the highly complex image characteristics. Indeed, in cytology imagery, the detection, segmentation and classification of nuclei are generally facilitated due to the well-separated nuclei and the absence of complicated tissue structures. In contrast, the detection, segmentation and classification of nuclei in histopathology imagery are relatively difficult since most of the nuclei are clustered, being parts of complex structures/architectures (tubules, blood vessels, nerves, muscles, DCIS) and zones/territories (neoplasm, fat, necrosis, connective tissue, hyperplasia, fibrosis) which provide a more comprehensive examination and understanding of the evolution of the disease. These complex structures formulate different challenges for quantitative image analysis. Nevertheless, recent works [158, 51, 17, 140, 11, 113, 176] show great potential for computer aided diagnostic of histopathological datasets for breast cancer grading.

Nuclei look different due to different tissue, nuclei type, cancer grade and nuclei life cycle. Having importance in cancer diagnosis and grading, these nuclei are broadly classified into two categories depending on nuclei type: lymphocyte and epithelial nuclei. Lymphocyte nuclei are inflammatory nuclei having regular shape and smaller size than epithelial nuclei [see in Fig. 1.6(a)]. Epithelial nuclei have nearly uniform chromatin distribution with smooth boundary [see in Fig. 1.6(b)]. In high grade cancer tissue, epithelial nuclei, often called cancer nuclei, are larger in size, having heterogeneous chromatin distribution, irregular boundaries and clearly visible nucleoli as compared to normal epithelial nuclei [see in Fig. 1.6(c)]. The variation in nuclei shape, size and texture during nuclei life cycle, mitotic nuclei, is another factor of complexity [see in Fig. 1.6(d)].

Nuclei detection, segmentation and classification are important steps in cancer diagnosis and grading. The presence of nuclei and their aspect are critical signs for evaluating the

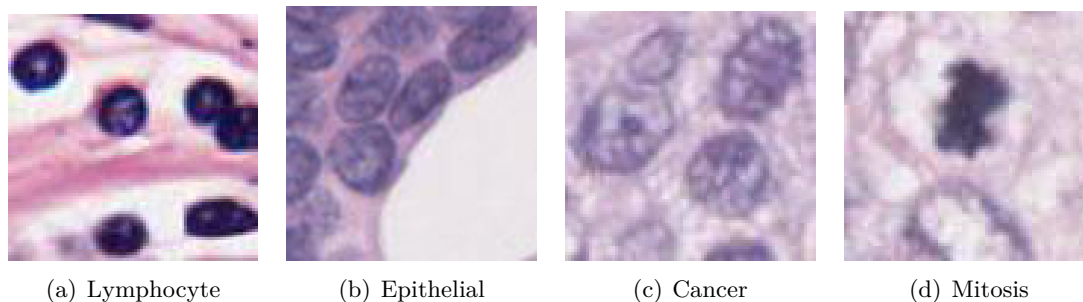


Figure 1.6: Different types of nuclei.

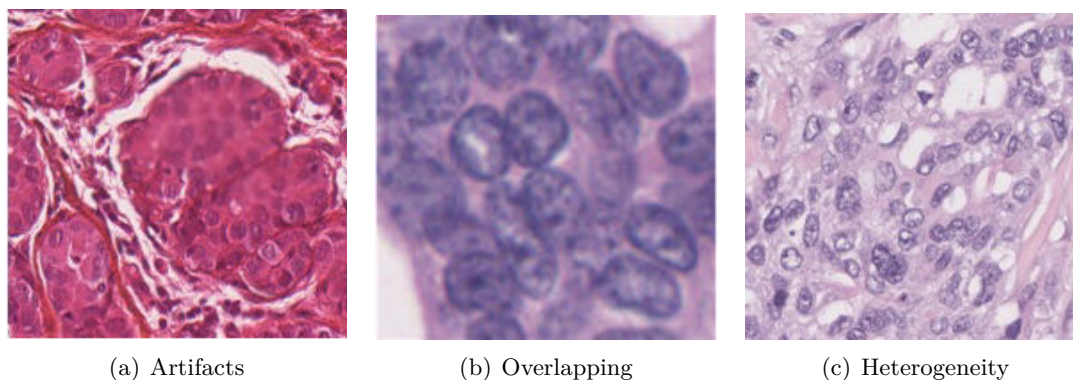


Figure 1.7: Examples of challenges in nuclei detection and classification.

existence of disease and its severity. For example, infiltration of lymphocyte in breast cancer histopathology images are related to patient survival and outcome [10]. Similarly, nuclei pleomorphism has diagnostic value for cancer grading [49, 163, 46]. Furthermore, mitosis count is also an important prognostic parameter in breast cancer grading [49]. Therefore, nuclei detection, segmentation and classification are prerequisites to cancer diagnosis and prognosis.

Automated nuclei detection, segmentation and classification is now a well-studied topic for which a large number of methods have been described in the literature and new methodologies continue to be investigated. Detection, segmentation and classification of nuclei in routinely stained histopathological images pose a difficult computer vision problem due to variations in dyes concentration, artefacts, noise and damaged nuclei boundaries during the slide preparation process, as imperfections in the staining and scanning of the slide. Furthermore, nuclei are clustered and heterogeneous in terms of both intensity gradient and color, even within the same nuclei. This may be due to uneven activation intensity leading to variable color intensity, the superposition of different colors on tissue layers and the variation of the illumination over the tissue specimen. All these problems (highlighted in Fig. 1.7) make the nuclei detection, segmentation and classification a challenging problem. A successful quantitative image analysis approach will have to overcome these issues in a robust way, in order to maintain a high level in the quality and accuracy of nuclei detection, segmentation and classification.

Multispectral imaging (MSI) has the advantage to retrieve spectrally resolved information of a tissue image scene at specific frequencies across the electromagnetic spectrum. MSI captures images with accurate spectral content, correlated with spatial information,

by revealing the chemical and anatomic features of histopathology [98, 100]. This modality provides option to biologists and pathologists to see beyond the RGB image planes to which they are accustomed. Recent publications [52, 101, 183, 89] have begun to explore the use of extra information contained in such spectral data. Specifically, a comparison of spectral methodologies demonstrate the advantage of multispectral data [99, 58]. The added benefit of MSI for analysis in routine H&E histopathology, however, is still largely unknown, although some promising results are presented in [148, 52, 89, 183].

1.7 Thesis Structure

This thesis is structured in six chapters. Chapter 2 gives a comprehensive review on state-of-the-art methodologies in nuclei detection, segmentation and classification in histopathology. Chapter 3 introduces two frameworks for mitosis detection in color images of breast cancer histopathology. Chapter 4 describes automated mitosis detection framework for multispectral images of breast cancer histopathology. Chapter 5 illustrates the orientable 2-Manifold surfaces and dynamic sampling algorithm for selection of frames on WSI. Thesis concludes in chapter 6 with its main contributions and future work.

1.8 Conclusion

This chapter described a brief overview of the histopathology and histology specimen preparation and imaging. It has been demonstrated that histopathological examination is based on the visual observation of chromatin texture, shapes and sizes of nuclei that can also be analyzed using quantitative image analysis techniques. Our study focuses on histopathology images of cancer, in particular breast cancer grading, with its associated breast cancer grading system. It has been discussed that quantitative image analysis in histopathology improves the diagnostic and prognostic capabilities and boosts the efficiency of histopathologists by giving a reliable second opinion. We explained the different types of nuclei in breast cancer and their importance in breast cancer grading, having also highlighted the challenges in nuclei detection, segmentation and classification. The next chapter will introduce a comprehensive review on state-of-the-art methodologies in nuclei detection, segmentation and classification in histopathology.

Chapter 2

Review of Quantitative Image Analysis Methods in Histopathology

Résumé du chapitre

Nous présentons un aperçu détaillé des méthodes d'analyse quantitative des images histopathologiques en général à travers un état de l'art des publications portant sur l'analyse d'images histopathologiques couvrant différentes modalités d'images et plusieurs types de cancer. Nous présentons brièvement les nombreux systèmes développés pour l'analyse d'images histopathologiques qui utilisent un seul ou une combinaison de différents algorithmes de traitement d'image. Nous commençons par introduire les méthodes de traitement d'image les plus couramment utilisées pour l'analyse d'images histopathologiques, puis nous expliquons brièvement les différentes approches utilisées par ces systèmes pour le prétraitement, la détection de noyaux, la segmentation, la séparation des noyaux accolés et la classification. Les méthodes similaires sont regroupées afin de fournir une description plus compacte des techniques générales qui sont utilisées dans la détection des noyaux, la segmentation, la séparation des noyaux accolés et la classification. Enfin, nous soulignons les limites et les défis non résolus par les systèmes existants pour l'analyse quantitative d'image avant de donner un aperçu du système que nous proposons et des nouveautés introduites.

2.1 Introduction

In this chapter, we present an extensive overview of quantitative image analysis methods in general histopathology. This literature review covers a range of image modalities, tissue and cancer type wherein we see a wide range of performance. We shortly explain numerous previous frameworks, which use single or combination of different image-processing algorithms. In the following subsections, we first introduce a short description of most commonly used image-processing methods, in order to later briefly explain different approaches used for preprocessing, nuclei detection, segmentation, separation and classification. We also make a deliberate attempt for grouping similar methods under same heading, in order to provide a more compact description of the general techniques that are used in nuclei detection, segmentation, separation and classification. In last, we point out the limitations and open challenges in existing frameworks for quantitative image analysis and give overview of proposed framework with a list of novelties.

2.2 Image-Processing Methods

We begin with basic definitions. An image I is a function:

$$I : U \longrightarrow [0, 1]^c \quad (2.1)$$

where $U = \llbracket 0; m - 1 \rrbracket \times \llbracket 0; n - 1 \rrbracket$, m and n are the numbers of rows and columns, and c is the number of channels (also called colors), usually $c \in \{1, 3\}$. $I(i)$ is the i th pixel value in the image I , where $i \in U$. A part of image I denoted I_i is a restriction of I to a connected subset of pixels.

2.2.1 Thresholding

Thresholding is a method used for converting intensity image I into a binary image I' by assigning all pixels to the value one or zero if their intensity is above or below some threshold T . Threshold T can be global or local. If T is a global threshold, then I' is a binary image of I as:

$$I'(i) = \begin{cases} 1 & \text{if } I(i) \geq T \\ 0 & \text{otherwise} \end{cases} \quad (2.2)$$

A threshold value can be estimated using computational methods like Ostu method, which determines an optimal threshold by minimizing the intra-class variance [130]. For a given image with L different gray levels, the intra-class variance is:

$$\sigma_\omega^2(T) = \omega_0(T)\sigma_0^2(T) + \omega_1(T)\sigma_1^2(T) \quad (2.3)$$

where ω_0 and ω_1 are probabilities of two classes separated by a threshold T and σ_0^2 and σ_1^2 are the variances of these classes, respectively. With $P(i)$ indicating the probability of the occurrence of gray level i in the image, ω_0 and ω_1 are defined as:

$$\omega_0 = \sum_{i=0}^{T-1} P(i), \quad \omega_1 = \sum_{i=T}^{L-1} P(i) \quad (2.4)$$

Another thresholding technique is local (adaptive) thresholding that handles non-uniform illumination. It can be determined by either splitting an image into sub-images and calculating thresholds for each sub-image or examining the image intensity in the pixel's neighborhood [178].

2.2.2 Morphology

Morphology is a set-theoretic approach that considers an image as the elements of a set [156] and process images as geometrical shapes [72]. The basic idea is to probe an image I with a simple, pre-defined shape, drawing conclusions on how this shape fits or misses the shapes in the image. This simple probe is called structuring element and is subset of the image. The typically used binary structuring elements are crosses, squares and open disks.

The two basic morphological operators are the erosion \ominus and the dilation \oplus . Let $I : U \longrightarrow \{0, 1\}$ be a binary image and $U_f = I^{-1}(\{1\})$ be the foreground pixels. The erosion and dilation of the binary image I by the structuring element $S \in \mathbb{Z} \times \mathbb{Z}$ are defined as:

$$\begin{aligned} \text{Erosion: } U_f \ominus S &= \{x | \forall s \in S, x + s \in U_f\} \\ \text{Dilation: } U_f \oplus S &= \{x + s | x \in U_f \wedge s \in S\} \end{aligned} \quad (2.5)$$

The basic effect of erosion (dilation) operator on a image is to shrink (enlarge) the boundaries of foreground pixels. Two other major operations in morphology are opening \circ

and closing \bullet . Opening is an erosion of an image followed by a dilation; it eliminates small objects and sharpens peaks in the object. Opening is mathematically defined as:

$$U_f \circ S = [U_f \ominus S] \oplus S \quad (2.6)$$

Closing is a dilation of an image followed by an erosion; it fuses narrow breaks and fills small holes and gaps in the image. Closing is mathematically defined as:

$$U_f \bullet S = [U_f \oplus S] \ominus S \quad (2.7)$$

White and black top-hat transforms are two other operations derived from morphology. They allow to extract small elements and details from given images. The white top-hat transform is defined as the difference between image I and its opening as:

$$T_w(I) = U_f - [U_f \circ S] \quad (2.8)$$

The black top-hat transform is defined as the difference between image I and its closing as:

$$T_b(I) = U_f - [U_f \bullet S] \quad (2.9)$$

In addition, morphological gradient, which is the difference between the dilation and the erosion of a given image, is useful for edge detection. It is defined as:

$$G(I) = [U_f \bullet S] - [U_f \circ S] \quad (2.10)$$

2.2.3 Region Growing

Region growing [189] is an image segmentation method consisting of two steps. The first step is selection of seed points and the second step is a classification of neighboring pixels to determine whether those pixels should be added to the region or not by minimizing a cost function. Let $Pr(I_i)$ is a logical predicate which measures the similarity of a region I_i . The segmentation results in a partition of I into regions (I_1, I_2, \dots, I_n) , so that the following conditions hold:

- i. $Pr(I_i) = \text{TRUE}, \forall i = 1, 2, \dots, n$
- ii. $Pr(I_i \cup I_j) = \text{FALSE}, \forall I_i, I_j (i \neq j, i, j = 1, 2, \dots, n)$, adjacent regions

The Pr that often used are grey level (average intensity and variance), color, texture and shape related.

2.2.4 Watershed

Watershed is a segmentation method that usually starts from specific pixels called markers and gradually floods the surrounding regions of markers, called catchment basin, by treating pixel values as a local topography. Catchment basins are separated topographically from adjacent catchment basins by maximum altitude lines called watershed lines. It allows classifying every point of a topographic surface as either belonging to the catchment basin associated with one of the local minima or to the watershed line. Details about watershed can be found in [146]. The basic mathematical definition contains lower slope $LS(i)$, that is the maximum slope connecting pixel i in the image I to its neighbors of lower altitude as:

$$LS(i) = \max_{j \in N(i)} \left(\frac{I(i) - I(j)}{d(i, j)} \right) \quad (2.11)$$

where $N(i)$ is neighbors of pixel i and $d(i, j)$ is the Euclidean distance between pixels i and j . In case of $i = j$, the lower slope is forced to be zero. The cost of moving from pixel i to j is defined as:

$$\text{cost}(i, j) = \begin{cases} LS(i) \cdot d(i, j) & \text{if } I(i) > I(j) \\ LS(j) \cdot d(i, j) & \text{if } I(i) < I(j) \\ \frac{1}{2}(LS(i) + LS(j)) \cdot d(i, j) & \text{if } I(i) = I(j) \end{cases} \quad (2.12)$$

The topographical distance between the two pixels i and j is expressed as:

$$\min_{(i_0, \dots, i_t) \in \Pi} \sum_{k=0}^{t-1} d(i_k, i_{k+1}) \cdot \text{cost}(i_k, i_{k+1}) \quad (2.13)$$

where Π is the set of all paths from i to j .

The catchment basin $CB(m_i)$ of a local minimum m_i is defined as the set of pixels, which have smaller topographical distances to m_i than any other local minimum. The set of pixels, which do not belong to any catchment basin, is referred as the watershed pixels. The watershed transformation is usually computed on the gradient image instead of the intensity image.

2.2.5 Active Contour Models and Level sets

Active contour models (ACMs) or deformable models, widely used in image segmentation, are deformable splines that can be used to delineate structures in an image using gradient information by seeking to minimize an energy function [83]. In nuclei segmentation, the contour points that yield the minimum energy level form the boundary of a cell. The energy function is often defined to penalize discontinuity in the curve shape and gray-level discontinuity along the contour [164]. The general ACM is defined using the energy function \mathbb{E} over the contour points c as:

$$\mathbb{E} = \oint_c (\alpha \mathbb{E}_{\text{Int}}(c) + \beta \mathbb{E}_{\text{Img}}(c) + \gamma \mathbb{E}_{\text{Ext}}(c)) dc \quad (2.14)$$

where E_{Int} controls the shape and length of the contour (often called internal force or energy), E_{Img} influences adjustment of local parts of the contour to the image values regardless of the contour geometry (referring as image force or energy) and E_{Ext} is user defined force or prior knowledge of object to control the contour (referring as external force or energy). α , β and γ are empirically derived constants.

There are two main forms of ACMs. An explicit parametric representation of the contour, called snakes, is robust to image noise and boundary gaps as it constrains the extracted boundaries to be smooth. However, in case of splitting or merging of contours, snakes are restricted for topological adaptability of the model. Alternatively, the implicit ACM, called level sets, is specifically designed to handle topological changes, but they are not robust to boundary gaps and have other deficiencies as well [165]. The basic idea is to determine level curves from a potential function.

2.2.6 K-means Clustering

The K-means clustering [110] is an iterative method used to partition an image into K clusters. The basic algorithm is:

- i. pick K cluster centers, either randomly or based on some heuristic
- ii. assign cluster label to each pixel in the image that minimizes the distance between the pixel and the cluster center

- iii. re-compute the cluster centers by averaging all the pixels in the cluster
- iv. repeat steps ii) and iii) until convergence is attained or no pixel change its cluster

The difference is typically based on pixel value, texture and location, or a weighted combination of these factors. Its robustness depends mainly on the initialization of clusters.

2.2.7 Probabilistic Models

Probabilistic models can be viewed as an extension of K-means clustering. Gaussian mixture models (GMM) are a popular parametric probabilistic model represented as weighted sum of Gaussian cluster densities. The image is modelled according to the probability distribution:

$$P(I(i)) = \sum_{k=1}^K w_k \mathcal{N}(I(i) | \mu_k, \sigma_k^2) \quad (2.15)$$

where K is the number of clusters (objects in the image), μ_k , σ_k^2 and w_k are mean, variance and weight of cluster k , respectively. The w_k are positive real values such that $\sum_{k=1}^K w_k = 1$.

The parameters of GMM are estimated from training data using computation method like Expectation Maximization (EM) [43] that iteratively finds maximum likelihood. The EM is based on the following four steps:

- i. Initialization: The parameters, $\mu_k^{(0)}$, $\sigma_k^{2(0)}$ and $w_k^{(0)}$, are randomly initialized for each cluster C_k .
- ii. Expectation: For each pixel $I(i)$ and cluster C_k , conditional probability $P(C_k|I(i))$ is computed as:

$$P(C_k|I(i))^{(t)} = \frac{w_k^{(t)} \mathcal{N}(I(i) | \mu_k^{(t)}, \sigma_k^{2(t)})}{\sum_{j=1}^K w_j^{(t)} \mathcal{N}(I(i) | \mu_j^{(t)}, \sigma_j^{2(t)})} \quad (2.16)$$

- iii. Maximization: The parameters $\mu_k^{(t)}$, $\sigma_k^{2(t)}$ and $w_k^{(t)}$ of each cluster C_k are now maximized using all pixels and the computed probabilities $P(C_k|I)^{(t)}$ from expectation step as:

$$\mu_k^{(t+1)} = \frac{\sum_i^U P(C_k | I(i))^{(t)} \cdot I(i)}{\sum_i^U P(C_k | I(i))^{(t)}} \quad (2.17)$$

$$\sigma_k^{(t+1)} = \frac{\sum_i^U P(C_k | I(i))^{(t)} \cdot (I(i) - \mu_k^{(t+1)})^2}{\sum_i^U P(C_k | I(i))^{(t)}} \quad (2.18)$$

$$w_k^{(t+1)} = \frac{\sum_i^U P(C_k | I(i))^{(t)}}{U} \quad (2.19)$$

with U , the total number of pixels in I .

- iv. Termination: steps (ii) and (iii) are repeated until parameters converge.

Instead of pixel values, other features can be used like texture. Carson *et al* [26] described the use of a new set of texture features polarity, anisotropy and contrast. Polarity is measure of gradient vector for all neighborhood pixels, anisotropy is a ratio of the eigenvalues of the second moment matrix, and contrast is a measure of homogeneity of pixels.

2.2.8 Graph Cuts

Graph cuts (Gcuts) refers to a wide family of algorithms, in which an image is conceptualized as weighted undirected graph $G(V, E)$ by representing nodes V with pixels, weighted edges E with similarity (affinity) measure between nodes $W : V^2 \rightarrow \mathbb{R}^+$. A similarity

measure is computed from intensity, spatial distribution or any features between two pixels. The Gcuts method partitions the graph into disjoint subgraphs so that similarity is high within the subgraphs and low across different subgraphs. The degree of dissimilarity between two subgraphs A and B can be computed as the sum of weights of the edges that must be removed to separate $A(V_A, E_A)$ and $B(V_B, E_B)$. This total weight is called a cut.

$$\text{cut}(A, B) = \sum_{u \in V_A, v \in V_B} w(u, v) \quad (2.20)$$

An intuitive way is to look for the minimum cut in the graph. However, the minimum cut criterion favors small isolated regions, which are not useful in finding large uniform regions. The normalized cut (Ncut) solves this problem by computing the cut cost as a fraction of total edge connections to all the nodes in the graph. It is mathematically defined as:

$$\text{Ncut}(A, B) = \frac{\text{cut}(A, B)}{\text{asso}(A, G)} + \frac{\text{cut}(A, B)}{\text{asso}(B, G)} \quad (2.21)$$

with $\text{asso}(A, G)$, $\text{asso}(B, G)$, associations of subgraphs A and B with all the nodes in graph, defined as:

$$\text{asso}(A, G) = \sum_{u \in V_A, t \in V} w(u, t), \quad (2.22)$$

$$\text{asso}(B, G) = \sum_{v \in V_B, t \in V} w(v, t) \quad (2.23)$$

Ncut value won't be small for the cut that partitions isolating points, because the cut value will be a large percentage of the total connection from that set to the others. The basic procedure used to find the minimum normalized cut [160] is as follow:

- i. Set up a weighted graph, compute the edge weights matrix W and the diagonal matrix D with size $U \times U$. The W is a similarity matrix where contains elements of $W = (w(i, j))_{i, j \in V}$ that denotes similarity between node i and j . The D is summarized information with $d(i)$ on its diagonal where $d(i)$ is total connection from node i to all the other nodes and defined as:

$$d(i) = \sum_{j \in V} w(i, j) \quad (2.24)$$

- ii. Let s be an indicator vector, $s_i = 1$ if node is in graph A and -1 , otherwise. Let t be the continuous approximation to s defined as:

$$t = (1 + s) - \frac{\sum_{s_i > 0} d_i}{\sum_{s_i < 0} d_i} (1 - s) \quad (2.25)$$

solve the system of equations:

$$(D - W)t = \lambda Dt \quad (2.26)$$

for eigenvectors t and eigenvalues λ [160].

- iii. Use eigenvector with the second smallest eigenvalue to bipartition the graph to find the splitting point such that Ncut is minimized
- iv. decide if the current partition is stable and check the value of the resulting Ncut
- v. Recursively repartition the segmented part if required

These image-processing methods are extensively used in recently proposed frameworks for preprocessing, nuclei detection, segmentation, separation and classification. Based on these image-processing methods, we compiled a list of existing frameworks for nuclei detection, segmentation, separation and classification in histopathology as shown in Table 2.1. In following subsections, we discussed how different image-processing methods have been used.

Ref.	Dataset	Object	Segmentation	Separation	Classification	Metrics
[177]	6 fluorescent prostate images	nuclei	H-maxima transform with watershed	distance transform	-	-
[67]	20 H&E brain images	cancer nuclei	Hysteresis thresholding & morphological operations	watershed transform	Morphology	OR > 0.9
[37]	208 H&E breast WSI	cancer nuclei	Hough transform & ACM	-	morphology & texture with SVM	TPR=0.92, FPR=0.8, F-Score=0.89, Kappa=0.72
[68]	8 H&E prostate images	nuclei	FCM with spatial constraint clustering & multiphase level set	iterative voting & oriented kernel	-	TPR=0.82, PPV=0.8
[186]	86 H&E blood images	nuclei & lymphocytes	gradient based color GVF snake	concavity detection & Dijkstra based edge path selection	-	Acc=0.89
[157]	17 H&E FL WSI	lymphocytes	K-means clustering & watershed	least square ellipse fitting	morphology, texture & topology with Bayesian	-
[39]	6 H&E breast WSI	cancer nuclei	gradient in polar space	-	morphology & texture	Acc Error=7.84
[166]	ki67 breast images	nuclei	Gcuts	-	-	-
[157]	10 H&E brain ROI	mitosis	2-step thresholding with likelihood based posterior probability	-	-	TPR=0.81
[79]	87 IHC breast images	nuclei	adaptive thresholding, H-maxima transform & watershed	H-minima transform based marker extraction	-	TPR=0.96
[80]	63 IHC breast images	nuclei	GMM & EM based topographic surface estimation	Bayesian based cluster analysis & separation using LDA	Bayesian	SDE=0.14
[158]	10 H&E FL WSI	lymphocytes	GMM & EM and adaptive thresholding	RST & spatial voting scheme	morphology & texture	Acc=0.81
[51]	9 H&E breast WSI	lymphocytes	GMM & EM and geodesic ACM	concavity detection & edge path selection	intensity with K-means clustering	TPR=0.8, PPV=0.86, HD=2.1, MAD=0.9, OR=0.72
[17]	41 H&E breast WSI	lymphocytes	region growing & Bayesian and MRF based MAP estimation	-	intensity, morphology & texture with SVM	HD=3.7, Acc > 0.9
[9]	10 H&E breast images	nuclei	Gcuts	-	-	F-Score=0.97
[74]	H&E liver images	cancer nuclei	marker controlled watershed & GVF ACM	-	intensity, morphology & texture using SVM with decision graph	-
[27]	10 IHC lung WSI	cancer nuclei	local adaptive thresholding	watershed	-	Acc=0.8
[8]	320 H&E brain images	nuclei	adaptive thresholding & morphological gradient	texture with Bayesian	texture	Acc=0.925
[128]	17 H&E prostate WSI	cancer nuclei	maximum object likelihood binarization	-	intensity, morphology & texture with SVM	TPR=0.78, FPR=0.6

... continued

Ref.	Dataset	Object	Segmentation	Separation	Classification	Metrics
[45]	95 H&E breast WSI	nuclei	dynamic thresholding & morphology	watershed based immersion simulation	-	-
[175]	19 H&E breast WSI	cancer nuclei	RST & marker controlled watershed	-	-	Acc=0.81
[90]	21 H&E FL images	lymphocytes	local Fourier transform based pixels classification	concavity detection & Fourier shape descriptor	-	Acc=0.76, ER=7.5
[91]	11 H&E brain ROI	nuclei	GVF ACM	distance transform & watershed	-	FPAR=0.23, FNAR=0.13, ER=.36, JI=0.72, CD=3, HD=7, AUC=0.73
[124]	H&E breast images	cancer nuclei	geodesic ACM	concavity detection & Dijkstra based edge path selection	-	-
[149]	8 IHC breast WSI	mitosis	multi-resolution Gcuts	-	texture	TPR> 0.7, TNR=0.8
[11]	80 H&E prostate WSI	nuclei & lymphocytes	watershed	shape priors based ACM in levelset formation	morphology with SVM	TPR=0.86, PPV=0.66, OR=0.91
[140]	234 TMA breast images	nuclei	-	single pass voting with mean shift & levelset	-	TPR=0.78, PPV=0.9, ER=6.63
[95]	6 H&E breast images	cancer nuclei	-	MPP with shape term based ACM	-	F-Score=0.7, JI=0.64
[31]	449 H&E brain images	nuclei	multi-reference Gcuts	convexity constraints	-	TPR=0.85, pPV=0.75
[173]	6 H&E breast images	cancer nuclei	texture based probability map & ACM	-	intensity & texture with SVM	-
[179]	H&E images	nuclei	contour based minimum a priori model	concavity detection	-	TPR=0.91, pPV=0.86
[176]	51 IHC breast WSI	nuclei	clustering & ACM	-	intensity & texture with Adaboost	Acc=0.95
[88]	MITOS [1]	mitosis	GMM & EM	-	intensity with SVM	TPR=0.57, PPV=0.47, F-Score=0.51
[113]	MITOS [1]	mitosis	adaptive thresholding & morphological operations	-	intensity, morphology & texture with SVM	TPR=0.75, PPV=0.59, F-Score=0.66

Table 2.1: A summary of state-of-the-art nuclei detection and segmentation frameworks in histopathology

2.3 Preprocessing

Preprocessing can be performed to compensate for adverse conditions such as the presence of batch effects. Batch effect refers to unevenness in illumination, color or other image parameters recurring across multiple images. Noise reduction and artefacts elimination can also be performed prior to detection and segmentation. Additionally, region of interest (ROI) detection can also be performed in order to reduce processing time.

2.3.1 Illumination Normalization

The illumination can be corrected either by using white shading correction or by estimating the illumination pattern from a series of images. In white shading correction, a blank (empty) image is captured and used to correct images pixel by pixel [117]. A common equation is:

$$\text{Transmittance} = \frac{\text{Specimen value} - \text{Background value}}{\text{White Reference value} - \text{Background value}} \quad (2.27)$$

A downside of this method is that a blank image must be acquired for each objective magnification whenever the microscope illumination settings are altered.

An alternative normalization method is based upon the intrinsic properties of the image which are revealed through Gaussian smoothing [97]. Another possible way is to estimate from background by exploiting the images of the specimen directly, even in the presence of the object [59, 135]. Can *et al* [25] introduced a method to correct nonuniform illumination variation by modeling the observed image $I(i)$ as product of the excitation pattern, $E(i)$, and the emission pattern, $M(i)$ as:

$$I(i) = E(i) \times M(i) \quad (2.28)$$

While the emission pattern captures the tissue dependent staining, the excitation pattern captures the illumination. From a set of J images, $I_j(i)$ is denoted an ordered set of pixels. Assuming that a certain percentage, g , of the image is formed from stained tissue (non-zero background), then a trimmed average of the brightest pixels can be used to estimate the excitation pattern:

$$E'_{AVE}(i) = \frac{1}{J - K + 1} \sum_{j=K}^J I_j(i) \quad (2.29)$$

where K is set to an integer closest to $J(1 - g) + 1$.

2.3.2 Color Normalization

Many color normalization techniques have been proposed [64, 112, 109, 93], including histogram or quantile normalization in which the distributions of the three color channels are normalized separately. Kothari *et al* [93] used histogram based normalization in histopathological images. They proposed a rank function that maps the intensity ranges across all pixels. Alternatively, Reinhard *et al* [143] proposed a method for matching the color distribution of an image to that of reference image by use of a linear transform in a perceptual color model (Lab). Magee *et al* [111] extended Reinhard's normalization approach to multiple pixel classes by using a probabilistic (GMM) color segmentation method. It applies a separate linear normalization for each pixel where class membership is defined by a pixel being coloured by a particular chemical stain or being uncoloured i.e., background.

In order to deal with stains colocalization, a very common phenomenon in histopathological images, color deconvolution is effective in separation of stains [153]. Ruifrok [153]

explains how virtually every set of three colors can be separated by color deconvolution and reconstructs for each stain separately. It requires prior knowledge of color vectors (RGB) of each specific stain. Later, Macenko *et al* [109] proposed the automatic derivation of these color vectors, a method further refined by Niethammer *et al* [129] and Magee *et al* [111]. Several nuclei detection and segmentation methods [37, 27, 175, 173, 87] are using color deconvolution based separation of stains in the histopathological images.

Different color models can be used. Several detection and segmentation methods [67, 68, 39, 74, 27, 8, 140] use the RGB color model, yet the RGB model is not a perceptually uniform color model. Other more perceptual color models such as HSV, Lab and Luv are sometimes used [185, 186, 17, 13, 45, 128, 91, 87, 88, 113].

2.3.3 Noise Reduction and Image Smoothing

Thresholding is used for noise reduction that usually follows filtering and background correction in order to minimize random noise and artefacts [15, 85]. The pixels that lie outside threshold values often determined using intensity histogram are considered to be noisy. Alternatively, applying threshold function on a group of pixels instead of an individual pixel eliminates a noisy region. While such techniques are successful to eliminate small spots of noise, they fail at eliminating large artefacts [69].

Alternatively, morphological operations can also be used for noise reduction. Noise and artefacts are eliminated using morphological operations like closings and openings [175]. Morphological grayscale reconstruction methods are used to eliminate noise while preserving the nuclei shape [74, 79, 80, 88]. While thresholding and filtering reduce noise according to pixel intensities, morphology reduces noise based on the shape characteristics of the input image, as characterized by a structuring element. Morphology cannot distinguish the cellular areas and artefacts having a cell-like shape but different intensity values. Thresholding (prior or subsequent to applying the morphological operations) removes such artefacts.

Adaptive filters [62], Gamma correction [39], and histogram equalization [157] have been used to increase the contrast between foreground (nuclei) and background regions. Anisotropic diffusion is used to smooth nuclei information without degrading nuclei edges [157, 87]. Gaussian filtering is also used to smooth nuclei regions [177, 17, 124].

2.3.4 Region Of Interest Detection

In some frameworks, noise reduction and ROI detection are performed at the same time. For instance, in the case of tissue-level feature extraction, the pre-processing step thresholds the image to identify the ROI by eliminating both noisy regions and those with little content [69]. While, in case of cellular-level feature extraction, noise reduction is followed by ROI detection to determine the nuclei region [87, 88].

Thresholding is popular for ROI detection. In follicular lymphoma (FL) tissue, there are five cytological components: nuclei, cytoplasm, extra-cellular material, red blood cells (RBCs) and background regions. Sertel *et al* [157] introduced the nuclei and cytological components as ROI for grading of FL. RBCs and background regions show uniform patterns as compared to other nuclei in FL tissue; thus thresholding is performed in RGB color model for elimination of RBCs and background. Similarly, Dalle *et al* [39] selected neoplasm ROI by using Otsu thresholding along with morphological operations.

Clustering is another method that commonly used for ROI detection. Cataldo *et al* [27] performed the automated separation of cancer from non-cancerous regions (stroma, blood vessels) using unsupervised clustering. Later, cancer and non-cancerous regions is refined using morphological operations. Dundar *et al* [45] proposed a framework for classification of intraductal breast lesions as benign or malignant using cellular component. The intraductal breast lesions contain four components: cellular, extra cellular, regions with hues of red

and illumina. The H&E stained image data is modeled into four components using GMM. Parameters of GMM model are estimated using EM [43]. The resulting mixture distribution is used to classify pixels into four categories. Those classified as cellular component are further clustered by dynamic thresholding to eliminate blue-purple pixels with relatively less luminance. The remaining pixels are considered cellular region and is used in lesion classification.

Using textural information, Khan *et al* [88] proposed a novel and unsupervised approach to segment breast cancer histopathology images into two regions; Hypo-Cellular Stroma (HypoCS) and Hyper-Cellular Stroma (HyperCS). This approach is employed magnitude and phase spectrum in the Gabor frequency domain to segment HypoCS and HyperCS regions, respectively. For mitosis detection in breast cancer histopathology images using this approach as ROI detection, it reduces the false positive rate (FPR) from four times [87].

2.4 Nuclei Detection, Segmentation and Classification Methods

2.4.1 Detection Methods

Identification of initial markers or seed points, usually one per nuclei and close to its center, is a pre-requisite for most nuclei segmentation methods. The subsequent frameworks use seeds points in order to delineate the spatial extent of each nuclei. Indeed, the accuracy of such segmentation methods depends critically on the reliability of the seed points. The early works in this field relies upon the peaks of the Euclidean distance map [39]. H-maxima transform detects local maxima as seed points [177, 166, 79, 80] but it is overly sensitive to texture and often results in overseeding. Hough transform detects seed points for circular shaped nuclei but requires heavy computation [37]. Centroid transform also detects seeds but limitations make it useful only for binarized images and unable to exploit additional cues.

The Euclidean distance map is commonly used for seeds detection and Laplacian of Gaussian (LoG) is a generic blob detection method. Using multiscale LoG filter with a Euclidean distance map offers important advantages, including computational efficiency and ability to exploit shape and sizes information. Al-kofahi *et al* [9] proposed a distance constrained multiscale LoG filtering method to identify the center of nuclei by exploiting shape and size cues available in the Euclidean distance map of the binarized image. The main steps of this methodology as follow:

- i. Initially, compute the response of the scale-normalized LoG filter ($\text{LoG}_{\text{norm}}(i; \xi) = \xi^2 \text{LoG}(i; \xi)$) at multiple scales $\xi = [\xi_{\min}, \dots, \xi_{\max}]$.
- ii. Use the Euclidean distance map $D_N(i)$ to constrain the maximum scale values when combining the LoG filtering results across scales to compute a single response surface $R_N(i)$ as:

$$R_N(i) = \arg \max_{\xi \in [\xi_{\min}, \xi_{\text{MAX}}]} \{\text{LoG}_{\text{norm}}(i; \xi) * I_N(i)\} \quad (2.30)$$

where $\xi_{\text{MAX}} = \max\{\xi_{\min}, \min\{\xi_{\max}, 2 \times D_N(i)\}\}$ and $I_N(i)$ is the nuclear channel image extracted by separating the foreground pixel from background pixel using automatic binarization.

- iii. Identify the local maxima of $R_N(i)$ and impose a minimum region size to filter out irrelevant minima.

This methodology improves the accuracy of seed locations. The main disadvantage of this methodology is its sensitivity to even minor peaks in the distance map that results in over segmentation and detection of tiny regions as nuclei.

Radial symmetry transform (RST) is also used for seeds detection. Loy and Zelinsky [107] proposed fast gradient based interest operator for detection of seed points having high radial symmetry. Although this approach is inspired by the results of the generalized symmetry transform, it determines the symmetrical contribution of each pixel around it, rather than considering the contribution of a local neighborhood to a central pixel. Veta *et al* [175] also employed RST for seeds detection.

Recently, several other approaches have been proposed to detect the seed points. Qi *et al* [140] proposed a novel and fast algorithm for seed detection by utilizing single-path voting with shifted Gaussian kernel. The shifted Gaussian kernel is specifically designed by amplifying the voting at the center of targeted object and resulted in low occurrence of false seeds in overlapping regions. First, a cone shape $(r_{\min}, r_{\max}, \Delta)$ with its vertex at (x, y) is used to define the voting area $A(x, y; r_{\min}, r_{\max}, \Delta)$ where r_{\min} is a minimum radius, r_{\max} is a maximum radius and Δ is aperture angle of cone. The voting direction $\alpha(x, y)$ is computed using the negative gradient direction $-(\cos(\theta(x, y)), \sin(\theta(x, y)))$ where θ is the angle of the gradient direction with respect to x axis. The voting image $V(x, y; r_{\min}, r_{\max}, \Delta)$ is generated using shifted Gaussian kernel with its means μ_x, μ_y and standard deviation σ located at the center (x, y) of the voting area A and oriented in the voting direction α using single path approach as:

$$V(x, y; r_{\min}, r_{\max}, \Delta) = \sum_{(u,v) \in A} \|\nabla I(x, y)\| \mathcal{N}(x, y, \mu_x, \mu_y, \sigma) \quad (2.31)$$

where $\|\nabla I(x, y)\|$ is magnitude of gradient image and $\mathcal{N}(x, y, \mu_x, \mu_y, \sigma)$ is a 2D shifted Gaussian kernel which is defined as:

$$\mathcal{N}(x, y, \mu_x, \mu_y, \sigma) = \frac{1}{2\pi\sigma^2} \exp\left(-\frac{(x - \mu_x)^2 + (y - \mu_y)^2}{2\sigma^2}\right) \quad (2.32)$$

where $\mu_x = x + \frac{\cos\theta}{2}(r_{\max} + r_{\min})$ and $\mu_y = y - \frac{\sin\theta}{2}(r_{\max} + r_{\min})$. Later, the seed points are determined by executing mean shift on the sum of voting images.

The counting of nuclei by types is highly important for grading purpose. However, manual counting of such nuclei is tedious and subject to considerable inter and intra reader variations. Fuchs and Buhmann [54] reported 42% disagreement between five pathologist on nuclei classification as normal or atypical. They also reported intra-pathologist error of 21.2%. This shows the high potential added value of automatic counting tools. This shows the considerable margin remaining to be fulfilled by a consensual use of digital tools combined with a more focused use of the pathologists skills (difficult cases, sensitive areas), in order to reach better clinical rates.

Mitosis count provides clues to estimate the proliferation and the aggressiveness of the tumor [149]. Anari *et al* [13] proposed fuzzy c -means clustering (FCM) method along with ultra-erosion operation in Lab color model for detection of mitotic nuclei in IHC images of Meningioma and reported detection accuracy nearly equal to manual annotation. The FCM method is based on the minimizing following objective function:

$$J_m(V, C) = \sum_{k=1}^c \sum_{i=1}^U v_{ki}^m \|I(i) - C_k\|^2 \quad (2.33)$$

with $m > 1 (m \in \mathbb{R})$, U is the total number of pixels in I , $C = \{C_1, C_2, \dots, C_c\}$ the cluster centers, and $V = [v_{ki}]$, a $c \times U$ matrix in which v_{ki} is the k^{th} membership value of i^{th} pixel, such that $\sum_{i=1}^U v_{ki} = 1$. The membership function v_{ki} is:

$$v_{ki} = \frac{1}{\sum_{j=1}^c \left(\frac{\|I(i) - C_k\|}{\|I(i) - C_j\|} \right)^{\frac{2}{m-1}}} \quad (2.34)$$

with the cluster center:

$$C_k = \frac{\sum_{i=1}^U v_{ki}^m \cdot I(i)}{\sum_{i=1}^U v_{ki}^m} \quad (2.35)$$

Recently, Roullier *et al* [149] proposed graph based multi-resolution framework for mitotic nuclei detection in breast cancer IHC images. This approach corresponds to unsupervised clustering at low-resolution followed by refinements at a higher resolution. At multi-resolution level, mitotic regions are initially segmented by using the discrete label regularization function following:

$$\min_{f \in \mathcal{H}(V)} \left\{ R(f) + \frac{\lambda}{2} \|f - f^0\|^2 \right\} \quad (2.36)$$

where the first term $R(f)$ is the regularizer and is defined as the discrete Dirichlet form of the function $f \in \mathcal{H}(V) : R_w(f) = \frac{1}{2} \sum_{u \in V} [\sum_{v \sim u} w(u, v)(f(v) - f(u))^2]^{\frac{1}{2}}$ and $\mathcal{H}(V)$ is the Hilbert space of real valued functions defined on the vertices V of a graph. The second term is the fitting term. $\lambda \geq 0$ is a fidelity parameter called the Lagrange multiplier which specifies the trade-off between the two competing terms. The Gauss-Jacobi method is used to approximate the solution of minimization (2.36) by following iterative algorithm:

$$\begin{cases} f^{(0)}(u) &= f^0(u) \\ f^{(t+1)}(u) &= \frac{\lambda f^0(u) + \sum_{v \sim u} w(u, v) f^{(t)}(v)}{\lambda + \sum_{v \sim u} w(u, v)}, \quad \forall u \in V \end{cases} \quad (2.37)$$

where $f^{(t)}$ is function at the iteration step t . More details on these definitions can be found in [149]. This discrete regularization is adapted for labeling the mitotic regions at higher resolution. The authors reported more than 70% TPR and 80% TNR.

The use of EM for GMM was recently proposed by Khan *et al* [87] for the detection of mitotic nuclei in breast cancer histopathological images. In this framework, pixel intensity is modelled as mitotic and non-mitotic region by a Gamma-Gaussian mixture model as:

$$f(I_i; \theta) = \rho_1 \Gamma(I_i; \psi, \xi) + \rho_2 \mathcal{N}(I_i; \mu, \sigma) \quad (2.38)$$

where ρ_1 and ρ_2 represent the mixing proportions (prior) of the intensities belonging to mitotic and non-mitotic regions. $\Gamma(I_i; \psi, \xi)$ represents Gamma density function for mitotic regions; it is parameterized by shape (ψ) and scale (ξ) parameters. $\mathcal{N}(I_i; \mu, \sigma)$ represents Gaussian density function for non-mitotic regions; it is parameterized by μ and σ .

Ciresan *et al* [35] used deep max-pooling convolutional neural networks (CNN) to detect mitotic nuclei and achieved highest F-Score (78%) during ICPR 2012 contest [150]. The CNN is used to compute a map of probabilities of mitosis over the whole image. Using ground truth (GT) mitosis in training dataset, CNN is trained to classify each pixel in the images, using as context a patch centered on the pixel. Their approach proved to be very efficient specifically having few false positives (FP) as compared to other contestants.

Grading of lymphocytic infiltration based on detection of large number lymphocyte nuclei in HER2+ breast cancer histopathology was reported by Basavanhally *et al* [17]. In this framework, lymphocyte nuclei are automatically detected by a region growing method, which uses contrast measures to find optimal boundary. This framework has reported high detection sensitivity, resulting in a large number of other nuclei being detected. In order to reduce the number of FP, size and luminance information based maximum a posteriori (MAP) estimation is applied to temporarily label candidates as either lymphocyte or cancer nuclei. Later, Markov random field (MRF) theory with spatial proximity is incorporated in order to finalize the labels. This framework is evaluated on 41 HER2+ WSI and reported 90.41% detection accuracy as compared to 94.59% manual detection accuracy.

2.4.2 Segmentation Methods

Nuclei features like size, texture, shape and other morphological appearance are important indicators for grading and prognosis of cancer. Consequently, classification and grading of cancer is highly depending on segmentation quality of nuclei. Depending on the type of the feature extraction method to be deployed, this may include determining the exact boundary points of nuclei [164] or determining their coarse locations [65]. In the former case, segmentation requires higher magnification images to resolve the exact details of nuclei and the success of the next steps becomes more sensitive to the success of the segmentation. A large number of publications on nuclei segmentation in histopathology use state-of-the-art image segmentation methods based on thresholding, morphology, region growing, watershed, ACMs, clustering and Gcuts separately or in combination.

The simplest way to detect and segment nuclei in histopathological images is based on thresholding and morphological operations, a simple methodology to segment nuclei [133, 67, 8]. This methodology reports higher performance on well-defined, preferably uniform background. The main parameters to tune this methodology are the size and shape of the structuring elements and threshold level. The difference between nuclei and background regions may be diffuse, making it harder to find a reliable threshold level. Even though this methodology is usually defined only on grayscale images, it can be extended to color images or stacks of images, using multi-dimensional kernels. This methodology actually suffers from its simplicity by including little object knowledge. In addition, this methodology lacks robustness on size and shape variations, as well as texture variations, which are very frequent in histopathological images. This methodology is not meant to segment clustered or overlapping nuclei.

Several frameworks have been using watershed transform for nuclei segmentation [177, 36, 79]. The main advantage of watershed is that there is no tuning to do before using it. However, it requires the prior detection of seed points. Edge map and distance transform are used for seeds detection [177, 79]. The reported results are suboptimal for ring-shaped nuclei having clear homogeneous regions. Furthermore, watershed transform does not include any prior knowledge, which may contribute to robustness.

ACMs can combine both shape characteristics (smoothness and shape model) with image features (image gradient and intensity distribution). However, the resulting segmentation is strongly dependent upon the initial hypothesis, a problem usually solved by the accurate detection of seed point. Cosatto *et al* [37] described an automated method for accurately and robustly measuring the size of neoplastic nuclei and providing an objective basis for pleomorphism grading. Initially, a Difference of Gaussian (DoG) filter is used to detect nuclei and Hough transform is used to pick up radially symmetric shapes. Finally an ACM with shape, texture and fitness parameters are used to extract nuclei boundaries. The authors claimed 90% TPR.

Huang and Lai [74] proposed watershed and ACM based framework for nuclei segmentation in hepatocellular carcinoma biopsy images. Initially, a dual morphological grayscale reconstruction method is employed to remove noise and accentuate the shapes of nuclei. Then, a marker-controlled watershed transform is performed to find the edges of nuclei. Finally, ACM is applied to generate smooth and accurate contours for nuclei. This framework achieves poor segmentation in case of low contrast, noisy background and damaged/irregular nuclei.

Dalle *et al* proposed gradient in polar space (GiPS), a novel nuclei segmentation method [39]. Initially, nuclei are detected using thresholding and morphological operations. Then, transformation into polar coordinate system is performed for every patch with the center of mass of the nucleus as the origin. Finally, a biquadratic filtering is used to produce a gradient image from which nuclei boundaries are delineated. GiPS reports overall 7.84% accuracy error.

Ta *et al* [166] proposed a method based on graph based regularization. A strong specificity of this framework is to use graphs as a discrete modeling of images at different levels (pixels or regions) and different component relationships (grid graph, proximity graph, etc.). Based on Voronoi diagram, a novel image partition (graph reduction) algorithm is proposed for segmentation of nuclei in serous cytological and breast cancer histopathological images. A pseudo-metric $\delta : V \times V \rightarrow \mathbb{R}$ is defined as:

$$\delta(u, v) = \min_{\rho \in P_G(u, v)} \sum_{i=1}^{m-1} (f(u_{i+1}) - f(u_i)) \sqrt{w(u_i, u_{i+1})} \quad (2.39)$$

where $w(u_i, u_{i+1})$ is a weight function between two pixels and $P_G(u, v)$ is a set of paths connecting two vertices. Given a set of K seeds $S = (s_i \subseteq V)$, where $i = 1, 2, \dots, K$, the energy $\delta : V \rightarrow \mathbb{R}$ induced by the metric δ for all the seeds of S can be expressed as:

$$\delta_S(u) = \min_{s_i \in S} \delta(s_i, u) \quad \forall u \in V \quad (2.40)$$

The influence zone z (also called Voronoi cell) of a given seed $s_i \in S$ is the set of vertices which are closer to s_i than to any other seeds with respect to the metric δ . It can be defined, $\forall j = 1, 2, \dots, K$ and $j \neq i$, as

$$z(s_i) = \{u \in V : \delta(s_i, u) \leq \delta(s_j, u)\} \quad (2.41)$$

Then, the energy partition of graph, for a given set of seeds S and a metric δ , is the set of influence zones $Z(S, \delta) = \{Z(s_i), \forall s_i \in S\}$.

Kofahi *et al* [9] proposed another Gcuts based method for segmentation of breast cancer nuclei. Initially, the foreground is extracted using Gcut based binarization. The pixel labeling $I'(i)$ is done by minimizing the following energy function:

$$\begin{aligned} \mathbb{E}(I'(i)) = & -\ln \mathcal{P}(I(i)) + \sum_i \sum_{j \in N(i)} \eta(I'(i), I'(j)) \\ & \times \exp\left(-\frac{I(i) - I(j)}{2\sigma_{I'}^2}\right) \end{aligned} \quad (2.42)$$

where $\mathcal{P}(I(i)|k)$, $k = 0, 1$ are Poisson distribution, $N(i)$ is a spatial neighborhood of pixel i and

$$\eta(I'(i), I'(j)) = \begin{cases} 1, & \text{if } I'(i) \neq I'(j) \\ 0, & \text{otherwise} \end{cases} \quad (2.43)$$

In (2.42), the first term is a data term that represented the cost of assigning a label to a pixel and the second term is a pixel continuity term that penalizes different labels for neighboring pixels when $|I(i) - I(j)| < \sigma_{I'}$. After binarization, nuclear seed points are detected by combining multi-scale LoG filtering constrained by a distance map based adaptive scale selection (2.30). These detected seed points are used to perform initial segmentation and later, refined using a second Gcuts based method with combination of alpha expansion and graph coloring to reduce computational complexity. The author reported 86% accuracy on 25 histopathological images containing 7400 nuclei. The framework often causes over-segmentation when chromatin is highly textured and the shape of nuclei is extremely elongated. In case of highly clustered nuclei with weak borders between nuclei, under-segmentation may occur.

For nuclei segmentation in Glioblastoma histopathology images, Chang *et al* [31] proposed a multi-reference Gcuts (MRGC) framework for solving the problem of technical and biological variations by incorporating geodesic constraints. During labeling, a unique label

$L(i)$ is assigned to each vertex $v \in V$ and the image cutout is performed by minimizing the energy:

$$\mathbb{E} = \sum_{v \in V} (\mathbb{E}_{gf}L(v) + \mathbb{E}_{lf}L(v)) + \sum_{(v,u) \in E} \mathbb{E}_{\text{smoothness}}(L(v), L(u)) \quad (2.44)$$

where \mathbb{E}_{gf} and \mathbb{E}_{lf} are the global and local data fitness term applying the fitness cost for assigning $L(v)$ to v , and $\mathbb{E}_{\text{smoothness}}(L(v), L(u))$ is the prior energy, denoting the cost when the labels of adjacent vertices, v and u are $L(v)$ and $L(u)$, respectively.

Recently, Nguyen et al. [128] proposed maximum object likelihood binarization (MOLB) algorithm for nuclei segmentation. First transformation of RGB image into Lab color model is performed and select 'b' channel for nuclei segmentation as it best represent the nuclei. By maximizing the average object likelihood of the nuclei, a binarized image is obtained using following threshold function:

$$t_O = \operatorname{argmax}_i \frac{1}{n^t} \sum_{i=1}^{n^t} g(f(B_i^t) | \hat{\Theta}) \quad (2.45)$$

where $f(B_i^t)$ is feature vector of blob B_i^t and $g(f(B_i^t))$ is the object likelihood of blob B_i^t since it estimated how similar the features of B_i^t are to the features of the object of interest O having density g and parameter $\hat{\Theta}$.

Vink *et al* introduced a deterministic approach using machine-learning, to segment epithelial, lymphocyte and fibroblast nuclei in IHC breast cancer images [176]. Initially, the authors report that one detector cannot cover the whole range of nuclei as diversity in appearance is too large to be covered by a single detector. They formulate two detectors (pixel based and line based) using modified AdaBoost. The first detector focuses on the inner structure of nuclei and second detector covers the line structure at the border of nuclei. The outputs of these two detectors are merged using an ACM to refine the border of the detected nuclei. The authors report 95% accuracy with computational cost of one second per field of view image.

These nuclei segmentation frameworks have reported good segmentation accuracy on lymphocyte, mitotic and epithelial nuclei having regular shape, homogeneous chromatin distribution, smooth boundaries and individual existence. However these frameworks have poor segmentation accuracy for cancer nuclei especially when cancer nuclei are clustered and overlapping. Furthermore, they are intolerant to chromatin variations, which are very common in cancer nuclei.

2.4.3 Separation Methods

A second generation of nuclei segmentation frameworks tackles the challenges of heterogeneity, overlapping and clustered nuclei by using machine-learning algorithms together with classical segmentation methods. In addition, statistical and shape models are also used to separate overlapping and clustered nuclei. As compared with nuclei segmentation methods, these methods are more tolerant to biological variation, partial occlusion and different staining.

Watershed transform is employed to address the problem of overlapping nuclei as a group of basins in the image domain, where ridges in-between basins are borders that isolate nuclei from each other [67, 79, 27, 45, 90]. Wahlby et al. [177] addressed the problem of clustered nuclei and proposed a methodology that combined the intensity and gradient information along with shape parameters for improved segmentation. Morphological filtering is used for finding nuclei seeds. Then, seeded watershed segmentation is applied on the gradient magnitude image to create the region borders. Later, the result of the initial segmentation is refined with gradient magnitude along the boundary separating neighboring objects, resulting in the removal of poorly contrasted objects. In final step, distance transform and

shape based cluster separation methodologies are applied keeping only the separation lines, which went through deep valleys in the distance map. The authors reported 90% accuracy for overlap nuclei. Cloppet and Boucher [36] presented a scheme for segmentation of overlapping nuclei in immunofluorescence images by providing a specific set of markers to the watershed algorithm. They defined markers as split between overlapping structures and resulted in 77.59% accuracy in case of overlapping nuclei and 95.83% overall accuracy. In [161], a similar approach is used for segmentation of clustered and overlapping nuclei in tissue micro array (TMA) and WSI colorectal cancers. First, combined global and local thresholding is used to select foreground regions then applied morphological filtering for seeds detection. Region growing is applied on detect seeds which produces initial segmented nuclei. Later, clustered nuclei are separated using watershed and ellipse approximation. The authors claimed 80.3% accuracy.

The main problem with most ACMs is their sensitivity to initialization. To solve this initialization problem, Fatakawala *et al* [51] proposed EM driven Geodesic ACM with overlap resolution (EMaGACOR) for segmentation lymphocyte nuclei in breast cancer histopathology and reported 86% TPR and 64% PPV. EM based ACM initialization allows the model to focus on relevant objects of interest. The magneto-static active contour (MAC) [184] model is used as a guiding force F for contour towards boundary. Based on contours enclosing multiple objects, high concavity points are detected on the contours and used in construction of edge-path graph. Then high concavity points and size heuristic based scheme is used to resolve overlapping nuclei. The degree of concavity/convexity is proportional to the angle $\theta(c_w)$ between contour points and computed as:

$$\theta(c_w) = \pi - \arccos \left(\frac{(c_w - c_{w-1}) \cdot (c_{w+1} - c_w)}{|c_w - c_{w-1}| |c_{w+1} - c_w|} \right) \quad (2.46)$$

where c_w is a point on the contour.

Yang *et al* [186] proposed a nuclei separation methodology in which concave vertex graph and Ncut algorithm are used. Initially, the outer boundary is delineated via robust estimation and color active model, and a concave vertex graph is constructed from automatically detected concave points on boundaries (2.46) and inner edges. By minimizing a morphological based cost function, the optimal path in graph is recursively calculated to separate the touching nuclei.

Mouelhi *et al* proposed an automatic separation method for clustered nuclei in breast cancer histopathology [124]. First, a modified GAC with Chan-Vese energy model is used to detect the nuclei region [29]. Second, high concavity points along touching nuclei regions are detected (2.46). Third, the inner edges are extracted by applying watershed transform on a hybrid distance transform image, which combines geometric distance and color gradient information. Four, concave vertex graph using high concavity points and inner edges is constructed. Last, the optimal separating curve is selected by computing the shortest path in the graph.

Moreover, for the recognition of single nuclei in nuclei cluster, Kong *et al* [90] integrated a framework consisting of a novel supervised nuclei segmentation and touching nuclei splitting method. For initial segmentation of nuclei, each pixel is classified into nuclei or background regions by utilizing color-texture in the most discriminant color model. The differentiation between clustered and separated nuclei is computed by the distance between the radial symmetry center and the geometrical center of the connected component. For splitting of clustered nuclei, the boundaries of touching clumps are smoothed out by Fourier shape descriptor and then carried out concave point detection. The author evaluated this framework on FL images and achieved average 77% TPR and 5.55% splitting ER.

Another adaptive ACM scheme that combines shape, boundary, region homogeneity and mutual occlusion terms in a multi-level set formulation was proposed by Ali *et al*

[12, 11]. The segmentation of K overlapping nuclei with respect to shape prior ψ is solved by minimizing following level set ϕ function:

$$\begin{aligned} \mathbb{E}(\Phi, \Psi, I_F, I_B) = & \underbrace{\beta_s \sum_{k=1}^{K=2} \int_{\varpi} (\phi_k(I) - \psi(I))^2 |\nabla \phi_k| \delta(\phi_k) dI}_{\text{Shape + boundary energy}} \\ & + \underbrace{\beta_r \int_{\varpi} (\Theta_F H_{\chi_1 \vee \chi_2}) dI + \int_{\varpi} (\Theta_B - H_{\chi_1 \vee \chi_2}) dI}_{\text{Region energy}} \\ & + \underbrace{\omega \int_{\varpi} H_{\chi_1 \wedge \chi_2} dI + \sum_{k=1}^{K=2} \int_{\varpi} (\phi_k - \psi_k)^2 dI}_{\text{Mutual occlusion energy}} \end{aligned} \quad (2.47)$$

where $\Phi = (\phi_1, \phi_2)$, $\Psi = (\psi_1, \psi_2)$, I_F and I_B are foreground and background regions, $\beta_s, \beta_r, \omega > 0$ are constants that balance contributions of the shape and boundary, region and mutual occlusion term, respectively, $\delta(\cdot)$ is the Dirac delta function, and $\delta(\phi_k)$ is the contour measure on $\{\phi = 0\}$, $H(\cdot)$ is the Heaviside function, $H_{\chi_1 \vee \chi_2} = (H_{\psi_1} + H_{\psi_2} - H_{\psi_1} H_{\psi_2})$, $H_{\chi_1 \wedge \chi_2} = H_{\psi_1} H_{\psi_2}$, and $\Theta_j = |I - I_j|^2 + \mu |\nabla I_j|^2$ and $j \in \{F, B\}$. Watershed transform is used for model initialization. The authors evaluated this framework on overlapping nuclei in prostate and breast cancer images and reported 86% TPR and 91% OR on breast images and 87% TPR and 90% OR on prostate images.

Qi *et al* [140] proposed a two-step method for the segmentation of overlapping nuclei in Hematoxylin stained breast TMA specimens that require very little prior knowledge. First, seed points are computed by executing mean shift on the sum of the voting images (2.31). Second, the following level set representation of the contours is used:

$$\begin{aligned} \mathbb{E} = & \alpha_N \sum_{k=1}^K \int_{\Lambda_k} |I - \mu_k|^2 di + \alpha_B \sum_{k=1}^K \int_{\Lambda_B} |I - \mu_b|^2 di \\ & + \beta \sum_{k=1}^K \int_0^1 g(|\nabla I(\varpi_k(z))|) |\varpi'_k(z)| dz \\ & + \lambda \sum_{k=1}^K \sum_{j=1, j \neq k}^K \Lambda_k \cap \Lambda_j \end{aligned} \quad (2.48)$$

where $\alpha_N, \alpha_B, \beta > 0$ are constants that balance contributions of each term, $\varpi_k (k = 1, \dots, K)$ is the nuclei contours that evolve towards boundaries, K is number of nuclei, Λ_k is region inside each contour ϖ_k , Λ_B is the background which represents the regions outside all the nuclei, μ_k and μ_b are mean intensities of nuclei and background regions, and g is a sigmoid function $g(x) = (1 + e^{\frac{x-\nu}{\zeta}})^{-1}$ with ν controls the slope of the output curve and ζ controls the window size. The last term in (2.48) is the repulsion term used to represent the repulsion energy between each touching nuclei and the λ is a regulation parameter. This repulsion term separates the touching nuclei to create smooth and complete contour of each nucleus. The authors claimed 78% TPR and 90% PPV in case of touching nuclei.

To overcome ACMs initialization sensitivity, Kulikova *et al* [95] proposed a method based on marked point processes (MPP). This methodology, a type of high order ACM, is able to segment overlapping nuclei as several individual objects, while no need to be initialized with the location of the nuclei to be detected. A shape prior term is used for handling overlapping nuclei. Fig. 2.1 shows a comparison of results using MPP, GiPS [39] and Levelset [123].

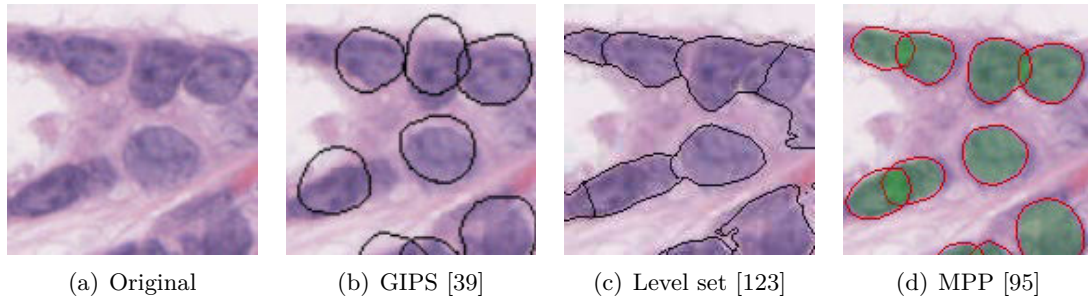


Figure 2.1: Results of segmentation and separation using different methods on same area of an image.

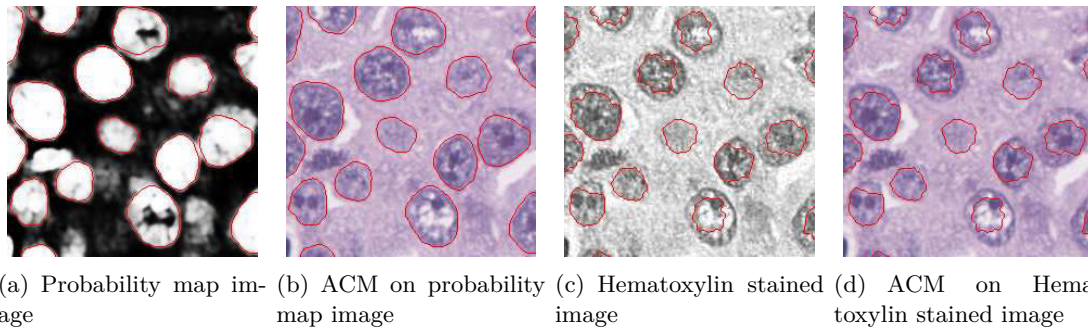


Figure 2.2: Segmentation results using ACM methods on probability and Hematoxylin stained image [173].

Recently, Veillard *et al* [173] proposed a method based on the creation of a new image modality consisting in a grayscale map where the value of each pixel indicated its probability to belong to a nuclei. This probability map is calculated from texture, scale information and simple pixel color intensities. The resulting modality has a strong object-background contrast and even out the irregularities within the nuclei and background. Later, segmentation is performed using an ACM with a nuclei shape prior [95] which resolve overlapping nuclei problem. Fig. 2.2 shows the result of ACM segmentation on probability map image and Hematoxylin stained image, produced using color deconvolution [153].

In general, model based approaches segment nuclei using a priori shape information, which may introduce a bias favoring the segmentation of nuclei with certain characteristics. To address this problem, Wienert *et al* [179] proposed a novel contour based minimum model for nuclei segmentation using minimal a priori information. This minimum model based segmentation framework consists of six internal processing steps. First, all possible closed contours are computed regardless of shape and size. Second, all initially generated contours are ranked using gradient fit. This gradient fit is computed using Sobel operator with its 3×3 convolution kernels K_x and K_y as:

$$\text{GradientFit}_i = \frac{\sum_j \varpi_i(j)^{\max}}{|\varpi_i|},$$

$$\varpi_i(j)^{\max} = \begin{cases} 1, & \text{if } \max\{|S(\varpi_u(v))|\} = |S(\varpi_i(j))| \\ 0, & \text{otherwise} \end{cases} \quad (2.49)$$

$$\forall u, v (x_i - 1 \leq u \leq x_i + 1) \cap (y_i - 1 \leq v \leq y_i + 1)$$

$$|S| = \sqrt{(I \times K_x)^2 + (I \times K_y)^2}$$

where ϖ_i is i th contour and $\varpi_i(j)$ its j th contour pixel. The mean gradient and the gradient fit are combined for evaluation of contour.

$$\text{Contour Value}_i = \text{MeanGradient}_i \cdot \text{GradientFit}_i \quad (2.50)$$

Third, non-overlapping segmentation is performed with ranked labeling in two dimensional map. Four, segmentation is improved using contour optimization. Five, cluster nuclei are separated using concavity point detection (2.46). Last, segmented regions are classified as nuclei or background using stained related information. This framework avoids a segmentation bias with respect to shape features. The authors managed to achieve 86% TPR and 91% PPV on a dataset of 7931 nuclei.

RST is an iterative algorithm attributed votes to pixels inside the region [107]. Maxima after the final iteration are used as marker of a nuclei segmentation algorithm such as watershed. Each boundary point contributes votes to a region defined by oriented cone-shape kernels as:

$$A(x, y ; r_{\min}, r_{\max}, \Delta) = \left\{ (x + r \cos \phi, y + r \sin \phi) \mid r_{\min} \leq r \leq r_{\max}, \theta(x, y) - \frac{\Delta}{2} \leq \phi \leq \theta(x, y) + \frac{\Delta}{2} \right\} \quad (2.51)$$

where the radial range is parametrized by r_{\min} and r_{\max} and the angular range by Δ . $\theta(x, y)$ is the angle between the positive x-axis and the voting direction. These parameters are updated using votes from the previous iterations.

Schmitt and Hasse [155] separated the clustered nuclei using RST based on the idea that center of mass in a nuclei is considered as a basic perceptual event that supports separation of clustered nuclei. They initialized iterative voting along the gradient direction where, at each iteration, the voting direction and shape of the kernel are refined iteratively. The voting area can be regulated by selecting the number of steps in the evolution of the kernel shape. Few number of steps resulted in fragmentation of the center of mass and large number of steps increased computational cost. They also proposed a way to deal with holes and sub holes in the region by processing boundaries iteratively. The major steps of RST algorithm are listed as:

- i. Initialization: $r_{\min}, r_{\max}, K, \Delta_k$ and \mathcal{B} are initialized where K is total number of iterations, Δ_k is angular range at k^{th} iteration such that $(\Delta_{\max} = \Delta_0 > \Delta_1 > \dots > \Delta_K), \Delta_k = \Delta_{\max}$ and $k = K$, and \mathcal{B} is set of all boundary points from external boundary of regions and holes inside regions in the binary image. The kernel radial range is initialized for each boundary point as $r_{\min} = 1.66 \times d$ and $r_{\max} = 0.33 \times d$ where d is the distance between the boundary point and the local maximum in the distance transformed image. The kernel direction θ is also initialized towards local maximum point d . The voting direction is initialized along gradient direction.
- ii. Determine the votes: Reset the vote image $V(x, y ; r_{\min}, r_{\max}, \Delta_k) = 0$ for all pixels (x, y) . For all points $(p, q) \in \mathcal{B}$ and $(u, v) \in A(p, q ; r_{\min}, r_{\max}, \Delta_k)$ update the vote image by:

$$V(u, v ; r_{\min}, r_{\max}, \Delta_k) \leftarrow V(u, v ; r_{\min}, r_{\max}, \Delta_k) + \varepsilon \quad (2.52)$$

where ε is the voting magnitude for each pixel of the image.

- iii. Update the voting direction for each boundary point $(i, j) \in \mathcal{B}$ along the maximum value in the voting area.
- iv. Update r_{\min} and r_{\max} .
- v. Refine the angular range $\Delta_k, k = k - 1$ and repeat steps ii – iv until $k = 0$.

- vi. To avoid over-segmentation the voting landscape arisen during the last iteration step is smoothed by median filtering.
- vii. In the last, local maxima are determined and used as marker for a marker based watershed transformation.

One limitation of RST is the prior knowledge of scale, which cannot be generalized. To overcome this limitation multi-scale extension of the RST seem to be reasonable. A similar method [155] is used in [68] to decompose regions of clustered nuclei in H&E stained prostate cancer biopsy images. They initially obtained regions of clustered nuclei by clustering and level-set segmentation. Recently, Veta et al [175] proposed a method similar to [74] that met the objective of nuclei segmentation in H&E stained breast cancer biopsy images by applying the fast RST [107] to produce markers for the watershed segmentation. Sertel *et al* [158] proposed adaptive likelihood-based nuclei segmentation for FL centroblasts. Initially, cellular components are clustered using GMM with EM. Using fast RST, spatial voting matrix is computed along the gradient direction. Finally, local maxima locations associated with individual nuclei are determined.

Alternatively, EM and GMM based unsupervised Bayesian classification scheme was used for segmentation of overlapping nuclei in IHC images [80]. The separation of overlapping nuclei is formulated as cluster analysis problem. This approach primarily involves applying the distance transform to generate topographic surface, which is viewed as a mixture of Gaussian. Then, a parametric EM algorithm is employed to learn the distribution of topographic surface (GMM). On the base of extracted regional maxima, cluster validation is performed to evaluate the optimal number of nuclei. The cluster validity index consists of a compactness measure φ (the smaller value means more compact) and a separation measure ε between the clusters. The main idea is to have nuclei as compact and as well separated as possible. Thus, cluster parameters are chosen to maximize $\frac{\varepsilon}{\varphi}$. A priori knowledge for the overlapping nuclei is incorporated to obtain separation line without jaggedness, as well as to reconstruct occluded contours in overlapping region. They achieved improvements of up to 6.80%, 5.70% and 3.43% with respect to classical watershed, conditional erosion and adaptive H-minima transform schemes in terms of separation accuracy. Overall, they achieved 93.48% segmentation accuracy for overlapping nuclei on specimens of cervical nuclei and breast invasive ductal carcinomas.

The novelty of these approaches are to use machine-learning and statistical methods to eliminate malformed nuclear outlines and thus to allow robust nuclei segmentation. The ability to manually train these models is constrained by the availability of expert annotations of the objects of interest. Datasets for training are difficult to define due to variability across images. Furthermore, such models may not be generalizable and have limited application due to the manual training step, sensitive to initialization, and potentially limited ability to segment multiple overlapping objects.

2.4.4 Nuclei Features and Classification Methods

Features computed from segmented nuclei are usually a prerequisite to nuclei classification that generate higher-level information regarding the state of the disease. The classifiers use nuclei features, which capture the deviations in the nuclei structures, to learn how to classify nuclei into different classes. In order to extract features, there are two different types of information available in the image: (i) the intensity values of pixels and (ii) their spatial interdependency [42]. Although all feature computation methods use the information on the intensity values, only a few use the spatial dependency between them. The use of intensity values only results in higher sensitivity to the noise that arises from the stain artefacts and the image acquisition conditions.

Category	Features
Cytology	Nucleoli
Intensity	Density, Hue, Hyperchromatism, mean, median, variance, skewness, kurtosis, etc
Morphology	Area, Area overlap ratio, Center of mass, Compactness, Concavity (Convexity), Density, Diameter, Inflection points, Minor axis, Major axis, Nucleocytoplasmic ratio, Peakiness, Perimeter, Radial ratio, Roundness, Shape Inertia, Smoothness, Symmetry
Texture	Co-occurrence, Fractal, Gabor, Markov random field, Run-Length, SIFT, Wavelets, Haar like features etc

Table 2.2: Summary of Nuclei Features used in Histopathology

We found a compilation of features for cytopathology imagery [145], but found relatively little such work for histopathology imagery. In histopathology, these features can be categorized into the following four categories: cytological, intensity, morphological and texture features. A summary of nuclei features is listed in Table 2.2; definition for all listed features can be found in [20, 113, 42].

In some frameworks, the computed features, like intensity and texture features, are explicitly used for segmentation of nuclei with K-means clustering [158, 51]. To address the problem of heterogeneity in cancer nuclei, Veillard *et al* [173] used intensity and textural features with support vector machine (SVM) classifier for the creation of a new image modality to segment cancer nuclei. Recently, Vink *et al* [176] constructed a large features set and modified AdaBoost to create two detectors that solved the problem of variations in nuclei segmentation. The first detector is formulated with intensity features, the second detector is constructed using Haar like features.

In addition to the morphological features computed from cytological regions, Huang *et al* [74] extracted intensity and Haralick co-occurrence (HC) features. They extracted a total of 14 features (intensity, morphological and texture features) from segmented nuclei in biopsy images, which comprise both local and global characteristics so that benignancy and different degrees of malignancy can be distinguished effectively. A SVM-based decision graph classifier with feature subset selection on each decision node of classifier is used in comparison with k-nearest neighbor and simple SVM, the accuracy rate of classification promoted from 92.88% to 94.54% with SVM-based decision graph classifier.

Intensity and morphological features are extensively used for nuclei classification as epithelial and cancer nuclei in [37, 39, 45]. An exhaustive set of features including morphological and texture features are explored to determine the optimal features for nuclei classification [104]. Their results of feature selection demonstrated that Zernike moment, Daubechies wavelets and Gabor wavelets are the most important features for nuclei classification in microscopy images. Malon and Cosatto [113] computed intensity, texture and morphological features and used these features with SVM for classification of segmented candidate regions into mitotic and non-mitotic regions.

Al-Kadi [8] presented that combining several texture measures instead of using just one might improve the overall accuracy. Different texture measures tend to extract different features each capturing alternative characteristics of the examined structure. They computed four different texture features, two of them are model-based: Gaussian Markov Random Field (GMRF) and Fractal Dimension (FD); the other two are statistically based: HC and RL features. Using selected features after excluding highly correlated features, Bayesian classifier was trained for meningioma subtype classification. They studied the variation of texture measure as the number of nuclei increased; the GMRF was nearly uniform, while

the RL and FD performed better in the high frequencies. They also studied the texture measures' response to additive texture distortion noise while varying cell nuclei shape densities. The GMRF was the least affected, yet the RL and FD performed better in high and low shape frequency, respectively. The combination of GMRF and RL improved the overall accuracy up to 92.50% with none of the classified meningioma subtypes achieving below 90%.

By observing the cancer detection procedure adopted by pathologists, Nguyen *et al* [128] developed a novel idea for cancer detection in prostate using cytological (nuclear) textural features. Prominent nucleoli (cytological feature) inside nuclei region is used to classify nuclei as cancerous or not. In addition, prostate cancer is detected using cytological, intensity, morphological, and textural features having 78% TPR on a dataset including six training and 11 test WSI.

2.5 Spectral and Spatial Characterization

MSI is a recent medical imaging technology, proven successful in increasing the segmentation and classification accuracy in histopathology [21, 183]. We found few methods in the MSI literature for spatial characterization of histopathological images. The main idea for extracting features from MSI is the use of combined spectral and spatial information for discrimination of regions or objects. Fernandez *et al* [52] coupled high-throughput Fourier transform infra-red spectroscopic imaging of tissue microarrays with statistical pattern recognition of spectra indicative of endogenous molecular composition and demonstrate histopathological characterization of prostatic tissue. They explicitly defined metrics consisting of spectral features that have a physical significance related to tissue biochemistry and facilitating the measurement of cell types.

We found few methods in the MSI literature for spatial characterization of histopathological images. Some of them employed single spectral band (SB) of MSI [118, 113] and other used multiple SBs of MSI [89, 183, 21]. Some methods computed one type of features on single SB for quantitative analysis. Masood and Rajpoot [118] proposed a colon biopsy classification method based on spatial analysis of hyperspectral images. First, SB 588 nm was selected, as it is the one that seemed to contain more textural information. Then, using circular local binary pattern algorithm, spatial analysis of patterns was represented by a feature vector in the selected SB. Later, classification was achieved using subspace projection methods like principal component analysis, linear component analysis and support vector machine.

Some methods computed different types of features on single SB for quantitative analysis. Malon and Cosatto [113] demonstrated a segmentation based features with CNN using the selected SB for identification of mitotic figures and achieved the best F-Measure (59%) on multispectral dataset during ICPR contest 2012 [150]. First, focal plane number five was selected as it was clearly focused. Second, two SBs were selected using PCA to extract the top two eigenvectors from a set of 10 SBs of hematoxylin and eosin images. Third, two step thresholding was applied on first eigenvectors (hematoxylin image) to obtain candidate blobs. Fourth, a set of shape, contour, pixel and texture features was computed on the selected SB only. Fifth, log likelihoods of class membership were computed using convolutional neural network classifier for each patch of candidate blob. In the last, the SVM classifier was used to classify each blob as either mitotic or non-mitotic blob using output of convolutional neural network along with feature vector. The previous approaches [118, 113] are limited to single SB. They discard additional potentially relevant information from other SBs.

Instead of limiting themselves to a single SB, some authors use multiple and sometime even all SBs, from a given dataset. Boucheron *et al* [21] presented a study in which the ad-

ditional SBs have additional useful information for nuclear classification in histopathology as compared to the three standard bands of RGB imagery. Using all SBs, they reported a 0.79% improvement in performance compared to the next best performing image type. Similarly, Wu et al. [183] proposed a multilayer conditional random field model using a combination of low-level cues and high-level contextual information for nuclei separation in high dimensional data set obtained through spectral microscopy. In this approach, the multilayer contextual information is extracted to interpret spectral data with dynamically imposed pairwise constraints along the neighboring spectral bands. It is an unsupervised process, which efficiently helps to suppress segmentation errors caused by intensity inhomogeneity and variable chromatin texture. Khelifi *et al* [89] proposed a spatial and spectral gray level dependence method in order to extend the concept of gray level co-occurrence matrix by assuming the presence of texture joint information between SBs. Some SBs have more relevant information for specific object or region classification than others. This approach is limited to a single spatial feature computed from all SBs. These approaches [21, 183, 89] used all available SBs but were limited to one type of features only. One possible improvement in object classification is multispectral spatial analysis using more types of features. Another possibility of additional improvement is the selection of SBs by minimization of the redundancy and maximization of the relevancy.

2.6 Performance Metrics

In order to compare the performance of a new approach to existing approaches, performance metrics are typically used that provides a ranking of the candidate algorithms (usually using numeric scores). Many performance metrics have been used to rank new algorithms, some measure similar features, but other measure drastically different quantities. For example, methods such as the root of the mean square error (RMSE) measure the distance between predicted preferences and true preferences over items, while the recall metric computes the portion of favored items that are suggested. Clearly, it is unlikely that a single approach would outperform all others over all possible methods. Therefore, we should expect different metrics to provide different rankings of approaches. Performance metrics are categorized into three classes; detection, segmentation and classification metrics.

2.6.1 Detection Metrics

The metrics used to evaluate nuclei detection include: false negative (FN), false positive (FP), true negative (TN), true positive (TP), precision or positive predictive value (PPV), specificity or true negative rate (TNR), recall or sensitivity or true positive rate (TPR), F-score or F-measure (FM) and error rate (ER). They are defined as: FN is the number of ground truth nuclei that have not been detected, FP is the number of detected nuclei that are not ground truth nuclei, TN is number of nuclei that are neither in ground truth nor in detected nuclei, TP is the number of detected nuclei that are ground truth nuclei, $\left(\text{TPR} = \frac{\text{TP}}{\text{TP}+\text{FN}}\right)$, $\left(\text{PPV} = \frac{\text{TP}}{\text{TP}+\text{FP}}\right)$ and $\left(\text{FM} = 2 \times \frac{\text{TPR} \times \text{PPV}}{\text{TPR} + \text{PPV}}\right)$.

2.6.2 Segmentation Metrics

Segmentation results are compared to manual segmentation performed by an pathologists (which serves as ground truth for segmentation evaluation) by computing boundary based metrics, namely TP (proportion of nuclei pixels that are correctly labeled as positive), TN (proportion of non-nuclei pixels that are correctly labeled as negative), centroid distance (CD, defined as distance between centroids of a corresponding pair of nuclear boundaries detected by algorithm and human), area overlap metrics like false-positive area ratios (FPAR,

defined as area detected by algorithm but not by human over human markup area), and false-negative area ratios (FNAR, defined as area detected by human but not by algorithm over human markup area), error-rate (ER, defined as the sum of FPAR and FNAR), overlap or Jaccard coefficient or Mean intersection to union ratio (MI2UR), TPR, PPV, TNR, FM, performance, overlap detection ratios (OR), overlap ratio (OV), pixel error (PE) and segmentation distortion evaluation (SDE). These measures are defined as follows:

$$\text{TPR} = \frac{|I_S \cap I_G|}{|I_S|} \quad (2.53)$$

$$\text{PPV} = \frac{|I_S \cap I_G|}{|I_G|} \quad (2.54)$$

$$\text{TNR} = \frac{U - |I_S \cup I_G|}{U - |I_G|} \quad (2.55)$$

$$\text{Performance} = \frac{TP + TN}{2} \quad (2.56)$$

$$\text{MI2UR} = \frac{|I_S \cap I_G|}{|I_S \cup I_G|} \quad (2.57)$$

$$\text{OR} = \frac{\text{Number of overlaps resolved}}{\text{Total number of overlaps}} \quad (2.58)$$

$$\text{OV} = \frac{TP}{TP + FP + FN} \quad (2.59)$$

where U is the total number of pixels, I_S is the segmented region, I_G is the GT region, $|I_S|$ is the number of pixels of segmented nuclei and $|I_G|$ is the number of pixels of GT nuclei.

The Hausdroff distance (HD) [76] and mean absolute distance (MAD), similarity measures, are used to compare the fidelity of automated segmentation region I_S against the GT region I_G . For each segmented region, HD and MAD are calculated as:

$$\text{HD} = \max_{v \in I_G} \min_{u \in I_S} d(u, v) \quad (2.60)$$

$$\text{MAD} = \frac{\sum_{v \in I_G} \min_{u \in I_S} d(u, v)}{|I_G|} \quad (2.61)$$

2.6.3 Classification Metrics

The metrics used for classification are TP, FP, TNR, TPR, PPV, PRC, accuracy (Acc), kappa, F-Score, correct classification rate (CCR), receiver operating characteristics (ROC) and precision recall curve (PRC). The ROC shows performance as a tradeoff between TPR and FPR or TNR [190]. The PRC shows performance as a tradeoff between PPV and TPR. While both curves measure the proportion of preferred items that are actually recommended, ROC emphasize the proportion of items that are not preferred and end up being recommended while PRC emphasizes the proportion of recommended items that are preferred. The Acc is defined as:

$$\text{Acc} = \frac{TP + TN}{TP + TN + FP + FN} \quad (2.62)$$

The CCR is defined as

$$\text{CCR} = \sum_{k=1}^C P(C_i) \frac{n_i}{N_i} \quad (2.63)$$

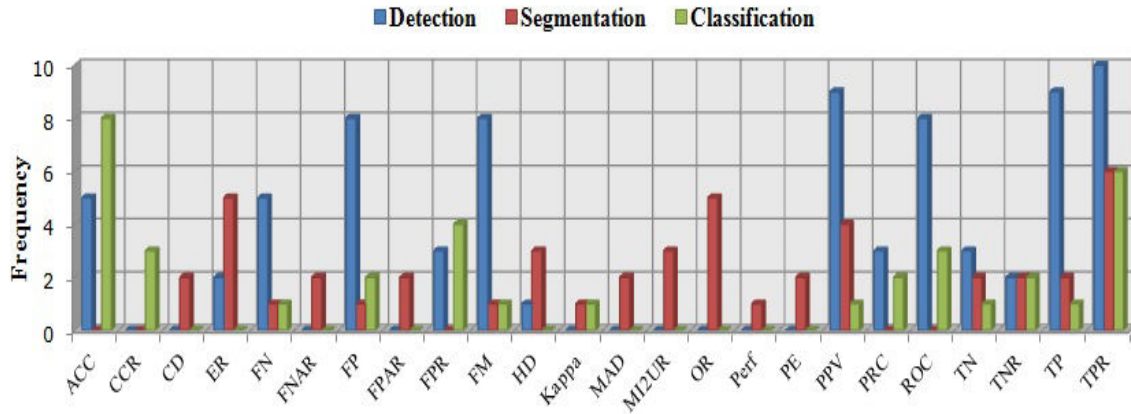


Figure 2.3: The count of performance metrics used in nuclei detection, segmentation and classification.

where n_i is the number of samples correctly classified to the i^{th} class by using classifier, C is the total number of classes, N_i is the total number of samples in the i^{th} class, $P(C_i)$ is the prior probability that an observed data falls in class C_i .

Figure 2.3 describes the count of different performance metrics used for nuclei detection, segmentation and classification methods.

2.7 Evaluation Methods

The supervised CAD systems need to be trained on manually annotated data. A medical expert, who labels the samples according to their class, usually provides these training data. Like many other biomedical applications, training data is not abundant either due to the cost involved in obtaining expert annotations or because of overall data scarcity. It generally requires two steps; first step is training the system to learn the parameters, and second step is testing or validating the system to evaluate the success of results. The amount of data in training and testing has critical impact on system performance. More data in training lead to better system designs, whereas more data in testing lead to more reliable evaluation of the system.

Cross-Validation (CV) is a statistical method of evaluating and comparing the systems according to the accuracy obtained on the training set. It brings the risk of memorization of data and obtaining over-optimistic error rates. To avoid the memorization problem, the system should be evaluated on a separate dataset (i.e. testing data), which is not used in training dataset. In typical CV, the training and testing sets must crossover in successive rounds such that each data point has a chance of being validated against. The basic form of CV is K-fold cross validation (Kf-CV). Other forms of CV are special cases of Kf-CV or involve repeated rounds of Kf-CV. In Kf-CV the data is first partitioned into K equally (or nearly equally) sized segments or folds. Subsequently K iterations of training and testing are performed such that a different fold of the data is holdout for testing while the remaining K-1 folds are used for training in each iteration.

Leave-one-out cross-validation (LOOCV) is a special case of Kf-CV where K equals the number of instances in the data. For each iteration, nearly all the data, except for a single observation, are used for training and the single observation is used to test the model. An accuracy estimate obtained using LOOCV is known to be almost unbiased but it has high

Table 2.3: List of Evaluation Techniques Used in Previous Studies

Techniques	Research Studies
Separate training and testing set	[134, 65, 187, 136, 183, 88, 113, 150]
HOV	[133, 39, 8, 74, 27, 22, 128, 33]
Kf-CV	[164, 37, 157, 17, 12, 45, 31]
LOOCV	[45, 138]

variance, leading to unreliable estimates [47]. In holdout validation (HOV), a subset of data is chosen randomly from the initial sample to form a testing set, and the remaining data are retained as the training set. This would generally not be considered to be CV since only a single partition of the data into training and testing sets is used. In Table 2.3, we provide the list of evaluation methods used in different studies.

2.8 Inspection and Editing Software

CellProfiler [82], ImageJ [142], ITK [77], LNKnet [105], Matlab [119], PLoS [159] and Weka [139] are well-known software packages, which are extensively used for preprocessing, segmentation, feature computation, feature selection and classification. These software packages integrate existing state-of-the-art methods like anisotropic diffusion, morphological operators, watershed, ACMs, thresholding, neural network, fuzzy, clustering, and other machine-learning methods into a modular software package. With LNKnet, Petushi *et al* [133] used LDA, SFFS and SFBS methods for feature selection and identified as accurately predicting the histologic grade significantly more frequently than other non-selected features.

2.9 Limitations and Challenges in Previous Frameworks

Since last decade, a huge number of articles have been published in the field of histopathology focusing on nuclei detection, segmentation and classification in different image modalities. Still there are some open research areas with little studies. These open research areas have unique challenges, which should be covered in future research. One of these is benchmark datasets. The results of previous studies are based on their own datasets. However, we believe that it is not straightforward to evaluate and numerically compare different studies solely based on their reported results as they used different datasets, evaluation methods and performance metrics. For numerical comparison of the studies, it is definitely necessary to develop benchmark datasets. These datasets should consist of samples that are taken from a large number of patients and annotated by different pathologists. Such an effort would make possible the numerical comparison of the results obtained by different studies and to identify the distinguishing features. To the best of our knowledge, we only found few benchmark datasets: UCSB Bio-Segmentation [57], the MITOS mitosis detection [1] benchmark, as well as a recent similar initiative AMIDA [3].

The UCSB Bio-Segmentation Benchmark dataset consists of 2D/3D images and time-lapse sequences that can be used for evaluating the performance of novel state-of-the-art computer vision methods. The data covers sub-cellular, cellular and tissue level. Tasks include segmentation, classification, and tracking.

The MITOS benchmark has been set up to provide a database of mitosis freely available to the research community. Mitotic count is an important parameter in breast cancer grading as it gives an evaluation of the aggressiveness of the tumor. Detection of mitosis

is a very challenging task, since mitosis are small objects with a large variety of shape configurations but it has not been addressed well in the literature, mainly because of the lack of available data. The MITOS benchmark has been set up as an international contest of mitosis detection in the framework of conference ICPR 2012. AMIDA benchmark reedited in 2013 the same type of challenge (mitosis detection from H&E images), as MITOS did in 2012.

Most of these benchmarks highlighted the fact that working on common professional digital image databases allow us to overcome any bias of the ones tested for the purpose of a precise publication. We all still more way to go to reach clinically acceptable results, at the image of the best results of MITOS, which from about 120 initial candidates (institutes, academia and major companies in the area) kept only 17 at the final round, with FM of about 0.7821 at best [1].

Preparing the GT, more specifically for nuclei, is another challenging problems. Fuchs and Buhmann [54] reported 42% disagreement between five pathologists on nuclei classification as normal or atypical. They also reported intra-pathologist error of 21.2%. In addition, the results of a very self-confident pathologist who was always very certain of his decisions but ended up with an error of 30% in the replication experiment. On the other hand, the results of a very cautious expert who is rather unsure of his decision, but with a misclassification error of 18% he performed significantly better than the previous one. This concludes that self-assessment is not reliable information to learn from. The intuitive notion, to select only those samples having high confidence by domain experts is not valid. A similar study by Malon *et al* [114] reported a moderate agreement between three pathologists for identifying mitotic cells on H&E stained breast cancer slides.

Other issues regarding standardization and experimental methods include: (i) different scanners used for image acquisition with different image resolution and pixel size, (ii) different staining characteristics, (iii) different lightening conditions, (iv) magnification levels, (v) different number and size of images (frames, whole slide images and regions of interest are examples of types of images used by the different authors).

Preprocessing consists in determining regions of interest, removing noise and enhancing image. Although different techniques such as histogram equalization, anisotropic diffusion, gamma correction, thresholding and morphology show different levels of success in noise and artefacts removal and image enhancement, yet the problem has not been entirely solved.

Segmentation methods like thresholding, region growing and watershed can locate the nuclei region but problems arise when they try to segment the touching and overlapping nuclei. They employ only local intensity information without any prior knowledge about the object to be segmented and produce inaccurate nuclei boundaries.

Dealing with overlapping and clustered nuclei is still a major challenge in the field of nuclei segmentation. While different methods have been developed with various levels of success in literature for the problem of overlapping and clustered nuclei, the problem has not yet been completely solved. A variety of schemes taking into account concavity [186, 51, 12, 90, 124, 179], distance transform [177, 79, 91], marker-controlled watershed [74, 67, 27, 45, 175], adaptive ACM with shape term and curvature information [95, 173, 11, 140], GMM & EM [80] and graphs [186, 9, 124] have been investigated to separate overlapping and clustered/touching nuclei. These methods have good results for nuclei that are slightly touching or overlapping each other, but they are not suitable for specimens containing larger numbers of nuclei with extensive overlapping and touching. These methods suffer from dependencies inducing instability.

For instance the computation of curvature is highly dependent on concavity point detection algorithm, region growing tends to rely on shape and size of nuclei, marker-controlled watershed needs true nuclei markers, and ellipse-fitting techniques are unable to accommodate the shape of most nuclei. Most of these methods also require prior knowledge. In spite

of the availability of few methods like clustering, GMM & EM and new image modality [173] able to deal with heterogeneity, accurate segmentation of touching or overlapping nuclei is still an open research area.

In MSI, the question is how to select the correct SBs from the spectral range to best characterize the problem. More specifically, there are strong correlations between SBs and some SBs cannot discriminate between nuclei and others; this is the reason why the effective dimension of SBs for classification is less than the total number of SBs. In literature, we found one possible solution that uses information theory, more specifically, by means of measures based on the mutual information (MI) to feature selection for pixel classification. Furthermore, Martinez-Usó *et al* [116] proposed a hierarchical clustering framework based on MI for SBs selection. Kamandar *et al* [81] used minimum redundancy maximum relevance (mRMR) technique [131] for SBs selection in AVARIS data.

To the best of our knowledge, only comparatively few supervised machine-learning techniques like Bayesian [17, 80], SVM [173] and AdaBoost [176] are used for nuclei segmentation. The basic philosophy of the machine-learning is that human provide examples of the desired segmentation (GT), and leave the optimization and parameter tuning tasks to the learning algorithm. They are supposed to adaptively extract domain specific knowledge and also to optimize tuning parameters. Although the context and domain information is utilized to improve the accuracy of nuclei segmentation, the nuclear characteristics are rarely used for nuclei segmentation. Overfitting, as the fact that two real clinical situations may have quite different characteristics, induces serious limitations to the machine-learning based methods.

This methodology review highlights an important gap to be fulfilled by all scientists in order to be able to reliably go to the next generation of important challenges, related to the "digital" exploration and the understanding of the WSI as an essential high-content imaging diagnostic biomarker and prognosis support. Consolidating, in the next few years these approaches with mining structured big data and analytics as with genomics and molecular imaging technologies, will certainly have the potential to lead to the next generation of healthcare technologies.

2.10 Overview of Proposed Framework and Scientific Contributions

In this thesis we aim at proposing novel frameworks for mitosis detection in color and multispectral images of breast cancer histopathology. In order to reach the aim, our main research directions and objectives are:

- i. A comprehensive analysis of different color spaces and color channels for mitosis discrimination
- ii. An extensive studies of intensity (first order statistical) features and texture (second order statistical) features in various color channels of different color models rather than single color model
- iii. A study of region and patch based texture features for mitosis classification
- iv. An automatic and unsupervised focal plane selection
 - v. An encyclopedic study of spectral absorption responses of different tissue components
- vi. A study of multispectral statistical features in selected SBs rather than single or all SBs
- vii. An extensive investigation of classifiers for mitosis classification
- viii. An inspection of over-sampling method for balancing the skewed dataset by increasing the number of minority class to improve the predictive accuracy of classification

- ix. An extension of *itk::QuadEdgeMesh* data structure to handle both primal and dual meshes, simultaneously, and illustrating two types of primal meshes: triangular / simplex meshes and Voronoi / Delaunay
- x. An efficient and robust strategy to explore WSI by combining computational geometry tools with local signal measure of relevance in a dynamic sampling framework
- xi. A real time evaluation of proposed frameworks in Cognitive Microscope (MICO) platform prototyping

This thesis proposes three frameworks for mitosis detection in breast cancer histopathology. The first is the Textural based Mitosis detection in Color images (TMC), second is Intensity, Textural & Morphology based Mitosis detection in Color images (ITM²C), and third is Multispectral Intensity, Textural and Morphology based Mitosis detection in Multispectral images (MITM³) frameworks.

In TMC framework, we investigate the various intensity and texture features using machine-learning techniques for mitosis detection. We also explore the features characteristics in three channels of RGB color model and blue-ration (BR) image, a new image modality.

In ITM²C framework, we comprehensively analyse the intensity and texture features in various color channels of different color models rather than a single color model and also combine selective intensity and textures features with morphological features in order to identify mitosis. In this framework, we also investigate the over-sampling method for balancing the unbalanced dataset by interpolating existing training samples.

In MITM³ framework, we address two important questions: First, does the multispectral statistical analysis on selected SBs (as opposed to single SB or all the SBs) suffice for efficient classification of mitotic and non-mitotic figures. An obvious advantage of using selected SBs is its reduced computational and storage complexity. Second, how effective are the multiple features for discrimination of mitotic and non-mitotic figures as compared to one type of features. The main novel contributions of this framework are:

- i. An automatic and unsupervised focal plane selection process
- ii. Three different methods for SBs selection including relative spectral absorption of different tissue components, spectral absorption of H&E stains and mRMR technique.
- iii. Computation of morphological & multispectral statistical features (MMSF) containing intensity, texture and morphological features which leverage discriminant information from a given candidate across selected SBs for classification of mitotic and non-mitotic figures.
- iv. An extensive investigation of classifiers and inference of the best one for mitotic figures classification.

We evaluate our proposed frameworks on MITOS dataset [150]. The dataset is made up of 50 high power fields (HPF) coming from five different patient slides scanned at $\times 40$ magnification. There are 10 HPF per slide. The pathologist has annotated all the mitosis nuclei manually in each selected HPF on the images generated by the Aperio scanner, Hamamatsu scanner and multispectral microscope as shown in Figure 2.4.

We also propose an extension of *itk::QuadEdgeMesh* data structure to handle both primal and dual meshes, simultaneously. The new data structure, *itk::QuadEdgeMeshWithDual*, already include by default the due topology, to handle dual geometry as well. Two types of primal meshes are specifically illustrated: triangular / simplex meshes and Voronoi / Delaunay.

Furthermore, we propose an innovative platform in which dynamic sampling method performed fast analysis of WSI. We test dynamic sampling method for real time evaluation of Cyto-nuclear atypia (CNA) score on breast cancer WSI in MICO platform. In the

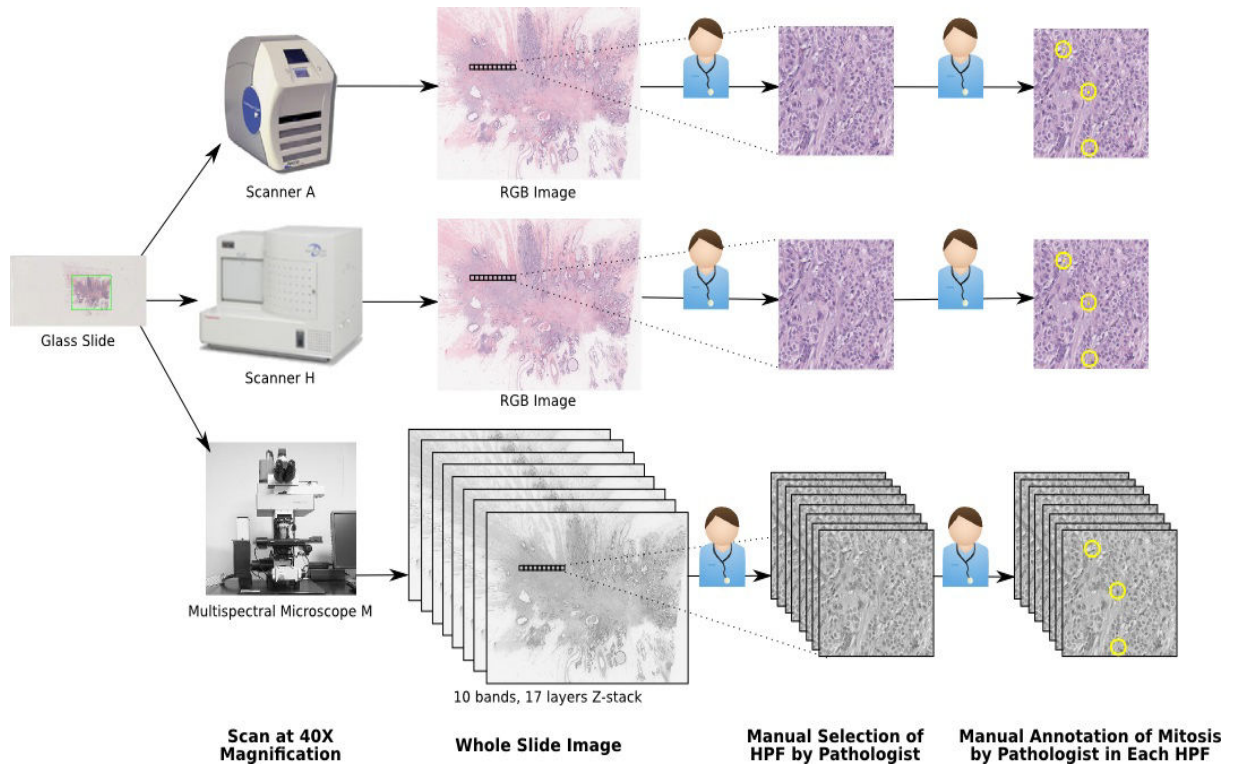


Figure 2.4: MITOS Dataset generated by Aperio scanner, Hamamatsu scanner and multispectral microscopy. These HPF are selected and annotated by senior pathologist.

medical application, more specifically analysing WSI, our dynamic sampling method has proved its ability to accurately find and measure the highest levels of CNA in a WSI within an acceptable time frame as well as to provide a useful, reliable visualization map for the end user. From a more global standpoint, this dynamic sampling method makes it possible to speed up the analysis, enhance the visualization and assist the exploration of high-content images.

2.11 Conclusion

In this chapter, we have briefly described the most commonly used image-processing methods. We have demonstrated the different steps of existing frameworks in quantitative histopathology. We have comprehensively described the state-of-the-art frameworks for nuclei detection, segmentation and classification, used in various types of tissue analysis and cancer grading. At last, we have identified the limitations and open challenges in existing frameworks and give overview of proposed framework with novelties. In next chapter, we will propose two frameworks for mitosis detection in color images of breast cancer histopathology.

Chapter 3

Automated Mitosis Detection in Color (RGB) Images

Résumé du chapitre

Dans ce chapitre, nous proposons deux systèmes pour la détection de mitoses dans des images couleur de cancer du sein en histopathologie. Nous expliquons les différentes étapes de chaque système proposé, à savoir le pré-traitement de l'image, la détection et la segmentation des candidats, le calcul de descripteurs pour obtenir une signature de chaque candidat, la sélection des paramètres les plus discriminants, la classification des candidats et la prise en compte de l'asymétrie entre le nombre d'exemples de mitoses (peu nombreux) et de non-mitoses (très nombreux) dans le jeu d'apprentissage. Dans le premier système, nous étudions les caractéristiques de texture pour la discrimination de mitoses. Nous explorons également la caractérisation des descripteurs dans les trois canaux rouge-vert-bleu (RVB) des images couleur et dans l'image rapport de bleu (blue ratio). Dans le second système, nous analysons l'apport de l'intensité et de la texture dans les canaux de couleur sélectionnés de plusieurs modèles de couleur ainsi que la combinaison de descripteurs morphologiques pour l'identification des mitoses. Nous introduisons le concept d'asymétrie des données pour prendre en compte le déséquilibre du jeu de données d'apprentissage, et nous comparons les résultats de nos deux systèmes avec ceux du concours MITOS [150]. Enfin, nous introduisons les stratégies utilisées pour la détection de mitoses et de l'atypie des noyaux dans les images d'histopathologie.

3.1 Introduction

In this chapter, we propose two frameworks for mitosis detection in color images of breast cancer histopathology. We explain different steps of the proposed frameworks in detail. These steps are pre-processing, candidate detection and segmentation, feature computation and selection, candidate classification and handling training set asymmetry. In the first framework, we investigate the texture features for mitosis discrimination. We also explore the features characterization in three channels of RGB color model and blue-ratio (BR) image. In the second framework, we comprehensively analyze the intensity and texture features in selected color channels of various color models rather than a single color model and also combine morphological features in order to identify mitosis. We introduce the concept of data asymmetry for handling imbalanced training set, and we compare the results of the proposed frameworks with the MITOS contest result [150]. Finally, we briefly explain the strategies used for mitosis detection and nuclei pleomorphism in WSI.

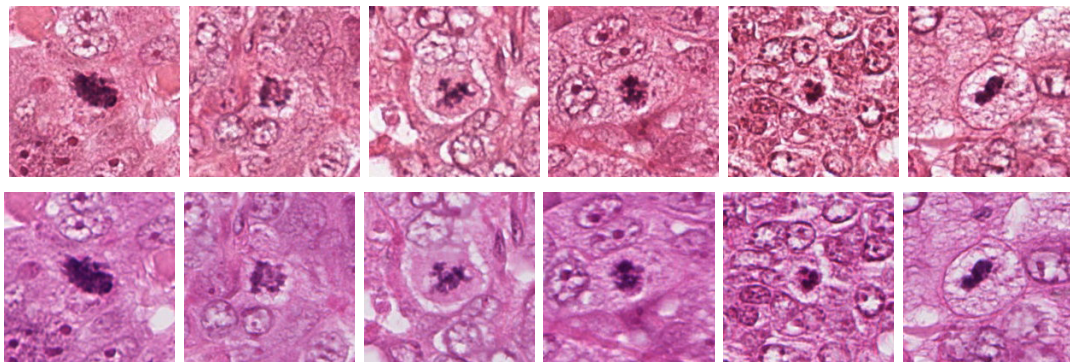


Figure 3.1: Example of ground truth mitosis nuclei for Aperio (first row) and Hamamatsu (second row) scanners.

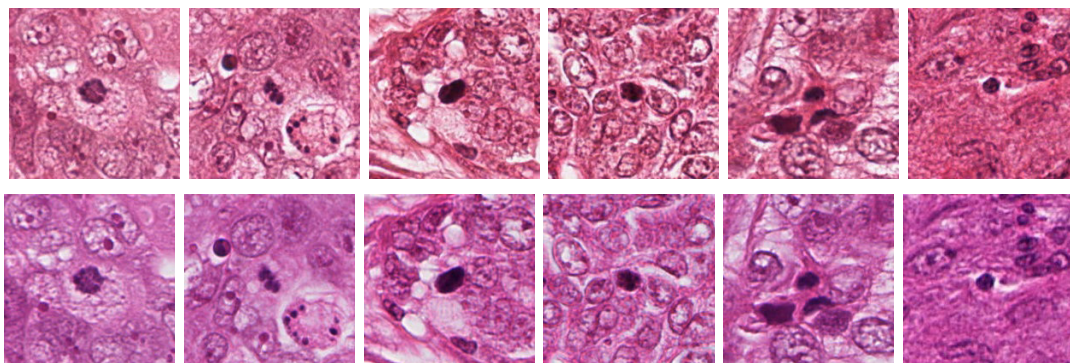


Figure 3.2: Some example of non mitosis nuclei for Aperio (first row) and Hamamatsu (second row) scanners. The non mitosis nuclei are located in the centre of each image.

3.2 Challenges in Mitosis Count

Mitosis count is an important parameter in breast cancer grading as it gives an evaluation of the aggressiveness of the tumor. Detection of mitosis is a very challenging task because they are small objects with a large variety of shape configurations, texture variation and low frequency of appearance. Some example of GT mitosis are shown in Figure 3.1. Mitosis have similarity with other types of nuclei, as shown in Figure 3.2 and other objects e.g., apoptosis and dust particles, as shown in Figure 3.3.

Mitosis count has not yet been addressed well in the literature. Only few works concern detection of mitosis. Belien *et al* [19] counted mitosis on Feulgen stained breast cancer sections. Liu *et al* [102] and Huh *et al* [75] proposed mitosis detection in time-lapse phase contrast microscopy image sequences of stem nuclei populations and Schlachter *et al* [154] performed detection of mitosis in fluorescence staining of colorectal cancer. Roullier *et al* [149] proposed detection of mitosis on breast cancer slides with an IHC staining that highlights specifically mitosis.

Few works concern mitosis counting on H&E stained slides. Malon *et al* [115, 113] proposed the use of CNN. Sertel *et al* [157] presented a method for mitosis and karyorrhexis nuclei (dying nuclei) counting all of them, without distinction. For breast cancer grading, only mitosis nuclei must be counted. Recently, Ciresan *et al* [35] used CNN to compute a map of probabilities of mitosis over the whole image. Their CNN has been trained with GT mitosis from the training dataset. Their approach proved to be very efficient as they had the best FM on Aperio dataset during ICPR contest [150].

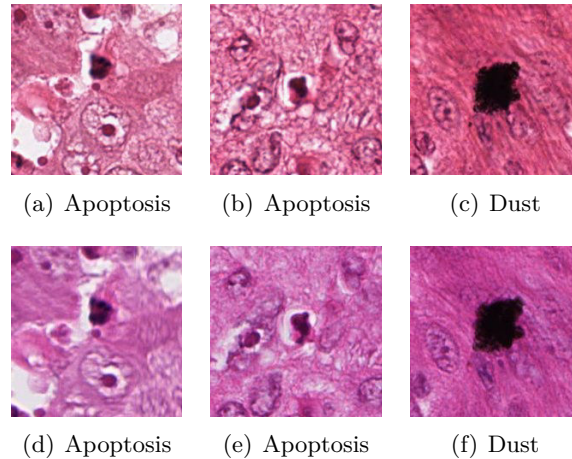


Figure 3.3: Example of apoptosis and dust particle that looks similar to mitosis nuclei for Aperio (first row) and Hamamatsu (second row) scanners.

Aperio and Hamamatsu scanners	
Training Data Set	226 mitotic cells
35 HPFs	69.3% of total
Evaluation Data Set	100 mitotic cells
15 HPFs	30.7% of total
TOTAL	326 mitotic cells

Table 3.1: Number of HPFs and mitosis nuclei in training and evaluation data sets.

3.3 Color Dataset

The dataset of MITOS contest is made up of 50 HPF coming from five different patient slides scanned at $\times 40$ magnification. There are 10 HPF per slide. The pathologist has annotated all the mitosis nuclei manually in each selected HPF on the images generated by the Aperio and Hamamatsu scanner. A HPF has a size of $512 \times 512 \mu\text{m}^2$ (that is an area of 0.262 mm^2), which is a surface equivalent to that of a microscope field diameter of 0.58 mm . These 50 HPFs contain a total of 326 mitotic cells on images of both scanners. Table 3.1 gives the number of mitosis nuclei in the training and the evaluation data sets.

Aperio scanner has a resolution of $0.2456 \mu\text{m}$ per pixel. Hamamatsu scanner has a slightly better resolution of $0.2273 \mu\text{m}$ (horizontal) and $0.22753 \mu\text{m}$ (vertical) per pixel. Note that a pixel of Hamamatsu scanner is not exactly a square. Table 3.2 shows the resolutions of the different scanners. For example, a mitosis having an area of $50 \mu\text{m}^2$ will cover about 830 pixels of the image produced by Aperio scanner and about 965 pixels of the image produced by Hamamatsu scanner.

3.4 Textural based Mitosis detection in Color images (TMC) Framework

We propose a TMC framework for mitosis detection in color images of breast cancer histopathology. This framework addresses the shortcomings of previous works which are:

Equipment	Resolution per pixel	HPF Dimension to cover an area of $512 \times 512 \mu\text{m}^2$
Aperio Scanner	$0.2456 \mu\text{m}$	2084×2084 pixels
Hamamatsu Scanner	$0.2273 \mu\text{m}$ horizontal $0.22753 \mu\text{m}$ vertical	2252×2250 pixels

Table 3.2: Resolution of the Aperio and Hamamatsu scanners.

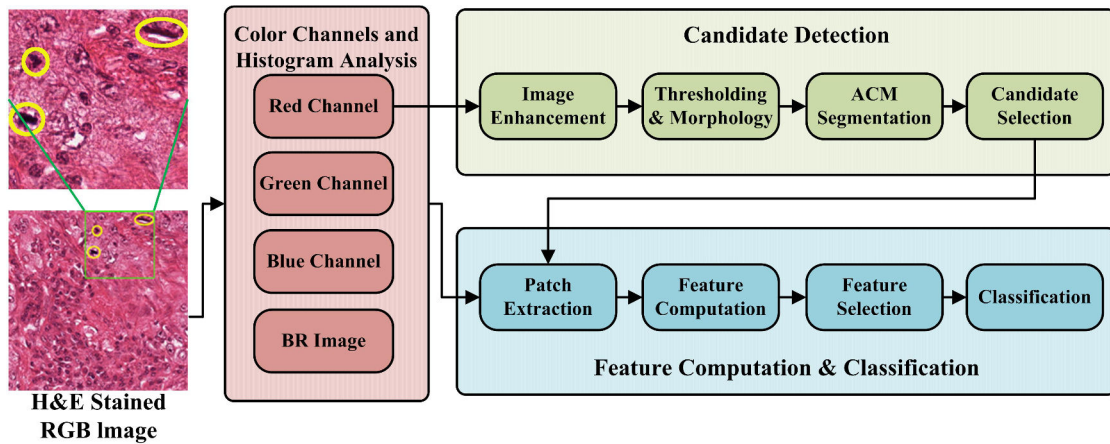


Figure 3.4: TMC Framework.

(1) by including comprehensive analysis of texture features (second order statistics features such as Haralick Co-occurrence (HC) and run-length (RL) features) in RGB color model and BR image, and (2) by exploring different classifiers to achieve a higher accuracy of mitosis detection. The aim is to improve the accuracy of mitosis detection by integrating the color channels that better capture the texture features, which discriminate mitosis from other objects. This framework is shown in Figure 3.4. First, we transform RGB image into BR images and channels of RGB color model. Second, we perform histogram analysis on selected regions in all color channels. Third, we perform smoothing, thresholding and morphological operations on selected color channel to generate candidate mitosis regions. The boundaries of these regions are refined using ACM segmentation. We select candidate regions using morphological rules; we calculate the centre of each region as seed point of candidate and extract a patch of size 70×70 pixels from BR image and red, green and blue channels. Fourth, we compute HC and RL for each candidate patch. Fifth, we select features having better discrimination of mitosis regions from others. Finally, a classification is performed to put the candidate patch in the mitosis or in the non-mitosis class. Four different classifiers have so been evaluated: DT, MLP, L-SVM and NL-SVM.

3.4.1 Blue-Ratio Image

In H&E stained images, nuclear and cytoplasm regions appear as hues of blue and purple while extracellular material have hues of pink. In order to reduce the extracellular regions responses, the RGB images are transformed into new image called Blue-Ratio (BR) image to accentuate the nuclear dye [31] as:

$$\text{BR} = \frac{100 \times B}{1 + R + G} \times \frac{256}{1 + B + R + G} \quad (3.1)$$

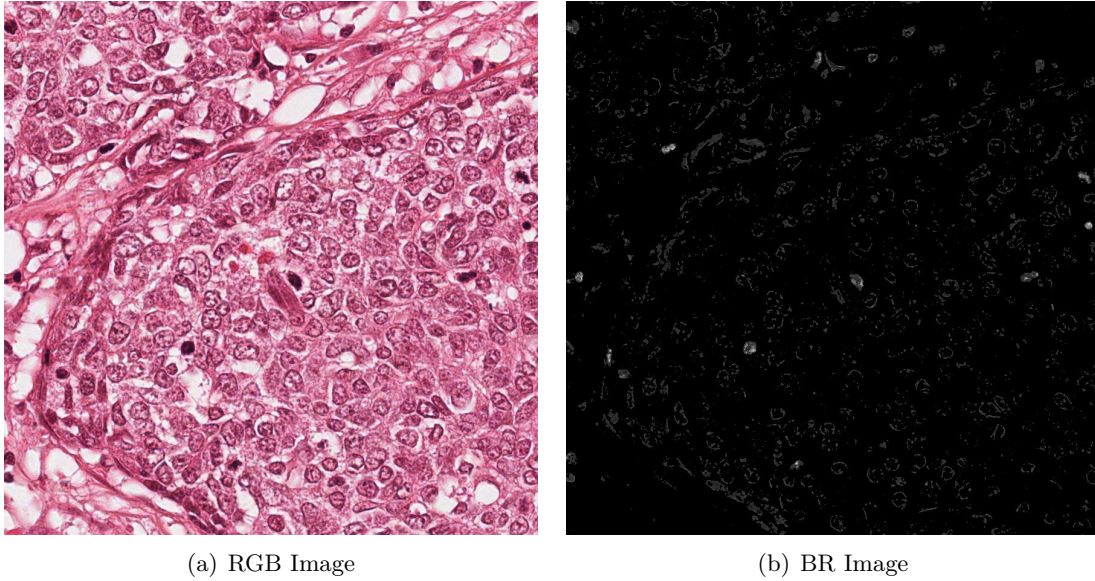


Figure 3.5: RGB and BR Image.

where B, R and G are blue, red and green channel of RGB, respectively. In a BR image, a pixel with a high blue intensity relatively to its red and green components is given a high value, whereas, a pixel with a low blue intensity as compared to its red and green components is given a low value. As we are interested in nuclei, which appear as blue-purple areas, a blue-ratio image is an efficient tool to have a first clue on the position of nuclei in the image. An example of blue-ratio image is shown in Figure 3.5. A histogram of absorption responses of mitosis, nuclei and background regions in BR images are shown in Figures 3.6(d) and 3.7(d) for Aperio and Hamamatsu images, respectively. These histograms describe the separation of mitosis, nuclei and background regions.

3.4.2 Color Channels Histogram Analysis & Importance of Red Channel

We compute absorption responses of mitosis and non-mitosis nuclei, and background regions for three color channels and BR image. The histogram analysis of these absorption responses for Aperio and Hamamatsu images are shown in Figures 3.6 and 3.7, respectively. The peaks of the mitosis and non-mitosis nuclei are almost similar in BR image and red, green and blue channels. Peaks of mitosis and non-mitosis nuclei are different from peaks of background regions in red, green and blue channels. As peaks of mitosis nuclei and background regions have the best separation in red channel as compared to BR image and green and blue channels, we select red channel for candidate detection. While red channel histogram analysis is thus able to differentiate between different tissue parts (i.e., nuclei, background) but absorption responses of mitosis and non-mitosis nuclei is not distinguishable.

The process of nuclei division has four different stages and each has different size, shape and textures. This motivates further textural analysis on different color channels to achieve reasonable classification of regions into mitosis and non-mitosis nuclei. For feature computation, we select all color channels including BR image.

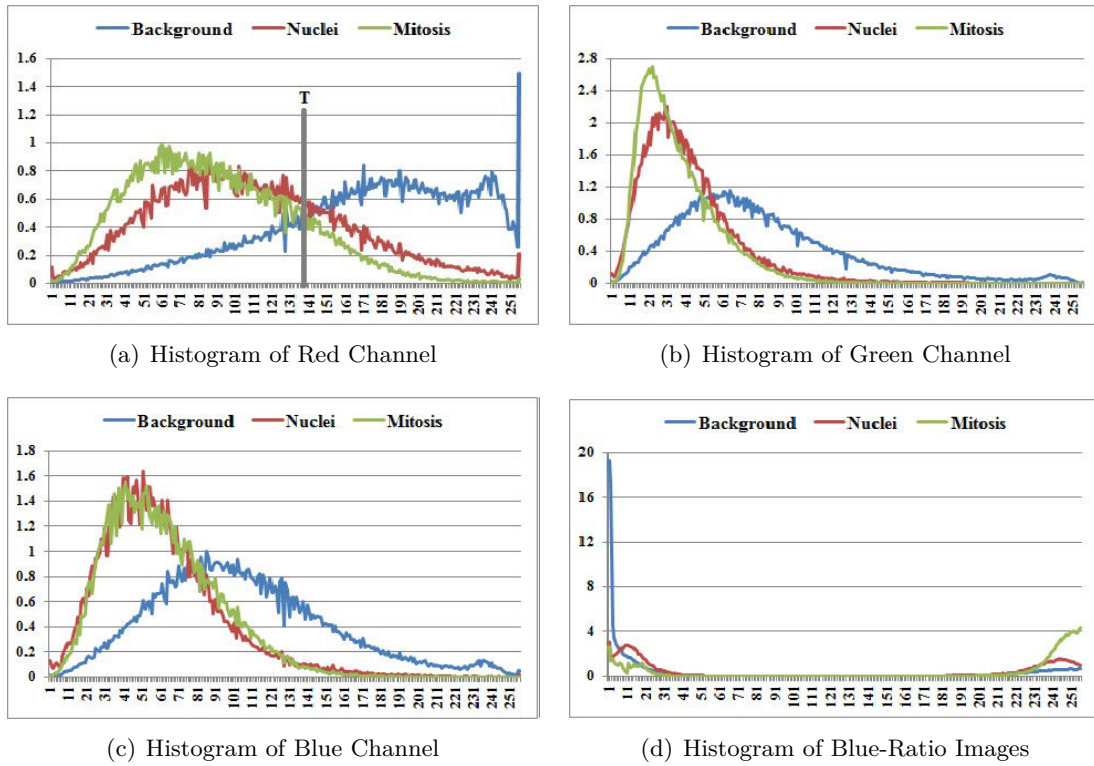


Figure 3.6: Histogram analysis of different channels on Aperio dataset.

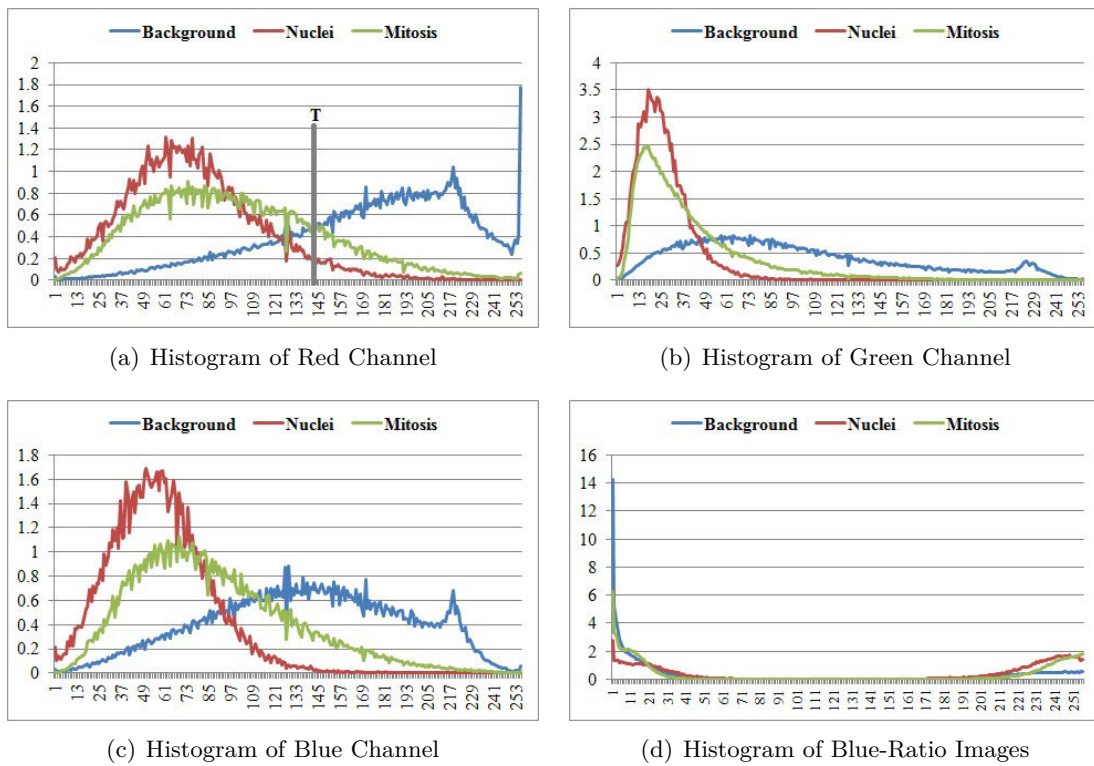


Figure 3.7: Histogram analysis of different channels on Hamamatsu dataset.

3.4.3 Candidate Detection in Red Channel

First, we smooth red channel image using median filter as shown in Figure 3.8(a) and 3.9(a). Then we perform binary thresholding on enhanced image using the threshold T as shown in Figure 3.8(b) and 3.9(b):

$$I^B(x, y) = \begin{cases} 1, & \text{if } I(x, y) < T \\ 0, & \text{otherwise} \end{cases} \quad (3.2)$$

The T value is selected from the histogram of red channel where mitotic and background regions are well separated.

Morphological opening and closing are applied to the binary image $I^B(x, y)$ to merge the clustered region into large regions, fill holes and eliminate too small regions. Then we segment the boundaries of candidates using active contour models with a level set implementation [123]. The key steps for the segmentation method are as follows:

- i. We present a given nuclei contour $\varpi(t)$ as the zero level of the signed distance function $\psi(x, t)$. Formally, $\varpi(t) = \{x : \psi(x, t)\}$
- ii. We use the active contour formation as:

$$\frac{\partial \psi}{\partial t} = f(I)(\alpha b + \beta k)|\nabla \psi| + \gamma \nabla f \cdot \nabla \psi \quad (3.3)$$

where α , β and γ are user-defined settings for the relative scaling of the terms, f refers to the image-based feature function that is minimized at nuclei boundary and remains high elsewhere, b is a balloon force that is added to evolve the curve outwards, k is the curvature along the normal to the level set contour and $\nabla f \cdot \nabla \psi$ is the boundary attraction term. The result of segmentation is shown in Figure 3.8(c) and 3.9(c).

Finally, we select candidates by filtering based on size of candidates and take a patch from BR image and red, green and blue channels. As Aperio and Hamamatsu scanners have different resolution per pixel, the size of window on Aperio dataset is $17.192\mu m \times 17.192\mu m$ and the size of window on Hamamatsu dataset is $15.911\mu m \times 15.927\mu m$. An example of candidate detection is shown in Figure 3.8(d) and 3.9(d).

3.4.4 Texture Features Computation

During mitosis (nuclei division), nuclei undergo four different stages and each has different shape, size and textures that are very distinguishable from shape or texture of nuclei not under division process. This motivates further texture analysis of these candidate regions. We compute two types of textural features for classification of these candidates as mitosis and non-mitosis regions.

Haralick Co-occurrence (HC) Features

The human eye cannot discriminate between texture pairs with matching second order statistics [78]. The first machine vision framework for calculating second order or grey level co-occurrence texture information was developed by analysing aerial photography images [70]. In this technique grey level co-occurrence matrix $GLCM(i, j ; d, \theta)$ is computed. This matrix is square with dimension N_g where N_g is the total number of grey levels in the image. The value at i^{th} row and j^{th} column in the matrix is produced by counting the total occasions a pixel with value i is adjacent to a pixel with value j at a distance d and angle θ . Then the whole matrix is divided by the total number of such comparisons that have been made. Alternatively we can say that each element of GLCM matrix is considered as

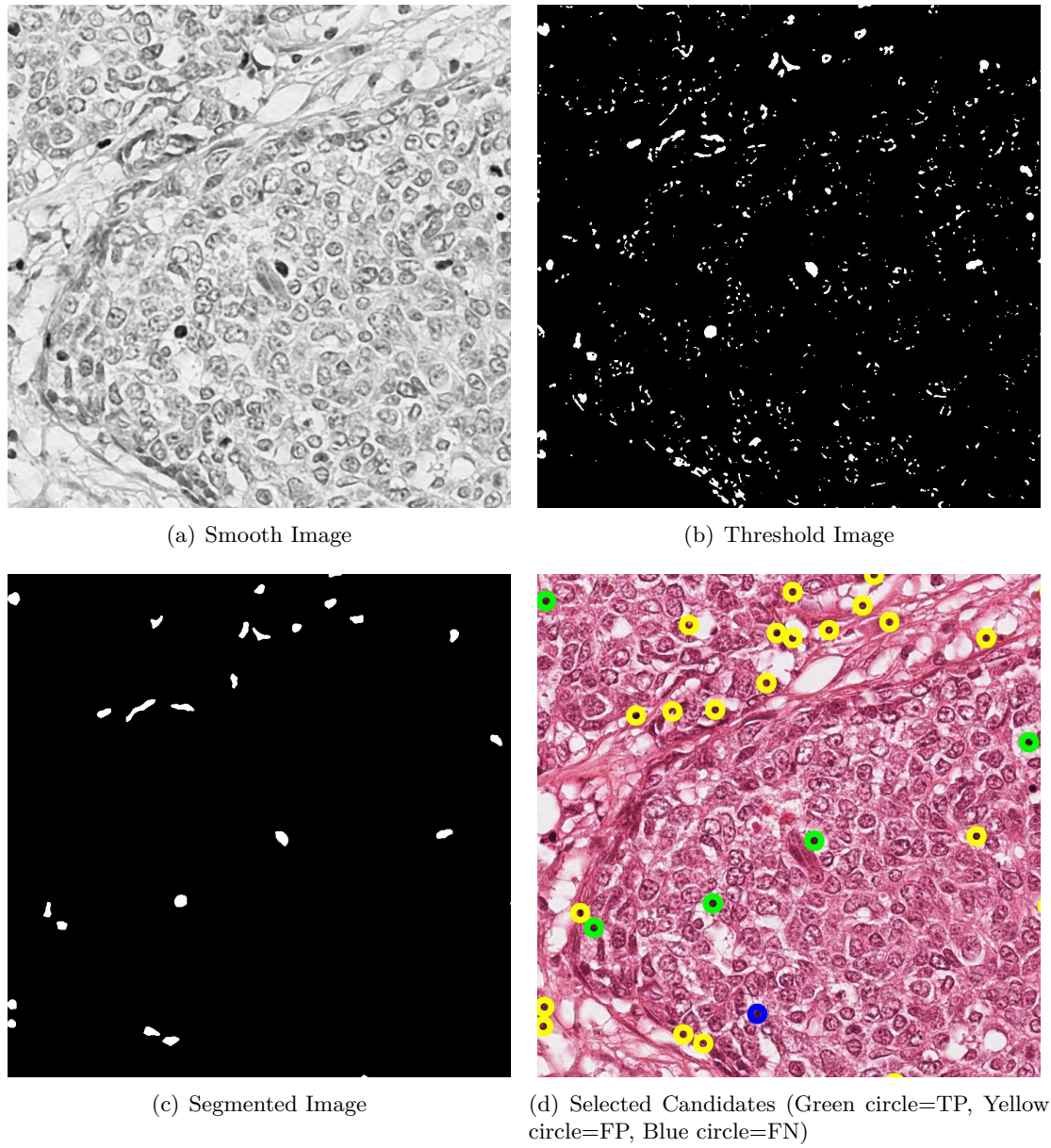


Figure 3.8: Different steps of candidate detection on Aperio image.

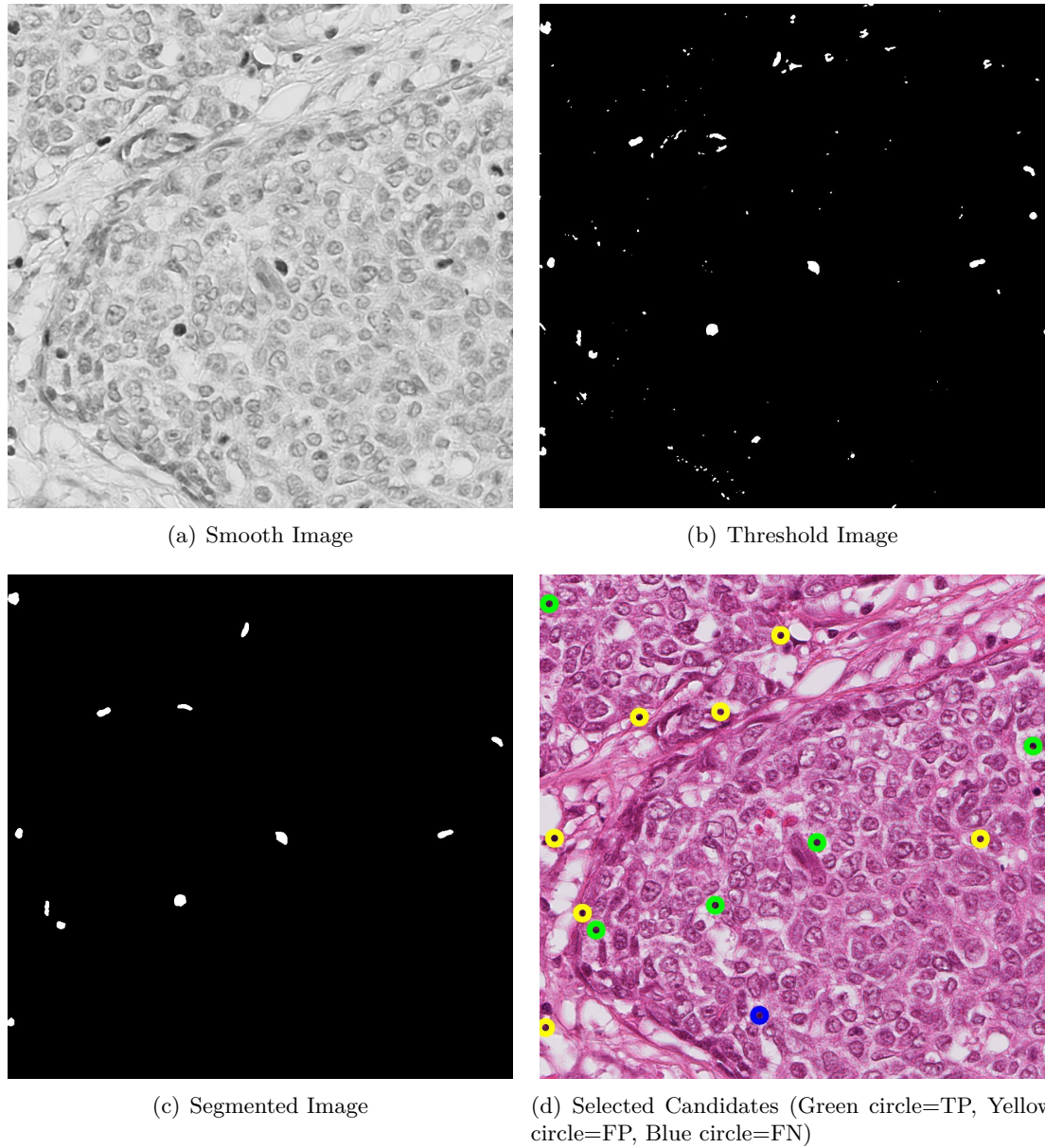


Figure 3.9: Different steps of candidate detection on Hamamatsu image.

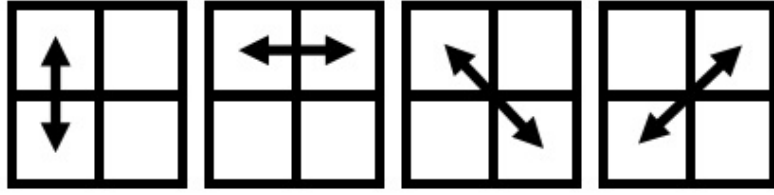


Figure 3.10: The four directions of adjacency that are defined for calculating texture features.

Table 3.3: Notation for HC Features

$P(i, j) = GLCM(i, j)/R$	$(i, j)^{\text{th}}$ entry in a normalized GLCM and R is the maximum number of resolution cells in the GLCM.
$P_x(i) = \sum_{j=1}^{N_g} GLCM(i, j)$	i^{th} entry in the marginal-probability matrix obtained by summing the rows of $P(i, j)$.
N_g	is the number of distinct grey-levels in the image.
μ_x and μ_y	are means of $P(i, j)$.
σ_x and σ_y	are standard deviations of $P(i, j)$.

the probability that a pixel with grey level i is to be found with pixel with grey level j at a distance d and angle θ .

$$GLCM = \begin{bmatrix} P(1, 1) & P(1, 2) & \cdot & \cdot & \cdot & P(1, N_g) \\ P(2, 1) & P(2, 2) & \cdot & \cdot & \cdot & P(2, N_g) \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ P(N_g, 1) & P(N_g, 2) & \cdot & \cdot & \cdot & P(N_g, N_g) \end{bmatrix} \quad (3.4)$$

By varying the displacement vector between each pair of pixels many GLCMs with different directions can be generated. We define adjacency in four directions (vertical, horizontal, left and right diagonals as shown in Figure 3.10) with one displacement vector, and as a result we compute four GLCMs. For this framework, we compute four GLCMs on each candidate patch in all color channels. In our problem, texture information is rotationally invariant. So, we take average in all four directions and result is one GLCM.

The GLCM captures the properties of a texture but they are not directly useful for further analysis, such as the classification using these matrices. To illustrate the computational requirements of this framework, we compute eight of the 14 numeric features proposed by Haralick [70] from the GLCM in order to represent the texture more compactly. These eight texture features are:

Correlation : the correlation of a pixel to its neighbor over the whole image

$$\text{Correlation} = \frac{\sum_{i,j}^{N_g} (i - \mu_x) (j - \mu_y) P(i, j)}{\sigma_x \sigma_y} \quad (3.5)$$

When correlation is high, the image will be more complex than when correlation is low.

Cluster shade : is a measure of the skewness of the GLCM, in other words the lack

of symmetry

$$\text{Cluster shade} = \sum_{i,j}^{N_g} (i + j - \mu_x - \mu_y)^3 P(i, j) \quad (3.6)$$

When cluster shade is high, the image is not symmetric.

Cluster prominence : is also a measure of the skewness of the GLCM

$$\text{Cluster prominence} = \sum_{i,j}^{N_g} (i + j - \mu_x - \mu_y)^4 P(i, j) \quad (3.7)$$

When cluster prominence is low, then there is a peak in the GLCM around the mean values that indicates little variations in grey values

Energy : also known as angular second moment, describes uniformity of the texture

$$\text{Energy} = \sum_{i,j}^{N_g} P(i, j)^2 \quad (3.8)$$

When energy is high, the image is homogeneous because of fewer entries of large magnitude in GLCM. Its range is [0,1].

Entropy : is a measure of the randomness

$$\text{Entropy} = - \sum_{i,j}^{N_g} P(i, j) \log_2 P(i, j) \text{ or } 0 \text{ if } P(i, j) = 0 \quad (3.9)$$

A homogeneous image has lower entropy than a heterogenous image. Its value is ≥ 0 . In fact, when energy gets higher, entropy should get lower.

Hara-correlation : is a measure of grey level linear dependence between pixels at the specified positions relative to each other.

$$\text{Hara-correlation} = \frac{\sum_{i,j}^{N_g} (i - j) P(i, j) - (\mu_x - \mu_y)}{\sigma_x \sigma_y} \quad (3.10)$$

Inverse Difference Moment (IDM) : also knows as local homogeneity. It measures the closeness of the distribution of elements in the GLCM to the GLCM diagonal.

$$\text{IDM} = \sum_{i,j}^{N_g} \frac{1}{1 + (i - j)^2} P(i, j) \quad (3.11)$$

When the IDM is low, then the image is inhomogeneous, and a relatively higher value for homogeneous images. Its range is [0,1].

Inertia : also knows as contrast. It measures the local variations.

$$\text{Inertia} = \sum_{i,j}^{N_g} (i - j)^2 P(i, j) \quad (3.12)$$

Its range is $[0, N_g^2]$.

Run-length (RL) Features

The set of consecutive pixels, with same grey level, collinear in a given direction, constitute a grey level run. The run length is the number of pixels in the run and the run length value is the number of times such a run occurs in an image. With investigation that, in a coarse texture, relatively long grey level runs would occur more often and that a fine texture should contain primarily short runs, Galloway proposed the use of a grey level run length matrix GLRLM for texture feature extraction [55]. $GLRLM(i, j; \theta)$ is defined as the number of runs with pixels of grey level i and run-length j in direction of θ . The dimension of GLRLM is $N_g \times R$, where N_g is the number of grey levels and R is the maximum run length. Similarly to the GLCM, we compute GLRLMs for four directions (vertical, horizontal, left and right diagonals as shown in Figure 3.10) and later average them.

We compute GLRLM for each candidate region, and then the following ten second order statistics features are derived:

Short Run Emphasis (SRE) :

$$SRE = \frac{1}{N_r} \sum_i^{N_g} \sum_j^R \frac{P(i, j)}{j^2} \quad (3.13)$$

Long Run Emphasis (LRE) :

$$LRE = \frac{1}{N_r} \sum_i^{N_g} \sum_j^R P(i, j) \cdot j^2 \quad (3.14)$$

Grey-level Nonuniformity (GLN) :

$$GLN = \frac{1}{N_r} \sum_i^{N_g} \left(\sum_j^R P(i, j) \right)^2 \quad (3.15)$$

Run Length Nonuniformity (RLN) :

$$RLN = \frac{1}{N_r} \sum_j^R \left(\sum_i^{N_g} P(i, j) \right)^2 \quad (3.16)$$

Low Grey-level Run Emphasis (LGRE) :

$$LGRE = \frac{1}{N_r} \sum_i^{N_g} \sum_j^R \frac{P(i, j)}{i^2} \quad (3.17)$$

High Grey-level Run Emphasis (HGRE) :

$$HGRE = \frac{1}{N_r} \sum_i^{N_g} \sum_j^R P(i, j) \cdot i^2 \quad (3.18)$$

Short Run Low Grey-level Emphasis (SRLGE) :

$$SRLGE = \frac{1}{N_r} \sum_i^{N_g} \sum_j^R \frac{P(i, j)}{i^2 \cdot j^2} \quad (3.19)$$

Short Run High Grey-level Emphasis (SRHGE) :

$$\text{SRHGE} = \frac{1}{N_r} \sum_i^{N_g} \sum_j^R \frac{P(i,j) \cdot i^2}{j^2} \quad (3.20)$$

Low Run Low Grey-level Emphasis (LRLGE) :

$$\text{LRLGE} = \frac{1}{N_r} \sum_i^{N_g} \sum_j^R \frac{P(i,j) \cdot j^2}{i^2} \quad (3.21)$$

Low Run High Grey-level Emphasis (LRHGE) :

$$\text{LRHGE} = \frac{1}{N_r} \sum_i^{N_g} \sum_j^R P(i,j) \cdot i^2 \cdot j^2 \quad (3.22)$$

where N_r is the total number of runs.

The eight HC features and ten RL features are computed for each candidate patch in BR and red, green and blue channels of RGB color model, which resulted in a total of 72 features.

3.4.5 Feature Normalization and Selection

Feature Normalization

In most cases, the features have different dynamic ranges. These different dynamic ranges of features affect the majority of classifiers which use the distance between two points. If one of the features has a broad range of values, the distance will be governed by this particular feature. Therefore, dynamic ranges of features are normalized so that each feature contributes proportionately to the final distance. We solve this problem by normalizing the features values so that they lie within similar dynamic ranges. The normalization formula is given as:

$$f' = \frac{f - f_{\min}}{f_{\max} - f_{\min}} \quad (3.23)$$

where f is original feature value, f' is the normalized feature value, f_{\min} is the minimum feature values and f_{\max} is the maximum feature value.

Feature Selection

Conceptually, a large number of descriptive features are highly desirable for classification of a patch as mitosis or non-mitosis. However, when we use all the extracted features for classification of candidate patch as mitosis and non-mitosis, the classification performance is poor. Some features are irrelevant for classification and some features are redundant that represents duplication of features and does not provide additional class discriminatory information, degrading the classification performance.

We use consistency subset evaluation method [103] to select a subset of features that maximize the consistency in the class values using the projection of subset of features from the training dataset. We evaluate the subset of features by looking for the combinations of features whose values divide the data into subsets containing a strong single class majority [92]. The search is biased in favor of small feature subsets with high class consistency. Our consistency subset evaluator uses the consistency metric proposed by Liu and Setiono [103] as:

$$\text{Consistency}_s = 1 - \frac{\sum_{j=0}^J |D_j| - |M_j|}{N} \quad (3.24)$$

where s is a feature subset, J is the number of the distinct combinations of features for s , $|D_j|$ is the number of occurrences of the j^{th} feature combination, $|M_j|$ is the cardinality of the majority class for the j^{th} feature combination and N is the total number of instances. The consistencies of these subsets are not less than that of the full set of features. We use these subsets in conjunction with a hill climbing search method, augmented with backtracking, which looks for the smallest subset with consistency equal to that of the full set of features.

3.4.6 Classification Techniques

This section presents a brief overview of the types of classifiers used for classification of mitosis and non-mitosis regions throughout this thesis. Since the last decade, classifiers based on a statistical learning theory have shown remarkable abilities to deal with both high-dimensional data and a limited training set [171, 172]. We select four well-known classification techniques which are briefly described below.

Decision Tree (DT)

The first classification technique used in this thesis is functional tree, a type of decision tree (DT) classifier. This classifier build decision tree in two phases [56]. In the first phase a large decision tree is constructed. In second phase, this tree is pruned back. We used divide-and-conquer approach to grow the tree. The most relevant aspects are the splitting rule, the termination criterion, and the leaf assignment criterion. We use logistic regression model at the inner nodes and leaves for the construction of new attributes [96]. It models the posterior class probabilities as:

$$P(C = c, X = x) = \frac{e^{F_c(x)}}{\sum_{i=1}^C e^{F_i(x)}} \quad (3.25)$$

where C is a label set, X is an instance set and $F_c(x) = \sum_{k=1}^K f_{kc}(x)$ and the f_{kc} are functions of the input variables.

Multilayer Perceptron (MLP)

The second classification technique used in this thesis is multilayer perceptron (MLP) classifier. Simple perceptron consists of a layer of input neurons, coupled with a layer of output neurons, and a single layer of weights between them. The learning process consists of selecting the appropriate weights between the input and output layer. This simple perceptron only solve linearly separable problems. To obtain a bilinear solution, more layers of weights are added to the simple perceptron model obtaining the MLP [167, 84]. In MLP classifier (Figure 3.11), each node of a hidden layer or output layer and the output $y(j)$ of node j is related to its input as:

$$y(j) = \frac{1}{1 + e^{-S(j)}} \quad (3.26)$$

where $S(j) = \sum_{k=1}^K y(k)w(k, j)$ and $w(k, j)$ are connections weights between the previous node k and the current node j ; $y(k)w(k, j)$ is the weighted output of the previous node k , which is used as input to node j ; K is number of inputs to node j ; and $S(j)$ is the sum of all weighted input $y(k)w(k, j)$ of the previous layer to node j .

The connection weights $w(k, j)$ between different layer nodes are calculated iteratively until they stabilize, by following:

$$w(k, j)^{t+1} = w(k, j)^t + \alpha \varepsilon(j) y(j) + \beta (w(k, j)^t - w(k, j)^{t-1}) \quad (3.27)$$

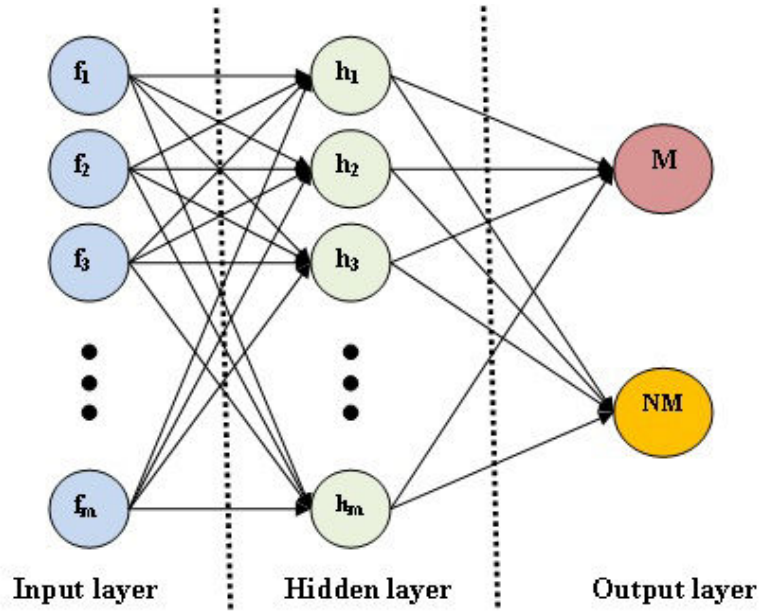


Figure 3.11: The used architecture of MLP contains one input layer with nodes equal to number of features, one hidden layer and one output layer with two classes.

where $(t + 1), t, (t - 1)$ correspond to next, current and previous weights, respectively, α, β are constants, $\varepsilon(j)$ is the error between the desired output $y(j)'$ and actual output $y(j)$, and is computed as:

$$\varepsilon(j) = (y(j)' - y(j))y(j)(1 - y(j)) \quad (3.28)$$

and error for a hidden layer node is computed as:

$$\varepsilon(j) = y(j)(1 - y(j)) \sum_l^L \varepsilon(l)w(j, l) \quad (3.29)$$

where l is associated with all layers nodes to the right of the current node j .

In our experiment, we used this MLP with backpropagation as learning model and sigmoid as activation function.

Linear Support Vector Machine (L-SVM)

The third classification technique used in this thesis is the Linear SVM (L-SVM) method [50]. Compared with conventional classification methods which minimize the empirical training error, the goal of SVM is to minimize the upper bound of the generalization error by finding the largest margin between the separating hyperplane and the data. The theory of L-SVM is briefly described as follows.

For a given set of instance-label pairs $(x_k, c_k), k = 1, \dots, K, x_k \in \mathbb{R}, c_k \in \{\text{Mitosis}, \text{NonMitosis}\}$, L-SVM solves the following unconstrained optimization problem with loss function $\varepsilon(w; x_k, c_k)$:

$$\min_w \frac{1}{2} w^T w + \alpha \sum_{k=1}^K \varepsilon(w; x_k, c_k) \quad (3.30)$$

$$\varepsilon(w; x_k, c_k) = (\max(0, 1 - c_k w^T x_k))^2$$

where $\alpha > 0$ is penalty parameter.

In the testing phase, we classify an instance x_k as mitosis if $w^T x_k > 0$, and non-mitosis otherwise.

Non-linear Support Vector Machine (NL-SVM)

Using kernel methods, it is possible to build a non-linear SVM (NL-SVM) in a very effective way. In NL-SVM, the separating hyperplanes in the transformed feature space are defined as:

$$z \cdot x_k + b = \pm 1 \quad (3.31)$$

where z is normal to the hyperplanes, x_k is instance, $+1$ is referred to mitosis class, -1 is referred to non-mitosis class and b is the bias which describes the distance of the decision hyperplane from the origin (that is equal to $\frac{b}{\|z\|}$). A NL-SVM can be formulated by following optimization problem:

$$\begin{aligned} \min_{z,b,\xi} \quad & \frac{1}{2} \|z\|^2 + \alpha \sum_{k=1}^K \xi_k \\ \text{subject to} \quad & c_k(z \cdot \mathcal{K}(x_k) + b) \geq 1 - \xi_k \end{aligned} \quad (3.32)$$

where $\xi_k \geq 0$ are real non-negative slack-variables, $\|\cdot\|$ represents the norm of a vector. Using this objective function, a training instance x_k is mapped to a higher dimensional space by a kernel function \mathcal{K} and a user defined penalty parameter $\alpha > 0$. By minimizing $\frac{1}{2} \|z\|^2$, we can get maximum margin between the separating hyperplane and the data. To reduce the number of training errors, the penalty term $\alpha \sum_{k=1}^K \xi_k$ consists of a number of positive-valued slack-variables ξ_k which can be used to construct a soft margin hyperplane.

In this thesis, we train all instances with the radial base function (RBF) kernel as:

$$\mathcal{K}(x_k, x_j) = e^{-\frac{\gamma \|x_k - x_j\|^2}{2\sigma^2}} \quad (3.33)$$

where $\gamma > 0$ is user defined constant.

3.4.7 Experiments and Results

This framework is evaluated on MITOS Aperio and Hamamatsu datasets [1]. The results of candidate detection and classification techniques are compared with GT information provided along with the dataset. The metrics used to evaluate the mitosis detection included: TP, FP, FN, TPR, PPV and FM.

Candidate Detection

We performed candidate detection on red, green and blue channels, and also on BR image. The candidate detection results are ranked according to FM and PPV as shown in Figure 3.12. The green channel detected maximum number of mitosis with a large number of non-mitosis as well, that result in a large number of candidates for classification.

On Aperio training dataset, red channel detects 194 mitosis out of 4386 detected candidates, green channel detects 207 mitosis out of 10103 detected candidates, blue channel detects 202 mitosis out of 9291 detected candidates and BR image detects 196 mitosis out of 8451 detected candidates. On Aperio evaluation dataset, red channel detects 90 mitosis out of 1780 detected candidates, green channel detects 91 mitosis out of 4991 detected candidates, blue channel detects 84 mitosis out of 5006 detected candidates and BR image detects 90 mitosis out of 4351 detected candidates.

Table 3.4: TMC Classification Results on MITOS Aperio Evaluation Dataset (GT = 100).

Classifiers	TP	FN	FP	TPR	PPV	FM
All Features						
DT	63	37	50	63%	56%	59.15%
MLP	65	35	51	65%	56%	60.19%
L-SVM	62	38	38	62%	62%	62.00%
NL-SVM	57	43	25	57%	70%	62.64%
Selected Features						
DT	59	41	26	59%	69%	63.78%
MLP	76	24	52	76%	59%	66.67%
L-SVM	68	32	40	68%	63%	65.38%
NL-SVM	59	41	21	59%	74%	65.56%

On Hamamatsu training dataset, red channel detects 203 mitosis out of 5336 detected candidates, green channel detects 208 mitosis out of 9996 detected candidates, blue channel detects 209 mitosis out of 10851 detected candidates and BR image detects 202 mitosis out of 7956 detected candidates. On Hamamatsu evaluation dataset, red channel detects 89 mitosis out of 2467 detected candidates, green channel detects 84 mitosis out of 4967 detected candidates, blue channel detects 86 mitosis out of 4708 detected candidates and BR image detects 88 mitosis out of 4032 detected candidates.

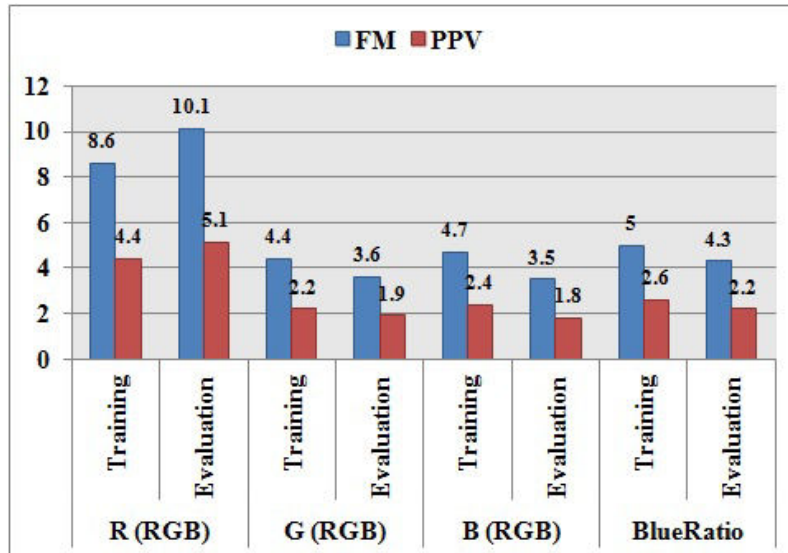
Overall, the green and blue channels detect more mitosis as compared to red channel and BR image, but also more non-mitosis. The red channel detects less non-mitosis than other color channels, on both datasets (Aperio and Hamamatsu). On Aperio dataset, it missed 32 and 10 GT mitosis from training and evaluation datasets, respectively. On Hamamatsu dataset, it missed 23 and 11 GT mitosis from training and evaluation datasets, respectively. Overall, the red channel outperformed during candidate detection with respect to FM and PPV.

Candidate Classification on MITOS evaluation dataset

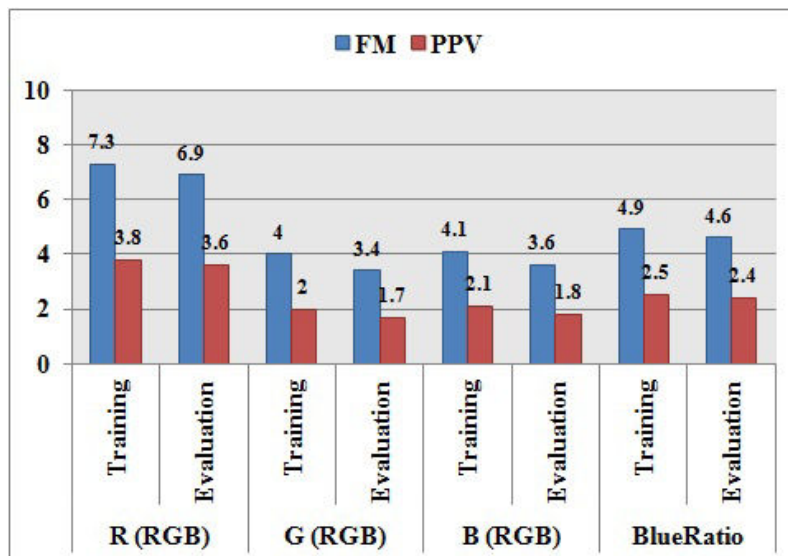
In this experiment, we evaluate this framework on MITOS evaluation dataset. First we use all computed features of each color channel separately for classification of detected candidates as mitosis and non-mitosis. The classification results on Aperio and Hamamatsu evaluation sets are shown in Figure 3.13. Overall, the red channel reports the highest mitosis detection on both datasets. The blue channel detects very few mitosis regions as compared to other color channels. L-SVM and MLP classifiers report higher classifications results using texture features of red channel on both datasets.

Using features from all color channels and BR image, the classification of mitosis regions is improved as shown in Tables 3.4 and 3.5. On Aperio dataset, the MLP classifier reports the highest TPR 65%, but low PPV 56% as well. As compared with other classifiers, the NL-SVM classifier detects few mitosis with also few FP and resulted highest FM 62.64%. When we select features from a set of all textural features computed in all color channels and BR image using feature selection technique (explained in section 3.4.5), the classification results are improved by reduction of FP. The MLP classifier reports the highest TPR 76% and FM 66.67% and NL-SVM reports the highest PPV 74%.

On Hamamatsu dataset, the L-SVM classifier reports the highest TPR 56%, but low PPV 51% as shown in Table 3.5. As compared with other classifiers, the NL-SVM classifier detects few mitosis with also few FP and resulted highest FM 53.89%. When we select

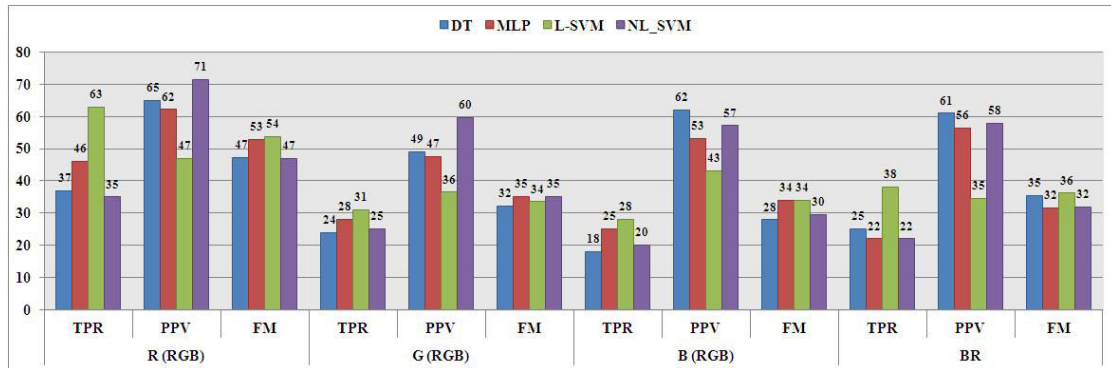


(a) Candidate detection on Aperio Dataset

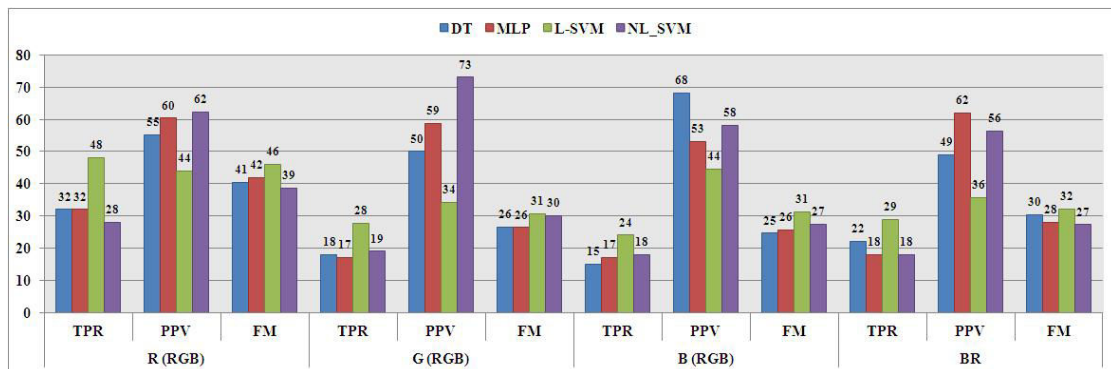


(b) Candidate detection on Hamamatsu Dataset

Figure 3.12: Candidate detection results (FM and PPV metrics) on four color channels.



(a) Classification results using single channel texture features on Aperio dataset



(b) Classification results using single channel texture features on Hamamatsu dataset

Figure 3.13: TMC classification results using single channel texture features with four classifiers.

Table 3.5: TMC Classification Results on MITOS Hamamatsu Evaluation Dataset (GT = 100).

Classifiers	TP	FN	FP	TPR	PPV	FM
All Features						
DT	45	55	23	45%	66%	53.57%
MLP	44	56	21	44%	68%	53.33%
L-SVM	56	44	53	56%	51%	53.59%
NL-SVM	45	55	22	45%	67%	53.89%
Selected Features						
DT	46	54	22	46%	68%	54.76%
MLP	48	52	28	48%	63%	54.55%
L-SVM	59	41	52	59%	53%	55.92%
NL-SVM	47	53	21	47%	69%	55.95%

Table 3.6: Classification Results on MITOS Aperio Full Dataset using 5-Fold CV (GT = 326).

Classifiers	TP	FN	FP	TPR	PPV	FM
All Features						
DT	168	158	78	52%	68%	58.74%
MLP	171	155	77	52%	69%	59.58%
L-SVM	167	159	63	51%	73%	60.07%
NL-SVM	164	162	51	50%	76%	60.63%
Selected Features						
DT	169	157	51	52%	77%	61.90%
MLP	189	137	79	58%	71%	63.64%
L-SVM	183	143	74	56%	71%	62.78%
NL-SVM	171	155	49	52%	78%	62.64%

features from a set of all textural features computed in all color channels and BR image using feature selection technique, the classification results are improved by more detection of mitosis and few FP. The L-SVM classifier reports the highest TPR 59% and NL-SVM reports the highest PPV 69% and FM 55.95%.

Candidate Classification on MITOS full dataset using 5-Fold CV

In this experiment, we evaluate proposed framework on MITOS full dataset using 5-Fold CV. Classification results on Aperio and Hamamatsu dataset are shown in Table 3.6 and 3.7, respectively. On Aperio dataset, the MLP and DT classifiers detect the more number of mitosis but with more FP, as well. The NL-SVM classifier reports the higher FM 60.63% and PPV 76% as compared to other classifiers. On selected features, the MLP classifiers detects more mitosis and overall reports the highest FM 63.64%.

On Hamamatsu dataset, the L-SVM classifier detects the more number of mitosis but with more FP, as well. The NL-SVM classifier reports the higher FM 50.49% and PPV 71% as compared to other classifiers. On selected features, the MLP classifiers detects more mitosis and overall reports the highest FM 51.61%.

Table 3.7: Classification Results on MITOS Hamamatsu Full Dataset using 5-Fold CV (GT = 326).

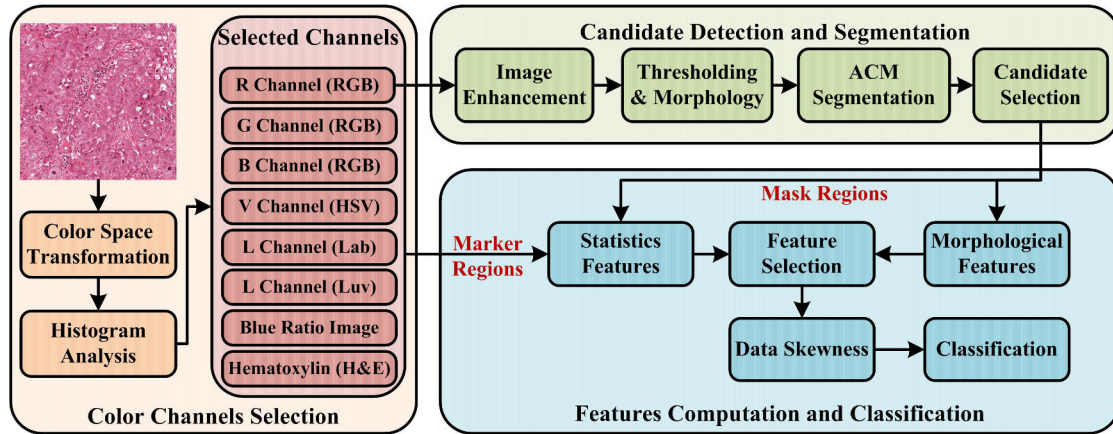
Classifiers	TP	FN	FP	TPR	PPV	FM
All Features						
DT	131	195	68	40%	66%	49.90%
MLP	139	187	89	43%	61%	50.18%
L-SVM	156	170	137	48%	53%	50.40%
NL-SVM	128	198	53	39%	71%	50.49%
Selected Features						
DT	133	193	63	41%	68%	50.96%
MLP	144	182	88	44%	62%	51.61%
L-SVM	158	168	131	48%	55%	51.38%
NL-SVM	131	195	53	40%	71%	51.37%

3.4.8 Discussion

In this framework, we have analyzed the RGB color model for mitosis detection on Aperio and Hamamatsu datasets. The histogram analysis of mitosis and background regions in all color channels showed that red channel has more information for mitosis regions discrimination as compared to other color channels. The candidate detection experiment supports the hypothesis of histogram analysis that red channels detect more mitosis with few non-mitosis candidates. An important finding is that green channel have more information than blue channel, which contain higher absorption response for mitosis nuclei as compared to other tissue components. The textural features computed from red channels contain more information for mitosis regions as compared to textural features computed from other color channels. Using textural features from all color channels, we manage to get FM 62.64% on Aperio dataset and 53.89% on Hamamatsu dataset with NL-SVM classifier. In case of selected textural features, we improve the FM up to 66.67% on Aperio dataset and 55.95% on Hamamatsu dataset.

3.5 Intensity, Textural and Morphology based Mitosis detection in Color images (ITM²C) Framework

The intensity, textural and morphology based mitosis count in color images (ITM²C) framework addresses some shortcomings in the previous framework like too low mitosis detection rate, too many false positives and having poor discrimination between mitosis and non-mitosis regions. More shortcomings are: (1) comprehensive analysis of statistical features in different color models, (2) combining statistical features with morphological features for mitosis discrimination and (3) exploring oversampling methods for balancing the unbalanced training set by increasing the minority class and decreasing the majority class to improve the predictive accuracy of mitosis classification. This improved framework contains (1) a robust multi-channels statistical features computation which integrates segmented nuclei features for variety of color models, (2) nuclei features describing nuclear morphology, pixel information and texture; and (3) studies of oversampling methods to increase minority (mitosis) class size by interpolating between several minority class examples that lie together, which makes classification more robust. This framework consists of three main components: color channels selection, candidate detection and segmentation, and feature computation and classification as shown in Figure 3.14.

Figure 3.14: ITM²C Framework.

3.5.1 Color Channels Selection and Color Deconvolution

In H&E stained images, hematoxylin is a purple-blue dye that binds to the nuclei chromatin and eosin is a pinkish dye that binds to the cytoplasmic components. In order to study the specific information carried by hematoxylin stain, which highlights different cellular structures in the tissue, we separate these two stains using color deconvolution [153]. After separation, we select hematoxylin stained component of image for further processing.

The color representation plays an important role in histological image analysis since it carries usually more information than other features of a given color image [121]. For instance, different color space transformations have been applied to increase the separability between nuclei and non-nuclei during nuclei detection, segmentation and classification. In addition, different color models are proposed to separate a color into more useful components that bring new information to the system. In this framework, our goal is to investigate the various color channels of different color models and select those channels having better pixels and texture information for mitosis detection. We convert RGB images into three other color models namely HSV (more intuitive for human perception), Lab and Luv (uniform color separation). These four color models present common color models studied in histopathology. By doing histogram analysis of mitosis nuclei, non-mitosis nuclei and background regions in all channels of RGB, HSV, Lab and Luv color models, we selected eight color channels which are R(RGB), G(RGB), B(RGB), V(HSV), L(Lab), L(Luv), BR image and Hematoxylin (H&E) image. The histograms of V(HSV), L(Lab), L(Luv) and Hematoxylin(H&E) image are shown in Figure 3.15 and 3.16 whereas the histograms of R(RGB), G(RGB), B(RBG) and BR image are shown in Figure 3.6 and 3.7.

3.5.2 Candidate Detection

This step is similar as explained in section 3.4.3.

3.5.3 Feature Computation and Classification

For each candidate, we extract two sets of quantitative features which are morphological and statistical features. The statistical features are intensity based (first order) and texture based (second order) statistical features.

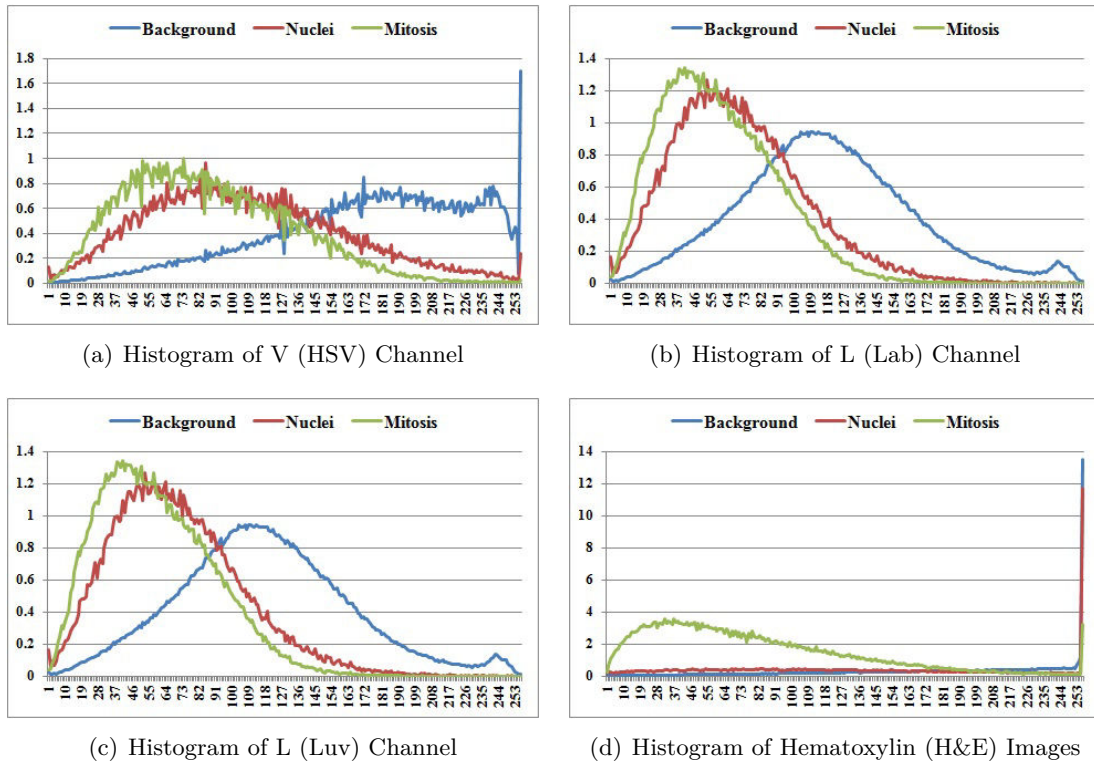


Figure 3.15: Histogram analysis of selected channels on Aperio dataset.

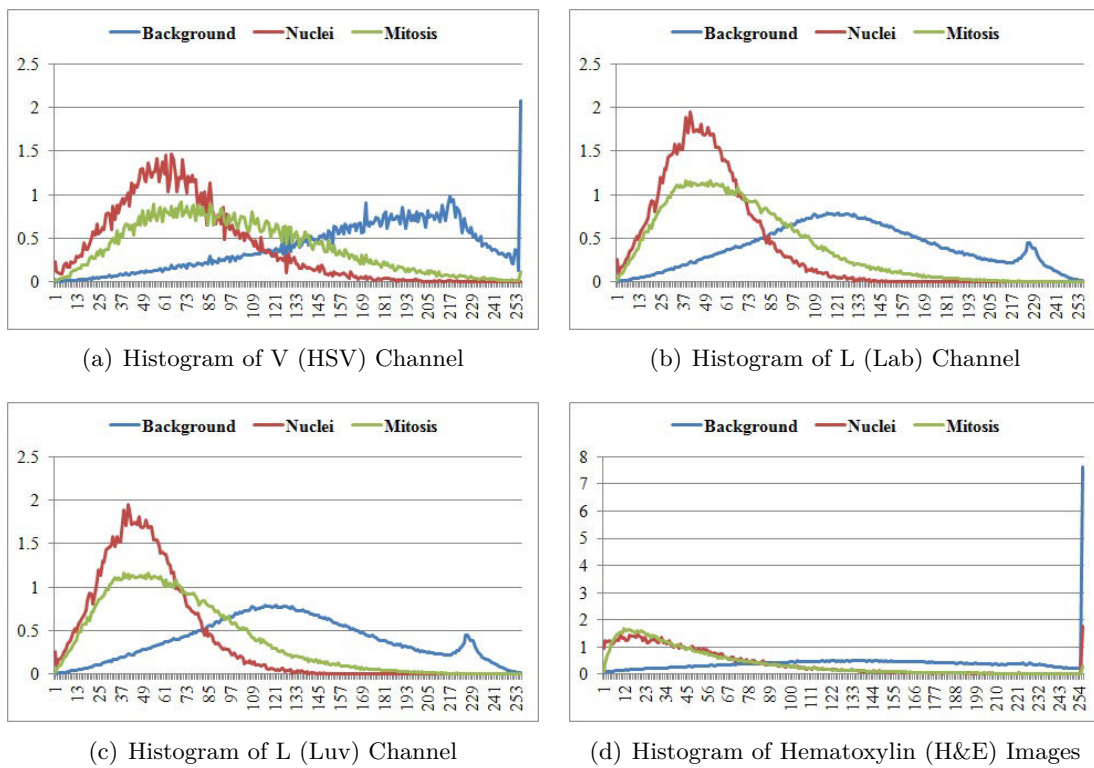


Figure 3.16: Histogram analysis of selected channels on Hamamatsu dataset.

Intensity Features

The intensity based features determine the distribution of grey level values within the candidate regions. These features are computed in all eight selected color channels. The computed features are:

Mean (Mn) :

$$\text{Mn} = \frac{\sum_i I(i)}{U} \quad (3.34)$$

where $I(i)$ is grey level values and U is the number of pixels in the segmented regions.

Median (Md) : describes the central tendency. After arranging all the grey level values in ascending order, the middle value is median of the candidate region.

Standard Deviation (SD) : represents the variation of grey level value in comparison with the mean value (Mn).

$$\text{SD} = \sqrt{\frac{(\sum_i I(i) - \text{Mn})^2}{U}} \quad (3.35)$$

Skewness : describes the degree of histogram asymmetry around the mean.

$$\text{Skewness} = \frac{1}{U} \frac{(\sum_i I(i) - \text{Mn})^3}{\text{SD}^3} \quad (3.36)$$

Kurtosis : describes the sharpness of the grey level histogram.

$$\text{Kurtosis} = \frac{1}{U} \frac{(\sum_i I(i) - \text{Mn})^4}{\text{SD}^4} \quad (3.37)$$

where U is total number of pixels in the segmented regions.

Morphological Features

Besides intensity and texture features of candidate regions, various shape and geometrical features are computed for candidate classification as mitosis or non-mitosis. These features are:

Area: is computed by counting the number of pixels in the segmented region.

$$\text{Area} = \sum_i I^B(i) \quad (3.38)$$

where $I^B(i)$ is the binary mask of a object consisting of ones within the object and zeros elsewhere.

Perimeter : is measured the distance around the boundary of the object, where boundary pixels are 8-connected:

$$\text{Perimeter} = \sum_{i=1}^B \sqrt{(x(i+1) - x(i))^2 + (y(i+1) - y(i))^2} \quad (3.39)$$

where x and y are the x- and y-coordinates of the B boundary pixels.

Roundness : measures shape irregularity as:

$$\text{Roundness} = \frac{\text{Perimeter}^2}{\text{Area}} \quad (3.40)$$

Elongation : is computed as the ratio of the largest axis and the smallest axis. Its value is greater or equal to 1.

$$\text{Elongation} = \frac{\text{Major Axis}}{\text{Minor Axis}} \quad (3.41)$$

Equivalent spherical perimeter : is the equivalent perimeter of the hypersphere of the same size.

These five morphological features reflect the phenotypic information of mitosis. Utilizing pixel intensity information of the selected color channels including BR and Hematoxylin image, we compute five intensity features of each segmented regions in all selected color channels. Using mask from candidate segmentation, HC and RL features are also computed for all selected color channels as explained in section 3.4.4.

Handling Imbalanced Dataset

There is a high degree of imbalance in the dataset as mitosis candidates are very few in number as compared to the non-mitosis candidates. In case of imbalanced dataset, the class boundary learned by the standard machine learning classifiers is biased towards the majority class resulting in a high false negative rate [94]. It is of utmost importance to balance the class distribution in the training set before training a classifier. To handle imbalanced dataset, we perform two things; (1) down sampling of non-mitosis instances and (2) oversampling of mitosis instances.

First, we remove borderline instances (having higher probability of being classified incorrectly) from non-mitosis instances. Second, we apply synthetic minority oversampling technique (SMOTE) [32] on mitosis instances of training set to increase number of mitosis in order to reduce bias of classifiers towards non-mitosis class. This oversampling approach creates extra synthetic training data for minority class by operating in feature space rather than data space. Depending upon the amount of over-sampling required, neighbors from the k nearest neighbors are randomly chosen. Two neighbors from the five nearest neighbors are chosen and one sample is generated in the direction of each. SMOTE provides more related minority class samples to learn from, thus allowing a classifier more coverage of the minority class due to broader decision regions.

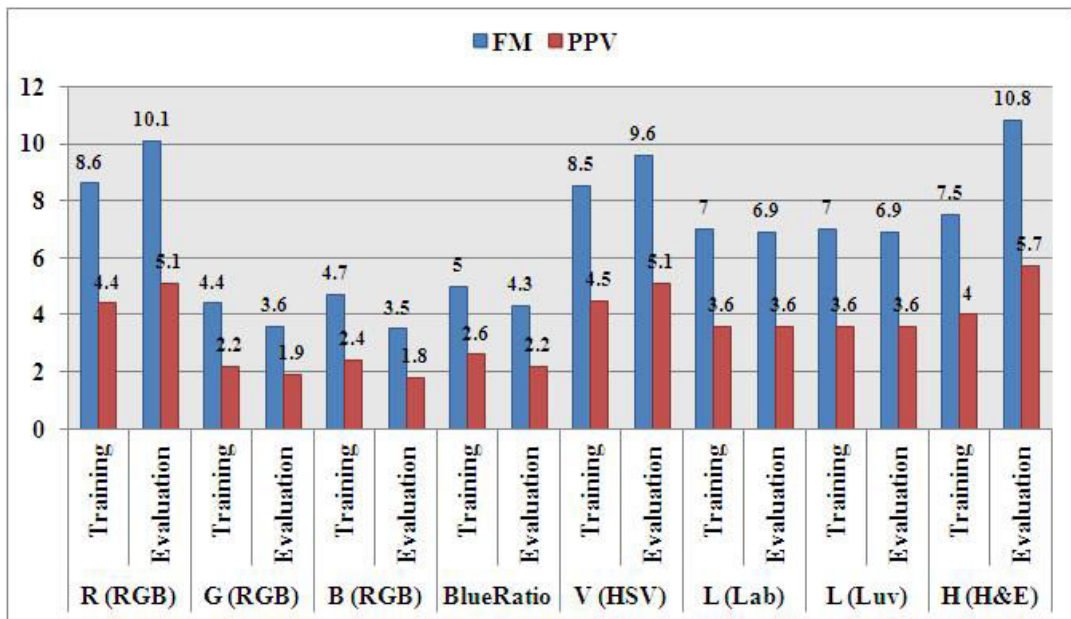
3.5.4 Experiments and Results

Candidate Detection

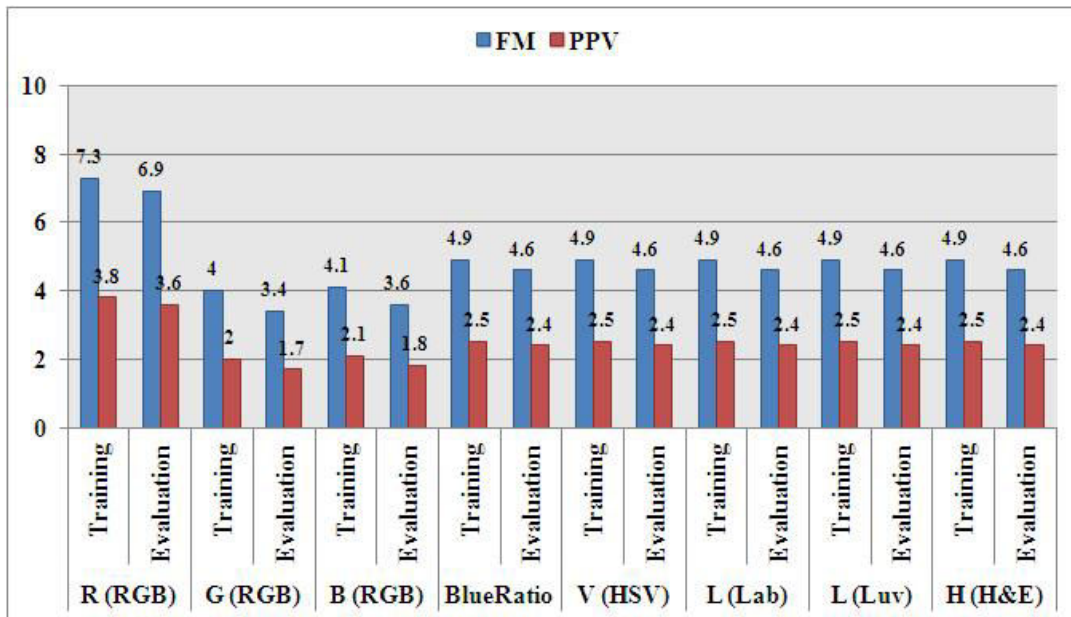
We performed candidate detection on selected 8 color channels. The candidate detection results are ranked according to FM and PPV as shown in Figure 3.17. Overall, the R(RGB) channel has higher FM on both datasets (Aperio and Hamamatsu). The G(RGB) channel detected maximum number of mitosis with a large number of non-mitosis as well, that result in a large number of candidates for classification. On Aperio dataset, R(RGB) and V(HSV) have almost similar FM and PPV, but R(RGB) channel has better FM and PPV than V(HSV) channel on Hamamatsu dataset. The R(RGB) channel outperformed overall during candidate detection with respect to FM and PPV.

Candidate Classification using features from Individual Color Channel

In this experiment, we evaluate this framework on MITOS evaluation dataset using computed features from individual color channels. The classification results on Aperio and Hamamatsu evaluation sets are shown in Figure 3.18. Overall, the R (RGB) and V (HSV) channels report the highest mitosis detection on both datasets. The green channel detects very few mitosis regions as compared to other color channels. The L-SVM classifier reports higher classifications results with more TP and FP. The NL-SVM classifier reports higher PPV with few FP but few TP as well.

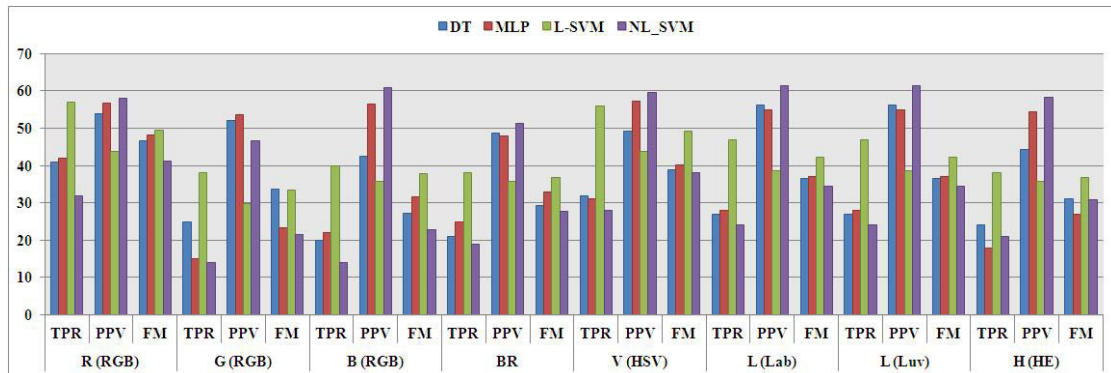


(a) Candidate detection on Aperio Dataset

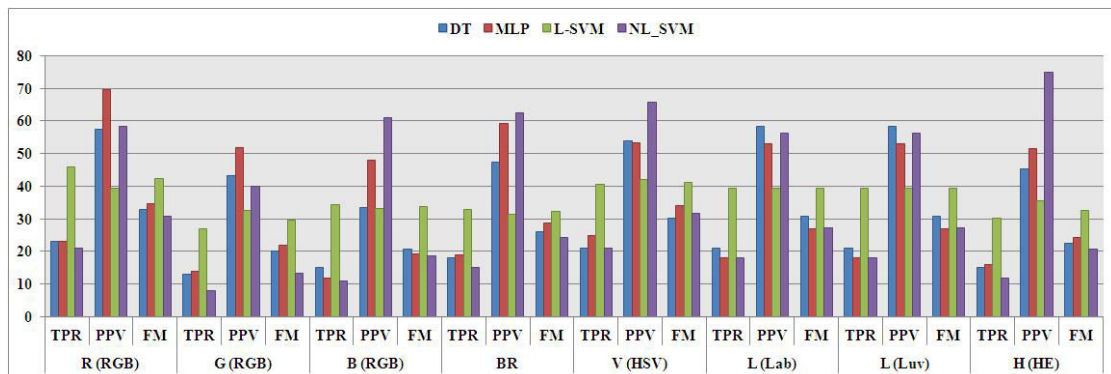


(b) Candidate detection on Hamamatsu Dataset

Figure 3.17: Candidate detection results on selected eight color channels.



(a) Classification result using single channel features on Aperio dataset.



(b) Classification result using single channel features on Hamamatsu dataset.

Figure 3.18: ITM²C classification results using single channel features with four classifiers.

Table 3.8: ITM²C Classification Results on MITOS Aperio Evaluation Dataset (GT = 100).

Classifiers	TP	FN	FP	TPR	PPV	FM
All Features						
DT	65	35	27	65%	71%	67.71%
MLP	68	32	31	68%	69%	68.34%
L-SVM	72	28	38	72%	66%	68.57%
NL-SVM	58	42	12	58%	83%	68.24%
Selected Features						
DT	67	33	25	67%	73%	69.79%
MLP	66	34	23	66%	74%	69.84%
L-SVM	74	26	30	74%	71%	72.55%
NL-SVM	59	41	11	59%	84%	69.41%

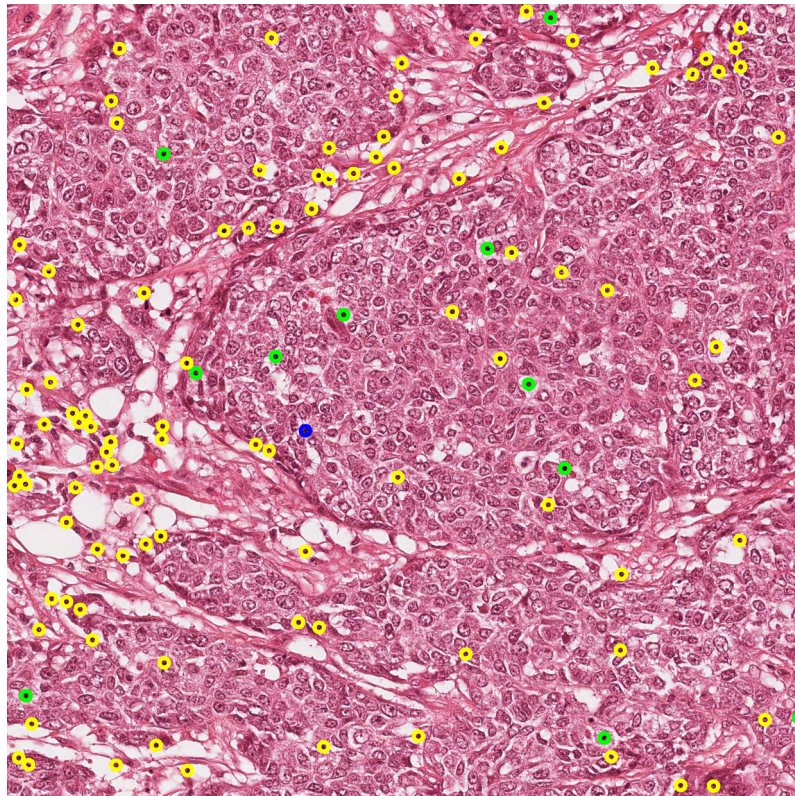
Candidate Classification using Features from Selected Color Channels

Using features from all selected color channels, the classification of mitosis regions is improved as shown in Table 3.8 and 3.9. On Aperio dataset, the L-SVM classifier detects more TP and also more FP as well. The L-SVM classifier with all computed features reports highest FM 68.57%, but also low PPV 66%. As compared with other classifiers, the NL-SVM classifier detects few mitosis with also few FP and results FM 68.24%. When we select features from a set of all computed features in selected color channels using feature selection technique (explained in section 3.4.5), the classification results are improved by detection of more mitosis and less FP. The L-SVM classifier reports the highest TPR 74% and FM 72.55% and NL-SVM classifier reports the highest PPV 84%.

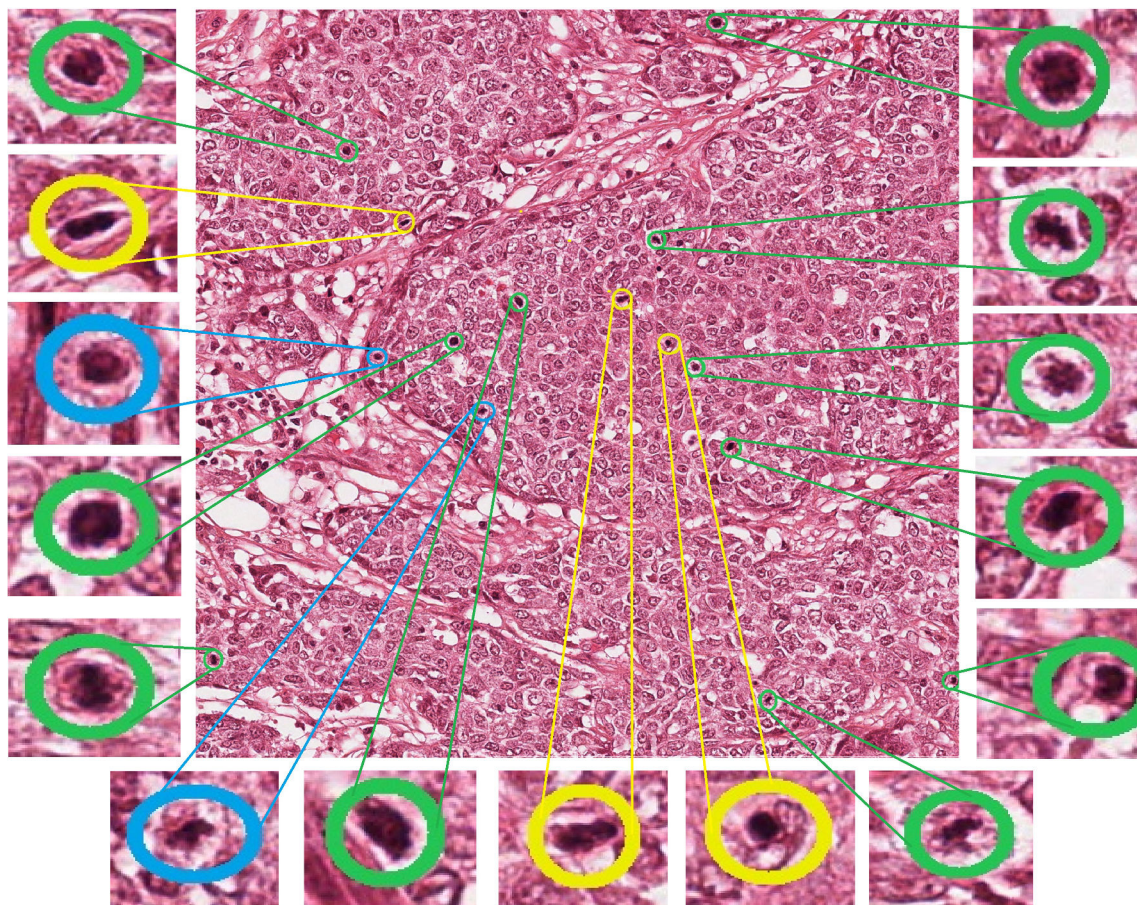
On Hamamatsu dataset, the L-SVM, MLP and DT classifiers detect more mitosis and more FP as well. The L-SVM classifier with all computed features reports highest FM 61.31%, but also low PPV 62%. As compared with other classifiers, the NL-SVM classifier detects few mitosis (53) with also few FP (20) and results FM 61.27%. When we select features from a set of all computed features in selected color channels, the L-SVM classifier reports the highest TPR 63% and FM 64.62% and NL-SVM classifier reports the highest PPV 74%. The visual results of candidate detection and candidate classification on Aperio and Hamamatsu sample image are shown in Figures 3.19 and 3.20, respectively, (green circle represents TP, blue circle represent FN and yellow circle represent FP).

Region vs Patch based features in ITM²C Framework

Besides segmented region based textural features, we also compute texture features on different patch size of detected candidates. The different patch sizes used in feature computation are shown in Table 3.10. The results of classification on different patch size features and region based features with L-SVM classifier are shown in Figure 3.21. In case of Aperio dataset, features computed on patch size $17.192\mu m \times 17.192\mu m$ show maximum FM and TPR as compared with region and other patch sizes. By reducing the patch size, we have fewer FP but fewer TP as well. With features computed on patch size $17.192\mu m \times 17.192\mu m$, we manage to achieve TPR 73%, PPV 76% and FM 74.49%. On Hamamatsu dataset, features computed on patch size $13.638\mu m \times 13.652\mu m$ reports maximum FM and TPR as compared with region and other patch sizes. With this patch size features, we manage to achieve TPR 71%, PPV 58% and FM 64.84%.

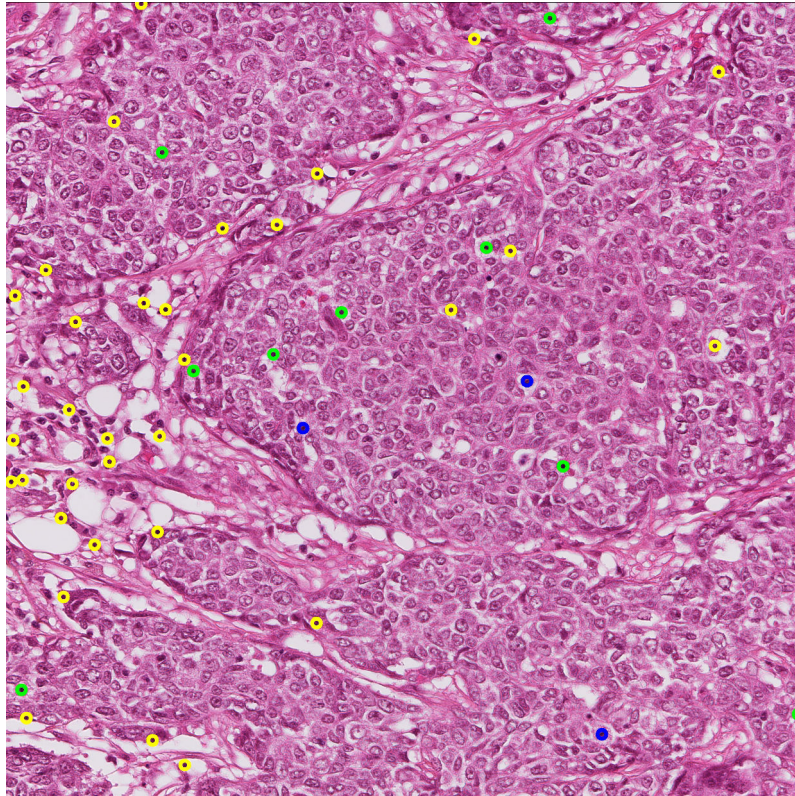


(a) Candidate Detection on Aperio Image

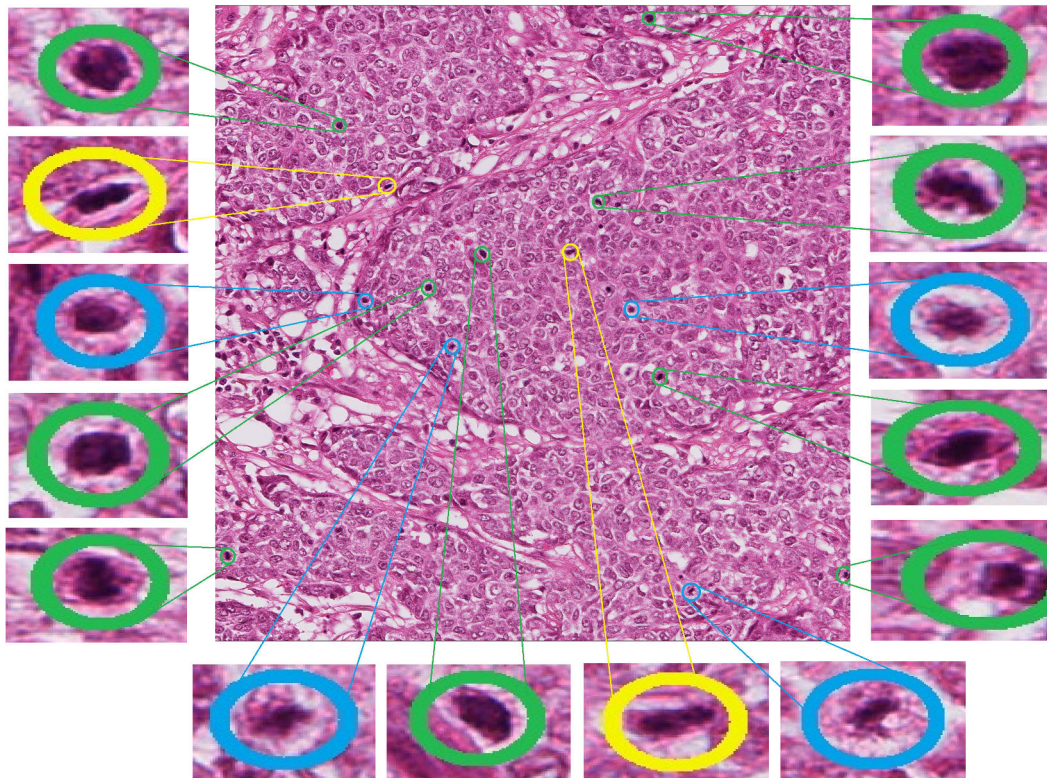


(b) Candidate Classification on Aperio Image

Figure 3.19: Visual results of mitosis detection on Aperio images (green circles represent TP, blue circles represent FN and yellow circles represent FP).

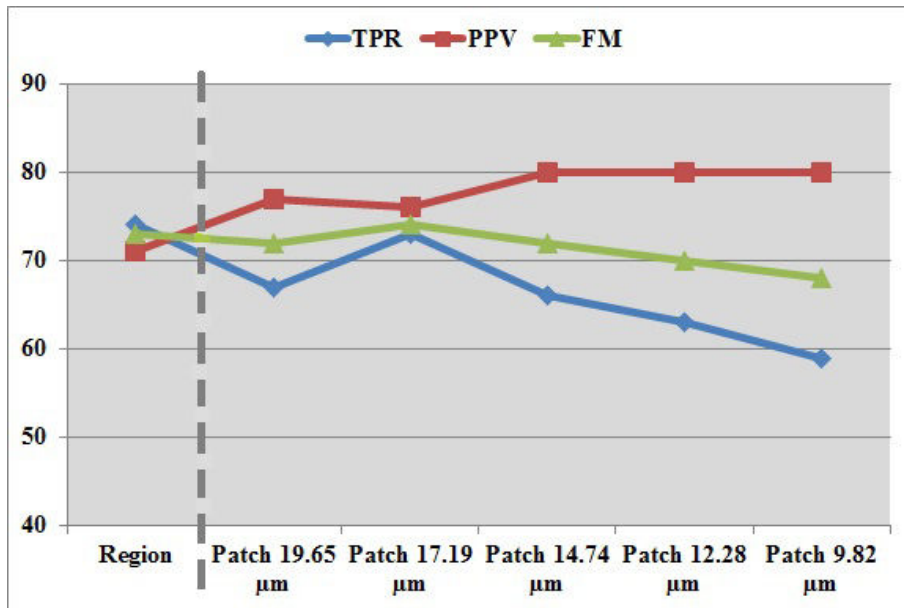


(a) Candidate Detection on Hamamatsu Image

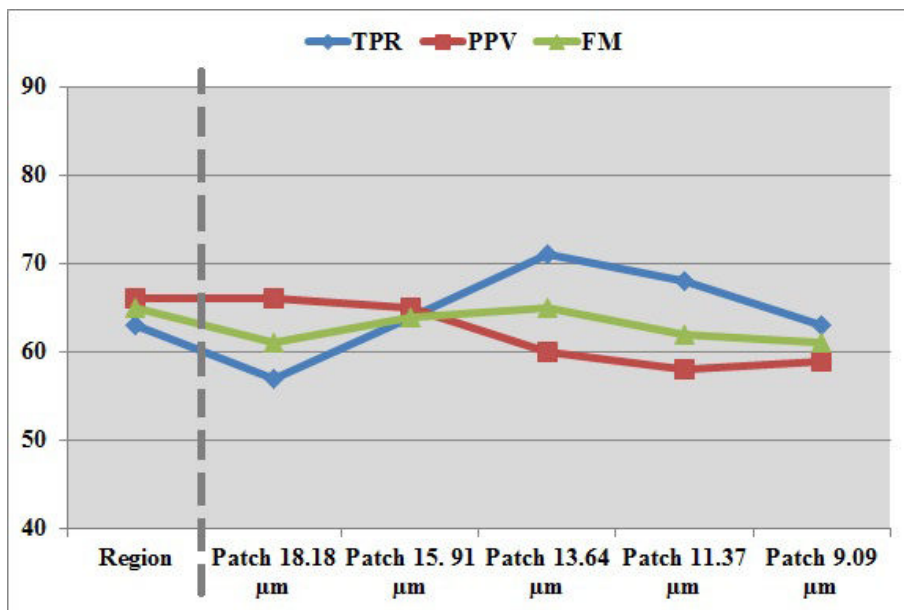


(b) Candidate Classification on Hamamatsu Image

Figure 3.20: Visual results of mitosis detection on Hamamatsu images (green circles represent TP, blue circles represent FN and yellow circles represent FP).



(a) Classification Results on Aperio Dataset



(b) Classification Results on Hamamatsu Dataset

Figure 3.21: Classification results of region and patches based features with L-SVM classifier.

Table 3.9: ITM²C Classification Results on MITOS Hamamatsu Evaluation Dataset (GT = 100).

Classifiers	TP	FN	FP	TPR	PPV	FM
All Features						
DT	60	40	37	60%	62%	60.91%
MLP	60	40	38	60%	61%	60.61%
L-SVM	61	39	38	61%	62%	61.31%
NL-SVM	53	47	20	53%	73%	61.27%
Selected Features						
DT	61	39	34	61%	64%	62.56%
MLP	60	40	31	60%	66%	62.83%
L-SVM	63	37	32	63%	66%	64.62%
NL-SVM	55	45	19	55%	74%	63.22%

Table 3.10: Patch sizes in pixels and μm on the Aperio and Hamamatsu dataset.

Patch sizes in pixels	Aperio Dataset (μm)	Hamamatsu Dataset (μm)
Patch 80×80	19.648×19.648	18.184×18.202
Patch 70×70	17.192×17.192	15.911×15.927
Patch 60×60	14.736×14.736	13.638×13.652
Patch 50×50	12.28×12.28	11.365×11.377
Patch 40×40	9.824×9.824	9.092×9.101

3.5.5 Discussion

ITM²C framework is proposed to count mitosis in H&E stained breast cancer histological images. This framework is based on multi-channel statistical and morphological features. Initially histogram analysis is performed on different color channels of various color spaces including BR and Hematoxylin (H&E) images to select the relevant color channels for mitosis discrimination. We perform candidate detection on all the selected (eight) color channels. The candidate detection results conclude that R(RGB) and V(HSV) are reported to have maximum FM and PPV on both Aperio and Hamamatsu datasets. This is also validated from histogram analysis of mitosis and background regions where both regions are well separated peaks.

In experiment of candidate classification using single channel statistical features, R (RGB) channel features show more discrimination as compared to other color channels features. By comparing classifiers, L-SVM classifier detects high number of mitosis with high number of FP as well. Overall, L-SVM classifier using R(RGB) computed features reports the highest FM 50% and 42% on Aperio and Hamamatsu datasets, respectively. The V(HSV) channel features shows slightly less performance as R(RGB) channels features. As compared to single channel patch features computed in TMC framework, the single channel region features computed in ITM²C framework reports low performance.

By combining all selected channels based regions features with morphological features, the ITM²C framework manages to improve the mitosis detection on Evaluation dataset from 66.67% to 71.29%. The improved framework detects more mitosis and less FP as compared to previous framework.

To compare the contribution of region and patch based features, we have made a study of mitosis classification using region and different patch sizes based features. By taking the

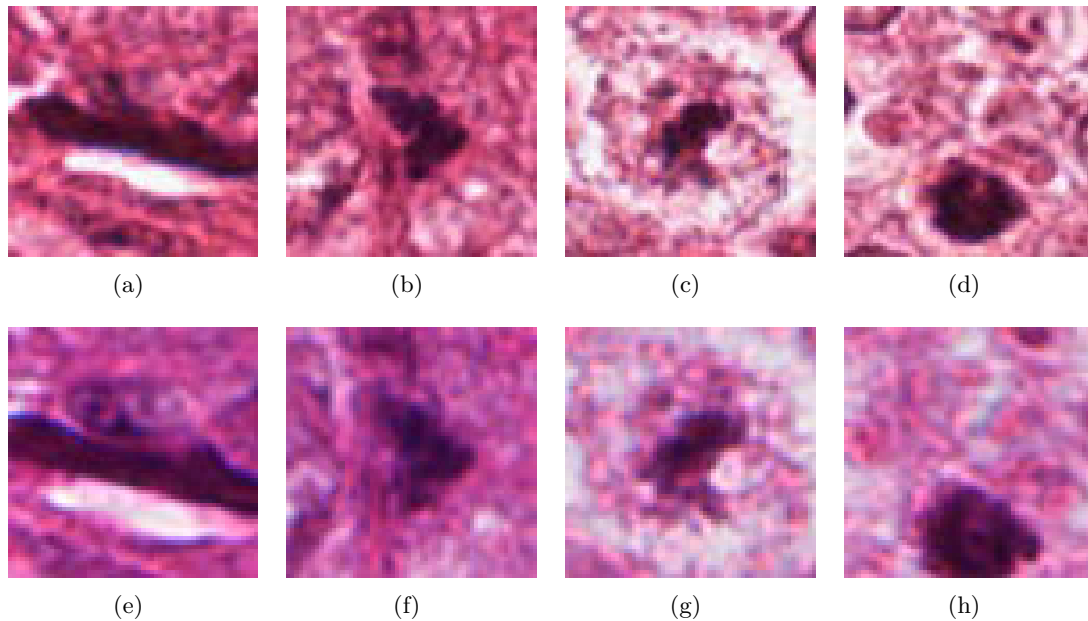


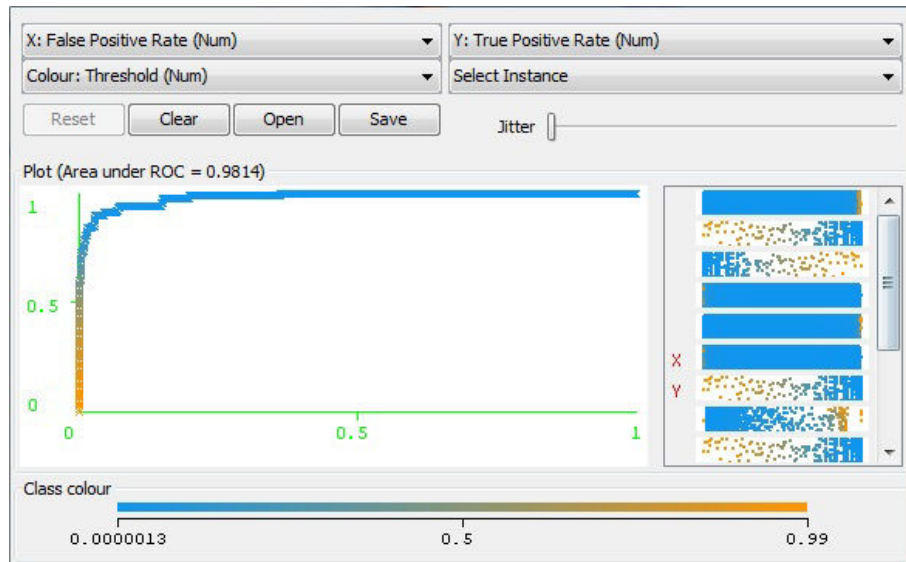
Figure 3.22: Aperio (first row) and Hamamatsu (second row) patches of mitosis nuclei on which texture features are computed.

advantage of neighboring regions of each candidate, the patch based features carries more discriminate information for mitosis detection. It is interesting to see that best results are achieved with different patch sized for both datasets (Aperio and Hamamatsu). The patch size $17.19\mu m \times 17.19\mu m$ gives best result on Aperio dataset and the patch size $13.64\mu m \times 13.65\mu m$ gives best result on Hamamatsu dataset. Examples of different patches of mitosis in Aperio and Hamamatsu datasets are shown in Figure 3.22. An important finding is that texture is different in Aperio and Hamamatsu datasets not only inside mitosis regions but also in neighboring mitosis regions.

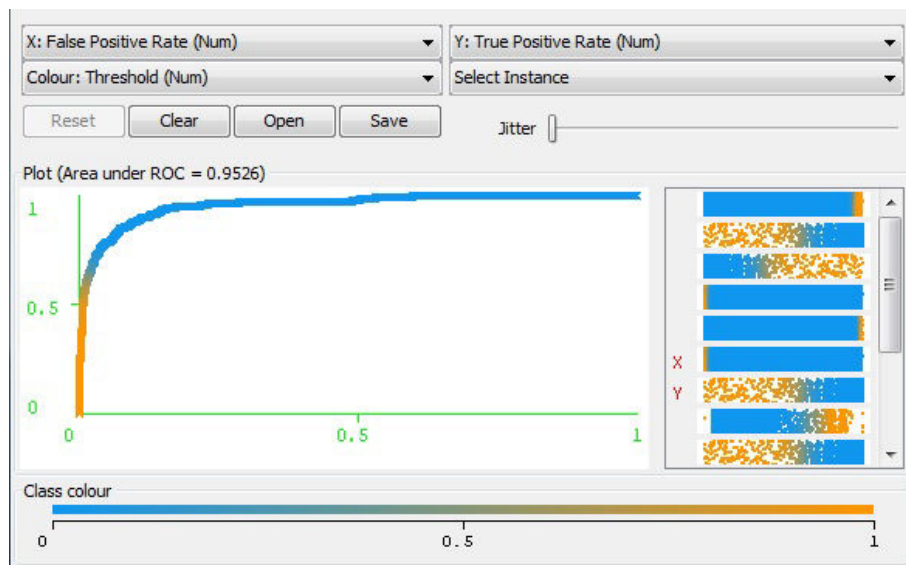
From the classifiers point of view, NL-SVM classifier always detects very few FP and results in high PPV. The L-SVM classifier always detects maximum number of mitosis, but with high number of FP as well. MLP classifier shows good results on selected features.

One of the parameters that mostly affect our experiments is the unbalanced training set having a huge number of non-mitosis compared to the small number of mitosis. Indeed, most classifiers are biased toward non-mitosis, which resulted high number of FP and low number of TP. Handling imbalanced dataset, by over-sampling of mitosis using SMOTE and non-mitosis cleaning, shows improvement in mitosis detection. The SMOTE method effectively forces the decision region of the minority class to become more general, thus eventually it reduces the bias of non-mitosis class and results in classification improvement. Figure 3.23 illustrates the ROC curve obtained with patches features. Figure 3.24 illustrates the margin curve between the probability predicted for mitosis class and the highest probability predicted for the non-mitosis class. This clearly demonstrates that our proposed framework results in an improved ability to distinguish mitosis from other objects.

To the best of our knowledge, MITOS is the only available de facto gold standard dataset of both multispectral and color images. It not only provides basis for comparison between our proposed framework and other previous frameworks, but also for comparison between multispectral and color images. Using this dataset, IPAL CNRS laboratory organized a contest during ICPR 2012 [150]. They also provided the performance metrics to evaluate the framework for mitosis count. The results of this framework has been submitted in this

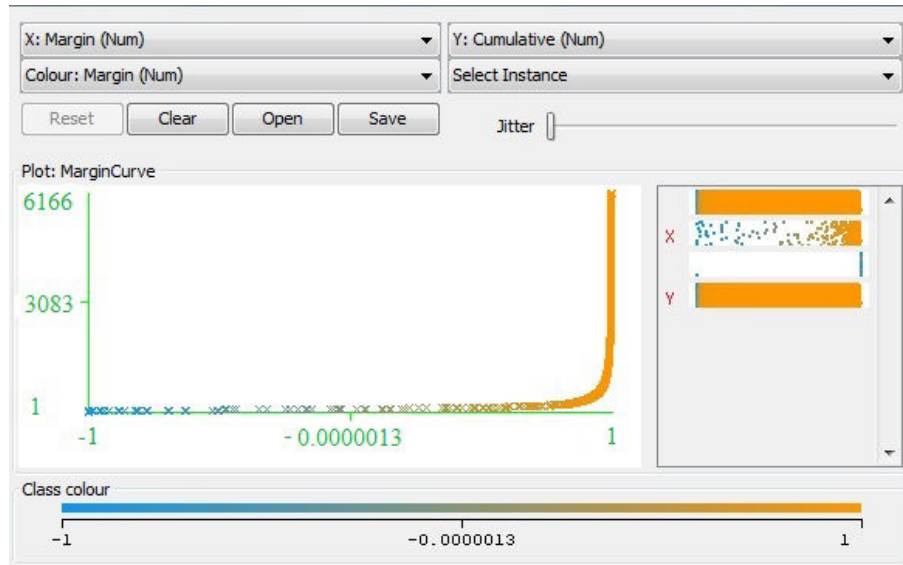


(a) Aperio Dataset

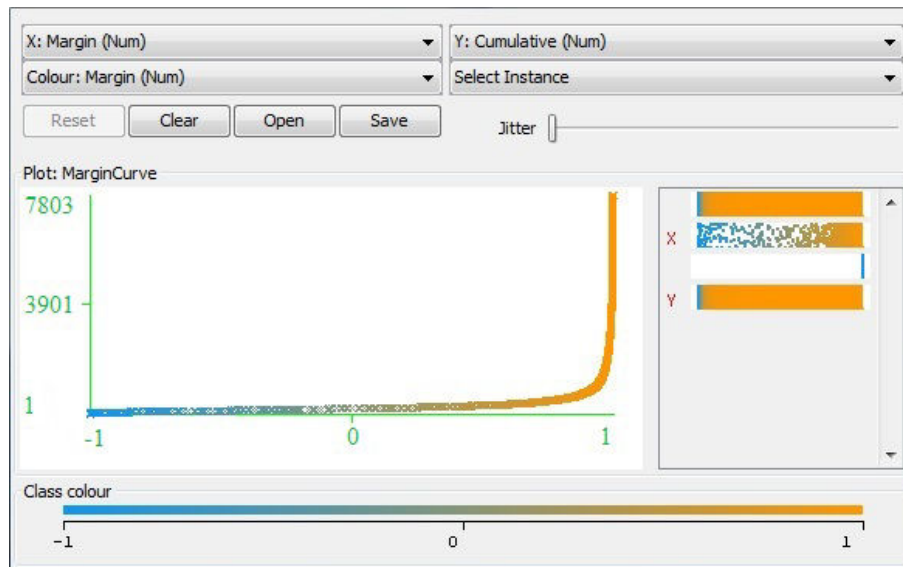


(b) Hamamatsu Dataset

Figure 3.23: The ROC curves of classification result using patch based features with L-SVM classifier.



(a) Aperio Dataset



(b) Hamamatsu Dataset

Figure 3.24: The margin curve illustrating the prediction margin between mitosis and non mitosis class.

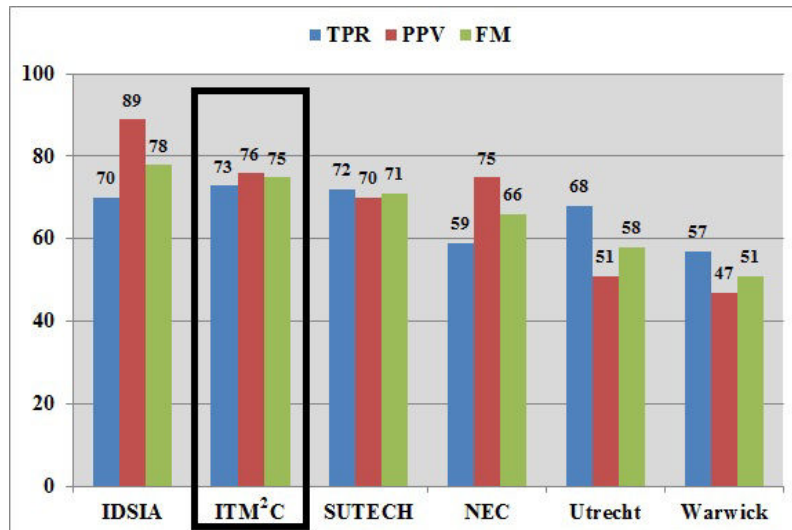


Figure 3.25: Comparison of ITM²C framework results with MITOS contest result on Aperio Dataset. IDSIA: Dalle Molle Institute for Artificial Intelligence Research [35], SUTECH: Shiraz university of technology, NEC: NEC Corporation [113], Warwick [87].

contest and ranked second on both Aperio and Hamamatsu dataset. The comparison of results on Aperio and Hamamatsu dataset are shown in Figures 3.25 and 3.26.

The missed mitosis normally are small in size and exhibit faint nuclear material as shown in Figure 3.27. These mitosis have also been missed by all the methods during MITOS contest 2012 [150]. A few examples of FPs are shown in Figure 3.28. The numerous FP are lymphocytes and artifacts that have dark regions and look like mitosis. In addition, some FP is in non-tumor region. One possibility of improvement can be selection of tumor region for mitosis detection.

The proposed frameworks are developed using Insight Segmentation and Registration Toolkit (ITK) [77] and Weka [139]. The candidate detection, segmentation and features computation are developed in ITK and feature selection and classification are developed using Weka.

3.6 MICO Platform Prototype

The Cognitive Microscope (MICO) project, funded by French National Research Agency (ANR), aims at enhancing the diagnosis process through a synergy between knowledge, context, cognition and experience based on a user-centered approach to provide visual prognosis assistance to pathologists [151, 106, 141]. The global architecture of MICO platform is shown in Figure 3.29. This project, launched in February 2011 for duration of three years, aims at developing a pertinent and robust tool for breast cancer grading, in order to give pathologists a second opinion on the state of a patient. MICO project raised the interest of both industrials and medical institutes. Figure 3.30 shows the list of the partners involved in this project.

3.6.1 Evaluation of ITM²C frameworks in MICO Platform

MICO platform addresses two criteria of NGS, i.e., mitosis count and cyto-nuclear atypia (CNA). Automated evaluation of mitosis count and CNA on WSI requires heavy image

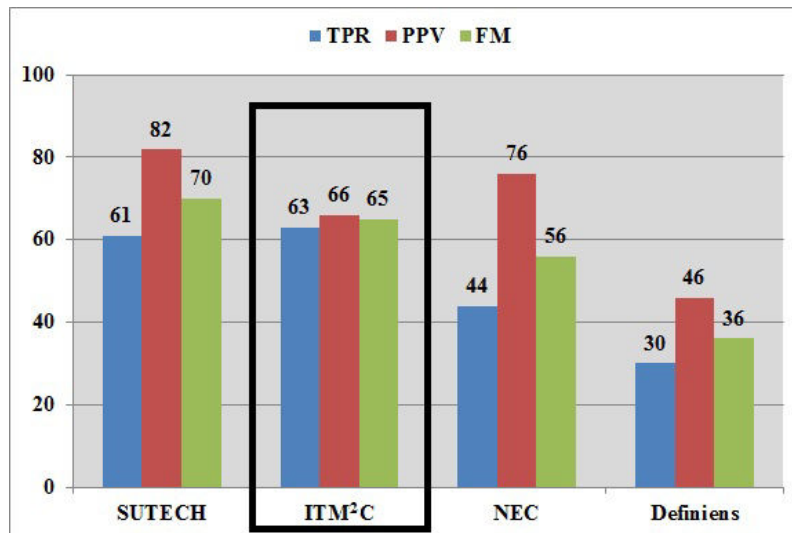


Figure 3.26: Comparison of ITM²C framework results with MITOS contest result on Hamamatsu Dataset. NEC: NEC Corporation [113].

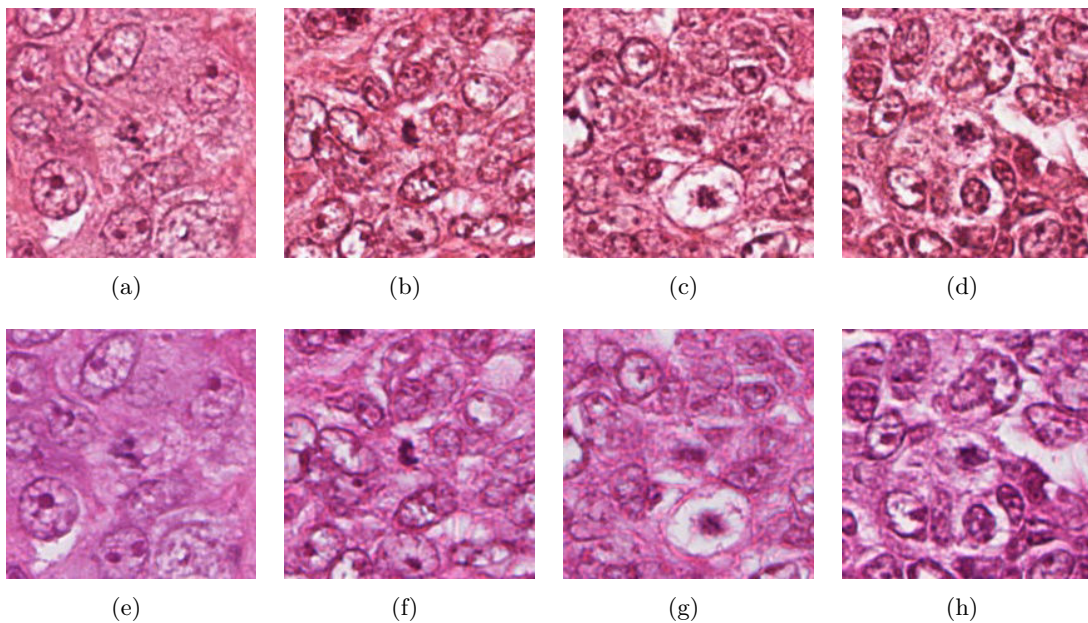


Figure 3.27: Some examples of FN. The missed mitotic nuclei are located in the center of each image. First row images (a-d) from Aperio and second row images (e-h) are from Hamamatsu Dataset.

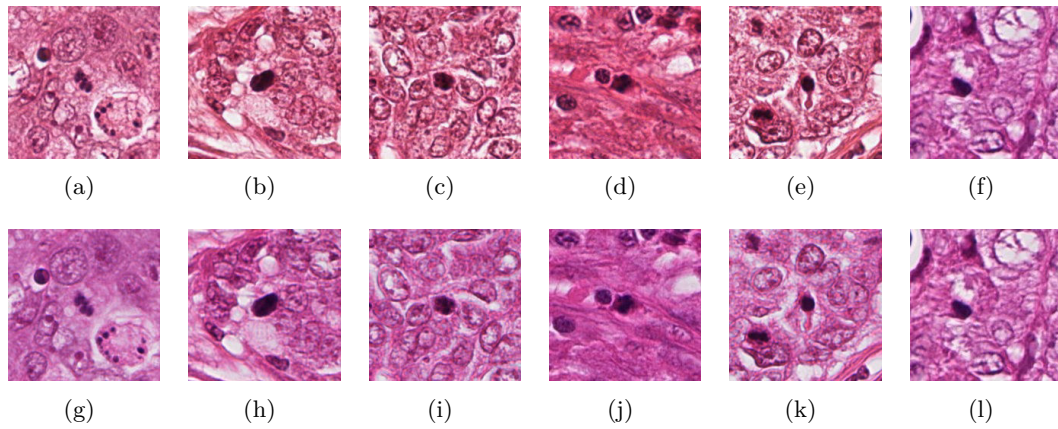


Figure 3.28: Some examples of FPs. The false mitotic nuclei are located in the center of each image. First row images (a-f) from Aperio and second row images (g-l) are from Hamamatsu Dataset.

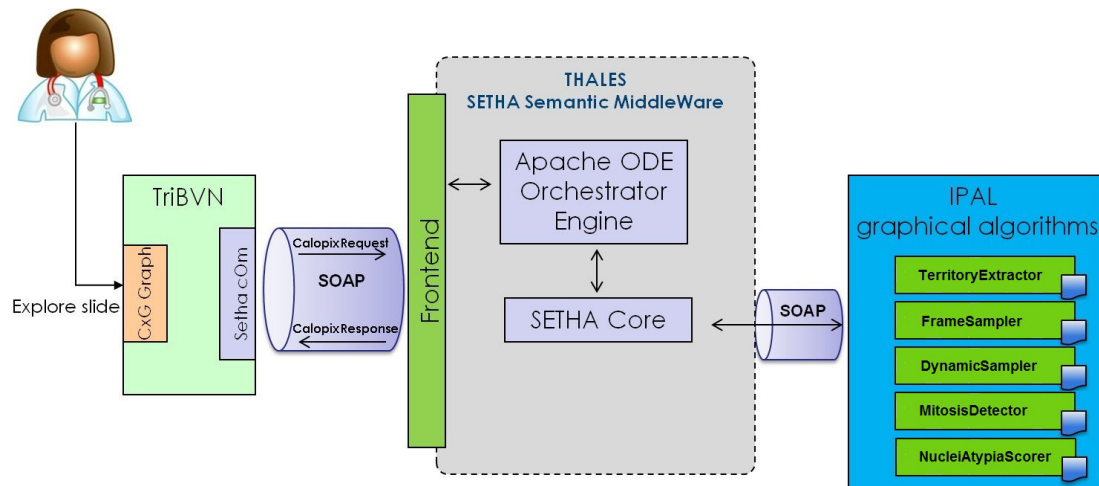


Figure 3.29: Architecture of MICO 2.0.







	IPAL — University Joseph Fourier Image Processing, Ontologies, Reasoning
	Pitié-Salpêtrière Hospital Pathologists, Expertise on breast cancer
	University Pierre and Marie Curie Contextual graphs, GPU
	TRIBVN User interface
	Agfa HealthCare Ontologies, Reasoning
	Thales Semantic middleware

Figure 3.30: MICO ANR TecSan project partners.

processing computations. However, it is not possible to compute the mitosis count and CNA on a ROI within reasonable time. Strategies are proposed to overcome the processing power limitations. As the histopathology expert doesn't need to watch every part of the WSI in order to grade it, MICO aims to understand and reproduce this expert behavior during automated grading. For mitosis count, we employ stereology flow for selection of frames in ROI as shown in Figure 3.31. First, the whole slide image is observed by a histopathologist, for the slide territories corresponding to tumor areas to be annotated using Calopix user interface [4]. Then the relevant territories are extracted from Calopix information storage system, for them to be split into rectangles called HPF frames. In the tumor, we are looking for the area having the highest concentration of mitotic cells. To perform this search faster, we operate a sampling on the set of frames covering the tumor (see Figure 3.32): the frames are grouped into blocks of 3×3 frames (see Figure 3.33). In each 3×3 block, only the top left frame is analyzed for mitosis detection. This sampling of one frame analyzed out nine give us a broad picture of the concentration of mitosis in each area of the tumor. To select the area having highest number of mitosis, we now consider blocks of 4×4 frames such that the frames at each corner of a block have already been analyzed for mitosis detection. We compute the sum of mitosis detected so far in each 4×4 block, and we select the block (or the blocks) having highest number of mitosis for further analysis. In this block of 16 frames, the four frames at the corners of the block have already been analyzed. We now proceed to detection of mitosis on the 12 remaining frames of the block. So all the 16 frames of the block are now analyzed. According to Nottingham Grading System, mitotic count is given for 10 consecutive frames. To comply with this definition, we select from our block of 16 frames the 10 frames with highest number of mitosis. The resulting mitotic count for the tumor territory will be equal to the sum of mitosis detected on these 10 frames. Of course, in case we had analyzed more than one 4×4 block, the final mitotic count will be the highest one among all the 4×4 blocks.

This whole strategy is split in several algorithm definitions that are properly designed towards MICO platform realisation. These algorithms are designed to communicate using Extensible Markup Language (XML) [24], a self-describe language, easy to read and to maintain. These algorithms are:

- i. **TerritoryExtractor** : extracts a labelled polygon from Calopix WFML file.

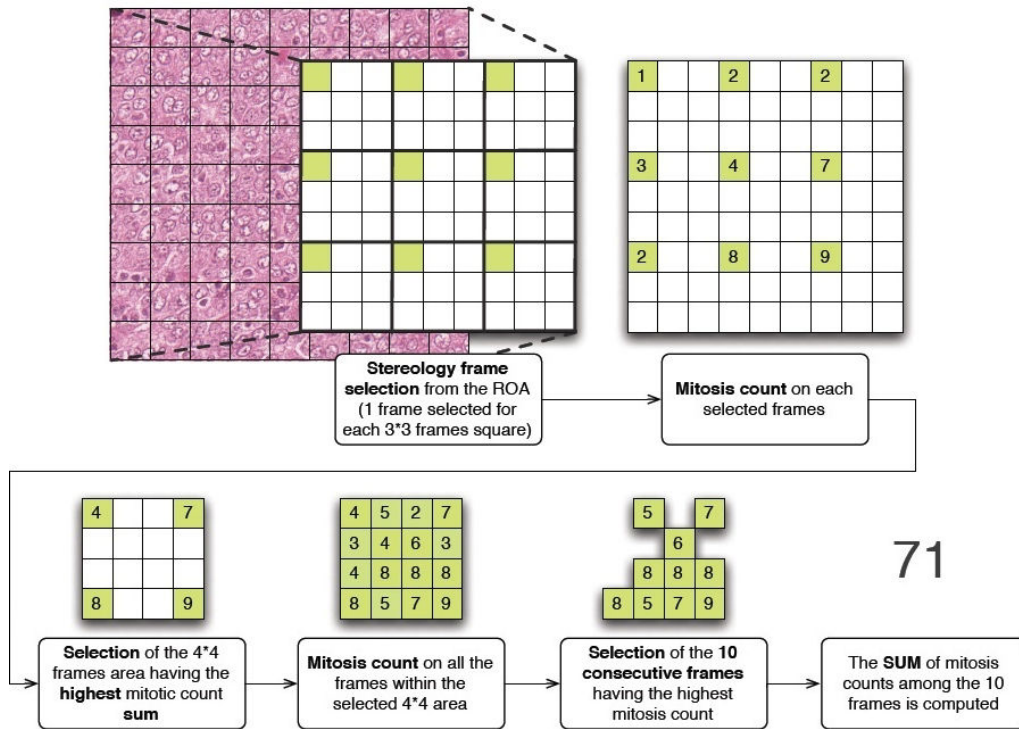


Figure 3.31: Stereology flow used for mitosis score over a ROI.

- ii. **FrameGenerator**: cuts a polygon into frames as shown in Figure 3.32.
- iii. **FrameSampler 3×3** : generates the blocks of 3×3 frames and launches the detection of mitosis on the top left frame of each 3×3 block.
- iv. **FrameSampler 4×4** : generates blocks of 4×4 frames such that the four frames on the corner of a block have already been analyzed for detection of mitosis; computes the number of mitosis detected on each 4×4 block; selects the 4×4 block having the highest number of mitosis; launch the detection of mitosis on the 12 frames not yet analyzed of the block so that all the 16 frames of the 4×4 block will be analyzed; among the 16 frames of the block, select the 10 frames having the highest number of mitosis and add these number of mitosis together: this is the resulting mitotic count for the whole tumor area.
- v. **ITM²C Framework**: count the mitosis on given frame.
- vi. **MitoticScorer**: selects 10 frames having the highest number of mitosis in 4×4 block and add these number of mitosis together to provide a mitotic score.

A list of snapshots of mitosis detection is shown in Figure 3.34. These snapshots are taken from video developed during MICO 2.0 platform evaluation (http://ipal.cnrs.fr/data/z/MICO_mitosis.mp4).

3.7 Conclusion

In this chapter, we have proposed two frameworks for mitosis count in breast cancer histopathology. These frameworks consist of candidate detection and segmentation, features computation and selection, classification and handling unbalanced training datasets. We select color channels based on histogram analysis of different color channels of various

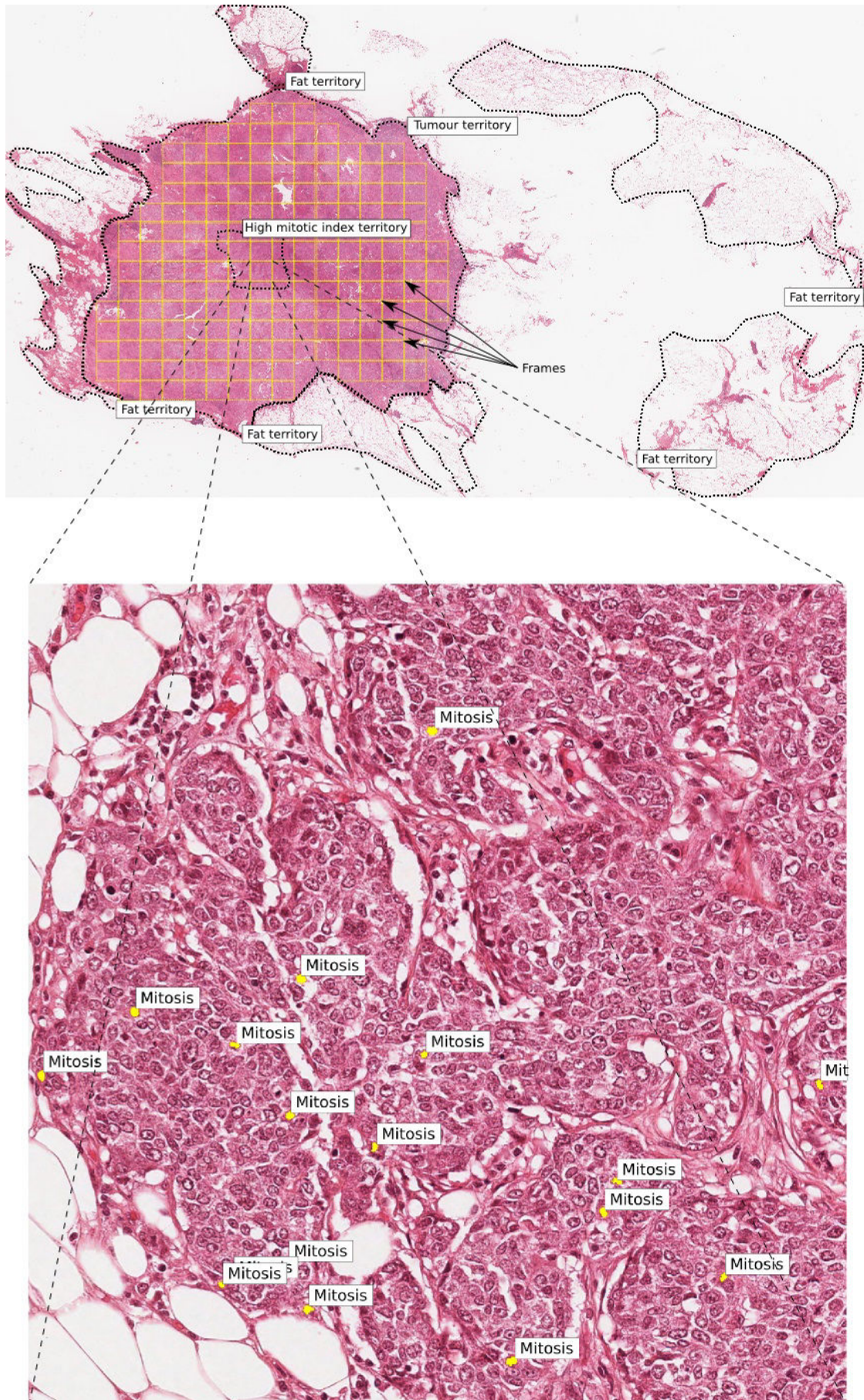


Figure 3.32: Territories and frames.

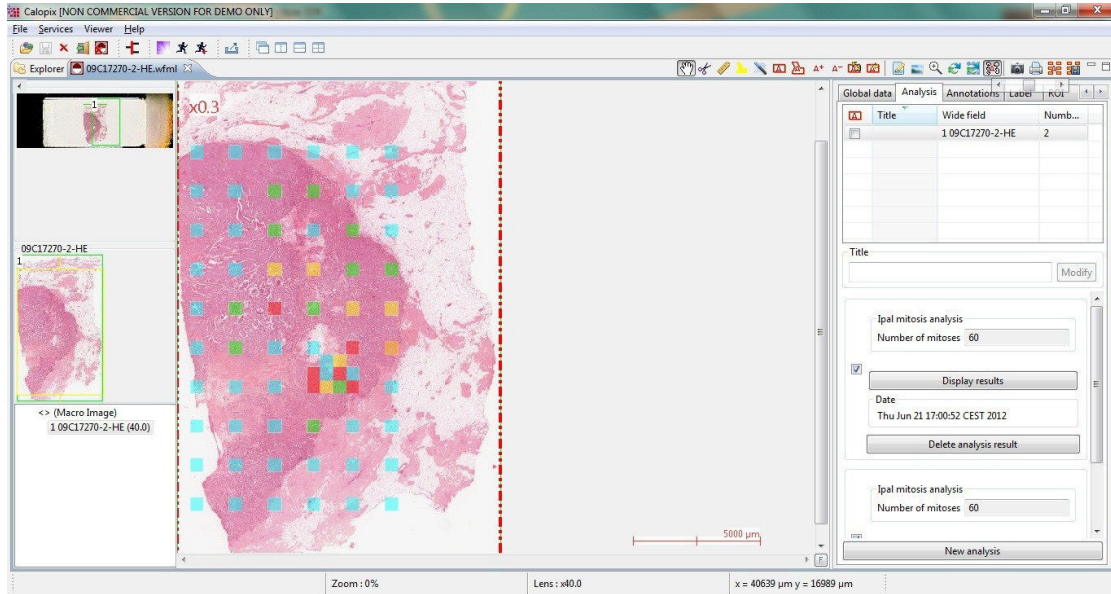
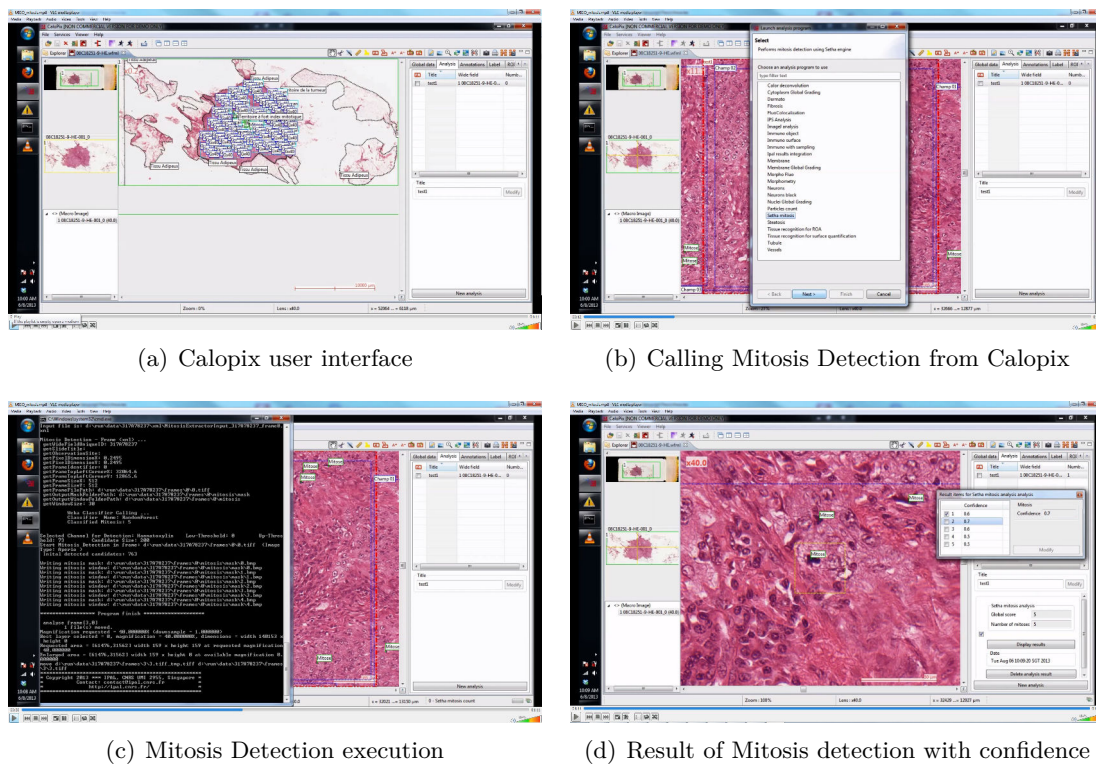


Figure 3.33: Frames analyzed by ITM²C Framework are displayed on TRIBVN Calopix platform. The color code is based on the number of mitosis detected in the frame (from blue for zero mitosis to red for 10 or more mitosis).



(a) Calopix user interface

(b) Calling Mitosis Detection from Calopix

(c) Mitosis Detection execution

(d) Result of Mitosis detection with confidence

Figure 3.34: Snapshot of Mitosis Detection video in MICO 2.0 platform [6].

color models using selected sampled of mitosis and non-mitosis nuclei and background regions. These frameworks are evaluated on MITOS benchmark, on Aperio and Hamamatsu datasets. Finally, we suggested an efficient generic strategy to explore large images like WSI with a local signal measure of relevance. The real time evaluations of these frameworks are done in the MICO project platform prototyping. In next chapter, we will propose an automated mitosis detection framework for breast cancer MSI based on multispectral spatial features.

Chapter 4

Automated Mitosis Detection in Multispectral Images

Résumé du chapitre

L'imagerie multispectrale présente l'avantage de fournir les caractéristiques d'un tissu pour différentes fréquences du spectre électromagnétique. L'imagerie multispectrale capture les images avec un contenu spectral précis, en corrélation avec l'information spatiale, en révélant les caractéristiques anatomiques chimiques de l'histopathologie [98, 100]. Cette modalité permet aux biologistes et aux médecins de voir au-delà des images couleur rouge-vert-bleu (RVB) auxquelles ils sont habitués. Des publications récentes [52, 101, 183, 89] ont commencé à explorer l'utilisation des informations supplémentaires contenues dans les images multispectrales. Plus précisément, la comparaison de méthodes spectrales montre l'avantage des données multispectrales [99, 58]. Cependant, l'avantage supplémentaire apporté par les images multispectrales pour l'analyse d'images colorées par hématoxyline et éosine (H&E) en histopathologie est encore largement inconnu, bien que certains résultats prometteurs soient présentés dans [148, 52, 89, 183]. Pour autant que nous sachions, il n'existe actuellement aucune étude portant sur l'utilisation d'images multispectrales pour la détection de mitoses en histopathologie.

Dans ce chapitre, nous présentons une extension aux images multispectrales du système précédent pour le comptage des mitoses dans le cas du cancer du sein. L'imagerie multispectrale est une technologie d'imagerie médicale récente qui a déjà montré dans d'autres domaines l'augmentation de la précision de la segmentation qu'elle permet d'obtenir. Le système proposé comprend la sélection de plan focal et de bandes spectrales, la détection de candidats mitoses et de calcul des caractéristiques spatiales multispectrales. Nous proposons trois méthodes différentes pour la sélection des bandes spectrales. Ce système est évalué sur le jeu de données multispectral du concours MITOS.

Chaque mitose présente différents niveaux d'informations pertinentes selon les bandes spectrales étudiées. Ce système répond également à deux questions importantes. Premièrement, l'analyse spatiale-spectrale sur les bandes spectrales sélectionnées (par opposition à l'analyse spatiale sur une seule bande spectrale ou l'analyse spatiale-spectrale sur toutes les bandes spectrales disponibles) est-elle suffisante pour la classification efficace des noyaux détectés en mitose ou non-mitose ? Un avantage évident de l'utilisation d'une sélection de bandes spectrales est la diminution de la complexité de calcul et du volume de données à manipuler. Deuxièmement, quelle est l'efficacité des multiples descripteurs pour la discrimination des noyaux des mitoses et des noyaux des non-mitoses par rapport à un seul type de descripteur ?

4.1 Introduction

Multispectral Imaging (MSI) has the advantage to retrieve spectrally resolved information of a tissue image scene at specific frequencies across the electromagnetic spectrum. MSI captures images with accurate spectral content, correlated with spatial information, by revealing the chemical and anatomic features of histopathology [98, 100]. This modality provides option to biologists and pathologists to see beyond the RGB image planes to which they are accustomed. Recent publications [52, 101, 183, 89] have begun to explore the use of extra information contained in such spectral data. Specifically, comparisons of spectral methodologies demonstrate the advantage of multispectral data [99, 58]. The added benefit of MSI for analysis in routine H&E histopathology, however, is still largely unknown, although some promising results are presented in [148, 52, 89, 183]. As far as we know, there is no existing study of the use of MSI for automation of mitosis detection in breast cancer histopathology.

In this chapter, we present another framework, an extension of previous framework, for counting of mitosis nuclei in breast cancer MSI. MSI is a recent medical imaging technology, proven successful in increasing the segmentation accuracy in other fields. The proposed framework includes a selection of SBs and focal plane, detection of candidate mitosis nuclei and computation of morphological & multispectral statistical features (MMSF). We propose three different methods for SBs selection. This framework is evaluated on MITOS MSI (multispectral) dataset.

Each mitosis region has different level of relevant information in different SBs. This framework also addresses two important questions: First, does the spatial-spectral analysis on selected SBs (as opposed to spatial analysis on single SB or spatial-spectral analysis of all the SBs) suffice for efficient classification of mitosis and non-mitosis nuclei. An obvious advantage of using selected SBs is its reduced computational and storage complexity. Second, how effective are the multiple features for discrimination of mitosis and non-mitosis nuclei as compared to one type of features.

4.2 Multispectral Dataset

The MITOS MSI dataset is made up of 200 images coming from five different slides scanned at 40X magnification using a 10 spectral bands (SBs) microscope. There are 40 images per slide and each image has a size of $251.6 \times 251.6 \mu\text{m}^2$ (that is an area of 0.063 mm^2). The 200 images contain a total 322 mitosis nuclei. The training data set consists of 140 images containing 224 mitosis and the evaluation data set consists of 60 images containing 98 mitosis [150].

The SBs are all in the visible spectrum. There is some spectrum overlapping for the SBs. In addition, for each spectral band, the digitization has been performed at 17 different focal planes (17 layers Z-stack), each focal plane being separated from the other by 500 nm. Therefore, for each image, there is a stack of 170 files (10 SBs and 17 focal planes for each SB). Figure 4.1 shows the spectral overage of each of the 10 SBs of the multispectral microscope.

4.3 Multispectral Intensity, Textural & Morphology-based Mitosis detection in Multispectral images (MITM³) Framework

In the proposed framework, we address the shortcomings of previous works, including (1) selection of focal plane and SBs; (2) analysis of multispectral statistical features in

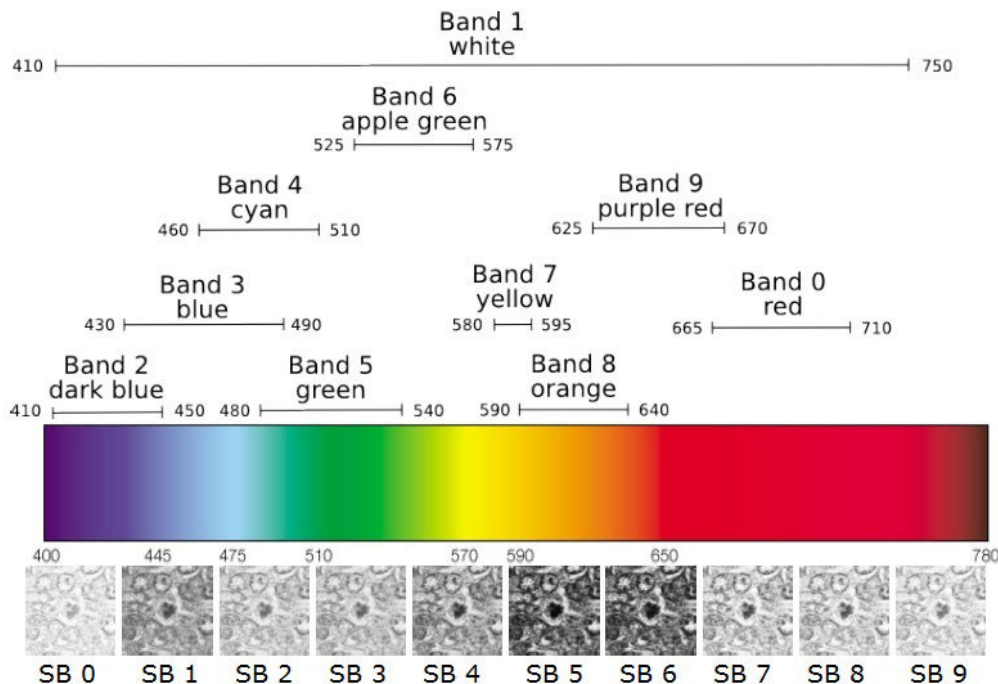
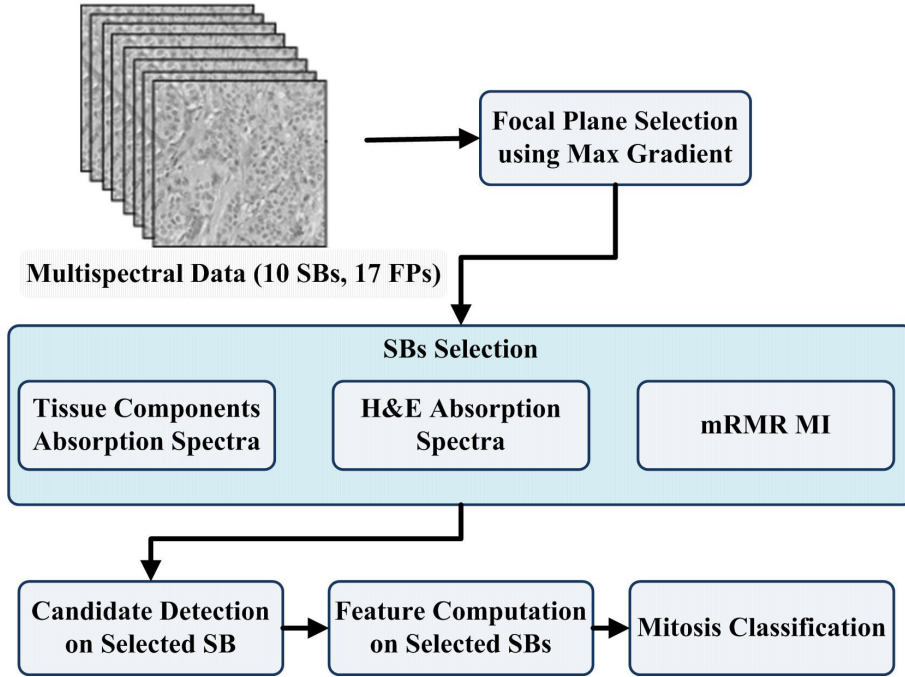


Figure 4.1: SBs of the multispectral microscope and examples for each SB.

selected SBs rather than single [118, 113] or all SBs [21, 183, 89] and (3) selection of the best classifier for discrimination of mitosis nuclei from other microscopic objects. The main novel contributions of this work are:

- i. An automatic and unsupervised focal plane selection process
- ii. Three different methods for SBs selection including relative spectral absorption of different tissue components, spectral absorption of H&E stains and mRMR technique.
- iii. Computation of morphological & multispectral statistical features (MMSF) containing intensity, texture and morphological features which leverage discriminant information from a given candidate across selected SBs for classification of mitosis and non-mitosis nuclei.
- iv. An extensive investigation of classifiers and inference of the best one for mitosis nuclei classification.
- v. Comparison of patch and region based features for mitosis classification.

The framework for mitosis detection in breast cancer MSI are shown in Figure 4.2. The proposed framework has five main steps. Step one performs a selection of the most informative focal plane based on maximum gradient information of mitosis nuclei from background. Step two is responsible for the selection of relevant SBs for the objective of mitosis detection. Candidates for mitosis nuclei are detected in step three. Then, in step four, a MMSF signature vector of intensity and texture information across selected SBs is computed for each detected candidate. In addition, using segmented regions of detected candidates, morphological features are also computed and added to the signature vector. During step five, candidates are classified into mitosis and non-mitosis classes using DT, MLP as well as L-SVM and NL-SVM classifiers. A side advantage of performing the spatial analysis on multiple SBs simultaneously is to investigate whether improvement in accuracy can be achieved with MMSF computation in selected SBs over those methods which use

Figure 4.2: MITM³ Framework.

single SB [118, 113] or all SBs [21, 183, 89]. In addition, we comprehensively analyse patch and region based features for mitosis discrimination.

4.3.1 Focal plane Selection using Maximum Gradient

For selection of focal plane, average gradient of mitosis nuclei from background regions is computed on all the focal planes. The computed gradient vector of image I :

$$\nabla I = \left[\frac{\partial I}{\partial x}, \frac{\partial I}{\partial y} \right] \quad (4.1)$$

where $\frac{\partial I}{\partial x}$ and $\frac{\partial I}{\partial y}$ are partial derivative of I with respect to x and y directions, respectively as:

$$\begin{aligned} \frac{\partial I(x, y)}{\partial x} &= \frac{I(x+1, y) - I(x-1, y)}{2} \\ \frac{\partial I(x, y)}{\partial y} &= \frac{I(x, y+1) - I(x, y-1)}{2} \end{aligned} \quad (4.2)$$

The maximum average gradient focal plane (i.e. having the best focus) is selected for the next steps of the framework.

4.3.2 Spectral Bands Selection

The main tissue components visible in the data set images can globally be categorized into fat, stroma and epithelial nuclei as shown in Fig. 4.3. As our purpose is the detection of mitotic nuclei only, we further subdivided epithelial nuclei into mitotic and non-mitotic nuclei. We selected 200 image patches, a patch being a region of interest of 150×150 pixels, for each tissue components and computed the spectral absorption responses of each tissue components for the available 10 spectral bands as shown in Fig. 4.4(a). In 4.4(a), fat tissue

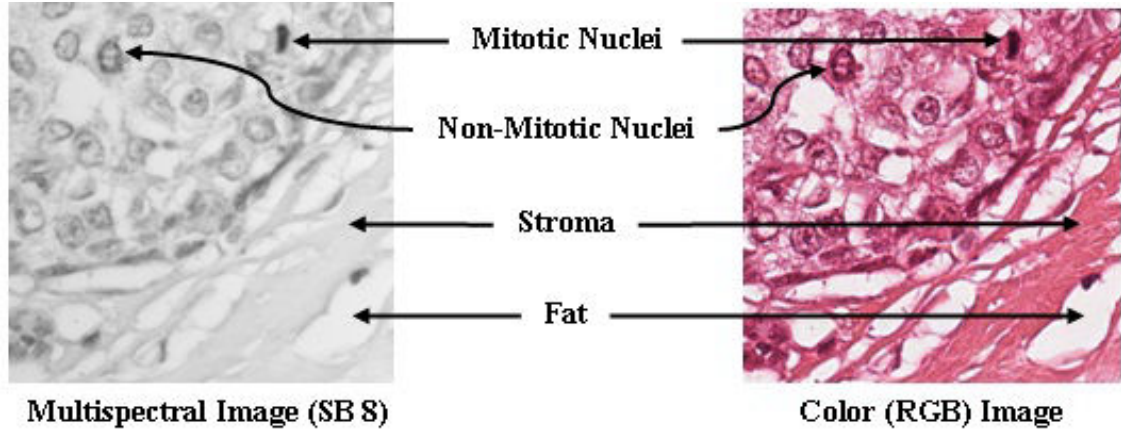


Figure 4.3: Example of different components of breast tissue in H&E stained histopathological image. Left image sample is taken from spectral band 8, focal plane 6 of multispectral microscope; right image is taken from Aperio Slide Scanner.

is negligible as it has very low absorption response. Moreover, mitotic and non-mitotic nuclei contributions are indistinguishable. We select mitotic nuclei and stroma curves to compute the maximum differentiation in contribution to pixel intensity. In 4.4(b), one can see that bands 7, 8 and 9 exhibit the biggest difference between mitotic nuclei and stroma contributions. They are the best candidates for maximum differentiation.

Method 2: Hematoxylin and Eosin Spectral Absorption

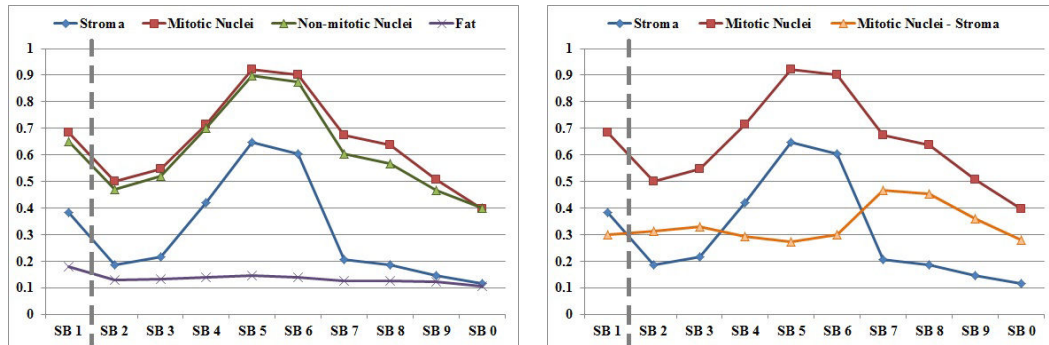
To illustrate the possible correlation between SBs and the staining characteristics of the spectral samples, the plot of hematoxylin and eosin dyes spectral absorptions are shown in Figure 4.5 (this plot is derived from the work of Bautista and Yagi [18]). Hematoxylin stains nuclei material, while eosin stains both nuclei and cytoplasm. The H–E plot in Figure 4.5 shows the difference of absorption between hematoxylin and eosin. The bands for which H–E is maximum are more suitable for discrimination between nuclei and cytoplasm. The absorption response of hematoxylin is maximum in SBs 7 and 8 with almost zero eosin response. Therefore, these bands should be good options for the task of mitosis detection. As SBs 4 and 5 have almost the same hematoxylin and eosin response, they should not provide much discriminating information between nuclei and cytoplasm, so they might be not suitable for mitosis detection. Hence, in this method, we reconstruct the spectrum of a pixel by using staining characteristics of tissue components for selection of the optimal number of SBs for mitosis discrimination in H&E stained MSI.

Method 3: mRMR Technique

In this method, mRMR technique [131] is used for SBs selection. Selection is based on two criteria; minimum redundancy $R(S, c)$ and maximum relevance $D(S, c)$. The relevancy of selected SBs to class labels has been measured by average of mutual information (MI) between each SB and class label. Their redundancy is measured by an average of MI between each pair of SBs. The average relevancy of selected SBs is defined as:

$$D = \frac{1}{|S|} \sum_{s_i \in S} \text{MI}(s_i; c_j) \quad (4.3)$$

where S denotes the selected SBs set, $|S|$ denotes the number of selected SBs, c_j denotes j^{th} class label in class set C , s_i denotes i^{th} SBs in S and MI is mutual information between



(a) Normalized absorption spectra of four tissue components in 10 SBs (b) Difference of mitotic nuclei and stroma absorption spectra in 10 SBs

Figure 4.4: Normalized absorption spectra of four tissue components in 10 spectral bands (SBs). Note that SB 1 (white band), in nature, is different from other SBs and may serve as reference as it covers the whole visible spectrum and contains all the information that other bands are containing, although at a lower resolution. It is separated from other SBs by a dotted line.

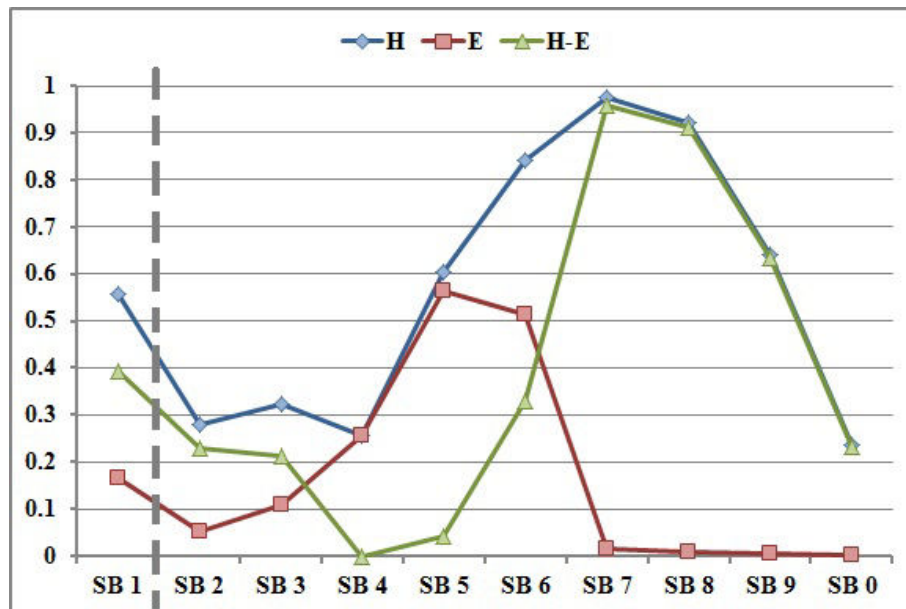


Figure 4.5: Normalized plot of the hematoxylin (blue line) and eosin (red line) dye absorption spectra in MSI and the difference of hematoxylin and eosin (green line).

SB s_i and class label c_j . MI is computed using entropy as

$$\text{MI}(S; C) = H(S) - H(S|C) \quad (4.4)$$

where

$$H(S) = - \sum_{s_i \in S} p(s_i) \log_2(p(s_i)) \quad (4.5)$$

and

$$H(S|C) = - \sum_{s_i \in S} \sum_{c_j \in C} p(s_i, c_j) \log_2(p(s_i|c_j)) \quad (4.6)$$

are entropy functions that calculates the uncertainty of the SBs and the class labels. In (4.5) and (4.6), $p(s_i)$ is the probability density function of s_i and $p(s_i|c_j)$ is the conditional probability density function of s_i and c_j . By maximizing D for full SBs set S_T , we can select a SBs set S having maximum relevance for mitosis discrimination by observing all SBs set S_T .

It is likely that selected SBs have rich redundancy. Therefore, the minimum redundancy $R(S, c)$ is added to select mutually exclusive SBs.

$$R = \frac{1}{|S|^2} \sum_{s_i, s_j \in S} \text{MI}(s_i; s_j) \quad (4.7)$$

$\text{MI}(s_i; s_j)$ is maximum when two SBs s_i and s_j have functional dependency and $\text{MI}(s_i; s_j) = 0$ if s_i and s_j are statistically independent. By minimizing R for selected SBs, we select SBs set with minimum redundancy.

Selection of Spectral Bands in Set (S) in Equations (4.3), (4.4), (4.5), (4.6) and (4.7):

The incremental search method is used to find the n SBs from the set $\{S_T - S_{n-1}\}$, maximizing the following condition expression:

$$\max_{s_i \in S_T - S_{(n-1)}} \left[\text{MI}(s_i; c) - \frac{1}{n-1} \sum_{s_j \in S_{(n-1)}} \text{MI}(s_i; s_j) \right] \quad (4.8)$$

The image samples, used in computation of spectral absorption of different tissue components, are divided into two classes. The non-mitosis class consists of three tissue components including adipose, cytoplasm and nuclei, and the remaining samples belong to mitosis class. We perform mRMR on these image samples and their MI, with ranking shown in Table 4.1. Figure 4.6 shows the relevant contribution of each SB in accumulated MI.

4.3.3 Candidate Detection on Selected SB

We perform candidate detection on the selected SB that has higher MI and difference between absorption response of mitosis and other tissue components. On the selected SB, we perform candidate detection as explained in section 3.4.3. Resulting candidates that are too small or too big to be mitotic nuclei were filtered out based on a minimum area of 200 pixels (that is $37 \mu\text{m}^2$, corresponding roughly to a circle of diameter $6.86 \mu\text{m}$) and a maximum area of 5,500 pixels (that is $1,017.5 \mu\text{m}^2$, corresponding roughly to a circle of diameter $36 \mu\text{m}$), computed on the segmented regions from MITOS multispectral dataset. An example of candidate detection is shown in Figure 4.7(d).

The process of nuclei division has four different stages, each one exhibiting different shape, size and textures. This motivates further spatial and morphological analysis on multispectral data to achieve reasonable classification of regions into mitosis and non-mitosis types.

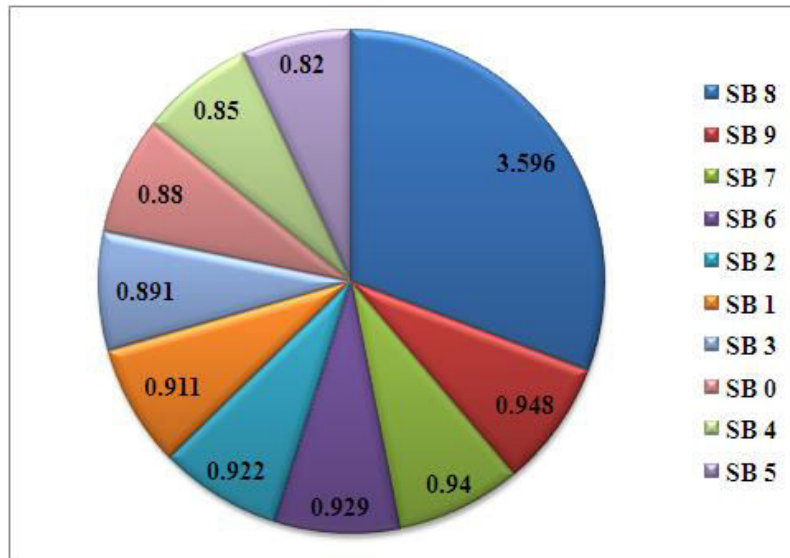


Figure 4.6: Relevant contribution of each SB in accumulated MI.

Table 4.1: SBs Mutual Information (MI) Measure.

SBs	MI	Accumulated MI	Accumulated MI%
SB 8	3.60	3.60	33%
SB 9	3.59	0.95	42%
SB 7	3.38	0.94	51%
SB 6	3.18	0.93	60%
SB 2	3.16	0.92	69%
SB 1	3.11	0.91	78%
SB 3	3.05	0.89	86%
SB 0	2.99	0.88	91%
SB 4	2.94	0.85	95%
SB 5	2.85	0.82	100%

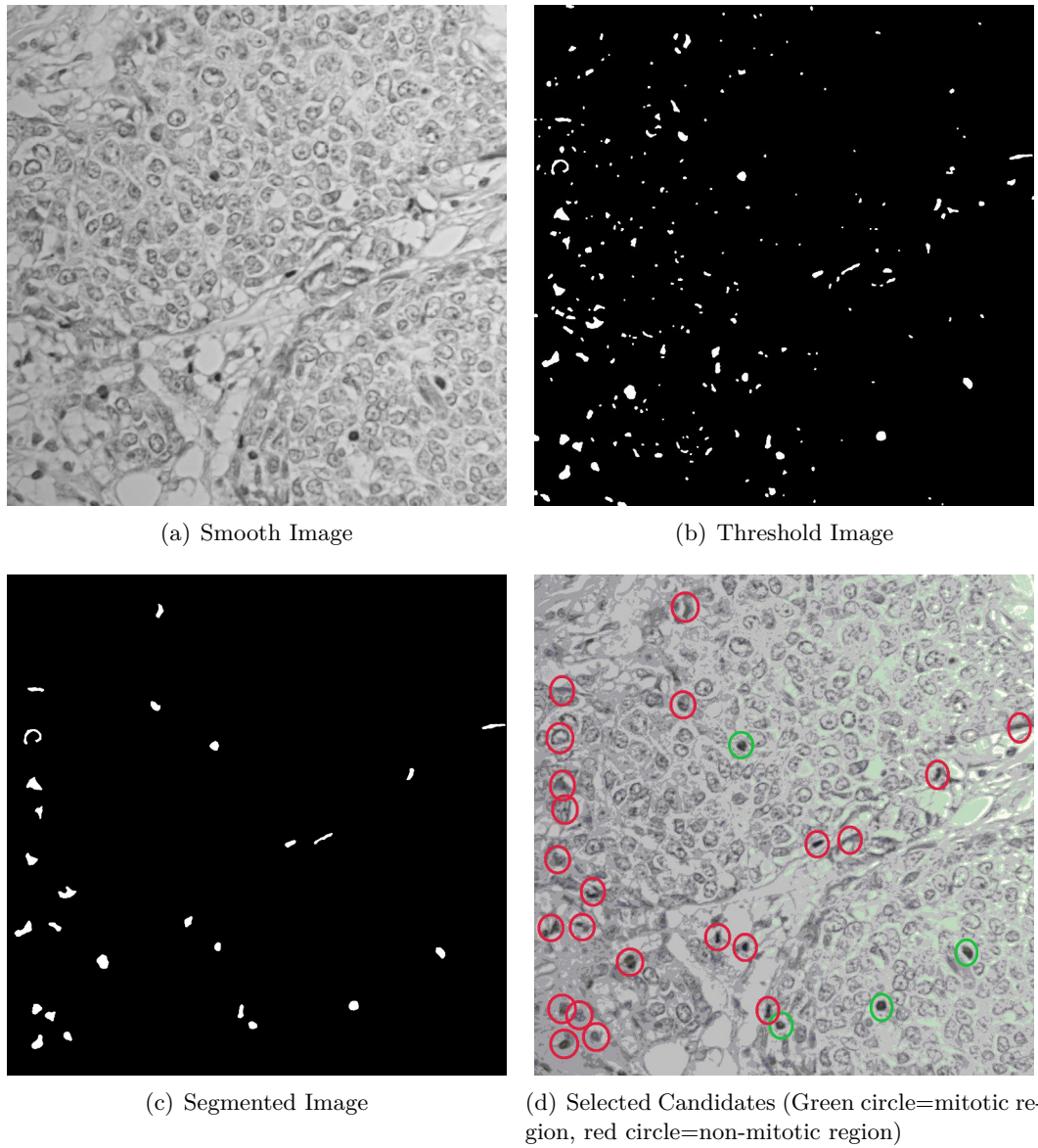


Figure 4.7: Different steps in candidate detection on breast cancer MSI histopathology.

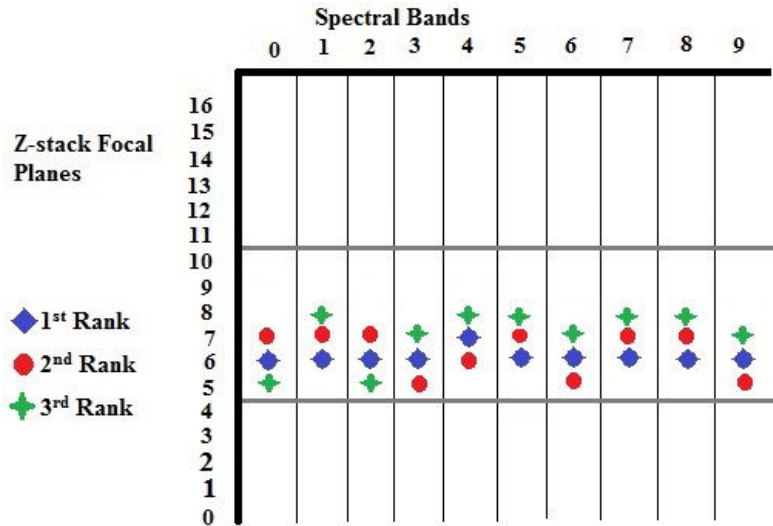


Figure 4.8: Top three ranked focal planes using candidate detection results in all SBs.

4.3.4 Multispectral Spatial Features (MMSF) Computation

We compute MMSF vector consisting of intensity and textural features in several selected SBs as explained in section 3.5.3 and 3.4.4. In addition, we also compute morphological features (such as area, roundness, elongation, perimeter and equivalent spherical perimeter) from segmented regions during candidate detection as explained in section 3.5.3.

4.4 Experiments and Results

The proposed framework is evaluated on MITOS MSI dataset [1]. The results of candidate detection and classification methods are compared with ground-truth (GT) information provided along with the dataset. The metrics use to evaluate the mitosis detection included: TP, FP, FN, TPR, PPV and FM. In addition to MITOS contest, the proposed framework is also evaluated with 5-fold CV [44] by merging the training and evaluation sets.

4.4.1 Focal Plane Selection

To gain a better understanding of the relative contributions of specific SBs, we perform candidate detection in all available SBs using six selected z-stack focal planes according to maximum gradient as explained in section 4.3.1. The results of candidate detection are ranked according to FM and reports top three ranked results in Figure 4.8. The focal plane 6 has more information for candidate detection as compared to other focal planes.

4.4.2 SBs Selection

How many SBs are necessary for a good detection of mitosis figures? Which SBs are relevant for mitosis figure detection? To discuss these two questions, we tried first to evaluate the contributions of each SBs using three different proposed methods as discussed in 4.3.2. The results are shown in Table 4.2. The SBs ranking in method one is based on difference between spectral absorption of mitosis nuclei and cytoplasm, while the SBs ranking in method two is based on difference between hematoxylin and eosin spectral absorption. These three rankings put the same SBs 7, 8 and 9 in top position. More specifically, the

Table 4.2: Different Rankings of SBs. The upper dotted line shows that SBs 7,8 and 9 are at top three positions in these ranking. The lower dotted line shows that SBs 4 and 5 are at bottom three positions.

Method 1		Method 2		Method 3	
SB	Mitosis–Cytoplasm	SB	H–E	SB	MI
7	0.47	7	0.96	8	3.6
8	0.45	8	0.91	9	3.59
9	0.36	9	0.64	7	3.38
3	0.33	1	0.39	6	3.18
2	0.31	6	0.33	2	3.16
6	0.30	0	0.23	1	3.05
1	0.30	2	0.23	3	3.05
4	0.29	3	0.21	0	2.99
0	0.28	5	0.04	4	2.95
5	0.27	4	0	5	2.83

top position in method one and two is occupied by SB 7 while method three gives the top position to SB 8 on the basis of highest MI.

At the bottom of the table, there are SBs 4 and 5 for all three rankings. According to method two ranking, the difference between absorption response of hematoxylin and eosin in SBs 4 and 5 are almost zero which represent that these two SBs are irrelevant for mitosis discrimination. Based on these analyses, we ignore SBs 4 and 5 for mitosis discrimination.

Considering the available SBs and their rankings, our selection of SBs contains the following eight bands: 8, 9, 7, 6, 2, 1, 3, and 0.

4.4.3 Candidate Detection

We perform candidate detection in top three SBs of the three proposed SBs selection methods. Those three SBs are 7, 8 and 9. In order to evaluate the ability of these SBs to provide adequate information for detection of mitosis, we also perform separately candidate detection on SB 1 only as this band covers the whole visible spectrum. The results of candidate detection step are ranked according to FM and reported in Figure 4.9.

SB 8 has higher FM than SBs 7, 9 and 1 with more TP and less FP. Although SB 1 covers the full spectrum of light, it reported poor results for candidate detection.

On training and evaluation sets, the candidate detection using SB 8 detects 3583 and 1655 candidates, containing 202 and 92 ground truth mitosis from a total 224 and 98 ground truth mitosis, respectively. Among all the detected candidates, there are 3381 and 1563 non-mitosis in the training and evaluation sets, respectively. The candidate detection step generates a large number of FP and missed 22 and 6 GT mitosis from training and evaluation sets, respectively.

4.4.4 Candidate Classification

Classification using Region and Patch based MMSF on MSI Evaluation Dataset

It is common practice that object are classified using information (intensity, texture and morphology) computed from object region. Besides region information, the information

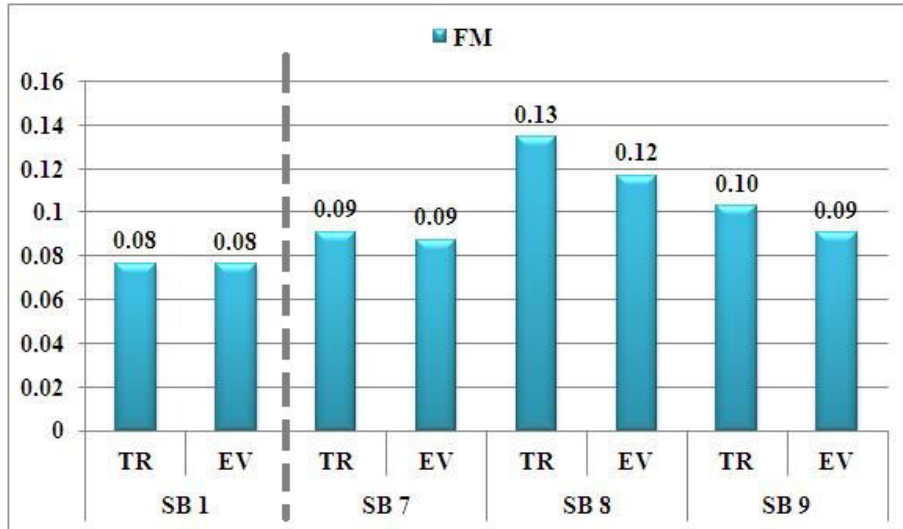


Figure 4.9: Candidate detection results on selected SBs.

Table 4.3: Patch sizes in pixels and μm on the MSI Dataset.

Patch sizes in pixels	Multispectral Microscope (μm)
Patch 110×110	20.35×20.35
Patch 100×100	18.50×18.50
Patch 90×90	16.65×16.65
Patch 80×80	14.80×14.80
Patch 70×70	12.95×12.95
Patch 60×60	11.1×11.1

computed from neighbouring region can also help to improve the classification. In order to validate this hypothesis and optimal size of patch for MMSF computation, we compute MMSF on different patch sizes as shown in Table 4.3.

The results of classification using all SBs MMSF of different patch size with L-SVM classifier are shown in Figure 4.10. MSI training dataset is used for training of L-SVM classifier and MSI evaluation dataset is used to test the classification accuracy of proposed framework with MMSF computed on different patch sizes. The classification results emphasize that the rate of mitosis detection increases by decreasing patch size and gets maximum mitosis detection with patch size 80×80 and after that reduces the detection of mitosis. We get maximum TPR 76% on patch 80×80 but minimum PPV. While patch 90×90 reports maximum FM and PPV. Now onward, we will use patch 90×90 for patch based MMSF for next experiments.

We also evaluate MITM³ framework by comparing segmented region and patch based MMSF as shown in Table 4.4. In this experiment, MSI training set is used to train the four selected classifiers and MSI evaluation set is used to test the classification accuracy of proposed framework with MMSF computed on region and patch. Classification is performed from MMSF computed both on selected eight SBs and on all the SBs. In case of region based MMSF computed using all SBs, NL-SVM classifier detects few FP as compared to other classifiers but few TP as well. While DT classifier detects more mitosis as compared to other classifiers, but more FP. Overall, L-SVM classifier reports the highest FM 61.69%. In case of region based MMSF using selected eight SBs, L-SVM detects more mitosis as compared to other classifiers with more FP. The NL-SVM classifier performs better by detecting few

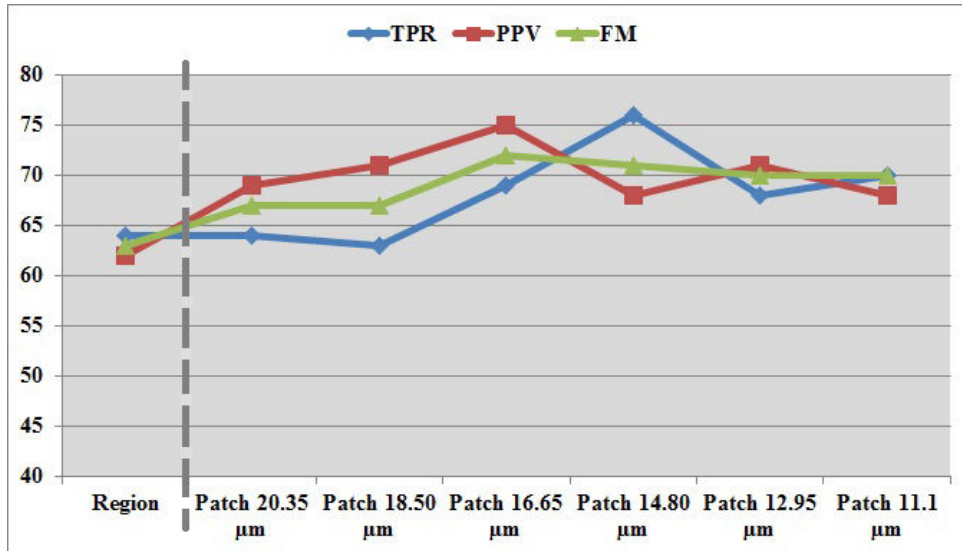


Figure 4.10: Classification results using different patch size based MMSF (L-SVM classifier).

FP and report the highest FM 63.74%. Using patch based MMSF, all classifiers report better results in comparison with region based MMSF. The L-SVM classifier, with MMSF computed on all SBs and selected eight SBs, reports the highest FM 71.96% and 73.74%, respectively. As patches carry more information for mitosis discrimination than region, we select patch 90×90 based MMSF for classification of candidates in coming experiments.

To consider potential over-fitting of classification, we have also tested our proposed framework using a training set made up from three out of the five slides and an evaluation set made up of the remaining two slides. LSVM classifier outperformed DT, MLP and NLSVM classifiers with TPR (90%), PPV (65%) and F-Measure (75.60%) on all spectral bands and TPR (84%), PPV (76%) and F-Measure (80%) on selected eight spectral bands. These results eliminate the potential risk of over-fitting of classification. In addition, by reducing the number of features, the classification results are of lower quality. This demonstrates that there is no over-fitting in our framework when using the complete set of features from selected spectral bands.

Classification using Single SB MMSF on MSI Evaluation Dataset

In this experiment, we evaluate MITM³ framework on MSI evaluation dataset using computed MMSF from single SB. The classification results achieved with L-SVM classifier are shown in Figure 4.11. This classification result also supports the result of SBs selection. We get poor results from SBs 4 and 5 that conclude these SBs are irrelevant for mitosis detection as also discussed in SBs selection. SB 8 contains more information as compared to rest of SBs which is also validated from SBs selection 4.2 and candidate detection results 4.9.

Using mRMR ranking of SBs, the different selection of SBs are also tested using all the classifiers introduced in Section 3.4.6. However, for both clarity and brevity, we only plot FM curves of classification results with all classifiers in Figure 4.12. Similar results are obtained in other cases. Figure 4.12 shows that FM increases while we add more SBs to the set of selected SBs. FM reaches a peak with a set of eight selected SBs, then it starts decreasing when adding more SBs. The sequence of SBs selection is according to ranking of MI from mRMR. In case of few SBs MMSF, L-SVM, NL-SVM and DT report

Table 4.4: Region vs Patch based MITM³ Classification Results on MSI Evaluation Dataset (GT = 98).

Classifiers	TPR	PPV	FM	TPR	PPV	FM
Region based MMSF						
	All SBs MMSF			Selected 8 SBs MMSF		
DT	67%	53%	59.19%	62%	62%	62.24%
MLP	64%	56%	59.72%	60%	66%	63.10%
L-SVM	63%	60%	61.69%	64%	62%	63.32%
NL-SVM	54%	68%	60.23%	59%	69%	63.74%
Patch based MMSF						
	All SBs MMSF			Selected 8 SBs MMSF		
DT	61%	71%	65.57%	65%	70%	67.72%
MLP	63%	67%	64.92%	66%	70%	68.06%
L-SVM	69%	75%	71.96%	74%	73%	73.74 %
NL-SVM	55%	77%	64.29%	59%	77%	67.05%

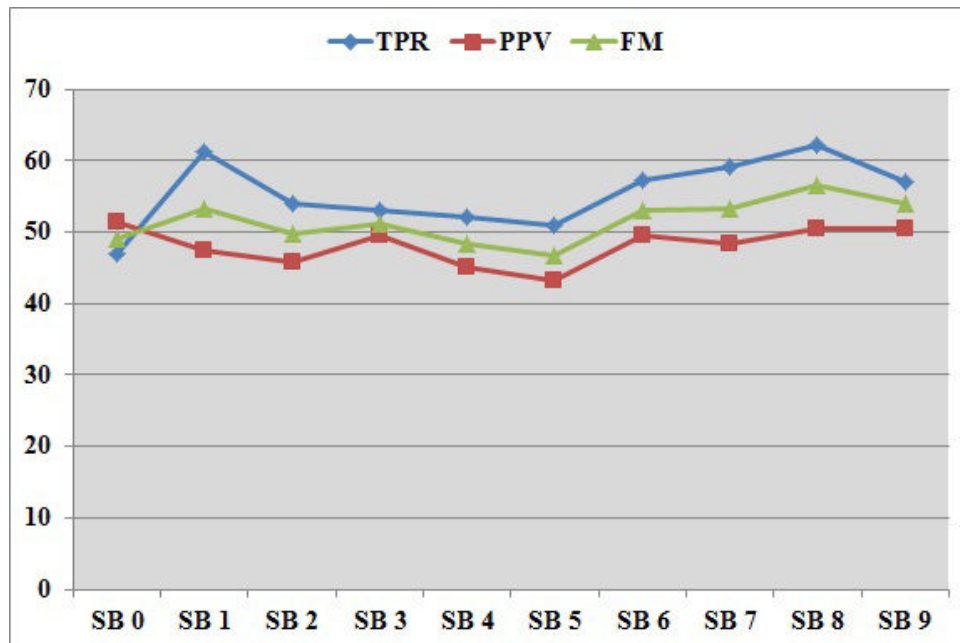


Figure 4.11: Classification results using single SB MMSF with L-SVM classifier.

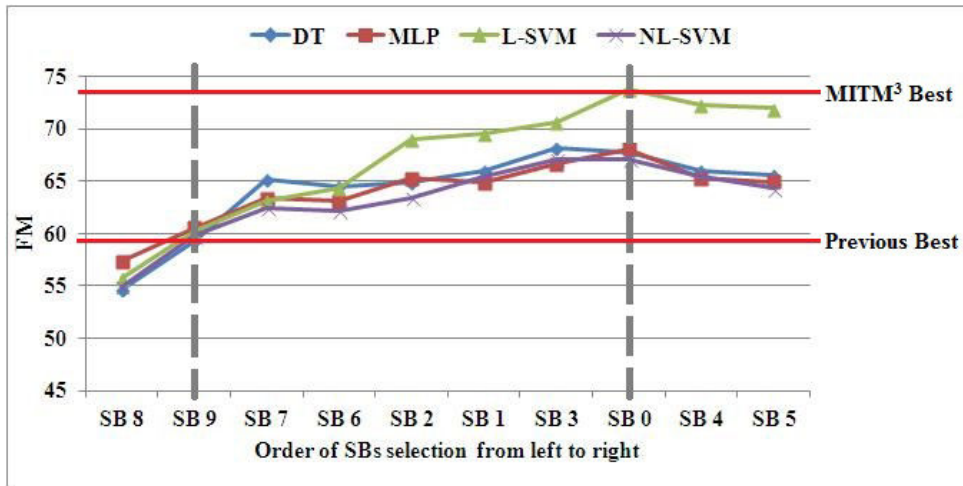


Figure 4.12: Plot of FM using SBs selection. Result from using all SBs from left to the current, e.g. SB 2 result uses SB 8, 9, 7, 6, 2. This order is taken from the mRMR ranking. First vertical dotted line shows that selecting first two SBs features matches the previous best result. Second vertical dotted line highlights the overall best result by selecting features up to SBs 0 which shows 25% increased in FM.

poor classification accuracy while MLP reports higher classification accuracy. As more SBs are selected L-SVM classifiers start performing better than other classifiers and reached maximum performance with first eight selected SBs. Figure 4.13 plot the TPR, PPV and FM of L-SVM classifier.

Classification on MSI Dataset using 5-Fold CV

In this experiment, the assessment of classification performance is made using 5-fold CV by combining both MSI training and evaluation sets. Classification results are shown in Table 4.5. In case of all SBs MMSF, L-SVM classifier outperforms the other classifiers and achieved the highest TPR (59%) and FM (65.74%). Overall, NL-SVM reported higher PPV (76%) but with few number of detected mitosis as well. In case of selected eight SBs MMSF, L-SVM classifier has its FM improved thanks to higher TPR and PPV.

Besides SBs selection, we also evaluate our framework on selected MMSF using feature selection technique as discussed in section 3.4.5. According to results of feature selection, most of the features are selected from SBs 1, 2, 6, 7, 8 and 9. None of feature is selected from SBs 4 and 5. A comparison of classification results with all SBs MMSF, selected 8 SBs MMSF and selected MMSF using feature selection technique are shown in Table 4.5. Overall, L-SVM classifier achieved the highest TPR (65%) and FM (67.75%) with selected 8 SBs MMSF and NL-SVM achieved the highest PPV (80%) with selected MMSF using feature selection technique.

Classification on MSI Dataset with White SB vs other SBs

To investigate the relative contribution of SBs, we also performed a comparative study of mitosis classification using spatial features from white SB (SB 1) with MMSF features from rest of SBs and the achieved results are shown in Table 4.6. It is important to note that the multispectral bands features (MSBF) excluding white SB outperforms with all classifiers and reports the highest FM 66.56% with L-SVM. The classification results obtained with all four classifiers are low using white spectral band features (WSBF). This experiment

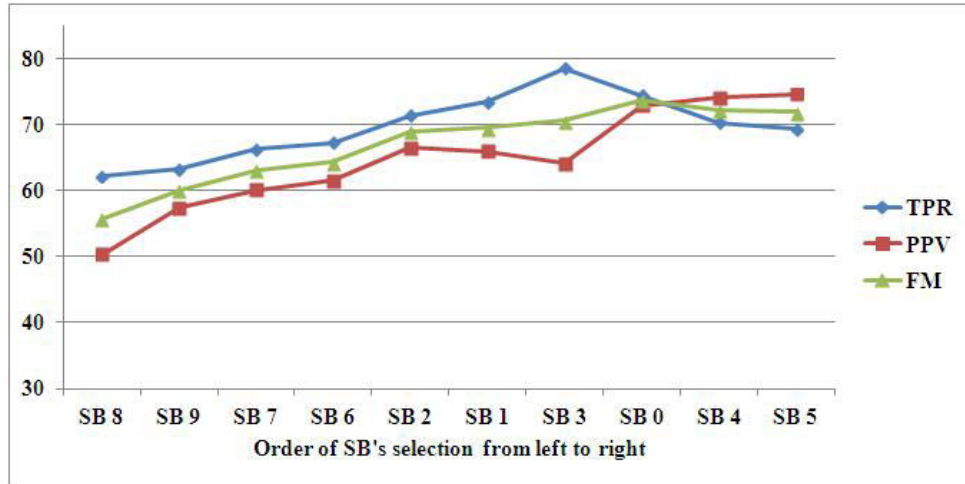


Figure 4.13: Plot of TPR, PPV and FM with L-SVM classifier using the order of SBs selection. Result from using all SBs from left to the current, e.g. SB 2 result uses SB 8, 9, 7, 6, 2. This order is taken from the mRMR ranking.

Table 4.5: Classification Results on MSI Dataset using 5-Fold CV (GT = 322).

Classifiers	TP	FP	TPR	PPV	FM
All SBs MMSF					
DT	178	95	55%	65%	59.83%
MLP	180	82	56%	69%	61.64%
L-SVM	190	66	59%	74%	65.74%
NL-SVM	171	53	53%	76%	62.64%
Selected eight SBs MMSF					
DT	186	81	58%	70%	63.16%
MLP	183	70	57%	72%	63.65%
L-SVM	208	84	65%	71%	67.75%
NL-SVM	173	51	54%	77%	63.37%
Selected MMSF using feature selection technique					
DT	178	75	58%	70%	63.16%
MLP	183	78	57%	70%	62.78%
L-SVM	202	82	63%	71%	66.67%
NL-SVM	168	43	54%	80%	63.04%

Table 4.6: Classification Result on MSI Dataset with SB 1 vs other SBs using 5-Fold CV (GT = 322).

Classifiers	TP	FP	TPR	PPV	FM
Using WSBF					
DT	136	72	42%	65%	51.32%
MLP	142	51	44%	74%	55.15%
L-SVM	181	166	56%	52%	54.11%
NL-SVM	141	42	44%	77%	55.84%
Using MSBF					
DT	175	102	54%	63%	58.43%
MLP	180	80	56%	69%	61.86%
L-SVM	209	97	65%	68%	66.56%
NL-SVM	164	54	51%	75%	60.74%

illustrates that multispectral bands have much more information for mitosis classification than white band alone.

Classification on MSI Dataset with Blue, Green and Red SBs

In this experiment, we explored the impact of different parts of visible spectrum on the mitosis classification. The range of visible spectrum is divided into three parts: blue spectrum (SB 2,3,4), green spectrum (SB 5,6,7) and red spectrum (SB 8,9,0). Each spectrum MMSF are used for mitosis classification and results are shown in Table 4.7. The red SBs MMSF reports the highest TPR (67%) with L-SVM classifier. The green SBs MMSF reports the highest PPV (78%) with NL-SVM classifier, while red SBs MMSF reports the highest FM (61.19%) with L-SVM classifier. These results illustrate that red SBs have more information for mitosis classification as compared to green and blue SBs. An important thing is that green SBs have more information as compared to blue SBs that might be due to SBs 6 and 7 having higher absorption response for mitosis figures.

Classification with Morphology, Intensity and Texture Features on MSI Dataset

In this experiment, we explored the impact of different features of MMSF on the mitosis classification. The MMSF contains three different types of features: multispectral intensity features (MSIF), multispectral texture features (MSTF) and morphology features (MorF). We also evaluate the combination of intensity and texture feature (MSITF). These feature set are used for mitosis classification using four classifiers and results are shown in Figure 4.14. The classification results with MorF are worst using all classifiers. The MSTF reports higher TPR, PPV and FM as compared to MSIF. In case of combining intensity and texture, MSITF improves the classification results. These results illustrate that texture have more information for mitosis classification as compared to intensity and morphology. An important thing is that alone morphological features are worst for mitosis classification. Mitosis classification improves in case of combining all features.

4.5 Discussion

The results seem to indicate the best scores are achieved using the selected focal plane, across selected SBs. The proposed focal plane selection is automatic and unsupervised. We

Table 4.7: Classification Result on MSI Dataset with blue, green and red SBs (GT = 322) using 5-Fold CV.

Classifiers	TP	FP	TPR	PPV	FM
Red SBs MMSF					
DT	163	96	51%	63%	56.11%
MLP	156	65	48%	71%	57.46%
L-SVM	216	168	67%	56%	61.19%
NL-SVM	159	52	49%	75%	59.66%
Green SBs MMSF					
DT	160	74	50%	68%	57.55%
MLP	162	86	50%	65%	56.84%
L-SVM	210	155	65%	58%	61.14%
NL-SVM	155	45	48%	78%	59.39%
Blue SBs MMSF					
DT	139	97	43%	59%	49.82%
MLP	149	66	46%	69%	55.49%
L-SVM	173	114	54%	60%	56.81%
NL-SVM	147	50	46%	75%	56.65%

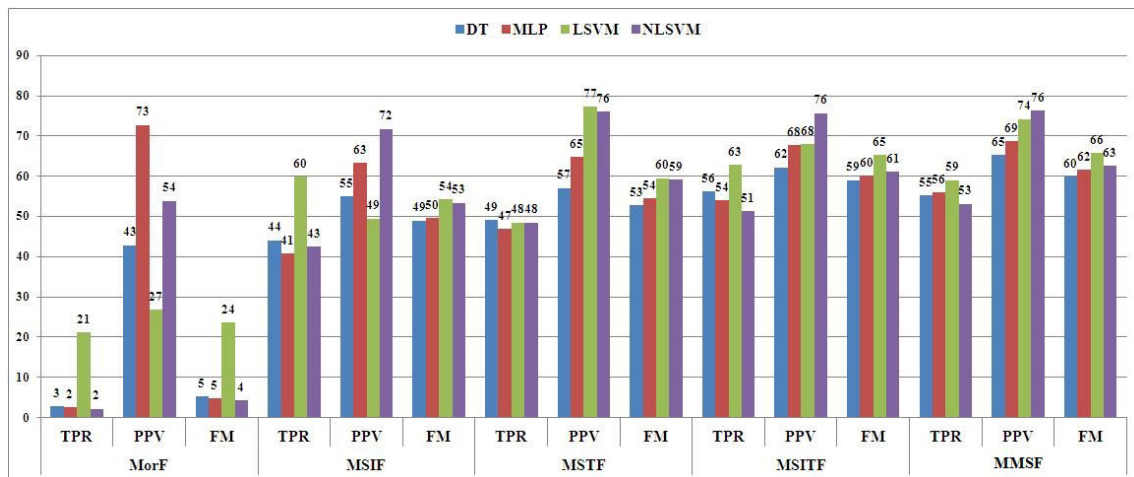


Figure 4.14: Classification results on different subset of MMSF using 5-Fold CV.

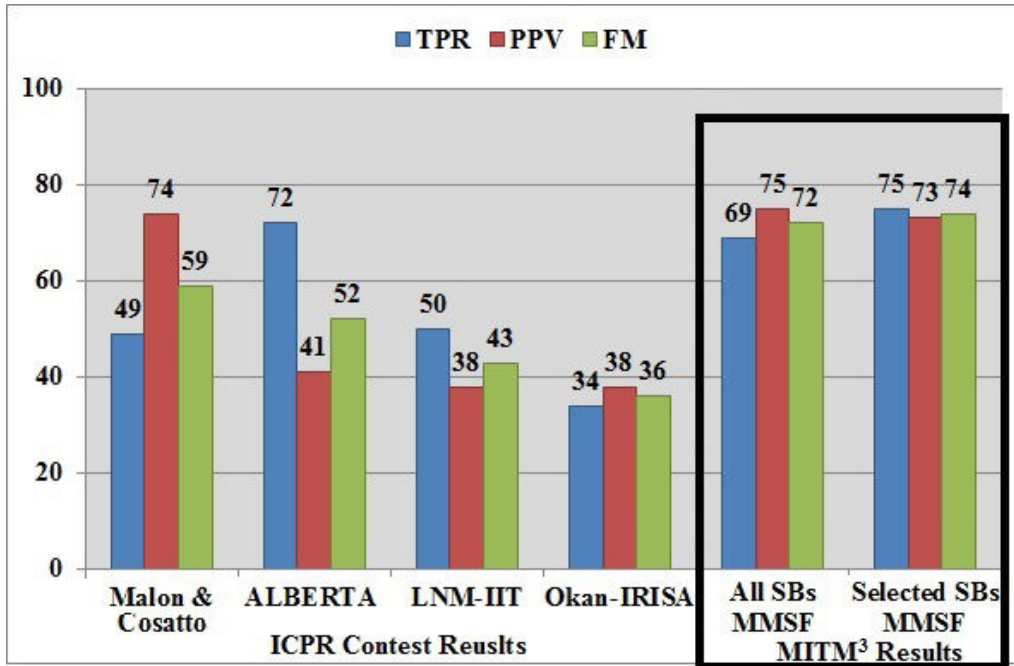


Figure 4.15: Comparison of (MITM³) framework results with MITOS contest result. Vertical dotted line is used to separate the result of contestant's method and proposed method.

performed selection of focal separately for each SB. However, as the best focal plane is the same for all of the SBs except SB 4, in future we could use the best focal plane computed on one SB only. It is not necessary to compute it for each SB separately. In other words, finding the best focal plane and finding the best SBs are separable problems.

The best FM for candidate detection was achieved on SB 8. The fact that the proposed framework achieved better results when using SBs 8, 9 and 7 for candidate detection than when using the full spectrum (SB 1) supports the claim that MSI improves the accuracy of the framework. As SBs 7, 8 and 9 actually overlap in terms of spectrum; it would be interesting to apply spectral unmixing between SBs 7, 8 and 9 to see if it can further improve the results. The results illustrate clearly the improved accuracy resulting of the SB selection process.

Separate training and evaluation sets of MITOS dataset [1] have been used for training and evaluation of the proposed framework. The comparison of proposed framework results with MITOS contest results [150] are shown in Figure 4.15. Malon and Cosatto [113] method ranked first during the contest with 58.90% FM. Recently, Tripathi *et al* [170] proposed a 2-sieve model for mitosis detection and reported 85% TPR, 60% PPV and 70.20% FM without balancing the training set and 82% TPR, 73% PPV and 77.20% FM with training set balancing. In comparison with MITOS contestants, the proposed framework compute features on selected SBs that have higher mitosis absorption response as compared to other tissue components. Using selected SBs MMSF and L-SVM classifier, we managed to achieve highest FM (73.74%) without balancing training set.

Figure 4.16 illustrates the ROC curve [190] obtained with selected SBs MMSF. Figure 4.17 illustrates the margin curve between the probability predicted for mitosis class and the highest probability predicted for the non-mitosis class. This clearly demonstrates that our new proposed framework results in an improved ability to distinguish mitosis from other objects.

In order to study the MMSF for classification of mitosis figures, we performed SBs selection by studying which SBs have minimum redundant and maximum relevant information

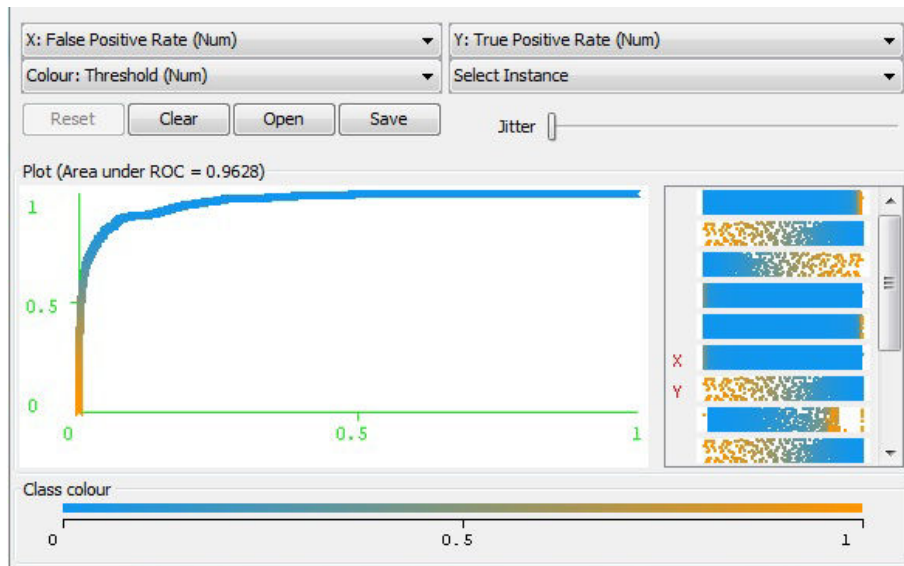


Figure 4.16: The ROC curve of classification result using selected SBs MMSF with L-SVM classifier.

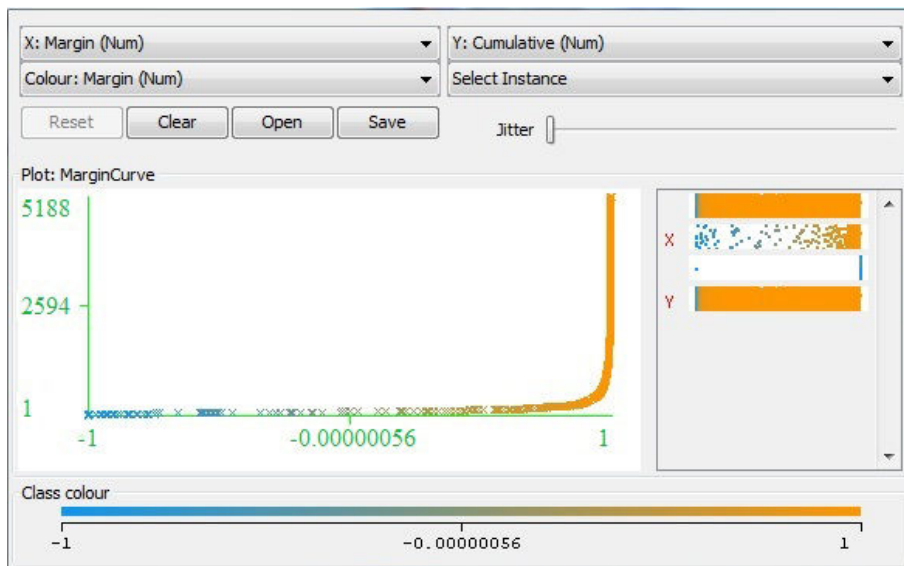


Figure 4.17: The margin curve illustrating the prediction margin between mitosis and non mitosis class.

for mitosis classification. Figure 4.12 show the results (FM) of mitosis classification using a selection of SBs based on mRMR ranking. With top two SBs, we match the best results achieved during MITOS contest [150]. With top eight SBs, we achieve 73.74% FM which is a 25% increase over the best know results. Adding more SBs (SBs 4 and 5), only has a negative impact on the results. This concludes that SBs 4 and 5 are irrelevant for mitosis discrimination. We also performed candidate classification using each SBs MMSF and results are shown in Figure 4.11. The results of classification also support the mRMR ranking of SBs and absorption response of SBs. To illustrate the possible correlation between selected SBs and the staining characteristics of the spectral samples, the shape of H–E plot in Figure 4.5 validated the proposed selection of SBs.

We also used the analysis of different subsets of multispectral features as complement to the analysis of performance on MMSF. Specifically, we have shown in experiment that multispectral data contains more discriminant information for detection of mitosis than white SB, which covers the full visible spectrum. According to experiment using red, green and blue SBs MMSF, red SBs (8,9,0) are more helpful for mitosis discrimination as compared to blue SBs (2,3,4) and green SBs (5,6,7). An important finding is that green SBs have more information than blue SBs, which might be reason of SB 6 and 7, which contain higher absorption response for mitosis nuclei as compared to other tissue components. In addition, multispectral texture features have more information for mitosis classification as compared to multispectral intensity features.

By comparing patch and regions features computed in MITM³ framework, the result of mitosis classification is better in patch based features computation. One limitation of region-based features is the inappropriate segmentation of candidates. The segmentation method is efficient and robust on nuclei having homogenous regions, while it is not robust for other type of regions and under segment most candidate regions. The computation of features on under segmented candidate regions might be reason for poor classification results. By taking the advantage of neighboring regions of each candidate, the patch based features carries more discriminate information for mitosis detection.

From the classifiers point of view, NL-SVM classifier always detects very few FP and results in high PPV. The L-SVM classifier always detects maximum number of mitosis, but with high FP as well. MLP classifier also shows good results after L-SVM classifier.

The majority of mitosis missed by the proposed framework is significantly different from the detected mitosis. Missed mitosis has very light nuclear material and a small size as compared to detected mitosis. In addition, some missed mitosis figures are clustered in small heterogeneous regions. Furthermore, most of FP is lymphocytes that look very similar to mitosis figures.

Our proposed framework achieve better performance on Aperio and Multispectral dataset as compared to Hamamatsu Dataset as shown in Figure 4.18. The important information is that our frameworks get better classification results on patch based texture features as compared to region based texture features. Examples of selective mitosis patches for each datasets are shown in Figure 4.19. In Hamamatsu dataset, texture is blurred, probably due to bad focus during scan of the slide. This might be reason for poor classification results on Hamamatsu Dataset. While Aperio and Multispectral patch size is almost similar but Hamamatsu patch size is smaller. It might be reason the background texture in Hamamatsu Dataset is not informative for mitosis classification as in case Aperio and Multispectral Dataset. Even we do not perform any training set balancing in Multispectral dataset; we get almost similar classification results in Multispectral Dataset as compared to Aperio Dataset. It would be interesting to perform training set balancing in Multispectral to see if it can improve the classification results and give better results than Aperio Dataset.

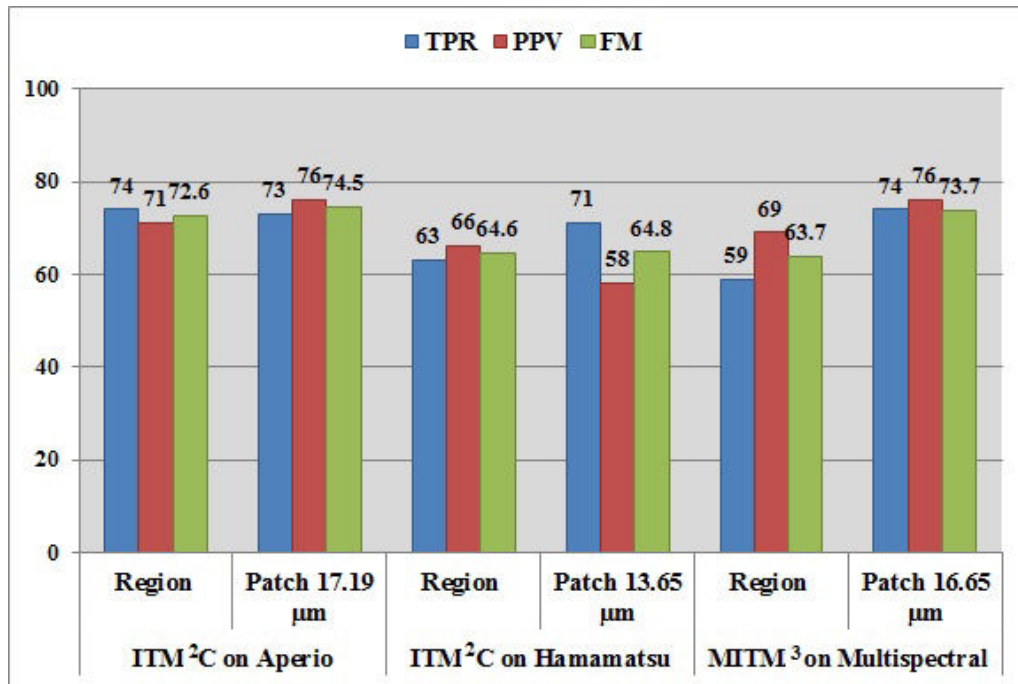


Figure 4.18: The proposed framework results on MITOS Aperio, Hamamatsu and Multispectral Datasets.

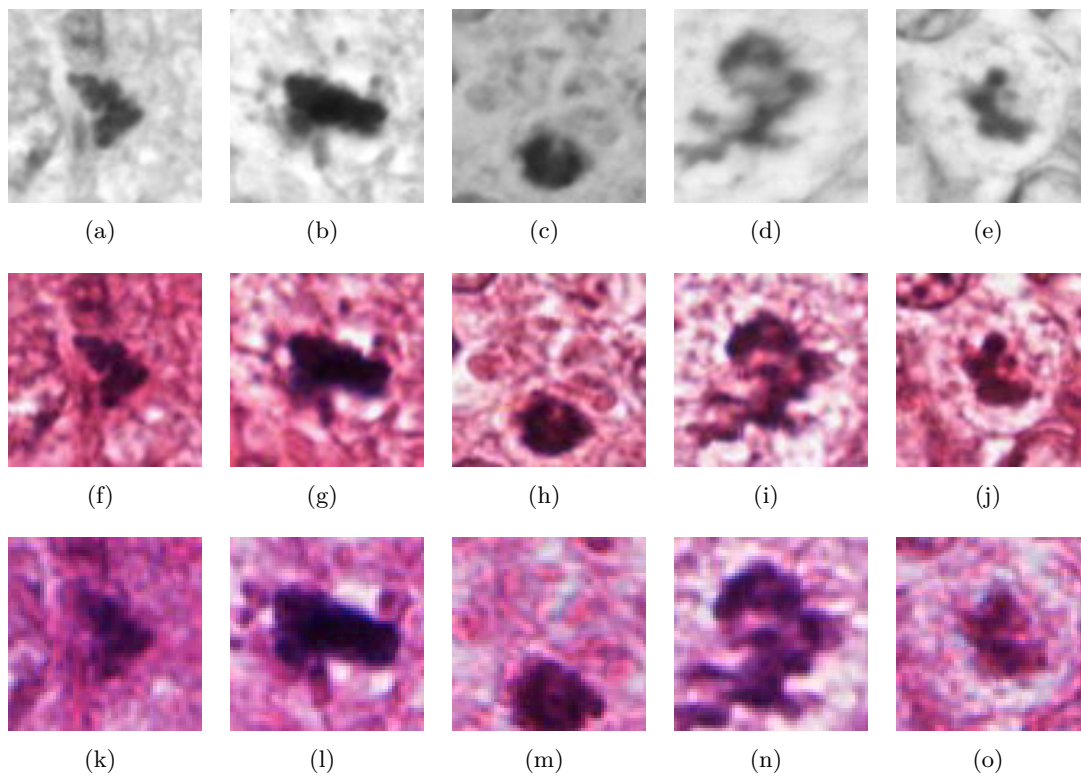


Figure 4.19: Multispectral (first), Aperio (second row) and Hamamatsu (third row) patches of mitosis on which texture features are computed.

4.6 Conclusion

In this chapter, we have proposed an automated mitosis detection framework for breast cancer MSI based on morphological & multispectral statistical features. First, focal plane selection is performed using maximum gradient information. Based on MI of SBs and spectral absorption of different tissue components and stains, SBs were selected for candidate detection and feature computation. Candidate detection was performed on the SBs that have relatively higher MI and mitotic absorption spectra. Then, MMSF are computed for each candidate in eight selected SBs, a highly efficient model for capturing spectral and spatial features for object discrimination. Multispectral texture features have more information for mitosis classification as compared to multispectral intensity features. The proposed framework outperformed the MITOS contest results with 25% improvement of FM. We expect to improve mitosis detection performance by selecting the feature set through the computation of MMSF of candidate regions in selected SBs. In future work, we plan to investigate unmixing of bands as most SBs have overlapping area, which increase redundancy. The pre-selection of the focal plane (or volumes) is also of great importance to reduce the complexity of the dataset and improve the actual performances to reach clinical operational acceptance expected by our professional consortia.

Due to the high-content nature of the WSI, in order to comply to the same operational clinical expectations, the image exploration needs dynamic tools, enabling efficient selection of the microscopic frames (HPF) which need to be analysed. The next chapter is therefore dedicated to describe an orientable 2-Manifold meshes and dynamic sampling framework for WSI analysis. This operational framework, inspired by [174], can be used to model the WSI and select the HPF for CNA and mitosis detection. We therefore propose an extension of existing data structure to handle duality of meshes simultaneously. In addition, we describe the dynamic sampling algorithm for CNA evaluation on WSI in MICO project prototype, in correlation with an existing CNA evaluation module [14], on the dynamically selected HPF.

Chapter 5

Orientable 2 - Manifold Meshes and Dynamic Sampling

Résumé du chapitre

Les images de lames entières sont des images à haute contenu. L'amélioration de l'efficacité de l'analyse est capable de stimuler, dans un proche avenir, un échantillon supplémentaire à analyser. Le volume de ces images de lames entières devient de plus en plus important, menant —à notre avis— à un défi lié au gros volume de données en pathologie numérique. Il est essentiel de soutenir cette tendance menant à une meilleure qualité des soins de santé aux patients. Garder l'exploration de ce gros volume de données dans les limites opérationnelles cliniques nécessite de nouveaux outils d'échantillonnage dynamique efficaces.

Dans ce sens, ce chapitre décrit brièvement des surfaces et des maillages multiples, en particulier les maillages 2-variété (variété topologique de dimension 2) orientables. Nous proposons une extension d'ITK [77] pour gérer simultanément des mailles primales et duales. Nous introduisons la structure de données, une extension de `itk::QuadEdgeMesh`, un filtre pour calculer et ajouter à une structure le double d'un maillage existant, et un adaptateur qui fournit au processus de pipeline le maillage dual comme s'il était un `itk::QuadEdgeMesh` natif ITK. La nouvelle structure de données `itk::QuadEdgeMeshWithDual` est une extension de la structure `itk::QuadEdgeMesh` déjà existante dans ITK [63], qui incluait déjà par défaut la topologie. Elle apporte en plus la gestion de la géométrie duale. Deux types de maillages primales ont été particulièrement illustrés : maillages triangulaires / simplex et Voronoi / Delaunay. Nous implémentons également une stratégie d'analyse de lames virtuelles pour l'évaluation de l'atypie nucléaire au moyen d'un échantillonnage dynamique utilisant les diagrammes de Voronoi.

5.1 Introduction

WSI is a high-content image. Increasing the efficiency of the analysis is able to stimulate, in the near future, additional sample to be analysed. The volume of these WSI is therefore increasingly important, leading - in our opinion - to an important big data challenge in future digital pathology. This trend is essential to be sustained, on the way to an increase quality of patient healthcare. Keeping the exploration of this big database still in clinical operational limits needs novel efficient dynamic sampling tools.

In clinical practice, mitosis and nuclear atypia score is determined by analysing the WSI or ROI. But in digital pathology, detection of mitosis and nuclear atypia is usually done on a HPF as shown in Figure 5.1(a). There are usually more than thousand HPF per WSI. So, computing

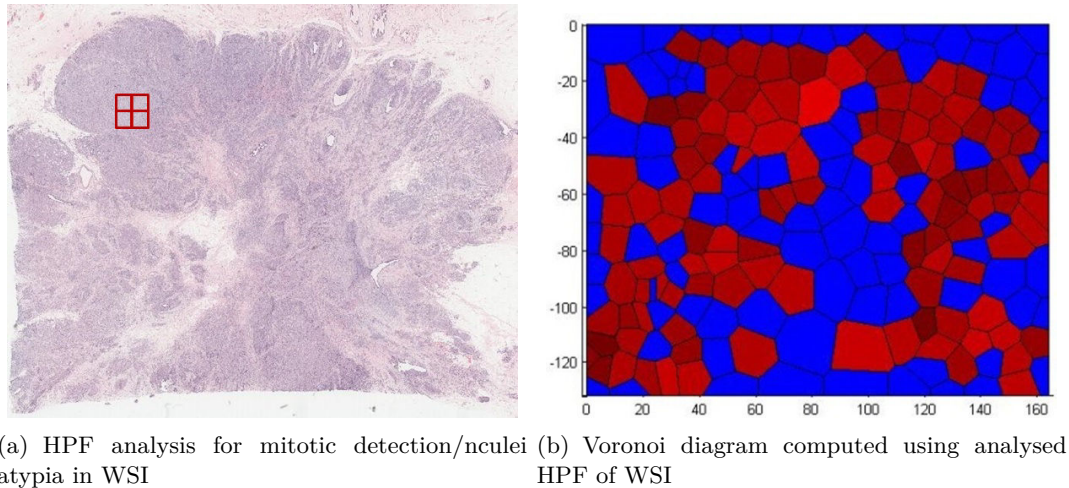


Figure 5.1: Switching from HPF to WSI Analysis.)

mitosis and nuclear atypia score on WSI after analysing more than thousand HPFs is a time consuming and tedious job. A dynamic sampling framework is proposed to select HPF for mitosis and nuclear atypia detection from Voronoi diagram that is computed using already analysed HPFs 5.1(b).

In this sense, this chapter briefly describes manifold surfaces and meshes, especially orientable 2-manifold mesh. We propose an extension of ITK [77] to handle both primal and dual meshes simultaneously. We present the data structure, an extension of *itk::QuadEdgeMesh*, a filter to compute and add to the structure the dual of an existing mesh, and an adaptor which let a downward pipeline process the dual mesh as if it is a native *itk::QuadEdgeMesh*. The new data structure, *itk::QuadEdgeMeshWithDual*, is an extension of the already existing *itk::QuadEdgeMesh* [63], which already included by default the due topology, to handle dual geometry as well. Two types of primal meshes have been specifically illustrated: triangular / simplex meshes and Voronoi / Delaunay. We also implement incremental scheme based dynamic sampling algorithm using Voronoi / Delaunay for real time evaluation of CNA on WSI.

5.2 Surfaces and Meshes

5.2.1 Notions of deformable surfaces

A deformable surface can be characterized by vector of shape parameters $q = (q_1, \dots, q_{n_q})^T$ and vector of deformation parameters $d = (d_1, \dots, d_{n_d})^T$ that controls the application of a global transformation T_d on the surface:

$$\begin{aligned} S(q, d) = T_d(S_q) : \mathbb{R}^{n_q} \times \mathbb{R}^{n_d} \times \Omega &\longrightarrow \mathbb{R}^3 \\ (q_1, \dots, q_{n_q}, d_1, \dots, d_{n_d}) &\longmapsto T_d(P_q(r, x)) \end{aligned} \quad (5.1)$$

where (r, x) denotes a point of surface parameter domain Ω .

A deformable surface can be represented by continuous and discrete models. With discrete representation, the geometry of surfaces is only known at a finite set of points. Continuous surfaces representation must be discretised for computational needs but they offer the ability to compute differential quantities such as surface norm or curvatures almost everywhere on the surface. Most discrete models are meshes defined as a set of points with some connecting relation that includes topological constraints.

5.2.2 Orientable 2-Manifold Mesh: A discrete real-world object

The surfaces of real world objects are oriented 2-manifolds. Those are usually represented in computer using meshes, which are the sampled, discrete version of the underlying, supposedly continuous surface. The definition of surface mesh is of combinatorial nature [86], that improves reasoning about data structure like the same facet cannot appear on both sides of an edge. The surface mesh is a union $C = V \cup E \cup F$ of three disjoint sets together with an incidence relation where V the vertices, E the edges and F the facets of the mesh. The incident relation on C must be symmetric. No two elements from the same set V, E, F are incident. There are four additional conditions: (1) every edge is incident to two vertices, (2) every edge is incident to two facets, (3) for every incident pair of vertices or facets, there are exactly two edges incident to both and (4) every vertex and every facets is incident to at least one other element. The neighbourhood of a vertex are edges and facets which are incident to that vertex. Thus, the neighbourhood decomposes into disjoint cycles, where each cycle is an alternating sequence of edges and facets.

A surface mesh is 2-manifold if (1) each edge is incident to only one or two facets and (2) the facets incident to a vertex form a close or an open fan i.e. for each point on a 2-manifold there exists a neighbourhood that is homeomorphic to the open disc. If every vertex has a closed fan, the given 2-manifold has no boundary. If a vertex has an open fan, then edges are incident to one facet; they are called border edges and they form the boundary of the 2-manifold mesh. A non-manifold example would be two tetrahedra glued together at a single vertex or common edge as shown in figure 5.2. A mesh is a 2-manifold if and only if the neighbourhood of each vertex decomposes into a single cycle. The next distinction is between orientable and non-orientable mesh. A surface mesh is oriented if each cycle around a facet is oriented and if, for each edge, the two cycles of its two incident facets are oriented in opposite direction. A 2-manifold mesh is orientable if there exists such an orientation. This new data structure only considers orientable 2-manifolds mesh representation with and without boundary.

Genus is a topologically invariant property of a surface defined as the largest number of non-intersecting closed curves that can be drawn on the surface without separating it. Also, it is a complete invariant in the sense that, if two orientable closed surfaces have the same genus, they must be topologically equivalent. The genus of a surface is related to the Euler characteristic χ . For an orientable surface such as a sphere (genus 0) or torus (genus 1), the relationship is

$$\chi = 2 - 2g - b \quad (5.2)$$

with g being the genus, and b being the number of borders (for non-closed surfaces).

Given an arbitrary polygonal mesh τ of a regular region $R \subset S$ of a surface S , we shall denote by F the number of polygonal faces, by E the number of sides (edges), and by V the number of vertices of the triangulation. Another way to compute the Euler characteristic is then

$$F - E + V = \chi \quad (5.3)$$

Special cases of discrete 2-manifolds of interest to us are triangular meshes and simplex meshes illustrated in figure 5.3.

5.2.3 Special Case A: Triangular Meshes

A common representation of discrete surfaces are triangulation τ for which the surface $\mathbb{R} \subset S$ is composed of a set of adjacent triangles $T_i, i = 1, \dots, n$, such that

- $\cup_{i=1}^n T_i = \mathbb{R}$.
- If $T_i \cap T_j \neq \phi$, then $T_i \cap T_j$ is either a common edge of T_i and T_j or a common vertex of T_i and T_j .

Each triangles of a triangulation shares at least one of its edge with a neighbouring triangle.

Triangles being the simplest polygon that can represent a surface, it has been used intensively in computer graphics and is still ubiquitous today in surface representations and corresponding data formats.

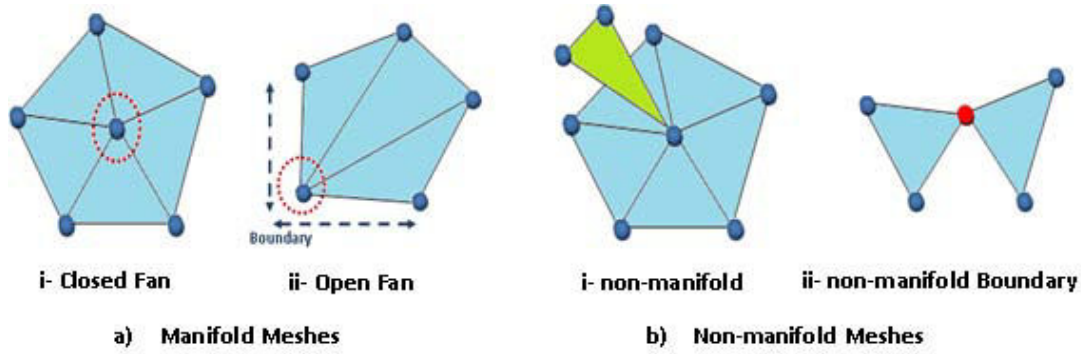


Figure 5.2: Examples of Meshes

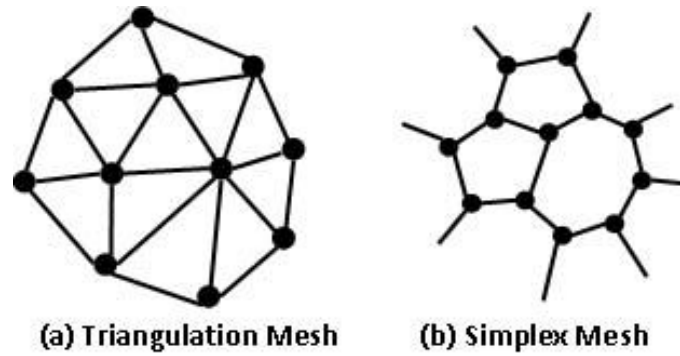


Figure 5.3: Triangulation and Simplex Mesh

5.2.4 Special Case B: Simplex Meshes

Simplex meshes are used for discrete surface representation. Simplex meshes have two main properties, (1) each vertex is adjacent to a fixed number of neighbouring vertices: 2 for a contour (1-simplex mesh), 3 for a surface (2-simplex mesh) and 4 for tetrahedron (3-simplex mesh, not treated here); and (2) the topology of a 2-simplex mesh is dual of a triangulation. A k -simplex can be referred to a $(k + 1)$ -connected mesh. For instance, a segment of non-zero length is a 1-simplex, a triangle (polygon) of non-zero area is a 2-simplex and a tetrahedron of non-zero volume is a 3-simplex mesh. Formally, a k -Simplex Mesh (kSM) of \mathbb{R}^3 is defined as a pair $(V(M), N(M))$ [40] where:

$$V(M) = \{P_i\}, \{i = 1, \dots, n\}, P_i \in \mathbb{R}^3 \quad (5.4)$$

$$N(M) : \{1, \dots, n\} \longrightarrow \{1, \dots, n\}^{k+1} \quad (5.5)$$

$$i \longmapsto (N_1(i), N_2(i), \dots, N_{k+1}(i))$$

$$\forall_i \in \{1, \dots, n\}, \forall_j \in \{1, \dots, k + 1\}, \forall_l \in \{1, \dots, k + 1\}, l \neq j \quad (5.6)$$

$$N_j(i) \neq i$$

$$N_l(i) \neq N_j(i) \quad (5.7)$$

$V(M)$ is the set of vertices of M and $N(M)$ is the associated connectivity function. Equations (5.6) and (5.7) present a mesh from exhibiting loops. It is important to make a distinction between the topological nature of a mesh represented by its connectivity function $N(M)$ and its geometric nature corresponding to the position of its vertices $V(M)$.

The structure of a simplex mesh is the one of a simply connected graph and does not in itself constitute a new surface representation. The simplex mesh representation has several desirable properties that make them well suited for the recovery of geometric models from range data. The characteristics of simplex mesh for discrete surfaces include generality (represent all types of

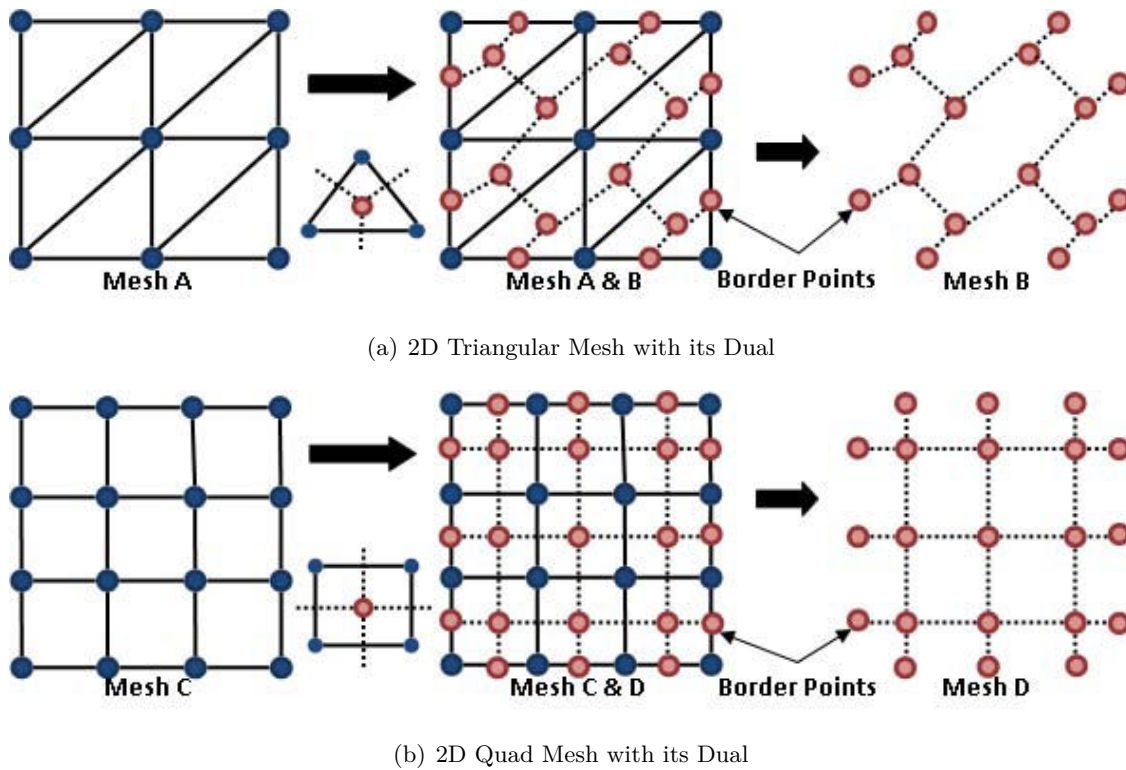


Figure 5.4: Examples of Dual Meshes. (We also sampled the border points.)

orientable surfaces regardless of their genus and end numbers), simplicity (minimum number of vertices to represents a surface or shape) and adaptability.

5.3 Duality

5.3.1 Notion of Duality

We define A and B to be dual surface meshes i.e., B is dual of A and vice versa, if the following conditions are satisfied.

- The number of vertices of A is the same as the number of face of B , so that they can be put into one-to-one correspondence.
- The number of vertices of B is the same as the number of face of A , so that they can also be put into one-to-one correspondence.
- Each pair of vertices of A that map to adjacent faces in B is joined by an edge which can be put into correspondence with the common edge of the associated pair of faces of B . The edges that join adjacent vertices of B can be put into the same correspondence with the common edges of the associated pairs of elements of A .

Figure 5.4 illustrates the duality of meshes. Each boundary edge of a face in mesh A is put into correspondence with a half-open edge in mesh B which starts at the corner corresponding to that face in A as shown in figure 5.4.

5.3.2 Triangulation - Simplex Duality

One of the most interesting way of considering simplex meshes is through duality of triangulations. The structure of a k -simplex mesh is indeed closely related to the structure of a k -triangulation. A k -triangulation of \mathbb{R}^d is composed of p -simplices ($1 \leq p \leq k$) which are the p -faces of the triangulation. We define a p -face of a k -simplex mesh as being the dual of a $(k - p)$ simplices of a k -triangulation. For instance, a 1-face of a 2-simplex mesh is an edge and a 2-face of

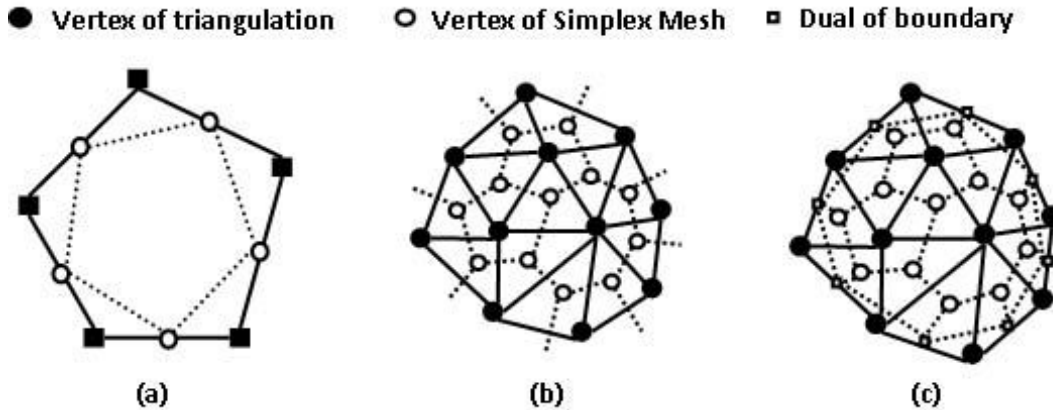


Figure 5.5: a) A 1-Simplex mesh and its dual; b) A 2-Simplex mesh and its dual triangulation; c) same as (b). The dual of the triangulation boundary is considered to extract the simplex mesh.

a 2-simplex mesh is a polygon. In general, a p -face of a k -simplex mesh is a $(p - 1)$ -simplex mesh and is, therefore, made of q -faces ($q < p$). A simplex mesh is said to be regular if all p -faces have the same number of vertices.

Simplex meshes are dual of triangulations. Thus, their connectivity functions $N(M)$ are mapped by an homeomorphism. Simplex meshes are topologically equivalent to triangulations but not geometrically equivalent. We can define a topological transformation that associates a k -simplex mesh to a k -triangulation. This transformation is pictured in figure 5.5 and considers differently the vertices and edges located at the boundary of the triangulation from those located inside.

5.3.3 Delaunay - Voronoi Duality

Taking a set of points \mathcal{P} in \mathbb{R}^3 , the Delaunay triangulation of \mathcal{P} is a specific triangulation of \mathcal{P} that respects the Delaunay criterion stating that no point of \mathcal{P} should be inside of the circumference circle of any triangle of the triangulation of \mathcal{P} . Taking a set of points \mathcal{P} in \mathbb{R}^3 , the Voronoi diagram (or tessellation) is the partition of \mathbb{R}^3 into n polyhedral regions such as each region T has a set of points in \mathbb{R}^3 which are closer to T than to any other region.

The Voronoi diagram is the dual of the Delaunay triangulation, and the Delaunay triangulation is the dual structure of the Voronoi diagram. By dual, we mean to draw a line segment between two Voronoi vertices if their Voronoi polygons have a common edge, or in more mathematical terminology: there is a natural bijection between the two which reverses the face inclusions. The duality between Delaunay triangulations and Voronoi diagram is geometric because it depends on the position of its vertices.

5.4 Implement Duality in ITK

5.4.1 Existing Data Structure for Meshes in ITK

The *QuadEdgeMesh* data structure in ITK, as depicted in figure 5.6, can handle discrete 2-manifold surfaces. It actually stores the geometry and both primal and dual topology. It has a constant complexity local access on modifications. The *QuadEdgeMesh* data structure is a 3 layers structure in which the bottom layer is called QuadEdge (*QE*) layer that represents the topology, the intermediate layer is called QE Geometric (*QEGeom*) layer that links topology and geometry and finally the upper layer is native to ITK called ITK layer. The *QE* data structure is presented in detail in [63]. For each edge, there are 4 *QEs* in the structure as illustrated in figure 5.6(b). It contains two primal *QEs* and two dual *QEs*. For the sake of simplicity, we only draw connection for one point and one face from *QE* to *QEGeom* and *QEGeom* to ITK layer as shown in figure 5.6(b),

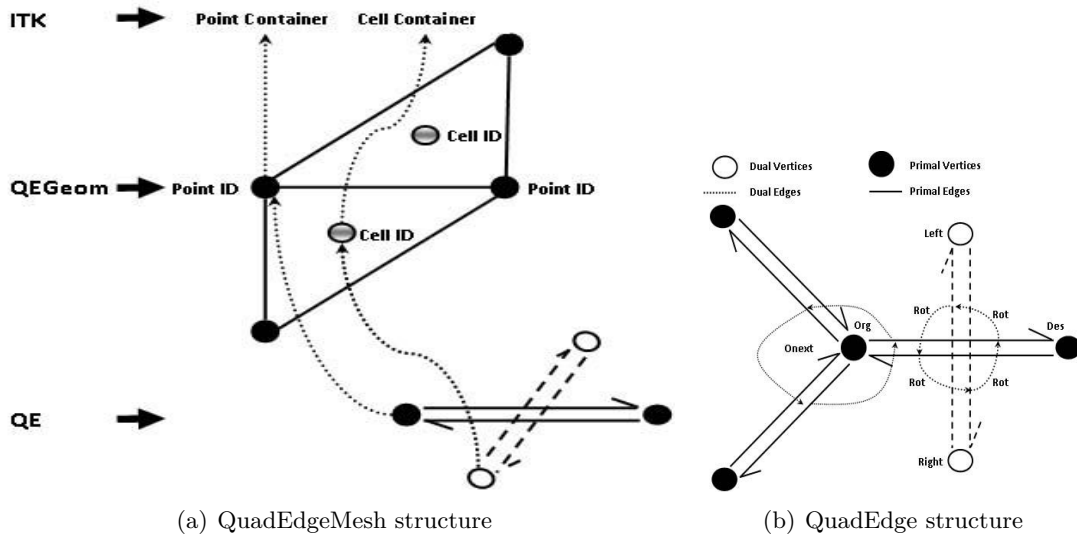


Figure 5.6: Existing data structures

Table 5.1: QuadEdgeMesh Data Structure

	Primal	Dual
Geometry	Yes	<u>No</u>
Topology	Yes	Yes

conversely both points and faces are equally linked in the data structure. This data structure only need three operators as *Rot*, *Onext* and *Splice* to implement all other modifications (Euler operator) and accessibility of the mesh.

Currently, *QuadEdgeMesh* data structure have topological duality but lack geometrical duality as represent in table 5.1. There are only few filters available in ITK that transform triangular mesh to simplex mesh but it is specific not generic to duality. Additionally, in many cases it is of much interest to have both representations of a discrete surface directly integrated in the structure. Our contribution includes an extension of data structure that contains both primal and dual mesh simultaneously, a filter that transforms primal mesh to primal/dual mesh just using single data structure and an adaptor for displaying dual mesh.

5.4.2 Extension in data structure, *QuadEdgeMeshWithDual* data structure

We create a new class *itk::QuadEdgeMeshWithDual* derived from *itk::QuadEdgeMesh*. This class now stores both primal and dual mesh simultaneously. The new design of *QuadEdgeMeshWithDual* data structure contains double reference i.e., one for primal point to dual cell and one for primal cell to dual point as depicted in figure 5.7(a). For the sake of simplicity, we only draw connection from QE layer to QEGeom layer and QEGeom layer to ITK layer for one point and one face instead of both points and both faces as shown in figure 5.7(a). The primal and dual overlapping structures of connections at QEGeom layer is shown in figure 5.7(b).

Furthermore, this class contains three new containers; *DualPointsContainer* for dual points, *DualCellsContainer* for dual cells and *DualEdgeCellsContainer* for boundary edges and three new functions; *AddDualPoint* for adding dual point, *AddDualFace* for dual cells (polygon) and *AddDualEdge* for boundary edges.

In order to keep the primal-dual references in a single data structure, we have two design options. In first design, we use to maintain two look up tables; one table for storing references

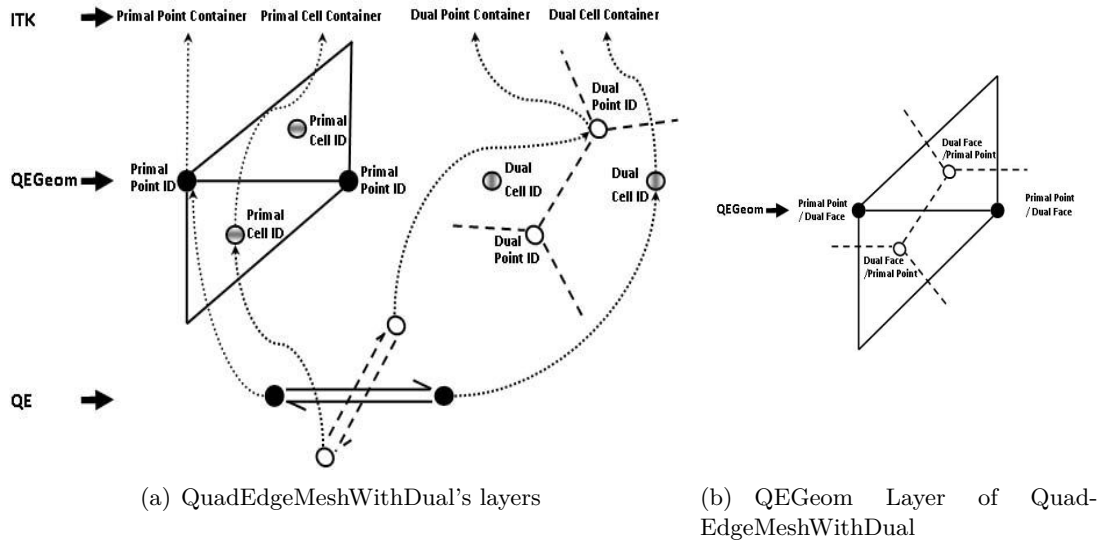


Figure 5.7: QuadEdgeMeshWithDual data structure

Table 5.2: Summaries of changes in new data structure

		Old Data Structure	New Data Structure
Changes	OriginRef Type	Point ID Cell ID	<i>Pair < Point, Cell ></i> <i>Pair < Cell, Point ></i>
Additions	Dual Containers	-	<i>DualPointsContainer</i> <i>DualCellsContainer</i> <i>DualEdgeCellsContainer</i>
Not yet implemented	Dual Data Containers	-	-

of primal cell to dual point and second table for primal point to dual cell. The advantage of this approach is backward compatibility of code and test cases. The bad side of this design is to maintain these tables that having the complexity $n \log(n)$ causing severe degradation of performance in case of large mesh. In second design, we modify the existing data structure by adding two reference pair; primal point to dual cell and primal cell to dual point as shown below. With this design, no look up table is required to maintain the primal and dual references. So it is very efficient approach but not compatible with respect to previous code and test cases.

```
typedef GeometricalQuadEdge<
    std::pair<PointIdentifier, CellIdentifier>,
    std::pair<CellIdentifier, PointIdentifier>,
    PrimalDataType,
    DualDataType
> QEPrimal;
```

A summaries of changes in new data structure can be depicted in Table 5.2

5.4.3 Primal to primal/dual filter

In order to transform primal mesh into dual mesh, we also create a new filter called *itk::QuadEdgeMeshToQuadEdgeMeshWithDualFilter*. This filter is templated with *QuadEdgeMeshWithDual* data structure. This filter generate dual mesh from primal mesh in three phases; first phase is computing dual point from primal cells, second phase is computing dual cells from primal points, and in last phase, primal borders edges are used to generate dual border cell.

We also implement a new adaptor for connecting the dual mesh natively to a downward pipeline.

5.4.4 Dual point functor

As explained before, there is no geometrical duality between the primal and the dual. Therefore any formula that computes points that satisfy the criteria of duality detailed in section 5.3.1 can be used. Not to restrict ourselves to a single option that may limit the application of the filter, a functor is used to compute the dual point. Depending on the case faced, the user is able to choose from the already two existing dual point functor, or use his own functor. Except from the classic ITK macro, *typedef* definition and constructor/destructor, the functor has only one method where the process is done.

```
template< class TInputMesh, class TOutputMesh=TInputMesh>
class DualPointFunctor
{
typedef typename TInputMesh::CellsContainer    CellsContainer;
typedef typename CellsContainer::ConstIterator CellIterator;
...
inline OutputPointType
operator() ( const TInputMesh* primalMesh, CellIterator cellIterator )
    {...}
};
```

Barycentre By default, the barycentre of each face is used to compute the location of the dual point. It has the advantage to be relatively straightforward to compute, to compute a dual point, which is always located within the face, and to work with any kind of face. The following equation is used to compute the centre

$$M = \frac{P_1 + P_2 + \dots + P_n}{n} \quad (5.8)$$

where P_1, P_2, \dots, P_n are the points retrieved from the current cell.

Circumcentre The circumcentre is a particular dual point of triangle mesh. It is the centre of the circumference circle of a triangle and is determined by the crossing point of the perpendicular bisectors. As such, it is not always within the face, and more costly to compute. The interest of this is in the case of the Delaunay triangulation in order to obtain its dual, the corresponding Voronoi tessellation. The following equation is used to compute the centre

$$M = P_1 + \frac{|P_3 - P_1|^2 [(P_2 - P_1) \times (P_3 - P_1)] \times (P_2 - P_1) + |P_2 - P_1|^2 [(P_3 - P_1) \times (P_2 - P_1)] \times (P_3 - P_1)}{2 |(P_2 - P_1) \times (P_3 - P_1)|^2} \quad (5.9)$$

where P_1, P_2, P_3 are the points retrieved from the current triangle, and \times the cross product. In order to simplify the calculus and avoid the use of square roots the edge length are squared and the coordinates of all the point relative to the first point P_1 are used. Due to the floating-point errors such solution may be unstable in the case of the denominator is close to 0. To prevent such case, the exact geometric predicate implemented for ITK [122] is used for the cross product calculation.

5.4.5 Dual borders calculation

As shown in figure 5.4, the dual of a primal mesh that contain a border is not a closed mesh. The dual edge point obtain from the primal border are not included into any faces of the dual. This representation may be problematic to some other process which may not take into account the *EdgeContainer* that store the edges in the *QuadEdgeMesh* structure, and therefore discard edges that are not part of any face (*i.e.* Mesh writing filter). If the dual edge points are not computed, the effect still occurs but is less important that in the previous case. Another option is to create a border by connecting the dual edge point, however this solution may lead to some flipped triangles in specific configuration. The *SetBorders()* methods allow the user to decide how the filter should manage the borders (Fig. 5.8). By default the filter will compute the dual edge point and create a border to the dual mesh.

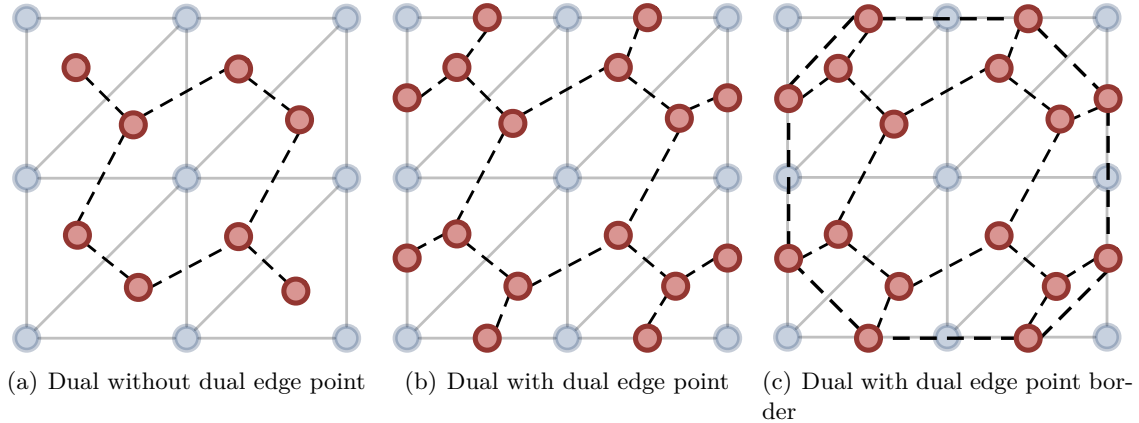


Figure 5.8: Dual borders management options

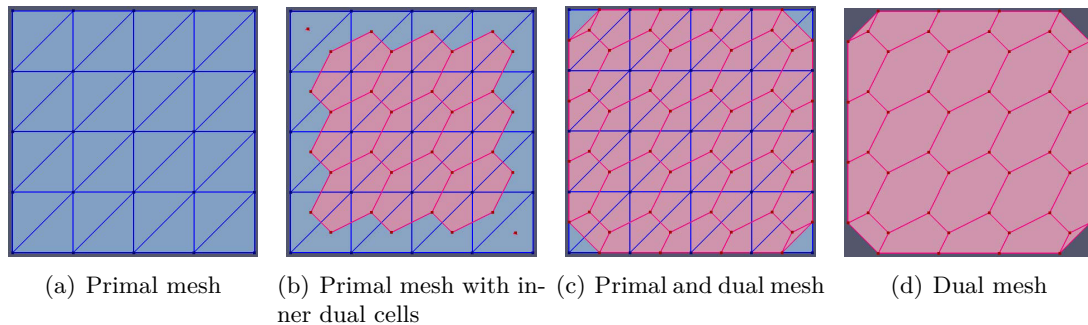


Figure 5.9: Primal to dual mesh

5.5 Validation

5.5.1 Test on planer triangular to simplex mesh with and without holes

We create a square triangulated (primal) mesh as shown in figure 5.9(a). Green color represents primal points and cells. From this primal mesh, we would try to generate dual mesh. First, we generate dual points using the *BarycentreDualPointFunctor* on the primal cells as shown in figure 5.9(b) with red points. We add these dual points in *m_DualPointsContainer* of *itk::QuadEdgeMeshWithDual* by using *AddDualPoint()*. Second, we iterate around each primal point to form dual cells and add dual cell in *m_DualCellsContainer* of *itk::QuadEdgeMeshWithDual* by using *AddDualFace()*. By doing this we generate all dual cells except boundary cells. Dual cells are represented by red color in figure 5.9(b).

In order to tackle borders, first we get boundary edges of primal mesh. Select one boundary edge from list; create a new point (dual) in the middle of edge and add in *m_DualPointsContainer* of *itk::QuadEdgeMeshWithDual* by using *AddDualPoint(...)*.

In figure 5.9(c), red points on border lines represent boundary points of dual mesh. Then, find the dual point associated with the face on the left and make an edge between these two dual points. Now iterate along left triangle to form dual cell and add this dual cell into *m_DualCellsContainer* of *itk::QuadEdgeMeshWithDual* by using *AddDualFace()*. In figure 5.9(c), red cells represent dual cells. The final dual mesh generated from primal mesh is shown in following figure 5.9(d).

For testing this data structure and filter, we deleted one primal edge and re-run the whole code for getting dual mesh. The snapshot of re-run is shown in figure 5.10.

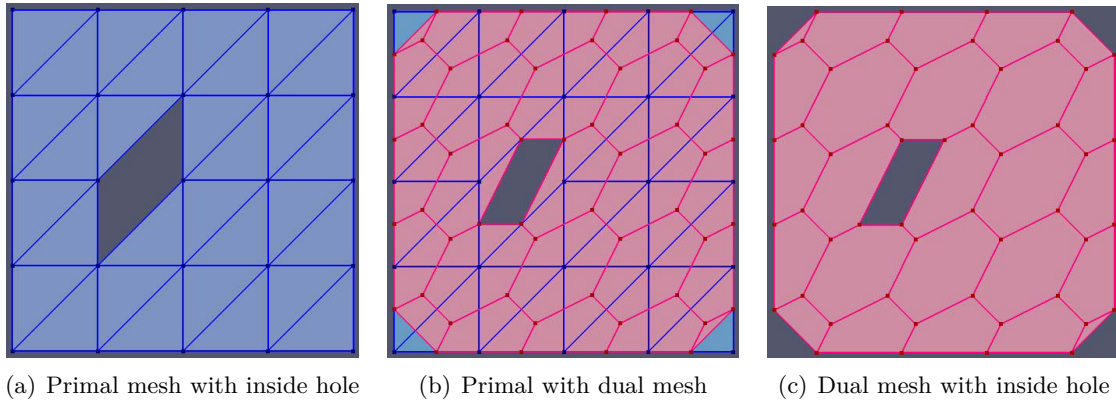


Figure 5.10: Primal to dual mesh with inside hole

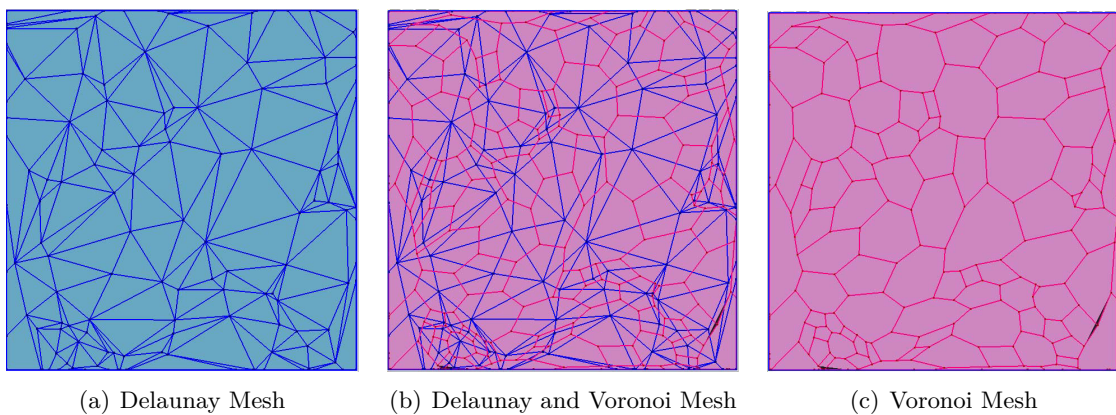


Figure 5.11: Delaunay to Voronoi Mesh

5.5.2 Test with Delaunay to Voronoi

Using the *PointSetToDelaunayTriangulationFilter* [144], we tested this data structure on Delaunay mesh to Voronoi diagram. We input a planer Delaunay mesh into new data structure as shown in figure 5.11(a) and generate the corresponding Voronoi diagram by using *QuadEdgeMeshToQuadEdgeMeshWithDualFilter* and the *CircumcentreDualPointFunctor* as shown in figure 5.11(b). Later, the Voronoi diagram is shown in figure 5.11(c) using new adaptor *itk::QuadEdgeMeshWithDualAdaptor*.

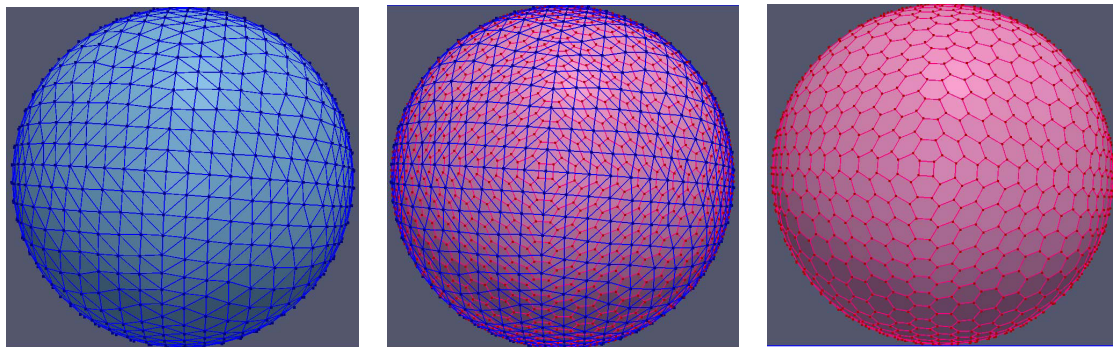
5.5.3 Test on non planar mesh

We perform last test on non-planer mesh. A spherical triangulation mesh can be seen in figure 5.12(a), generated simplex (dual) mesh along with triangulation (primal) mesh can be seen in figure 5.12(b) and finally, simplex (dual) mesh generated with new adaptor can be seen in figure 5.12(c).

5.5.4 Usage

An example of code is provided here. The filter *QuadEdgeMeshToQuadEdgeMeshWithDualFilter* is templated on float and 3 dimensions *itk::QuadEdgeMeshWithDual*.

```
// Typedef definition
typedef itk::QuadEdgeMeshWithDual< float, 3 > MeshType;
```



(a) Non-Planer Triangulation Mesh (b) Non-Planer Triangulation and Simplex Mesh (c) Non-Planer Simplex Mesh

Figure 5.12: Non-Planer Mesh containing (Triangulation and Simplex Mesh)

```

typedef itk::QuadEdgeMeshToQuadEdgeMeshWithDualFilter< MeshType >
                                                    FillDualFilterType;
typedef itk::QuadEdgeMeshWithDualAdaptor< MeshType >  AdaptorType;

typedef itk::VTKPolyDataWriter< MeshType >          MeshWriterType;
typedef itk::VTKPolyDataWriter< AdaptorType >       DualMeshWriterType;

// Create primal mesh
MeshType::Pointer myPrimalMesh = MeshType::New();
CreateSquareTriangularMesh< MeshType >( myPrimalMesh );

// Create dual mesh
FillDualFilterType::Pointer fillDual = FillDualFilterType::New();
fillDual->SetInput( myPrimalMesh );
fillDual->Update();

AdaptorType* adaptor = new AdaptorType();
adaptor->SetInput( fillDual->GetOutput() );

// Write dual mesh
DualMeshWriterType::Pointer writer = DualMeshWriterType::New();
writer->SetInput( adaptor );
writer->SetFileName( "TestSquareTriangularSimplexMesh.vtk" );
writer->Write();

```

5.6 Dynamic Sampling for Cyto-Nuclear Atypia Score in MICO Platform

Automated evaluation of CNA on WSI requires heavy image processing computations. However, it is not possible to compute the CNA on a ROI within reasonable time. Strategies are proposed to overcome the processing power limitations. As the histopathology expert doesn't need to watch every part of the WSI in order to grade it, MICO aims to understand and reproduce this expert behavior during automated grading. The strategy used for CNA evaluation to avoid exhaustive analysis of WSI is dynamic sampling method based on computational geometry. Initially, the WSI is observed by an histopathologist, for slide territories to be annotated using Calopix user interface [4]. Then, the relevant territories are extracted from Calopix information storage system, for them to be split into several HPF frames. Then, 50 of these frames are randomly selected, and

CNA scores are computed within the selected frames using MPP with simple and complex shape objects [14]. These scores are used for the initialization step of a Voronoi diagram [174]. This geometric construction is aimed at approximating the score within a whole Voronoi cell by the score of the frame at its centre, which results in a nearest neighbor approximation. Accordingly, the most undermined areas are at the intersection of multiple cells, i.e., frames containing a vertex. Next frame is selected based on following two criteria: (1) at least one of its neighboring Voronoi cells has a high score that controls the convergence of the method towards areas with high score, and (2) the distance between the new sample and its neighbors is not too short that prevents oversampling. The final overall CNA grade is the grade of the most atypia frame. This strategy is shown in Figure 5.13. Figures 5.14 and 5.15 shows the evolution (200 frames, 300 frames and 400 frames) of a Voronoi driven CNA analysis on two WSIs.

5.6.1 Dynamic Sampling Algorithm

At each iteration, given E the frames already computed nuclear atypia in the WSI, we construct the Voronoi diagram of the centroids of the frames in E denoted VD_E . VD_E is a collection of Voronoi cells $\{v_x \mid x \in E\}$, defined by $v_x = \{p \in I \mid \forall y \in I - \{x\}, \text{Dist}(p, x) \leq \text{Dist}(p, y)\}$. The set of Voronoi vertices, later referred as V_E are the vertices of the planar graph representation of VD_E . V_E share the modesty to be locally the farthest position from their nearest neighbor in E , therefore in the case of our algorithm from already computed frames.

The geometric construction of Voronoi diagram VD_E is performed by approximating the score S within a whole Voronoi cell. This score is computed from score of the frame at its center which results in a nearest neighbor approximation. Accordingly, the most undetermined areas are at the intersection of multiple cells, i.e., frames containing a vertex from V_E . We select our next frame x out of V_E by two criteria; (1) at least one of its neighboring cells has a high score, and (2) the distance between the new sample and its neighbors is not too short. The first condition controls the convergence of the algorithm towards areas with high scores and latter condition prevents over sampling.

The pseudo-code for one iteration of dynamic sampling algorithm is given below:

Input : current frames E , Voronoi diagram VD_E , p , d , \max_E

Output : updated values E , VD_E , \max_E

- i. Compute V_E
- ii. Sort V_E according to decreasing distance to E
- iii. **for** every $x \in V_E$ **do**
- iv. **if** $\text{Dist}(x, E) \geq d$ **then**
- v. **if** $\text{MaxScore}(x) \geq p \times \max_E$ **then**
- vi. $E = E \cup \{x\}$
- vii. update VD_E
- viii. $\max_E = \max(S(x), \max_E)$
- ix. break loop
- x. **end if**
- xi. **else**
- xii. break loop
- xiii. **end if**
- xiv. **end for**

5.7 Conclusion

We have proposed an extension of *itk::QuadEdgeMesh* data structure to handle both primal and dual meshes, simultaneously. The new data structure, *itk::QuadEdgeMeshWithDual*, already include by default the due topology, to handle dual geometry as well. Two types of primal meshes

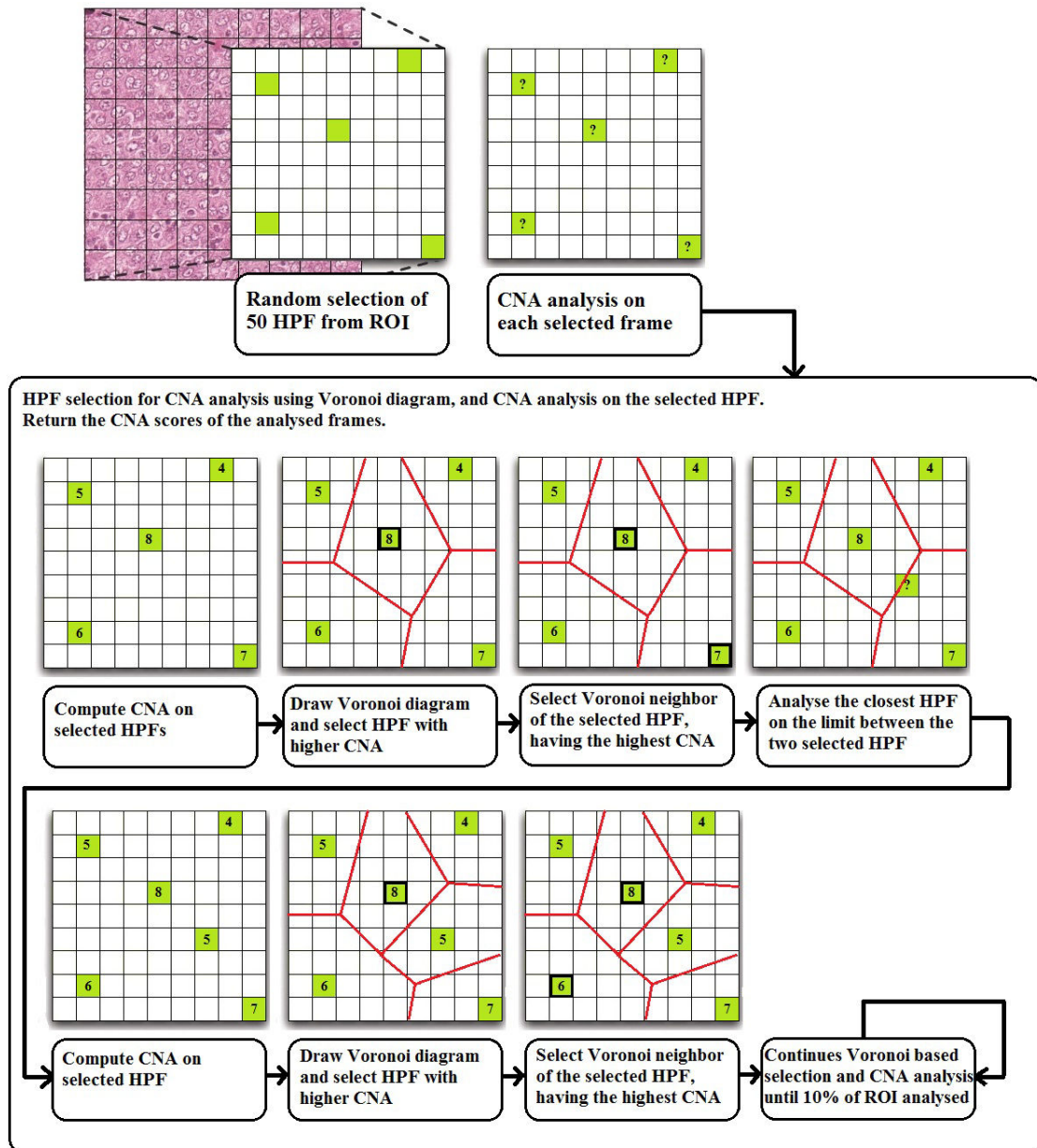


Figure 5.13: Dynamic sampling method applied over a ROI for CNA score.

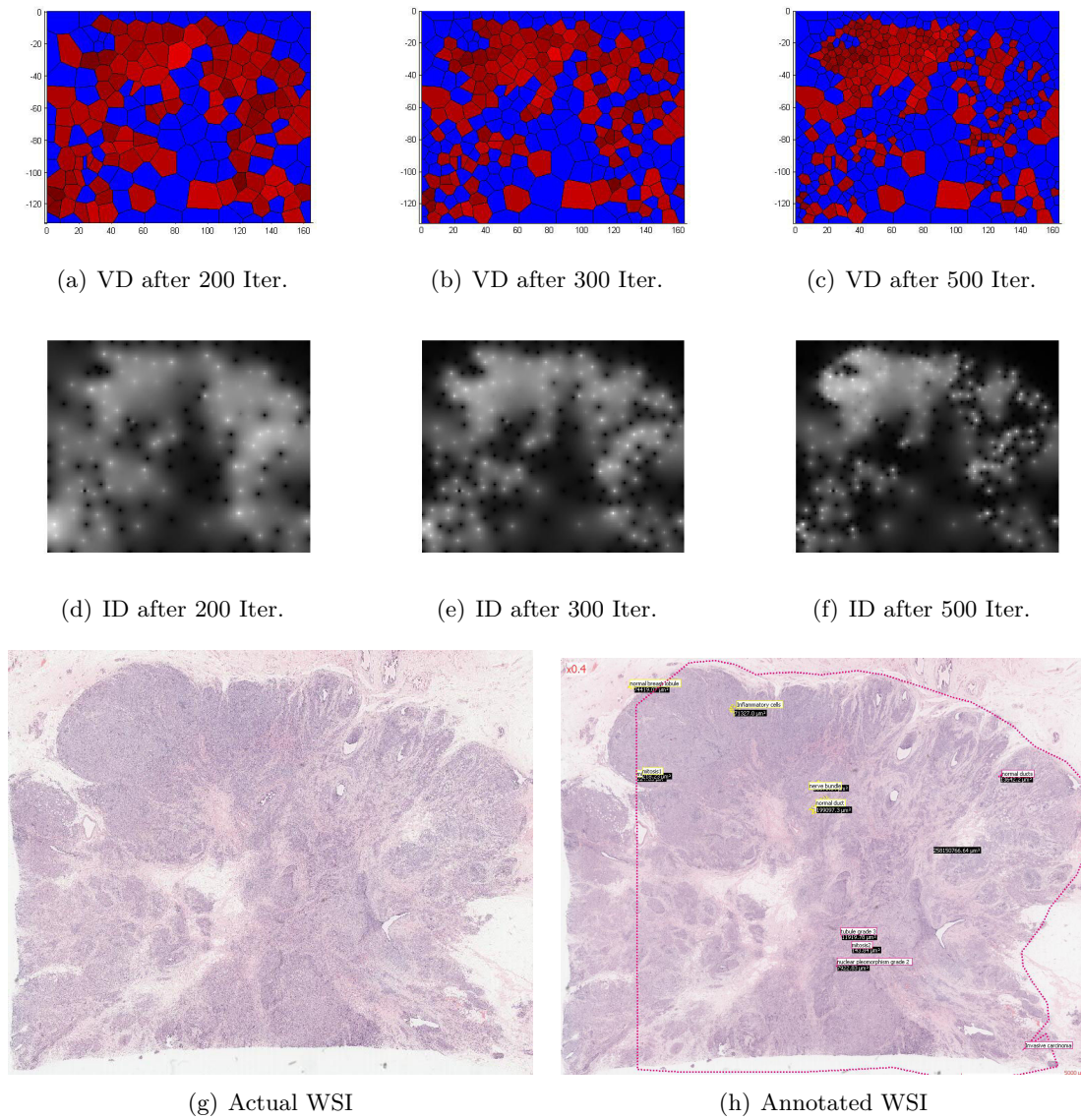


Figure 5.14: Dynamic sampling method applied over WSI. The incrementally construction of Voronoi Diagram (VD) (first row) and its Intensity Map (IM) (second row) of score are shown. Each cell contains a single frame at its center. The bright color in IM represents higher CNA score (mean higher degree of malignancy). After 200 iteration, the whole WSI area is being explored. No area seems favored. After 300 iteration, the algorithm converges towards a high CNA. After 500 iteration, the sampling is very dense around this area and remains sparse in others.

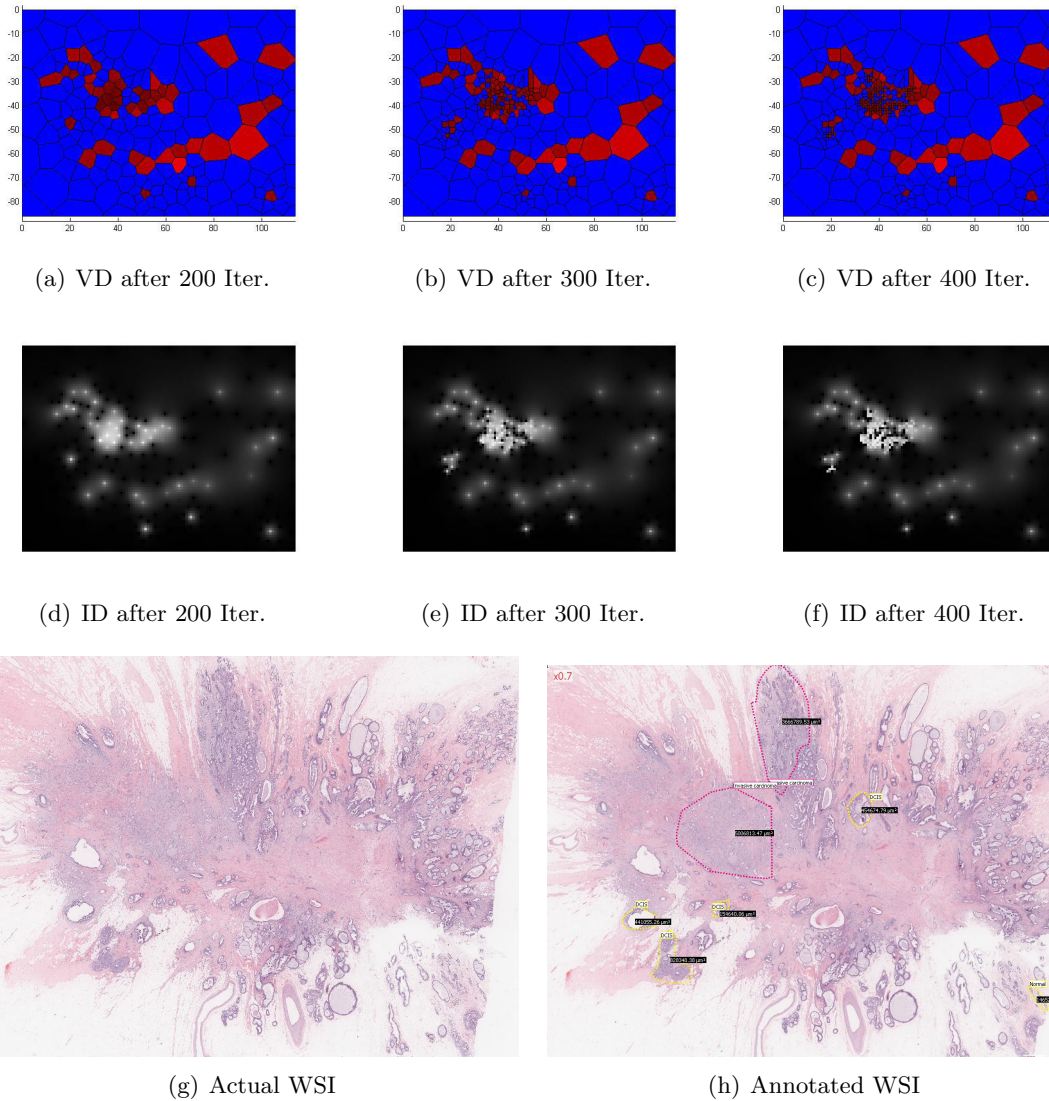


Figure 5.15: Dynamic sampling method applied over WSI. The incrementally construction of Voronoi Diagram (VD) (first row) and its Intensity Map (IM) (second row) of score are shown. Each cell contains a single frame at its center. The bright color in IM represents higher CNA score (mean higher degree of malignancy). After 200 iteration, the whole WSI area is being explored. No area seems favored. After 300 iteration, the algorithm converges towards a high CNA. After 400 iteration, the sampling is very dense around this area and remains sparse in others.

have been specifically illustrated: triangular / simplex meshes and Voronoi / Delaunay. We also test our proposed data structure without and with hole in meshes.

Finally, we modeled WSI as 2-manifold mesh. We have implemented dynamic sampling method for selection of HPF in WSI. In addition, we have tested dynamic sampling method for the evaluation of CNA score on breast cancer WSI in MICO platform. In the medical application, more specifically analysing WSI, our dynamic sampling method has proved its ability to accurately guide the finding of the highest levels of CNA in a WSI within an acceptable time frame as well as to provide a useful, reliable visualization map for the end user. From a more global standpoint, this dynamic sampling method makes it possible to speed up the analysis, enhance the visualization and assist the exploration of high-content images.

Chapter 6

Overall Conclusion and Future Perspectives

Conclusion

L'objectif général de cette thèse portait sur l'étude des défis relatifs aux techniques robustes d'analyse quantitative d'images en histopathologie. Depuis la dernière décennie, un important volume de recherches a porté sur l'histopathologie, notamment pour la détection, la segmentation et la classification de noyaux dans différentes modalités d'imagerie. Détection, segmentation et classification de noyaux sont des étapes importantes pour le diagnostic et la gradation du cancer. La présence de noyaux et leur aspect sont les signes essentiels pour l'évaluation de la présence de la maladie et de sa gravité. La détection, la segmentation et la classification des noyaux dans les images histopathologiques posent de nombreux et difficiles problèmes de vision par ordinateur en raison de la forte variabilité dans les images causée par un certain nombre de facteurs, notamment les différences dans la préparation des lames, l'acquisition de l'image et la structure complexe des tissus observés.

Nous avons présenté une étude exhaustive de l'état de l'art des méthodes de détection, de segmentation et de classification des noyaux limitée à deux types de modalités d'images largement disponibles : hématoxyline et éosine (H&E) d'une part et immunohistochimie (IHC) d'autre part. Cette revue de la littérature met en lumière des domaines de recherche ouverts peu étudiés. Ces domaines de recherche ouverts présentent des défis uniques qui devraient être abordés par de futures recherches. L'un de ces défis concerne la disponibilité de jeux de données communs. Les résultats des études actuelles sont basés sur leurs propres jeux de données qui ne sont pas publics. Nous pensons qu'il n'est pas satisfaisant ni même possible d'évaluer et de comparer numériquement les différentes études uniquement en fonction de leurs résultats puisqu'aucune de ces études n'ont utilisé des jeux de données communs, des méthodes d'évaluation identiques ou encore des indicateurs de performance similaires. Pour comparer les résultats numériques de chaque étude, il est nécessaire de disposer de jeux de données de référence publics pouvant être utilisés par tous. Ces jeux de données doivent être validés médicalement, constitués d'échantillons pris à partir d'un grand nombre de patients et annotés par différents médecins, comme l'a fait le projet MICO avec le jeu de données du concours MITOS lors de la conférence ICPR 2012, ou encore avec le jeu de données du concours MITOS & ATYPIA pour la conférence ICPR 2014. Un tel effort rendrait possible la comparaison numérique des résultats obtenus par différentes études et l'identification de leurs caractéristiques distinctives.

Parmi les différentes études, la détection et la classification automatique de noyaux est une tâche récurrente, particulièrement difficile pour les images histopathologiques. Le compte du nombre de mitoses est un paramètre pronostique important pour la gradation du cancer, en particulier dans le cas de la gradation du cancer du sein. Nous

avons proposé trois systèmes de détection automatique de mitoses pour différents types de scanners et un microscope multispectral, à savoir les systèmes TMC et ITM²C pour les images couleur et le système MITM³ pour les images multispectrales. Ces systèmes effectuent la détection et la segmentation de candidats, le calcul et la sélection de descripteurs, la classification et enfin la gestion de l'asymétrie des jeux de données. Les principales contributions apportées par les systèmes TMC et ITM²C sont au nombre de sept. Tout d'abord, nous avons analysé de manière exhaustive les informations statistiques et morphologiques concernant les noyaux mitotiques dans différents canaux de couleurs de plusieurs modèles de couleurs qui améliorent la détection de mitoses dans les images couleur produites par les scanners Aperio et Hamamatsu. Deuxièmement, nous avons effectué une étude approfondie sur les descripteurs calculés au niveau de la région englobant une mitose ou du seul noyau segmenté d'une mitose pour le classement des mitoses. Troisièmement, nous avons étudié des méthodes de suréchantillonnage pour augmenter le nombre d'instances de la classe minoritaire (mitose) par interpolation entre plusieurs exemples proches de la classe minoritaire, ce qui rend la classification plus robuste. Quatrièmement, nous avons effectué une étude approfondie sur de nombreux classifieurs pour proposer celui qui est le meilleur pour la classification des mitoses. Cinquièmement, nous avons évalué notre système sur les jeux de données du concours MITOS organisé à l'occasion de la conférence internationale ICPR 2012. Notre système a été classé en deuxième position sur 17 finalistes pour les jeux de données Aperio et Hamamatsu. Sixièmement, nous avons proposé une stratégie efficace et générique pour explorer les grandes images que sont les images de lame entière en combinant des outils de géométrie algorithmique avec une mesure de signal local de pertinence dans le cadre d'un échantillonnage dynamique. Septièmement, nous avons également procédé à l'évaluation de ces systèmes dans le prototype de plateforme du projet MICO.

Nous avons également proposé le système MITM³ pour les données multispectrales. Les principales contributions de ce système sont au nombre de six. Tout d'abord, nous avons proposé une sélection automatique non supervisée du meilleur plan focal des données multispectrales. Deuxièmement, nous avons proposé trois méthodes différentes pour la sélection des bandes spectrales, à savoir l'absorption spectrale relative des différents composants des tissus, l'absorption spectrale des colorants hématoxyline et éosine et la technique de mRMR. Troisièmement, nous avons calculé les descripteurs spatiaux multispectraux comprenant des informations au niveau du pixel, de la texture et de la morphologie pour les bandes spectrales sélectionnées qui exploitent l'information discriminante pour la classification des mitoses dans les images multispectrales. Quatrièmement, nous avons effectué une étude approfondie sur les descripteurs calculés au niveau de la région englobant une mitose ou du seul noyau segmenté d'une mitose pour la classification des mitoses. Cinquièmement, nous avons effectué une étude approfondie sur de nombreux classifieurs pour proposer celui qui est le meilleur pour la classification des mitoses. Sixièmement, nous avons évalué notre système sur le jeu de données multispectrales du concours MITOS et nous avons réussi à obtenir le F-score le plus élevé.

Parmi les résultats du concours MITOS sur les trois types de jeux de données, notre système pour les données multispectrales a amélioré de manière significative les résultats du concours selon le F-score. En ce qui concerne les jeux de données des images couleur des scanners Aperio et Hamamatsu, nous avons réussi à atteindre la deuxième place du concours. Nos systèmes parviennent à atteindre le même niveau de précision dans la détection des mitoses sur les images du scanner Aperio et sur les images multispectrales.

Par ailleurs, nous avons également proposé une extension de la structure de données `itk::QuadEdgeMesh` capable de gérer à la fois les maillages primales et duales. La nouvelle structure de données, `itk::QuadEdgeMeshWithDual`, apporte en plus la gestion de la géométrie duale. Deux types de maillages primales ont été spécialement illustrés : maillages triangulaires / simplex et Voronoi / Delaunay.

Finally, we have proposed an innovative platform in which a dynamic sampling method allows for the rapid analysis of a full slide image. We tested the dynamic sampling method for the evaluation of the nuclear atypia score in breast cancer on full slide images. Our dynamic sampling method has proven its ability to detect and precisely measure the highest levels of nuclear atypia in a full slide image within an acceptable time frame and to provide a reliable and useful visualization for the user. From a more global point of view, this dynamic sampling method allows for accelerating the analysis, improving the visualization and assisting in the exploration of very large images.

Perspectives

In future work, we envisage optimizing the detection of candidates by reducing the initial number of candidates for mitosis classification. A possible way to achieve this objective is the automatic selection of tumor zones to circumscribe the detection of mitoses only to zones containing tumor cells. We also plan to study the calculation of other descriptors for mitosis detection. In the present study, we analyzed in detail the descriptors calculated at the level of the single segmented mitosis nucleus as well as descriptors based on the size of the region encompassing a mitosis for mitosis classification. It would be interesting to calculate texture characteristics separately inside the segmented nuclei and in the region surrounding the segmented nuclei in order to study whether this allows for improving the classification results.

There is an important degree of imbalance in the learning set, as mitosis instances are very few compared to non-mitosis examples. In this case, the learned boundary is biased towards the majority class, which leads to high rates of false negatives and false positives. This is due to the presence in histopathological images of many types of nuclei and other objects that resemble mitoses. A possible way to improve the precision of the classification is to make mitosis classification a multi-class problem.

Concerning multispectral imaging, we plan to study deconvolution of spectral bands as most of the spectral bands of the data have overlapping support, which increases redundancy. The pre-selection of the focal plane is also of great importance to reduce the complexity of the data and improve the performance of mitosis detection in order to reach an acceptable quality level.

6.1 Conclusion

The overarching goal of this dissertation is to investigate challenges in robust quantitative image analysis techniques in histopathology. Since last decade, a significant amount of research has been done in the field of histopathology, focusing on nuclei detection, segmentation and classification in different image modalities. Nuclei detection, segmentation and classification are important steps in cancer diagnosis and grading. The presence of nuclei and their aspect are critical signs for evaluating the existence of disease and its severity. In routinely stained histopathological images, detection, segmentation and classification of nuclei pose a difficult computer vision problem, due to high variability in images. This issue is caused by a number of factors, including differences in slide preparation, image acquisition and complex tissue structure.

We have presented a comprehensive review on state-of-the-art methodologies in nuclei detection, segmentation and classification restricted to two widely available types of image modalities: H&E and IHC. This literature review highlights open research areas with few existing studies.

These open research areas are characterized by unique challenges, which should be covered in future research. Evaluating and comparing the existing studies is objectively impossible to do solely based on their reported results, as they used different (often unbalanced, inconsistent or even medically irrelevant) datasets, evaluation methods and performance metrics. For a relevant numerical comparison of the studies, it is definitely necessary to develop representative, referenced benchmark datasets. These datasets should be medically validated and consist of samples that are taken from a large number of patients and annotated by different pathologists - like our MICO project did with MITOS and will soon do with MITOS & ATYPIA @ ICPR. Such an effort will make possible the numerical comparison of the results obtained by different studies and to identify the distinguishing features.

This methodology review also highlights an important gap to be fulfilled by all scientists in order to be able to reliably go to the next generation of important challenges, related to the “digital” exploration and the understanding of the WSI as an essential high-content imaging diagnostic biomarker and prognosis support. Consolidating, in the next few years these approaches with mining structured big data and analytics as with genomics and molecular imaging technologies, will certainly have the potential to lead to the next generation of healthcare technologies.

Among the various studies, automated nuclei detection and classification is a recurring task, particularly difficult on histopathology images. Mitosis count is an important prognostic parameter for cancer grading particularly in breast cancer grading. Automated mitosis detection frameworks have been proposed for different types of scanners and multispectral microscope datasets. We have proposed three frameworks; TMC and ITM²C frameworks for color datasets and MITM³ framework for multispectral dataset. These frameworks consist of candidate detection and segmentation, features computation and selection, classification and handling unbalanced training datasets. The main contributions in TMC and ITM²C frameworks are seven fold. First, we have comprehensively analysed the statistical and morphological information concerning mitotic nuclei on different color channels of various color models that improve the mitosis detection in color datasets (Aperio and Hamamatsu scanners). Second, we have performed a comprehensive study on region and patch based features for mitosis classification. Third, we have studied oversampling methods to increase the number of instances of the minority class (mitosis) by interpolating between several minority class examples that lie together, which makes classification more robust. Fourth, we have performed an extensive investigation of classifiers and inference of the best one for mitosis classification. Fifth, we have evaluated our framework on MITOS datasets during ICPR 2012 contest and ranked second from 17 finalists on Aperio and Hamamatsu datasets. Sixth, we have proposed an efficient and generic strategy to explore large images like WSI by combining computational geometry tools with a local signal measure of relevance in a dynamic sampling framework. Seventh, we have also performed real time evaluations of these frameworks in MICO platform prototyping.

We have also proposed an automated MITM³ framework for multispectral dataset. The main contributions of this framework are six fold. First, we have proposed an automatic and unsupervised focal plane selection for multispectral dataset. Second, we have proposed three different methods for spectral bands selection including relative spectral absorption of different tissue components, spectral absorption of H&E stains and mRMR technique. Third, we have computed multispectral spatial features containing pixel, texture and morphological information on selected spectral bands which leverage discriminant information for mitosis classification on multispectral dataset. Fourth, we have performed a comprehensive study on region and patch based features for mitosis classification. Fifth, we have performed an extensive investigation of classifiers and inference of the best one for mitosis classification. Sixth, we have evaluated this framework on MITOS multispectral dataset and have managed to achieve the highest FM.

Compared to MITOS contest results on three types of datasets, our multispectral framework have outperformed significantly the results of the contest according to FM, while for the color framework for Aperio and Hamamatsu datasets, we managed to rank in second position of the contest. Our frameworks manage to reach the same level of accuracy in mitosis detection on Aperio and Multispectral datasets.

Furthermore, we have also proposed an extension of *itk::QuadEdgeMesh* data structure to handle both primal and dual meshes, simultaneously. The new data structure, *itk::QuadEdgeMeshWithDual*, already include by default the dual topology, to handle dual geometry as well. Two types of primal meshes have been specifically illustrated: triangular / simplex meshes and Voronoi / Delaunay.

Finally, we proposed an innovative platform in which dynamic sampling method performed fast analysis of WSI. We tested dynamic sampling method for real time evaluation of CNA score on breast cancer WSI in MICO platform. In the medical application, more specifically analysing WSI, our dynamic sampling method has proved its ability to accurately find and measure the highest levels of CNA in a WSI within an acceptable time frame as well as to provide a useful, reliable visualization map for the end user. From a more global standpoint, this dynamic sampling method makes it possible to speed up the analysis, enhance the visualization and assist the exploration of high-content images.

6.2 Future Perspectives

In future work, we plan to optimize the candidate detection by reducing the number of initial candidates for mitosis classification. One possible way is the automated selection of tumor areas that restrict the detection of mitosis to areas containing tumor cells only and avoid potentially misleading results from analysis of stromal regions. We plan to investigate other model-based features computation for mitosis detection. In the current study, we comprehensively analysed region based features and also different patch size based features for mitosis classification. It would be interesting to compute separate texture features for segmented regions (in region) and neighboring region (out region) to see if it can further improve the classification results.

There is a high degree of imbalance in the training set, mitotic instances being very few in number as compared to the non-mitotic instances. In this case, the class boundary learned by the standard machine learning algorithms is biased towards the majority class resulting in high false negative and false positive rates. It might be reasonable to consider many types of nuclei and other mitosis-like objects in the classification, as it is practically the case in histopathological images. A possible way to improve this classification accuracy is to migrate the mitosis classification into a multi-class problem.

Specifically for multispectral imaging, we plan to investigate unmixing of SBs as most SBs have overlapping area, which increase redundancy. The pre-selection of the focal plane is also of great importance to reduce the complexity of the dataset and improve the actual performance to reach clinical operational acceptance expected by our professional consortia.

Appendix A

Glossary

Adenocarcinoma : A carcinoma originating in glandular tissue.

Aspirative cytology : Cytology specimens extracted via syringe.

Atypia : Cells or tissue displaying some characteristics of a malignancy, but not considered either malignant or benign. The diagnosis of atypia generally requires a more comprehensive (and possibly invasive) follow-up to determine the true diagnosis.

Benign : A condition which will not metastasize and is not harmful in and of itself.

Brightfield microscopy : Microscopy techniques using a broad spectrum light source to visualize the specimen.

Carcinoma : A cancer of the epithelium.

Chromatin : Nuclear material that is readily stained, consisting of the nucleic acids and associated proteins.

Confocal : Confocal microscopy images different focal planes through the specimen.

Counterstain : A stain used as contrast to another, generally more specific, stain.

Cytology : The study of cells at a microscopic level, generally via a light microscopy technique.

Cytopathology : The study of diseased cells at the microscopic level.

Densitometry : Measurements related to the optical density of a sample.

Ductal carcinoma : Carcinoma originating in ductal structures.

Eosin : A pink-staining acidic dye that stains membranes and fibers.

Epithelium : The internal and external lining of cavities within the body; also the external covering (skin).

Feulgen : A stain specific to DNA which lends a purple color.

Fibroadenosis : A benign cause of many breast lumps.

Fine needle aspiration : A procedure using a small needle inserted into the lesion and drawing a small amount of cellular material into a syringe; a form of aspirative cytology.

Fluorescence imagery : Fluorescent dyes are attached to antibodies specific to some feature of interest (e.g., certain proteins) and imaged by exciting the fluorescence of the dyes with appropriate incident light. This method can very specifically target certain molecular attributes of a biological specimen.

Gleason grading : A grading for prostate cancer, characterizing the tumor into one of 5 categories based on tumor differentiation.

Hematoxylin : A blue-staining basic dye that stains genetic material; this is mainly seen in nuclear material, although some components of cytoplasmic and extracellular material is also stained.

Histology : The study of tissue at a microscopic level, generally via a light microscopy technique.

Histopathology : The study of diseased tissue at the microscopic level.

Hyperchromasia : An overall increase in staining intensity.

Hyperplasia : Abnormalities in the characteristics of cells and tissues, generally including an increase in cellularity and/or mitosis; often used interchangeably with dysplasia.

Immunostain : Immunostains use antibodies to specifically target molecules of interest, similar to fluorescence imaging, but use standard dyes for viewing with light microscopy.

in situ : Within normal boundaries, not invading surrounding tissues.

in vivo : Living tissue in its natural environment.

Karyometry : Nuclear characteristics, generally texture.

Lobular carcinoma : A type of adenocarcinoma.

Malignant : A condition which will eventually lead to death if untreated. Malignant conditions tend to metastasize, grow uncontrollably, and lack proper tissue differentiation.

Metastasis : The spread of cancer from the originating tissue to other parts of the body.

Microarray : Tissue microarrays align many (hundreds or thousands) of tissue core samples on a single slide; this allows for simultaneous analysis of all samples and is commonly used in high-throughput operations.

Nucleolus : A small, round sub-organelle within the cell nucleus.

Pathology : The study of disease, with emphasis on disease structure and the effects on the body as a whole.

Pleomorphic : Containing more than one stage of the life cycle.

Premalignancy : A diseased state that, while not considered cancerous, will progress to cancer if left untreated.

Stroma : Connective tissue.

Bibliography

- [1] *MITOS*. <http://ipal.cnrs.fr/ICPR2012/?q=node/5>, 2012.
- [2] *Participant ISC*. http://scanner-contest.charite.de/en/participants/2nd_isc/vendors/, 2012.
- [3] *AMIDA13*. <http://amida13.isi.uu.nl/>, 2013.
- [4] *Calopix - TRIBVN Software Informer*. <http://calopix.software.informer.com/>, 2013.
- [5] *Gastric Breast Cancer Network Center*. <http://www.gastricbreastcancer.com>, 2013.
- [6] *MICO 2.0 - Mitosis Detection Video, IPAL UMI CNRS Lab*. http://ipal.cnrs.fr/data/z/MICO_mitosis.mp4, 2013.
- [7] *US Breast Cancer Statistics*. http://www.breastcancer.org/symptoms/understand_bc/statistics, 2013.
- [8] Omar S Al-Kadi. Texture measures combination for improved meningioma classification of histopathological images. *Pattern Recognition*, 43(6):2043–2053, 2010.
- [9] Yousef Al-Kofahi, Wiem Lassoued, William Lee, and Badrinath Roysam. Improved automatic detection and segmentation of cell nuclei in histopathology images. *IEEE Transactions on Biomedical Engineering*, 57(4):841–852, 2010.
- [10] Gabriela Alexe, Gul S Dalgin, Daniel Scanzfeld, Pablo Tamayo, Jill P Mesirov, Charles DeLisi, Lyndsay Harris, Nicola Barnard, Maritza Martel, Arnold J Levine, et al. High expression of lymphocyte-associated genes in node-negative HER2+ breast cancers correlates with lower recurrence rates. *Cancer research*, 67(22):10669–10676, 2007.
- [11] Sahirzeeshan Ali and Anant Madabhushi. An integrated region-, boundary-, shape-based active contour for multiple object overlap resolution in histological imagery. *IEEE Transactions on Medical Imaging*, 31(7):1448–1460, 2012.
- [12] Sahirzeeshan Ali, Robert Veltri, Jonathan Epstein, Christhunesa Christudass, and Anant Madabhushi. Adaptive energy selective active contour with shape priors for nuclear segmentation and gleason grading of prostate cancer. In Gabor Fichtinger, Anne Martel, and Terry Peters, editors, *Medical Image Computing and Computer-Assisted Intervention*, volume 6891 of *Lecture Notes in Computer Science*, pages 661–669. Springer Berlin / Heidelberg, 2011.
- [13] V Anari, P Mahzouni, and R Amirfattahi. Computer-aided detection of proliferative cells and mitosis index in immunohistochemically images of meningioma. In *6th Iranian Conference on Machine Vision and Image Processing*, pages 1–5, 2010.
- [14] Christophe Avenel and Maria S Kulikova. Marked point processes with simple and complex shape objects for cell nuclei extraction from breast cancer h&e images. In *SPIE Medical Imaging*, pages 86760Z–86760Z. International Society for Optics and Photonics, 2013.
- [15] S Baheerathan, F Albregtsen, and H E Danielsen. New texture features based on the complexity curve. *Pattern Recognition*, 32(4):605–618, 1999.
- [16] P H Bartels, D Thompson, M Bibbo, and J E Weber. Bayesian belief networks in quantitative histopathology. *Analytical and Quantitative Cytology and Histology*, 14(6):459–473, 1992.
- [17] 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. *IEEE Transactions on Biomedical Engineering*, 57(3):642–653, March 2010.

- [18] Pinky A Bautista and Yukako Yagi. Digital simulation of staining in histopathology multi-spectral images: enhancement and linear transformation of spectral transmittance. *Journal of Biomedical Optics*, 17(5):0560131–05601310, 2012.
- [19] J A M Belien, J P A Baak, P J Van Diest, and A H M Van Ginkel. Counting mitoses by image processing in feulgen stained breast cancer sections: the influence of resolution. *Cytometry*, 28(2):135–140, 1997.
- [20] L E Boucheron. *Object and Spatial-level quantitative analysis of multispectral histopathology images for detection and characterization of cancer*. PhD thesis, University of California Santa Barbara, CA, 2008.
- [21] Laura E Boucheron, Zhiqiang Bi, Neal R Harvey, BS Manjunath, and David L Rimm. Utility of multispectral imaging for nuclear classification of routine clinical histopathology imagery. *BMC Cell Biology*, 8(Suppl 1):S8, 2007.
- [22] Laura E Boucheron, B S Manjunath, and Neal R Harvey. Use of imperfectly segmented nuclei in the classification of histopathology images of breast cancer. In *Acoustics Speech and Signal Processing (ICASSP), 2010 IEEE International Conference on*, pages 666–669. IEEE, 2010.
- [23] Peter Boyle, Bernard Levin, et al. *World cancer report 2008*. IARC Press, International Agency for Research on Cancer, 2008.
- [24] Tim Bray, Jean Paoli, C Michael Sperberg-McQueen, Eve Maler, and François Yergeau. Extensible markup language (xml). *World Wide Web Journal*, 2(4):27–66, 1997.
- [25] A Can, M Bello, H E Cline, T Xiaodong, F Ginty, A Sood, M Gerdes, and M Montalto. Multi-modal imaging of histological tissue sections. In *5th IEEE International Symposium on Biomedical Imaging: From Nano to Macro*, pages 288–291, 2008.
- [26] Chad Carson, Serge Belongie, Hayit Greenspan, and Jitendra Malik. Blobworld: Image segmentation using expectation-maximization and its application to image querying. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 24(8):1026–1038, 2002.
- [27] S Di Cataldo, E Ficarra, A Acquaviva, and E Macii. Automated segmentation of tissue images for computerized ihc analysis. *Computer Methods and Programs in Biomedicine*, 100(1):1–15, 2010.
- [28] Santa Di Cataldo, Elisa Ficarra, Andrea Acquaviva, and Enrico Macii. Achieving the way for automated segmentation of nuclei in cancer tissue images through morphology-based approach: A quantitative evaluation. *Computerized Medical Images and Graphics*, 34(6):453–461, 2010.
- [29] T F Chan and L A Vese. Active contours without edges. *IEEE Transactions on Image Processing*, 10(2):266–277, 2001.
- [30] Douglas E Chandler and Robert W Roberson. *Bioimaging: Current Techniques in Light and Electron Microscopy*. Jones & Bartlett Publishers, 2009.
- [31] Hang Chang, Leandro A Loss, and Bahram Parvin. Nuclear segmentation in H&E sections via multi-reference graph cut (MRGC). In *9th IEEE International Symposium on Biomedical Imaging: From Nano to Macro*, pages 614–617, 2012.
- [32] N V Chawla, K W Bowyer, L O Hall, and W P Kegelmeyer. Smote: Synthetic minority over-sampling technique. *J. Art. Int. Res.*, 16:321–57, 2002.
- [33] Cheng Chen, J A Ozolek, Wei Wang, and G K Rohde. A pixel classification system for segmenting biomedical images using intensity neighborhoods and dimension reduction. In *8th IEEE International Symposium on Biomedical Imaging: From Nano to Macro*, pages 1649–1652, April 2011.
- [34] Edmund S Cibas and Barbara S Ducatman. *Cytology: Diagnostic principles and clinical correlates*. Saunders, 2009.
- [35] D C Ciresan, A Giusti, L M Gambardella, and J Schmidhuber. Mitosis detection in breast cancer histology images with deep neural networks. In *MICCAI*, 2013.
- [36] F Cloppet and A Boucher. Segmentation of overlapping/aggregating nuclei cells in biological images. In *19th International Conference on Pattern Recognition*, pages 1–4, December 2008.

- [37] E Cosatto, M Miller, H P Graf, and J S Meyer. Grading nuclear pleomorphism on histological micrographs. In *19th International Conference on Pattern Recognition*, pages 1–4, December 2008.
- [38] David J Dabbs. *Diagnostic Immunohistochemistry: Theranostic and Genomic Applications*. Saunders Elsevier, 2010.
- [39] Jean-Romain Dalle, Hao Li, Chao-Hui Huang, Wee Kheng Leow, Daniel Racoceanu, and Thomas C Putti. Nuclear pleomorphism scoring by selective cell nuclei detection. In *IEEE Workshop on Applications of Computer Vision*, 2009.
- [40] H Delingette. Simplex meshes: a general representation for 3d shape reconstruction. In *Computer Vision and Pattern Recognition. Proceedings CVPR '94., IEEE Computer Society Conference on*, pages 856–859, jun 1994.
- [41] Vincenzo Della Mea, Giampiero Duglio, Filippo Crivelli, Pierluigi Banfi, and Giancarlo Chiovini. Preliminary slide scanner throughput evaluation in a intensive digitization facility setting. *Diagnostic Pathology*, 8(Suppl 1):S45, 2013.
- [42] C Demir and B Yener. Automated cancer diagnosis based on histopathological images: a systematic survey. Technical report, Rensselaer Polytechnic Institute, Department of Computer Science, 2005.
- [43] Arthur P Dempster, Nan M Laird, and Donald B Rubin. Maximum likelihood from incomplete data via the em algorithm. *Journal of the Royal Statistical Society. Series B (Methodological)*, pages 1–38, 1977.
- [44] Pierre A Devijver and Josef Kittler. *Pattern recognition: A statistical approach*. Prentice/Hall International Englewood Cliffs, NJ, 1982.
- [45] Murat Dundar, Sunil Badve, Gökhan Bilgin, Vikas C Raykar, Rohit K Jain, Olcay Sertel, and Metin N Gurcan. Computerized classification of intraductal breast lesions using histopathological images. *IEEE Transactions on Biomedical Engineering*, 58(7):1977–1984, 2011.
- [46] B Dunne and J J Going. Scoring nuclear pleomorphism in breast cancer. *Histopathology*, 39(3):259–265, 2001.
- [47] Bradley Efron. Estimating the error rate of a prediction rule: improvement on cross-validation. *Journal of the American Statistical Association*, 78(382):316–331, 1983.
- [48] C W Elston and I O Ellis. Pathological prognostic factors in breast cancer. i. the value of histological grade in breast cancer: experience from a large study with long-term follow-up. *Histopathology*, 19(5):403–410, 1991.
- [49] C W Elston and I O Ellis. Pathological prognostic factors in breast cancer. the value of histological grade in breast cancer experience from a large study with long-term follow-up. *Histopathology*, 41:151, 2002.
- [50] Chang K W Hsieh C J Wang X R Lin C Fan, R E. Liblinear: A library for large linear classification. *The Journal of Machine Learning Research*, 9:1871–1874, 2008.
- [51] H Fatakdawala, Jun Xu, A Basavanhally, G Bhanot, S Ganesan, M Feldman, J E Tomaszewski, and A Madabhushi. Expectation maximization-driven geodesic active contour with overlap resolution (EMaGACOR): Application to lymphocyte segmentation on breast cancer histopathology. *IEEE Transactions on Biomedical Engineering*, 57(7):1676–1689, 2010.
- [52] Daniel C Fernandez, Rohit Bhargava, Stephen M Hewitt, and Ira W Levin. Infrared spectroscopic imaging for histopathologic recognition. *Nature biotechnology*, 23(4):469–474, 2005.
- [53] H Fox. Is H&E morphology coming to an end? *Journal of Clinical Pathology*, 53:38–40, 2000.
- [54] Thomas J Fuchs and Joachim M Buhmann. Computational pathology: Challenges and promises for tissue analysis. *Computerized Medical Imaging and Graphics*, 35(7):515–530, 2011.
- [55] M M Galloway. Texture analysis using gray level run lengths. *CGIP*, 4:172–9, 1975.

- [56] J Gama. Functional trees for classification. In *Data Mining, 2001. ICDM 2001, Proceedings IEEE International Conference on*, pages 147–154, 2001.
- [57] Elisa Drelie Gelasca, Boguslaw Obara, Dmitry Fedorov, Kristian Kvilekval, , and BS Manjunath. A biosegmentation benchmark for evaluation of bioimage analysis methods. *BMC Bioinformatics*, 10:1–12, 2009.
- [58] Stephen M Gentry and Richard M Levenson. Biomedical applications of the information-efficient spectral imaging sensor (ISIS). In *International Biomedical Optics Symposium (BiOS)*, pages 129–142. International Society for Optics and Photonics, 1999.
- [59] Alessandro Gherardi, Alessandro Bevilacqua, and Filippo Piccinini. Illumination field estimation through background detection in optical microscopy. In *Computational Intelligence in Bioinformatics and Computational Biology (CIBCB), 2011 IEEE Symposium on*, pages 1–6. IEEE, 2011.
- [60] Joan Gil, Haishan Wu, and Beverly Y Wang. Image analysis and morphometry in the diagnosis of breast cancer. *Microscopy Research and Technique*, 59(2):109–118, 2002.
- [61] Yun Gong. Breast cancer: Pathology, cytology, and core needle biopsy methods for diagnosis. In Mahesh K Shetty, editor, *Breast and Gynecological Cancers*, pages 19–37. Springer New York, 2013.
- [62] R C Gonzalez and R E Woods. *Digital Image Processing*. Pearson Prentice Hall, 2008.
- [63] A Gouaillard, L Florez-Valencia, and E Boix. Itkquadedgemesh: A discrete orientable 2-manifold data structure for image processing. *Insight Journal*, Sep 2006.
- [64] Leonard E Grenier, Brian V Funt, Paul H Orth, and Donald MF McIntosh. Transillumination method apparatus for the diagnosis of breast tumors and other breast lesions by normalization of an electronic image of the breast, January 7 1992. US Patent 5,079,698.
- [65] Cigdem Gunduz, Bulent Yener, and S Humayun Gultekin. The cell graphs of cancer. *Bioinformatics*, 20:145–151, 2004.
- [66] M N Gurcan, L E Boucheron, A Can, A Madabhushi, N M Rajpoot, and B Yener. Histopathological image analysis: A review. *IEEE Reviews in Biomedical Engineering*, 2:147–171, 2009.
- [67] M N Gurcan, T Pan, H Shimada, and J Saltz. Image analysis for neuroblastoma classification: Segmentation of cell nuclei. In *28th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, pages 4844–4847, 2006.
- [68] A Hafiane, F Bunyak, and K Palaniappan. Clustering initiated multiphase active contours and robust separation of nuclei groups for tissue segmentation. In *19th International Conference on Pattern Recognition*, pages 1–4, December 2008.
- [69] Peter W Hamilton, Peter H Bartels, Deborah Thompson, Neil H Anderson, Rodolfo Montironi, and James M Sloan. Automated location of dysplastic fields in colorectal histology using image texture analysis. *Journal of Pathology*, 182:68–75, 1997.
- [70] R M Haralick, K Shanmugam, and I H Dinstein. Textural features for image classification. *IEEE Trans. on Systems, Man and Cybernetics*, 3:610–21, 1973.
- [71] X He and Q Liao. A novel shape prior based segmentation of touching or overlapping ellipso-like nuclei. In *SPIE*, 2008.
- [72] H J A M Heijmans. Mathematical morphology: basic principles. In *Proceedings of Summer School on §Morphological Image and Signal Processing*. Citeseer, 1995.
- [73] Chao-Hui Huang, Antoine Veillard, Ludovic Roux, N. Loménie, and Daniel Racoceanu. Time-efficient sparse analysis of histopathological whole slide images. *Computerized Medical Imaging and Graphics*, 35:579–591, November 2011.
- [74] Po Whei Huang and Yan Hao Lai. Effective segmentation and classification for hcc biopsy images. *Pattern Recognition*, 43(4):1550–1563, 2010.
- [75] Seungil Huh, Dai Fei Elmer Ker, Ryoma Bise, Mei Chen, and Takeo Kanade. Automated mitosis detection of stem cell populations in phase-contrast microscopy images. *Medical Imaging, IEEE Transactions on*, 30(3):586–596, 2011.

- [76] Daniel P Huttenlocher, Gregory A Klanderman, and William J Rucklidge. Comparing images using the hausdorff distance. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 15(9):850–863, 1993.
- [77] L Ibáñez, W Schroeder, L Ng, and J Cates. *The ITK Software Guide*. Kitware, Inc. ISBN 1-930934-15-7, <http://www.itk.org/ItkSoftwareGuide.pdf>, 2 edition, 2005.
- [78] Bela Julesz. Experiments in the visual perception of texture. *Scientific American*, 232:34–43, 1975.
- [79] Chanho Jung and Changick Kim. Segmenting clustered nuclei using h-minima transform-based marker extraction and contour parameterization. *IEEE Transactions on Biomedical Engineering*, 57(10):2600–2604, 2010.
- [80] Chanho Jung, Changick Kim, S Wan Chae, and Sukjoong Oh. Unsupervised segmentation of overlapped nuclei using bayesian classification. *IEEE Transactions on Biomedical Engineering*, 57(12):2825–2832, 2010.
- [81] Mehdi Kamandar and Hassan Ghassemian. Maximum relevance, minimum redundancy band selection for hyperspectral images. In *Electrical Engineering (ICEE), 2011 19th Iranian Conference on*, pages 1–5. IEEE, 2011.
- [82] Lee Kametsky, Thouis R Jones, Adam Fraser, Mark-Anthony Bray, David J Logan, Katherine L Madden, Vebjorn Ljosa, Curtis Rueden, Kevin W Eliceiri, and Anne E Carpenter. Improved structure, function and compatibility for cellprofiler: modular high-throughput image analysis software. *Bioinformatics*, 27(8):1179–1180, 2011.
- [83] Michael Kass, Andrew Witkin, and Demetri Terzopoulos. Snakes: Active contour models. *International journal of computer vision*, 1(4):321–331, 1988.
- [84] Vojislav Kecman. *Learning and soft computing: support vector machines, neural networks, and fuzzy logic models*. MIT press, 2001.
- [85] Stephen J Keenan, James Diamond, W Glenn McCluggage, Hoshang Bharucha, Deborah Thompson, Peter H Bartels, and Peter W Hamilton. An automated machine vision system for the histological grading of cervical intraepithelial neoplasia (cin). *Journal of Pathology*, 192(3):351–362, 2000.
- [86] Lutz Kettner. Using generic programming for designing a data structure for polyhedral surfaces. *Comput. Geom. Theory Appl*, 13:65–90, 1999.
- [87] Adnan M Khan, Hesham El-Daly, and Nasir M Rajpoot. A gamma-gaussian mixture model for detection on mitotic cells in breast histology images. In *21th International Conference on Pattern Recognition*, October 2012.
- [88] Adnan M Khan, Hesham El-Daly, E Simmons, and Nasir M Rajpoot. A hybrid magnitude-phase approach to unsupervised segmentation of tumor areas in breast cancer histology images. *Journal of Pathology Informatics*, March 2013.
- [89] Riad Khelifi, Mouloud Adel, and Salah Bourennane. Multispectral texture characterization: application to computer aided diagnosis on prostatic tissue images. *EURASIP Journal on Advances in Signal Processing*, 2012(1):1–13, 2012.
- [90] Hui Kong, Metin Gurcan, and Kamel Belkacem-Boussaid. Partitioning histopathological images: An integrated framework for supervised color-texture segmentation and cell splitting. *IEEE Transactions on Medical Imaging*, 30(9):1661–1677, 2011.
- [91] Jun Kong, L Cooper, T Kurc, D Brat, and J Saltz. Towards building computerized image analysis framework for nucleus discrimination in microscopy images of diffuse glioma. In *Engineering in Medicine and Biology Society, 33rd Annual International Conference of the IEEE*, pages 6605–6608, September 2011.
- [92] Igor Kononenko. Estimating attributes: analysis and extensions of relief. In *Machine Learning: ECML-94*, pages 171–182. Springer, 1994.
- [93] Sonal Kothari, John H Phan, Richard A Moffitt, Todd H Stokes, Shelby E Hassberger, Qaiser Chaudry, Andrew N Young, and May D Wang. Automatic batch-invariant color segmentation of histological cancer images. In *Biomedical Imaging: From Nano to Macro, IEEE International Symposium on*, pages 657–660, 2011.

- [94] Sotiris Kotsiantis, Dimitris Kanellopoulos, Panayiotis Pintelas, et al. Handling imbalanced datasets: A review. *GESTS International Transactions on Computer Science and Engineering*, 30(1):25–36, 2006.
- [95] Maria Kulikova, Antoine Veillard, Ludovic Roux, and Daniel Racoceanu. Nuclei extraction from histopathological images using a marked point process approach. In *SPIE Medical Imaging*, 2012.
- [96] Niels Landwehr, Mark Hall, and Eibe Frank. Logistic model trees. *Machine Learning*, 59(1-2):161–205, 2005.
- [97] F J W M Leong, M Brady, and J OŠD McGee. Correction of uneven illumination (vignetting) in digital microscopy images. *Journal of clinical pathology*, 56(8):619–621, 2003.
- [98] Richard M Levenson. Spectral imaging perspective on cytomics. *Cytometry Part A*, 69(7):592–600, 2006.
- [99] Richard M Levenson, Paul J Cronin, and Kirill K Pankratov. Spectral imaging for brightfield microscopy. In *Biomedical Optics 2003*, pages 27–33. International Society for Optics and Photonics, 2003.
- [100] Richard M Levenson, Alessandro Fornari, and Massimo Loda. Multispectral imaging and pathology: seeing and doing more. *Expert Opinion on Medical Diagnostics*, 2(9):1067–1081, 2008. PMID: 23495926.
- [101] Richard M Levenson and James R Mansfield. Multispectral imaging in biology and medicine: slices of life. *Cytometry Part A*, 69(8):748–758, 2006.
- [102] An-An Liu, Kang Li, and Takeo Kanade. A semi-markov model for mitosis segmentation in time-lapse phase contrast microscopy image sequences of stem cell populations. *Medical Imaging, IEEE Transactions on*, 31(2):359–369, 2012.
- [103] H Liu and R Setiono. A probabilistic approach to feature selection - a filter solution. In *ICML*, pages 319–27, 1996.
- [104] Song Liu, P A Mundra, and J C Rajapakse. Features for cells and nuclei classification. In *33rd Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, pages 6601–6604, September 2011.
- [105] LNKnet. LNKnet software package, 2012.
- [106] Nicolas Loménie, Daniel Racoceanu, and Ludovic Roux. The mico platform: Cognitive virtual microscopy for breast cancer grading. In *10th European Congress on Telepathology and 4th International Congress on Virtual Microscopy*, Vilnius, Lithuania, 07/2010 2010.
- [107] G Loy and A Zelinsky. Fast radial symmetry for detecting points of interest. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 25(8):959–973, 2003.
- [108] Wen W Ma and Alex A Adjei. Novel agents on the horizon for cancer therapy. *CA: A Cancer Journal for Clinicians*, 59(2):111–137, 2009.
- [109] Marc Macenko, Marc Niethammer, JS Marron, David Borland, John T Woosley, Xiaojun Guan, Charles Schmitt, and Nancy E Thomas. A method for normalizing histology slides for quantitative analysis. In *Biomedical Imaging: From Nano to Macro, 2009. ISBI'09. IEEE International Symposium on*, pages 1107–1110. IEEE, 2009.
- [110] James MacQueen et al. Some methods for classification and analysis of multivariate observations. In *Proceedings of the fifth Berkeley Symposium on Mathematical Statistics and probability*, volume 1, page 14. California, USA, 1967.
- [111] D Magee, D Treanor, P Chomphuwiset, and P Quirke. Context aware colour classification in digital microscopy. In *Proc. Medical Image Understanding and Analysis*, pages 1–5, 2010.
- [112] Derek Magee, Darren Treanor, Doreen Crellin, Mike Shires, Katherine Smith, Kevin Mohee, and Philip Quirke. Colour normalisation in digital histopathology images. In *Proc. Optical Tissue Image analysis in Microscopy, Histopathology and Endoscopy (MICCAI Workshop)*, pages 100–111, 2009.
- [113] C D Malon, Eric Cosatto, et al. Classification of mitotic figures with convolutional neural networks and seeded blob features. *Journal of Pathology Informatics*, 4:9, May 2013.

- [114] Christopher Malon, Helena Brachtel, Eric Cosatto, Hans Peter Graf, Atsushi Kurata, Masahiko Kuroda, John S. Meyer, Akira Saito, Shulin Wu, and Yukako Yagi. Mitotic figure recognition: Agreement among pathologists and computerized detector. *Analytical Cellular Pathology*, 35:97–100, 2012.
- [115] Christopher Malon, Matthew Miller, Harold Christopher Burger, Eric Cosatto, and Hans Peter Graf. Identifying histological elements with convolutional neural networks. In *Proceedings of the 5th international conference on Soft computing as transdisciplinary science and technology*, pages 450–456. ACM, 2008.
- [116] A Martinez-Uso, F Pla, J M Sotoca, and P Garcia-Sevilla. Clustering-based multispectral band selection using mutual information. In *Pattern Recognition, 2006. ICPR 2006. 18th International Conference on*, volume 2, pages 760–763, 2006.
- [117] Gary D Marty et al. Blank-field correction for achieving a uniform white background in brightfield digital photomicrographs. *BioTechniques*, 42(6):716, 2007.
- [118] Khalid Masood and Nasir Rajpoot. Texture based classification of hyperspectral colon biopsy samples using clbp. In *IEEE International Symposium on Biomedical Imaging: From Nano to Macro ISBI*, pages 1011–1014. IEEE, 2009.
- [119] MATLAB. *version 7.10.0 (R2010a)*. The MathWorks Inc., Natick, Massachusetts, 2010.
- [120] Mike May. A better lens on disease. computerized pathology slides may help doctors make faster and more accurate diagnoses. *Scientific American*, 302:74–77, 2010.
- [121] M Mete and U Topaloglu. Statistical comparison of color model-classifier pairs in hematoxylin and eosin stained histological images. In *IEEE Symposium on CIBCB*, pages 284–91, 2009.
- [122] B Moreau and A Gouaillard. Exact geometrical predicate: Point in circle. *Insight Journal*, Nov 2011.
- [123] Kishore Mosaliganti, Lee Cooper, Richard Sharp, Raghu Machiraju, Gustavo Leone, Kun Huang, and Joel Saltz. Reconstruction of cellular biological structures from optical microscopy data. *IEEE Transactions on Visualization and Computer Graphics*, 14(4):863–876, 2008.
- [124] A Mouelhi, M Sayadi, and F Fnaiech. Automatic segmentation of clustered breast cancer cells using watershed and concave vertex graph. In *International Conference on Communications, Computing and Control Applications*, pages 1–6, March 2011.
- [125] T Mouroutis, S J Roberts, and A A Bharath. Robust cell nuclei segmentation using statistical modeling. *Bioimaging*, 6:79–91, 1998.
- [126] Douglas B Murphy and Michael W Davidson. *Fundamentals of light microscopy and electronic imaging*. Wiley. com, 2012.
- [127] David L Nelson, Albert Lester Lehninger, and Michael M Cox. *Lehninger principles of biochemistry*. Macmillan, 2008.
- [128] Kien Nguyen, Anil K Jain, and Bikash Sabata. Prostate cancer detection: Fusion of cytological and textural features. *Journal of Pathology Informatics*, 2(2):3, 2011.
- [129] Marc Niethammer, David Borland, JS Marron, John Woosley, and Nancy E Thomas. Appearance normalization of histology slides. In *Machine Learning in Medical Imaging*, pages 58–66. Springer, 2010.
- [130] Nobuyuki Otsu. An automatic threshold selection method based on discriminate and least squares criteria. *Denshi Tsushin Gakkai Ronbunshi*, 63:349–356, 1979.
- [131] Long F Ding C Peng, H. Feature selection based on mutual information: criteria of max-dependency, max-relevance, and min-redundancy,. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 27 (8):1226–1238, 2005.
- [132] Norman Zerbe Peter Hufnagl. *VIRTUAL MICROSCOPY Ū STATE OF THE ART*. 2012.
- [133] S Petushi, F U Garcia, M M Haber, C Katsinis, and A Tozeren. Large-scale computations on histology images reveal grade-differentiating parameters for breast cancer. *BMC Medical Imaging*, 6:14, 2006.

- [134] Sokol Petushi, Constantine Katsinis, Chip Coward, Fernando Garcia, and Aydin Tozeren. Automated identification of microstructures on histology slides. In *Biomedical Imaging: Nano to Macro, 2004. IEEE International Symposium on*, pages 424–427. IEEE, 2004.
- [135] F Piccinini, E Lucarelli, A Gherardi, and A Bevilacqua. Multi-image based method to correct vignetting effect in light microscopy images. *Journal of Microscopy*, 248(1):6–22, 2012.
- [136] M E Plissiti and C Nikou. Overlapping cell nuclei segmentation using a spatially adaptive active physical model. *Image Processing, IEEE Transactions on*, 21(11):4568–4580, 2012.
- [137] Marina E Plissiti and Christophoros Nikou. A review of automated techniques for cervical cell image analysis and classification. In *Biomedical Imaging and Computational Modeling in Biomechanics*, pages 1–18. Springer, 2013.
- [138] Marina E Plissiti, Christophoros Nikou, and Antonia Charchanti. Automated detection of cell nuclei in pap smear images using morphological reconstruction and clustering. *Information Technology in Biomedicine, IEEE Transactions on*, 15(2):233–241, 2011.
- [139] Weka Machine Learning Project. Weka. URL <http://www.cs.waikato.ac.nz/~ml/weka>, 2012.
- [140] Xin Qi, Fuyong Xing, D J Foran, and Lin Yang. Robust segmentation of overlapping cells in histopathology specimens using parallel seed detection and repulsive level set. *IEEE Transactions on Biomedical Engineering*, 59(3):754–765, 2012.
- [141] Daniel Racoceanu, Nicolas Loménie, and Ludovic Roux. Cognitive virtual microscopy: a cognition-driven visual explorer for histopathology—the mico anr tecsan 2010 initiative. In *BMC Proceedings*, volume 5, page P77. BioMed Central Ltd, BioMed Central Ltd, 01/2011 2011.
- [142] W S Rasband. Imagej, 1997-2008.
- [143] Erik Reinhard, Michael Adhikhmin, Bruce Gooch, and Peter Shirley. Color transfer between images. *Computer Graphics and Applications, IEEE*, 21(5):34–41, 2001.
- [144] S Rigaud and A Gouaillard. Incremental delaunay triangulation. *Insight Journal*, Jul 2012.
- [145] Karsten Rodenacker and Ewert Bengtsson. A feature set for cytometry on digitized microscopic images. *Journal of Cellular Pathology*, 25:1–36, 2003.
- [146] Jos BTM Roerdink and Arnold Meijster. The watershed transform: Definitions, algorithms and parallelization strategies. *Fundamenta Informaticae*, 41(1):187–228, 2000.
- [147] Marcial García Rojo, Ana M Castro, Luis Gonçalves, et al. Cost action "eurotelepath": digital pathology integration in electronic health record, including primary care centres. *Diagnostic Pathology*, 6(Suppl 1):S1–6, 2011.
- [148] M A Roula, A Bouridane, F Kurugollu, and A Amira. A quadratic classifier based on multi-spectral texture features for prostate cancer diagnosis. In *Seventh International Symposium on Signal Processing and Its Applications*, volume 2, pages 37–40. IEEE, 2003.
- [149] V Roullier, O Lézoray, V T Ta, and A Elmoataz. Multi-resolution graph based analysis of histopathological whole slide images: Application to mitotic cell extraction and visualization. *Computerized Medical Imaging and Graphics*, 35:603–615, 2011.
- [150] Ludovic Roux, Daniel Racoceanu, Nicolas Lomenie, Maria Kulikova, Humayun Irshad, Jacques Klossa, Frederique Capron, Catherine Genestie, Gilles Le Naour, and Metin N Gurcan. Mitosis detection in breast cancer histological images an icpr 2012 contest. *Journal of Pathology Informatics*, 4:8, May 2013.
- [151] Ludovic Roux, Adina Tutac, Nicolas Loménie, Didier Balensi, Daniel Racoceanu, Antoine Veillard, Wee-Kheng Leow, Jacques Klossa, and Thomas C Putti. A cognitive virtual microscopic framework for knowledge-based exploration of large microscopic images in breast cancer histopathology. In *Engineering in Medicine and Biology Society, 2009. EMBC 2009. Annual International Conference of the IEEE*, pages 3697–3702. IEEE, 2009.
- [152] Raphael Rubin and David S Strayer. *Rubin's Pathology: Clinicopathologic Foundations of Medicine*. Lippincott Williams & Wilkins, fourth edition, April 2004.
- [153] Arnout C Ruifrok, Dennis A Johnston, et al. Quantification of histochemical staining by color deconvolution. *Analytical and Quantitative Cytology and Histology / The International Academy of Cytology [and] American Society of Cytology*, 23(4):291, 2001.

- [154] Matthias Schlachter, Marco Reiser, Corinna Herz, Fabienne Schlurmann, Silke Lassmann, Martin Werner, Hans Burkhardt, and Olaf Ronneberger. Harmonic filters for 3d multichannel data: rotation invariant detection of mitoses in colorectal cancer. *Medical Imaging, IEEE Transactions on*, 29(8):1485–1495, 2010.
- [155] Oliver Schmitt and Maria Hasse. Radial symmetries based decomposition of cell clusters in binary and gray level images. *Pattern Recognition*, 41(6):1905–1923, 2008.
- [156] Jean Serra. *Image Analysis and Mathematical Morphology*. Academic Press, Inc., Orlando, FL, USA, 1983.
- [157] O Sertel, U V Catalyurek, H Shimada, and M N Gurcan. Computer-aided prognosis of neuroblastoma: Detection of mitosis and karyorrhexis cells in digitized histological images. In *31st Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, pages 1433–1436, September 2009.
- [158] O Sertel, G Lozanski, A Shana’ah, and M N Gurcan. Computer-aided detection of centroblasts for follicular lymphoma grading using adaptive likelihood-based cell segmentation. *IEEE Transactions on Biomedical Engineering*, 57(10):2613–2616, 2010.
- [159] Lior Shamir, John D Delaney, Nikita Orlov, D Mark Eckley, and Ilya G Goldberg. Pattern recognition software and techniques for biological image analysis. *PLoS computational biology*, 6(11):e1000974, 2010.
- [160] Jianbo Shi and Jitendra Malik. Normalized cuts and image segmentation. In *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pages 731–737, 1997.
- [161] J Shu, H Fu, G Qiu, P Kaye, and M Ilyas. Segmenting overlapped cell nuclei in digital histopathology images. In *35th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, pages 5445–5448, July 2013.
- [162] C J Stewart, J A Duncan, M Farquharson, and J Richmond. Fine needle aspiration cytology diagnosis of malignant lymphoma and reactive lymphoid hyperplasia. *Journal of clinical pathology*, 51(3):197–203, 1998.
- [163] Michael Stierer, Harald Rosen, and Renate Weber. Nuclear pleomorphism, a strong prognostic factor in axillary node-negative small invasive breast cancer. *Breast cancer research and treatment*, 20(2):109–116, 1991.
- [164] W N Street, W H Wolberg, and O L Mangasarian. Nuclear feature extraction for breast tumor diagnosis. *IS&T/SPIE 1993 International Symposium on Electronic Imaging: Science and Technology*, 1905:861–870, 1993.
- [165] Jasjit S Suri, Kecheng Liu, Sameer Singh, Swamy N Laxminarayan, Xiaolan Zeng, and Laura Reden. Shape recovery algorithms using level sets in 2-D/3-D medical imagery: a state-of-the-art review. *IEEE Transactions on Information Technology in Biomedicine*, 6(1):8–28, 2002.
- [166] Vinh-Thong Ta, Olivier L  zoray, Abderrahim Elmoataz, and Sophie Sch  upp. Graph-based tools for microscopic cellular image segmentation. *Pattern Recognition*, 42(6):1113–1125, 2009.
- [167] Sergios Theodoridis and Konstantinos Koutroumbas. *Pattern recognition*. San Diego, CA: Academic Press, 2006.
- [168] J P Thiran and B Macq. Morphological feature extraction for the classification of digital images of cancerous tissues. *IEEE Transactions on Biomedical Engineering*, 43(10):1011–1020, 1996.
- [169] Peter T  r  k and Fu-Jen Kao. *Optical Imaging and Microscopy*. Springer, 2007.
- [170] Ardhendu Shekhar Tripathi, Atin Mathur, Mohit Daga, Manohar Kuse, and Oscar C Au. 2-simdom: A 2-sieve model for detection of mitosis in multispectral breast cancer imagery. In *20th IEEE International Conference on Image Processing (ICIP)*, 2013.
- [171] V Vapnik. *Statistical learning theory*. 1998, 1998.
- [172] Vladimir Vapnik. *The nature of statistical learning theory*. springer, 2000.

- [173] Antoine Veillard, Maria Kulikova, and Daniel Racoceanu. Cell nuclei extraction from breast cancer histopathology images using color, texture, scale and shape information. In *11th European Congress on Telepathology and 5th International Congress on Virtual Microscopy*, 2012.
- [174] Antoine Veillard, Nicolas Loménie, and Daniel Racoceanu. An exploration scheme for large images: application to breast cancer grading. In *Pattern Recognition (ICPR), 2010 20th International Conference on*, pages 3472–3475. IEEE, 2010.
- [175] M Veta, A Huisman, M A Viergever, P J van Diest, and J P W Pluim. Marker-controlled watershed segmentation of nuclei in H & E stained breast cancer biopsy images. In *8th IEEE International Symposium on Biomedical Imaging: From Nano to Macro*, pages 618–621, April 2011.
- [176] J P Vink, M B Van Leeuwen, C H M Van Deurzen, and G Haan. Efficient nucleus detector in histopathology images. *Journal of Microscopy*, 249(2):124–135, 2013.
- [177] C Wahlby, I M Sintorn, F Erlandsson, G Borgefors, and E Bengtsson. Combining intensity, edge and shape information for 2D and 3D segmentation of cell nuclei in tissue sections. *Journal of Microscopy*, 215(1):67–76, 2004.
- [178] Pierre D Wellner. Adaptive thresholding for the digital desk. *Xerox, EPC1993-110*, 1993.
- [179] Stephan Wienert, Daniel Heim, Kai Saeger, Albrecht Stenzinger, Michael Beil, Peter Hufnagl, Manfred Dietel, Carsten Denkert, and Frederick Klauschen. Detection and segmentation of cell nuclei in virtual microscopy images: a minimum-model approach. *Scientific Reports*, 2, 2012.
- [180] William H Wolberg, W Nick Street, and Olvi L Mangasarian. Breast cytology diagnosis via digital image analysis. *Analytical and Quantitative Cytology and Histology*, 15(6):396–404, 1993.
- [181] Franco Woolfe, Mauro Maggioni, Gustave Davis, Frederick Warner, Ronald Coifman, and Steven Zucker. Hyper-spectral microscopic discrimination between normal and cancerous colon biopsies. *IEEE Transactions on Medical Imaging*, 99(99), 1999.
- [182] Richard Wootton and David R Springall. *Image analysis in histology: Conventional and confocal microscopy*, volume 2. CUP Archive, 1995.
- [183] Xuqing Wu, Mojgan Amrikachi, and Shishir K Shah. Embedding topic discovery in conditional random fields model for segmenting nuclei using multispectral data. *IEEE Transactions on Biomedical Engineering*, 59(6):1539–1549, 2012.
- [184] Xianghua Xie and Majid Mirmehdi. Correction to "MAC: Magnetostatic active contour model" [apr 08 632-646]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 30(5), 2008.
- [185] Lin Yang, Peter Meer, and David J Foran. Unsupervised segmentation based on robust estimation and color active contour models. *IEEE Transactions on Information Technology in Biomedicine*, 9(3):475–486, 2005.
- [186] Lin Yang, Oncel Tuzel, Peter Meer, and David Foran. Automatic image analysis of histopathology specimens using concave vertex graph. In *Medical Image Computing and Computer-Assisted Intervention*, volume 5241 of *Lecture Notes in Computer Science*, pages 833–841. Springer Berlin / Heidelberg, 2008.
- [187] Xiaobo Zhou, Fuhai Li, Jun Yan, and Stephen TC Wong. A novel cell segmentation method and cell phase identification using markov model. *Information Technology in Biomedicine, IEEE Transactions on*, 13(2):152–157, 2009.
- [188] Xiaobo Zhou and STC Wong. Informatics challenges of high-throughput microscopy. *Signal Processing Magazine, IEEE*, 23(3):63–72, 2006.
- [189] Steven W Zucker. Region growing: Childhood and adolescence. *Computer Graphics and Image Processing*, 5(3):382–399, 1976.
- [190] Mark H Zweig and Gregory Campbell. Receiver-operating characteristic (roc) plots: a fundamental evaluation tool in clinical medicine. *Clinical chemistry*, 39(4):561–577, 1993.

Index

- 2-Manifold Mesh, 129
- Acc, 53
- ACM, 32, 42, 45, 46
- Anisotropic Diffusion, 38
- Aperio Scanner, 59, 63
- BR, 64
- Breast Cancer, 20
- CAD, 15
- Calopix, 99, 140
- CCD, 17
- CCR, 53
- CD, 52
- CNA, 58, 97
- CNN, 41, 51
- Color Deconvolution, 37, 81
- Color Microscopy, 18
- Color Normalization, 37
- Concave Point Detection, 45
- CV, 54
- Delaunay, 133
- DoG, 42
- DT, 74
- Duality, 132
- Dynamic Sampling, 140
- EM, 33, 41, 49
- ER, 52
- FCM, 40
- FD, 50
- Feature Normalization, 73
- Feature Selection, 73
- FM, 52
- FN, 52
- FNAR, 53
- Focal Plane Selection, 114
- FP, 52
- FPAR, 52
- Gcut, 33, 43
- GiPS, 42
- GLCM, 67
- GLRLM, 72
- GMM, 33, 41, 49
- H-Maxima Transform, 39
- H&E, 16, 109
- Hamamatsu Scanner, 59, 63
- HC, 50, 67
- HD, 53
- HES, 16
- Histogram Equalization, 38
- Histopathology, 15
- HPF, 23, 58, 63, 99
- IHC, 16
- Illumination, 37
- ITM²C, 58, 81
- K-means, 32, 50
- LoG, 39
- LOOCV, 54
- Lymphocytic Infiltration, 41
- MAD, 53
- MAP, 41
- Margin Curve, 93, 123
- MI, 57, 109
- MI2UR, 53
- MICO, 58, 97, 100, 140
- MITM³, 58, 107
- MITOS ICPR Contest 2012, 93, 123
- MLP, 74
- MMSF, 58, 106, 107, 114
- MOLB, 44
- Morphology, 30, 38, 42, 67
- MPP, 46
- MRF, 41, 50
- MRGC, 43
- mRMR, 109
- MSI, 26, 51, 57, 106
- Multispectral, 51
- Multispectral Microscopy, 18, 59, 106
- Ncut, 34
- NGS, 23, 97
- OR, 53

OV, 53

PE, 53
PPV, 52
PRC, 53
Probabilistic Models, 33

Region Growing, 31
RL, 50, 72
RMSE, 52
ROC, 53, 93, 123
ROI, 38
RST, 40, 48

SB, 51
SDE, 53
SIFT, 50
Simplex Mesh, 132, 133
Slide Preparation, 16
Slide Scanner, 17
SMOTE, 85
Spectral Bands Selection, 114
SVM, 50, 51, 75, 76

Thresholding, 30, 38, 42, 67
TMA, 46
TMC, 58, 63
TN, 52
TNR, 52
TP, 52
TPR, 52
Triangulation Mesh, 131, 132

Voronoi, 43, 133, 140

Watershed, 31, 42, 44
WSI, 18, 97, 140