

**MACHINE LEARNING FOR CANCER DETECTION, GRADING, AND PROGNOSIS:
AUTOMATED SEGMENTATION, PATTERNING RECOGNITION, AND AI-BASED
ANALYSIS**

by

Nilgoon Zarei

B.Sc., Shiraz University, 2009

M.Sc, Simon Fraser University, 2013

A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF
THE REQUIREMENTS FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

in

THE FACULTY OF GRADUATE AND POSTDOCTORAL STUDIES
(Interdisciplinary Oncology)

THE UNIVERSITY OF BRITISH COLUMBIA
(Vancouver)

December 2019

© Nilgoon Zarei, 2019

The following individuals certify that they have read, and recommend to the Faculty of Graduate and Postdoctoral Studies for acceptance, the dissertation entitled:

Machine Learning for Cancer Detection, Grading, and Prognosis: Automated Segmentation, Patterning Recognition, and AI-Based Analysis

submitted by Nilgoon Zarei in partial fulfillment of the requirements for

the degree of Doctorate of Philosophy

in Interdisciplinary Oncology

Examining Committee:

Dr. Calum MacAulay, Pathology and Laboratory Medicine

Supervisor

Dr. Pierre Lane, Engineering Science

Supervisory Committee Member

Dr. Haishan Zhang, Physics and Astronomy

Supervisory Committee Member

Dr. Stefan Reinsberg, Physics and Astronomy

University Examiner

Dr. Shahriar Mirabbasi, Electrical and Computer Engineering

University Examiner

Additional Supervisory Committee Members:

Dr. Mira Keyes, Radiation Oncology

Supervisory Committee Member

Abstract

In this thesis, we presented the design steps for developing new, reliable, and cost-effective diagnostic and prognostic tools for cancer using advanced Machine Learning (ML) techniques.

We proposed tools to improve the diagnostic, prognostic and detection accuracy of quantitative digital pathology by incorporating advanced image analysis, image processing, and classification methods.

In this thesis we presented our ML image-based analytic approaches for three cancer types (prostate, cervix, and kidney) with different scale ranges from the sub-micron to multiple centimeters.

In this thesis, we demonstrated the full workflow to design an automated prognostic and grading system specially designed for prostate cancer. We started with demonstrating techniques for prostate glandular structures detection. Next, we introduced an automated cell segmentation method along with an interactive segmentation correction method requiring minimum user-interaction and finally, we introduced our ML classification algorithms. We trained our ML method on the features extracted from cells/nuclei that were segmented via our proposed techniques.

Next, we studied renal carcinoma. We presented the workflow of renal carcinoma classification from image processing to feature selection and development of machine learning classification techniques. We extracted the features from renal vessel structures and demonstrated the design steps of machine learning classifiers to discriminate between different renal carcinoma subtypes using these features.

The last cancer site studied was the cervix. We applied our techniques for cervical pre-cancer abnormality detection. We showed the whole pipeline of designing an automated classification method starting from tissue imaging to the development of ML classifiers using both classical and deep-learning methods.

Although we conducted these studies on specific cancer types, a modified version of our algorithm could be applied to other cancers and disease sites.

These techniques have great potential to improve the healthcare environment by providing extra information/second opinion to the medical experts or to be used as a part of the first line of a screening program.

Lay Summary

Cancer is a worldwide health problem and is the second leading cause of death in the North America. The number of patients diagnosed with cancer increases daily and despite advances in medical technology and extensive research, the incident rate of cancer is still very high.

Over the past few decades, Machine Learning (ML) methods have been applied successfully in the healthcare domain. ML offers promising avenues for prediction of disease progression, extraction of clinically relevant information from the data, therapy planning, patient management and much more.

In this thesis, we presented methods for automated detection and diagnosis using advanced ML techniques to classify tissues into aggressive/non-aggressive and detect abnormalities. These methods, if used in clinical settings, could improve patients life quality and increase the cancer survival rate.

In this thesis, we showed the design steps for developing new, reliable, and cost-effective grading/diagnostic/prognostic tools for prostate, renal and cervical cancer.

Preface

This thesis presents research conducted by Nilgoon Zarei, under the guidance and supervision of Dr. Calum MacAulay.

Chapter 2: A version of this material has been published in [C1]. Nilgoon was the lead investigator, responsible for all major areas of concept formation, and analysis, as well as manuscript composition. Amir Bakhtiari helped with algorithm development, Paul Gallagher was responsible for image preparations using the Hyperspectral imaging system. The entire work was conducted under the supervision of Dr. Calum MacAulay and support and input from Dr. Pierre lane, and Dr. Mira Keys.

Chapter 3 and Chapter 4: A version of this material has been published in [J1] and [J2]. Nilgoon was the research lead on this study. She was responsible for the study design, algorithm development and validation as well as manuscript composition. The entire work was conducted under the supervision of Dr. Calum MacAulay and support and input from Dr. Pierre Lane, Dr. Mira Keys and Dr. Martial Guillaud.

Chapter 5: A version of this material has been published in [C2]. Nilgoon was the research lead on this study. Nilgoon was responsible for the design and coding. Amir Bakhtiari helped with algorithm development. The entire work was conducted under the supervision of Dr. Calum MacAulay.

Chapter 6: The content and patient material of this chapter is based on a collaboration with Inria France, with Dr. Xavier Descombes' guidance. Dr Ambrosetti was responsible for data collection and image annotation. Nilgoon was responsible for the literature review, feature implementation and data analysis.

Chapter 7 and Chapter 8: A version of this material has been published in [C3] and [C4]. The development and construction of the Multispectral Imaging System (developing the device and collecting the patient images) was performed as a part of a NIH funded collaboration with US researchers Drs. Michele Follen, Rebecca Richards-Kortum, Dennis Cox, Neely Atkinson, Drs Pierre Lane, Matrial Guillaud and Calum MacAulay at the BCCRC. Cantor SB., Atkinson EN., Jose-Miguel Yamal, Leonid Fradkin, Daniel Serachitopol, Sylvia Lam, Drik Niekerk, Dianne Miller, Jessica McAlpine , Kayla Castaneda and Felipe Castaneda helped with the collection, creation and annotation of TMA images. Nilgoon was the research lead on the study. She was responsible for the literature review, designing and implementing the proposed algorithms, performing all experiments, analyzing the results and writing the manuscript. The entire work was conducted under the supervision of Dr. Calum MacAulay and support and input from Dr. Pierre Lane, and Dr. Michele Follen.

The thesis was written by Nilgoon Zarei, with editing assistance from Dr. Calum MacAulay.

[C1] Zarei N.*, Bakhtiari A., Gallagher P., Keys M., MacAulay C. (2017) “Automated Prostate Glandular and Nuclei Detection Using Hyperspectral Imaging”. IEEE Biomedical Imaging 1028-1031.

[J1] Zarei N. *, Keyes M., Guillaud M., MacAulay C.(2016) “Automated Localized Region-based Prostate Cancer (PCs) Cell Segmentation as a Prognostic Modality Tool for PCs Patients” Analytical and Quantitative Cytology and Histology 38(2):59

[J2] Zarei N. *, Keyes M., Guillaud M., MacAulay C.(2016) “Introducing an Interactive Method to Improve Digital Pathology Image Segmentation: Case Study on Prostate Cancer” Analytical and Quantitative Cytology and Histology, 39(1), 1-16.

[C2] Zarei N.*, Bakhtiari A., MacAulay C. (2016) “Designing an Automated Classification Method Using Latent Support Vector Machine Algorithms as a Predictive Tool for Prostate Cancer Lesions”. SPIE Medical Imaging, San Diego, USA.

[C3] Zarei N.*, Cox DD., Lane P., Cantor SB., Atkinson EN., Yamal JM., Fradkin L., Serachitopol D., Lam S., Niekerk D., Miller D., McAlpine J., Castaneda K., Castaneda F., Follen M., MacAulay C. (2017) “Boosted Tree Classifier for in Vivo Identification of Early Cervical Cancer using Multispectral Digital Colposcopy”. Bioimaging 2017,85-91.

[C4] Zarei N.*, Lane P., Lam S., Follen M., MacAulay C. (2018) “Deep learning techniques for identification of abnormalities in early cervical cancer using multispectral digital colposcopy”, SPIE, Multimodal Biomedical Imaging, San Fran. USA.

Table of Contents

Abstract	iii
Lay Summary.....	v
Preface	vi
Table of Contents	ix
List of Tables	xiv
List of Figures.....	xvii
List of Abbreviations.....	xxxiv
Acknowledgments.....	xxxvi
Dedication.....	xxxviii
Chapter 1 Introduction	1
1.1 Cancer Statistics.....	1
1.2 Personalized Medicine – Oncology	2
1.3 Cellular and Tissue Level Analysis	3
1.4 Digital Pathology	4
1.5 The Necessity for Development of Automated Diagnostic Methods	4
1.6 Challenges	5
1.6.1 Staining and Imaging Issue	5
1.6.2 Lack of Gold Standard	5
1.6.3 Diagnoses of Atypia.....	6
1.7 Computer-Aided Diagnosis (CAD)	6
1.7.1 Localization and Segmentation	7
1.7.2 Feature Extraction and Classification	7
1.7.3 Disease Diagnosis and Prognosis	7
1.8 Application of Machine Learning (ML) in Medicine	8
1.8.1 Supervised Learning	9
1.8.2 Unsupervised Learning	10
1.8.3 Classical ML vs Deep Learning Methods	11
1.9 State of the Art ML for Cancer Diagnosis/ Prognosis	12
1.9.1 Cancer Detection/ Diagnosis	13

1.9.2	Prediction of Individual Cancer Susceptibility.....	14
1.9.3	Prediction of Cancer Recurrence	15
1.9.4	Prediction of Cancer Survival	15
1.9.5	Deep Learning and Cancer	15
1.10	Motivation	17
1.11	Structure of This Thesis	18
Chapter 2	Automated Prostate Glandular Detection Using Hyper-Spectral Imaging	24
2.1	Introduction	24
2.1.1	Prostate Cancer Epidemiology, Screening and Diagnostic Techniques	24
2.1.2	Automated Gland Detection	26
2.2	Hypothesis	27
2.3	Objective	27
2.3.1	Specific Objective.....	27
2.4	Materials and Instrumentation	28
2.4.1	Clinical Data	28
2.4.2	Instrumentation	28
2.4.3	Data Preparation	30
2.5	Methodology.....	30
2.5.1	Principal Component Analysis	30
2.5.2	K-means Clustering and Morphological Cleaning	32
2.6	Results	35
2.7	Discussion.....	39
2.8	Conclusion.....	41
Chapter 3	Design of an Automated Technique for Digital Pathology Cell Segmentation	42
3.1	Introduction	42
3.1.1	Medical Image Segmentation Methods.....	43
3.2	Hypothesis	46
3.3	Objective	47
3.3.1	Specific Objective.....	47
3.4	Materials and Data Preparation	48
3.5	Methodology.....	49

3.5.1	Automated Prostate Cancer Nuclei Segmentation.....	49
3.5.2	Algorithm Performance and Experiment Results	62
3.6	Discussion.....	63
Chapter 4	Interactive Cell Segmentation for Digital Pathology	64
4.1	Introduction	64
4.2	Motivation	64
4.3	Objective/Hypothesis	66
4.4	Methodology.....	66
4.5	Results	74
4.6	Discussion.....	76
4.6.1	Generalization to Other Cancer Sites and Comparison with Ilastik.....	76
4.6.2	Reproducibility Analysis.....	87
4.7	Conclusion.....	89
Chapter 5	Design of Machine Learning Methods as a Predictive Tool for Prostate Cancer ..	91
5.1	Introduction	91
5.2	Hypothesis	91
5.3	Objective	92
5.3.1	Specific Objective.....	92
5.4	Materials and Instrumentation	93
5.4.1	Clinical data.....	93
5.5	Methodology.....	95
5.5.1	Feature Selection.....	95
5.6	Experiment and Result	116
5.6.1	ML for Prostate Gleason Score Grading.....	117
5.6.2	ML for Prostate Prognoses.....	117
5.7	Conclusion.....	121
Chapter 6	Machine Learning for Renal Cell Carcinoma (RCC) Detection	122
6.1	Background-Renal Carcinoma.....	122
6.2	Hypothesis	123
6.3	Objective	124
6.3.1	Specific Objective.....	125

6.1	Material	126
6.2	Methodology.....	127
6.3	Feature Extraction and ML Algorithm Design.....	130
6.3.1	Feature Extraction.....	131
6.3.2	Design/Select the Classifier.....	132
6.4	Results	135
6.5	Discussion and Future Directions	139
Chapter 7	Cervical Cancer Epidemiology, Screening and Diagnostic Techniques.....	142
7.1	Objective	144
7.1.1	Specific Objective.....	145
7.2	Materials and Methods.....	146
7.2.1	Instrumentation	146
7.2.2	Clinical Data.....	148
7.3	Methodology-Conventional ML from Data Preparation to Data Analysis	148
7.3.1	Data Preparation	148
7.3.2	MDC Images and Pathological Map.....	150
7.3.3	Feature Extraction and ML Algorithm Design.....	153
7.4	Experiment and Result	155
7.4.1	Improving the Performance.....	158
7.5	Discussion and Future Directions	167
7.5.1	Future Direction and More Evaluation of the Study Procedure	168
7.6	Conclusion.....	173
Chapter 8	Deep Learning for Cervical Cancer Detection	174
8.1	Introduction	174
8.2	Objective	175
8.2.1	Specific Objectives	175
8.3	Methodology-Convolutional Neural Network Architecture	176
8.3.1	Patch-based CNNs	176
8.3.2	Deep CNNs on Patches	178
8.3.3	Problem Formulation	179
8.3.4	Data Preparation	184

8.4	Experiment and Results.....	185
8.4.1	Per MDC Channel Performance Analysis.....	192
8.4.2	Improving the Performance.....	197
8.5	Discussion.....	199
Chapter 9	Conclusion and Future Work.....	201
9.1	Prostate Cancer: Gland Detection, Image Segmentation, and Prediction.....	201
9.2	Renal Carcinoma: Vascular Structure Segmentation, Feature Extraction, and Classification.....	203
9.3	Cervical Cancer: Abnormalities Detection Using Classical and Convolutional Neural Network	204
9.4	Future Directions	205
9.5	Significance and Application of This Study.....	208
References	209
Appendix	221

List of Tables

Table 1-1 Selected summary of deep-learning publication for cancer detection/diagnosis/prognosis	16
Table 3-1 Automated PCa nuclei segmentation algorithm	61
Table 3-2 Processed nuclei segmentation results with our proposed algorithm.	63
Table 4-1 Interactive nuclei segmentation algorithm steps	73
Table 4-2 Prostate tissue (H&E) segmentation results for 10 randomly selected images.....	81
Table 4-3 Oral tissue (Thionin) segmentation results for 10 randomly selected images.	83
Table 4-4 Cervix tissue (Thionin) segmentation results for 10 randomly selected images...85	
Table 4-5 Lung tissue (Thionin) segmentation results for 10 randomly selected images.	87
Table 5-1 Number of patients with different Gleason scores in failure and no- failure groups	93
Table 5-2 The classification statistics for classifying aggressive vs non-aggressive on the test dataset.....	118
Table 5-3 The classification statistics for classifying 6-7 vs 8-9 Gleason score cases on the test set.	120

Table 6-1 classification results for CCRCC and other tissue sites using a boosted tree classifier on the training set	136
Table 6-2 potential feature list for renal carcinoma study	141
Table 7-1 Tiles used for MDC device calibration.....	147
Table 7-2 CIN level of patients for this study.....	149
Table 7-3 The statistics results using a boosted tree classifier on the training set, using histogram intensities as the features.....	156
Table 7-4 The statistics results using a boosted tree classifier on the test set, using histogram intensities as the features.	157
Table 7-5 Table of MDC channel names and acronyms.	159
Table 7-6 The statistics results using a boosted tree classifier on the training set.....	162
Table 7-7 The statistics results using a boosted tree classifier on the test set.	163
Table 7-8 The statistics results using a boosted tree classifier on the training set.....	165
Table 7-9 The statistics results using a boosted tree classifier on the test set.	166
Table 7-10 CIN level of the second half of the patients recruited for further studies.....	168
Table 7-11 The statistics results using a boosted tree classifier on the training set.	169

Table 7-12 The statistics results using a boosted tree classifier on the test set.....	170
Table 7-13 The statistics results using a boosted tree classifier trained on the old-data and tested on the new data.	172
Table 8-1 Hyper-parameter table for Adam optimizer	184
Table 8-2 Simple model summary table, shows input to each layer and number of parameters.....	187
Table 8-3 VGG16 summary table, shows input to each layer and number of parameters	189
Table 8-4 VGG16 summary table, shows input to each layer and number of parameters	190
Table 8-5 comparison between different networks	192

List of Figures

Figure 1-1 This figure shows the distribution of the number of publications related to cancer and machine learning per year. As we can see in this histogram, the number of publications in this domain increased over the years. This histogram is normalized by the number the total number of publications on Machine Learning.	8
Figure 1-2 Flowchart of dissertation, chapters 2, 3, 4 and 5.....	21
Figure 1-3 Flowchart of dissertation, chapter 6.....	22
Figure 1-4 Flowchart of dissertation, chapters 7 and 8.	23
Figure 2-1. This figure shows the pipeline of our study. The focus of this chapter is on the Glandular Segmentation. In the later chapters, we will explain the classification and validation processes.....	28
Figure 2-2. This figure shows our hyperspectral image acquisition set-up (on the left) and a representation of hyperspectral image construction that could be acquired by this system (on the right). 16 images with different wavelengths (420-720 nm) were captured with this imaging system (only 8 shown for illustration), the images are then combined and PCA images are constructed.....	29
Figure 2-3. Process of constructing the RGB images from PCA. The image in the red-box is the first three PCA components and is used for further analysis. The second image (with orange background) is second, third and fourth PCA components and the last image is the	

third fourth and fifth PCA components. No satisfactory result is obtained using the last two images. They are presented here only for illustration.32

Figure 2-4 This figure shows the steps required for prostate glandular detection as well as detection of its nuclei. First, the PC RGB image transferred into LAB space and then using K-means clustering, gland/stroma, stroma, and glass are segmented. Morphological cleaning is used to remove undesired stroma in the gland/stroma image. At the last step, another K-means clustering sequence is applied to detect nuclei within each detected gland.34

Figure 2-5 This figure shows examples of “good” glandular structure detection, the top row shows the PC RGB images. GL stands for glass, S stands for stroma and N stands for nerve trunk. The second row shows the gland detection results where G stands for gland; we showed an example of nerve trunk detection and removal in the third image. The third row shows the nuclei segmentation results.36

Figure 2-6 This figure shows an example of “moderate” glandular detection. Some glands are missed due to poor sectioning/scanning/imaging processes.37

Figure 2-7 This figure shows an example of “moderate” glandular detection. This is an example of a core with densely packed tumor glands and small luminal spaces. In this example, the tumor has glandular structures with slit-like luminal spaces. The gland image is not shown here.37

Figure 2-8 This figure shows examples of “bad” glandular detection in which the stroma is infiltrated by inflammatory cells. (A) the PCA image with a zoom-out of a section of the core

with glands. (B) shows the failure of the gland detection. In the red box, multiple glands were mislabeled and were not detected. The nuclei image is not shown here.....38

Figure 2-9 This figure shows examples of “bad” glandular detection, image (A) and image (B) show examples of cores in which the stroma is infiltrated by inflammatory cells.39

Figure 2-10 Example of gland detection result using the code presented by (Alexe, Deselaers et al. June 2010). The input image is an RGB image from the same core as shown in image 2 before the PCA transformation. The superimposed rectangles show the locations with a high probability of the presence of glands (mostly in the center.)40

Figure 3-1 This figure shows the pipeline of our study. In the previous chapter, we discussed our algorithm for glandular segmentation. The focus of this chapter is on the cell segmentation. In the later chapters, we will describe the classification and validation processes.48

Figure 3-2 Thionin-stained prostate cancer core from a tissue microarray, a) 16 sub-images. b) the whole prostate core image obtained from (a).49

Figure 3-3 Gaussian filter. a) Perspective plot of Gaussian filter b) filter display as an image c) filter structure for different values of D_050

Figure 3-4 Process of image thresholding by Mean Gradient histogram and Gradient weight. an) input image b) binary image obtained from GW, c) binary image obtained from MG. The red circles enclose cells where the difference between the two thresholding methods is obvious.52

Figure 3-5 Morphological image cleaning and labeling. a) binary image b) cleaned image c) labeled image, each color represents a connected component. The red circles enclose areas where debris existed and where debris removed upon applying the proposed algorithm. ...	53
Figure 3-6 Global thresholding versus local thresholding (a) whole image (b) histogram of the whole image (c) region of interest (d) histogram of (c).	54
Figure 3-7 An example of segmentation failure with simple thresholding. Failure is due to clumped nuclei. (a) whole image (b) cluster of nuclei, (c) segmentation mask from (b).....	56
Figure 3-8 Localization procedure for clustered nuclei. This method finds the skeleton of a cluster and uses the endpoints as a possible location for the cell center. As we can see in the rightmost figure, the initial four nuclei resulted in six potential nuclei centers, we will explain how to find the four nuclei out of these six spots.....	57
Figure 3-9 Nuclei segmentation for three different object types (isolated nucleus, a single nucleus with overlapping neighbor, and a cluster of nuclei).	58
Figure 3-10 Nuclei segmentation results for different nucleus arrangement.....	59
Figure 3-11 Nuclei detection result for prostate TMA tissue. Red dots represent detected nuclei.....	60
Figure 3-12 Gallery of segmented nuclei using our segmentation algorithm.....	62

Figure 3-13 A carton that represents the whole processes used in our algorithm for the detection of prostate cancer nuclei in TMA slides.62

Figure 4-1 Faint nuclei resulted in segmentation failure. a) a section of prostate tissue (Thionin stained), blue objects are cell nuclei, light green dots (not very visible) are cell locations predicted by our algorithm. b) zoomed-in section of image(a) red circle circumscribes a pale nucleus that our algorithm missed.65

Figure 4-2 Failure of segmentation method on a cluster of nuclei. a) blue objects are cell nuclei and the light green is the algorithm prediction result for the location of nuclei (missing nuclei). b) incorrect localization of cell nuclei by our algorithm (over-segmentation) for a cluster of nuclei.65

Figure 4-3 Interactive correction method to add/remove the nuclei. a) original image, the red cross represents the interactive method on the location where we plan to apply some correction. b) extracted the nucleus and surrounding neighbors.....67

Figure 4-4 Thresholding based on local information a) masked image. b) extracted edges in the area c) dilated edges using Canny d) extracted mask using Otsu thresholding.....68

Figure 4-5 Applying region-growing methods a) original image, selected nucleus and zoomed nucleus. b) original region growing method c) region growing result using hysteresis.69

Figure 4-6 Clustering flowchart, shows the steps necessary to select relevant clusters. This algorithm consists of two clustering methods (K-means clustering and density-based clustering) in the first run, pixels are selected if they belong to the same cluster (generated

from two clustering methods), on the second run we feed the remaining pixels to the density-based clustering and choose those with large area.....72

Figure 4-7 Comparison between clustering methods (Density-based clustering, K-means clustering and hybrid method) a) input image. b) clustering result of Density-based clustering. c) clustering result of K-means clustering. d) results of our hybrid clustering method.....73

Figure 4-8 Section of Prostate TMA core, selected for comparing the segmentation results with/without our interactive method.....75

Figure 4-9 Automated segmentation result with our interactive method.75

Figure 4-10 Automated segmentation result without our interactive method.....76

Figure 4-11 Comparison between Ilastik nuclei segmentation and our algorithm. (a) a selected image from prostate TMA which is Thionin stained (column b) Ilastik nuclei segmentation required steps (manually selection of first label, nuclei, (I) manual selection of background, stroma, (II) final results is presented in (III) each color in (III) represents an object) (column c) our algorithm nuclei segmentation steps (automated localization of nuclei with minor adjustments(I), final results presented in (II), again each color in (II) represents an object)78

Figure 4-12 This figure shows the comparison between Ilastik nuclei segmentation and our algorithm. (a) a selected image from prostate TMA, Thionin stained (column b) Ilastik nuclei segmentation steps (manually selection of first label, nuclei, (I) manual

selection of background, stroma, (II) and final results presented in (III) each color in (III) represents an object) (column c) our algorithm nuclei segmentation steps (automated localization of nuclei with minor adjustments(I), final results presented in (II), again each color in (II) represents an object)80

Figure 4-13 This figure shows the comparisons between ILASTIK nuclei segmentation and our algorithm performance on oral tissue. (a) a selected image from oral tissue, which is Thionin stained, (b) our segmentation results where each color represents an object. (c) gallery of selected nuclei from our segmentation method. (d) original oral image (the same as (a)), and segmentation result from ILASTIK nuclei segmentation is presented in (e) where each color represents an object.....82

Figure 4-14 Comparison between ILASTIK nuclei segmentation and our algorithm for cervix tissue. (a) a selected image from cervix tissue, Thionin stain, we manually select the epithelial layer for segmentation purposes. (b) Automated nuclei localization with our algorithm, (c) the segmentation results from our algorithm where each color represents an individual object. (d) the first and second steps of ILASTIK segmentation, manual selections of two labels (red for cells and green for background/stroma). (e) segmentation results from ILASTIK where each color represents an object.....84

Figure 4-15 Comparison between ILASTIK nuclei segmentation and our algorithm for lung tissue. (a) a selected image from lung tissue, Thionin stain, (b) Automated nuclei localization with our algorithm, (c) the segmentation results from our algorithm where each color represents an individual object. (d) original image presented in (a). (e) the first and second steps of ILASTIK segmentation, manual selections of two labels (red for cells and green for

background/stroma). (f) segmentation results from ILASTIK where each color represents an object.	86
Figure 4-16 Reproducibility analysis (segmentation results and the dependency of the algorithm to user input/selection). The blue dot represents the original location selected by the segmentation algorithm and other dots are potential locations that the user could possibly select.....	88
Figure 4-17 Reproducibility analysis of five cores. Statistic results on segmented cells obtained from cells/nuclei features.....	89
Figure 5-1 This figure shows the pipeline of our study. The focus of this chapter is on the classification step.....	93
Figure 5-2 This figure shows the steps required from segmentation, (A) shows a part of an image used for segmentation (B) shows segmentation results and some statistics (area of cells) on the left panel and (C) shows some relevant features, such as area, nucleus density, etc.	94
Figure 5-3 An image of the Voronoi Tessellation feature on a part of TMA tissue.....	112
Figure 5-4 A Simplified cartoon representing an overview of our LSVM algorithm.....	114
Figure 5-5 In this figure we present the ROC result of Gleason grade classifications between different scores. The black line represents a random line. The purple line is the baseline using SVM. LSVM is used to classify between 6-7 vs 8-9 red line, 6 vs 7 shown with the blue line and 8 vs 9 shown with the green line.	119

Figure 5-6 In this figure we present the ROC result of aggressive vs non-aggressive classification for patients with Gleason score 7. We used both SVM and LSVM as our classification method, the results are shown with green and blue lines respectively.....120

Figure 6-1 Microenvironment features. (a) shows a clinical image of renal carcinoma where lots of large cells with clear cytoplasm and a very large vessel in the tissue are present. (b) shows the cell/nuclei sociology/ tree-pattern. (c) shows the vessel and detected branches which is an indication of angiogenesis.123

Figure 6-2 Renal tissue, pathological features. In this image, many pathological features are illustrated, although, in this study, we are interested in angiogenesis hallmark, other pathological indications such as infiltration of inflammatory cells are worth mentioning. Other features such as abundant clear cell carcinoma, although is not discussed in this chapter, were the topic of previous chapters where we showed the importance of nuclei features and how they can be used for classification purposes.....124

Figure 6-3 This figure shows the pipeline of our renal carcinoma study. The processes consist of imaging and pre-processing of the captured images. The next step is post-processing and extracting/segmenting clinical important features and the last step is ML techniques implementation for classification of normal and abnormal cases.125

Figure 6-4 This image shows the Sliding window implementation. This image illustrates the process required to slice the large image to smaller pieces so that our computing instrument could handle the pre/post-processing step (required image processing). (a) an image of renal

carcinoma tumor, (b) manually annotated image where the black area represents the tumor and other areas represent the healthy tissue. (c) a sliding window is shown on the image.126

Figure 6-5 Pipeline of the pre/post-processing. This workflow starts with an RGB image, pathological image (input) and the resulted output is the skeleton of the vessels where all the nuclei and other structures are removed.....129

Figure 6-6 The whole process of the pre-post processing step. This workflow starts with an RGB image, the pathological image of renal carcinoma (input) and ends with the output which is the skeleton of the vessels where all the nuclei and other structures are removed.130

Figure 6-7 Diagram of a simple decision tree. In each node, a selection is made as shown in red.....134

Figure 6-8 Diagram of the Random Forest algorithm, this algorithm is based on an ensemble decision making where each tree can have many branches. The final class selection would be based on the majority votes of all trees.134

Figure 6-9 Diagram of Boosted Tree algorithm. Boosted tree is an ensemble algorithm that uses the majority votes of some stump trees, in each step the weights for the next decision making would be updated as shown with the green arrows above.135

Figure 6-10 This figure represents the ROC of all 5 classifiers (linear discriminate, logistic regression, decision tree, random forest, boosted tree classifier) for classification between CCRCC and other tissue sites.....136

Figure 6-11 This figure represents the ROC of all 5 classifiers (linear discriminate, logistic regression, decision tree, random forest, boosted tree classifier) for classification between CCRCC and other tissue sites when the features are normalized.....	137
Figure 6-12 This figure represents the ROC of all 5 classifiers (logistic regression, decision tree, random forest, boosted tree classifier) for classification between CCP and other tissue sites.	138
Figure 6-13 This figure represents the ROC of all 5 classifiers (logistic regression, decision tree, random forest, boosted tree classifier) for classification between PRCC and other tissue sites.	139
Figure 7-1. This figure shows the pipeline of our study. The processes mentioned in the blue boxes are completed and fully developed by our group. The processes illustrated in green boxes are the focus of this chapter.....	145
Figure 7-2 This figure shows A) a schematic diagram of the Multispectral Digital Colposcope (MDC) device, B) a photo of the “in house” device, and C) the table of the illumination color and excitation wavelengths.	147
Figure 7-3. This figure shows the study design. Patients were consented for an IRB-approved study, images were acquired, the Loop Electrical Excision Procedure (LEEP) was carried out, the specimen was processed and annotated, and the dataset was subjected to analysis for training and testing classification	149

Figure 7-4. The protocol to select regions of interest. (a) pathology map, where the green yellow and blue annotations represent high grade, low grade, and squamous columnar junction respectively. (b) reviewers selected the region of interest (each color boundaries represents a reviewer’s drawing) on the RGB image of MDC using the image provided in (a). (c) shows the heatmaps where the hot or glowing areas represent the most agreement between reviewers’ selections of ROIs..... 150

Figure 7-5 This figure shows four patients' results. In each column, we presented the clinical impression of the colposcopy physician, the histopathologic map, the tracing by the 11 reviewers, the agreement map, and the mask generated by the calculation of 60% agreement. 152

Figure 7-6 This figure shows the strategy for the analysis. Eighty percent of the data was used to train and validate a classifier, shown in A. The classifier was tested on the remaining data (B). Output from a sample patient is shown in C1 and C2. The probability of disease is shown in C1 and the Receiver Operating Characteristic (ROC) curve is shown in C2. 154

Figure 7-7 This figure represents the ROC of the boosted tree classifier on the training set (using 80% of the data (for 39 patients)) and using histogram intensities as the features. 155

Figure 7-8 This figure represents the ROC of the boosted tree classifier on the test-set (using 20% of the data (for 10 patients)) using histogram intensities as the features. 156

Figure 7-9 The algorithm performance per image. The first row is the label, where white, gray and black represents abnormal, normal and don’t-care regions. The second row

represents the probability prediction for each pixel within an image. Red drawing illustrates the fuzzy regions such as edges where even the reviewers might be unsure of the selection.

.....158

Figure 7-10 This image shows the importance of each feature using our algorithm when only ratio features are used.....161

Figure 7-11 This figure represents the ROC of the boosted tree classifier on the training set and (using 80% of the data (for 39 patients)), only ratio features are used.162

Figure 7-12 This figure represents the ROC of the boosted tree classifier on the test-set (using 20% of the data (for 10 patients)), only ratio features are used.163

Figure 7-13 This image shows the importance of each feature using our algorithm when all features are used.....164

Figure 7-14 This figure represents the ROC of the boosted classifier on the training set (using 80% of the data (for 39 patients)), using all features(ratio and histogram intensities).165

Figure 7-15 This figure represents the ROC of the boosted tree classifier on the test-set (using 20% of the data (for 10 patients)), using all features(ratio and histogram intensities).166

Figure 7-16 This figure shows the results of the Ensemble (MDC images and only white light) and the Boosted-Tree classifier being applied on the test set.167

Figure 7-17 This figure represents the ROC of the boosted tree classifier on the training set on the second half of LEEP patients (using 80% of the data (for 32 patients)), using all features (ratio and histogram intensities).....	169
Figure 7-18 This figure represents the ROC of the boosted tree classifier on the test-set on the second half of LEEP patients (using 20% of the data (for 9 patients)), using all features(ratio and histogram intensities).	170
Figure 7-19 This figure represents the ROC of the boosted tree classifier on the new dataset (the second half of LEEP patients), all 50 images are used as the input. The classifier is trained on the first half of the images (49 patients).	171
Figure 7-20 ROC, 10-fold cross-validation, using an image of 90 patients (old and new dataset).	172
Figure 8-1. Workflow of our process from tissue imaging to hyperparameter optimization. The first step is imaging the tissue with MDC technology, next, we will prepare the data and labels according to the type of CNN that we plan to use (we might also need to center the data, by using normalization procedures) and finally, we will optimize some parameters to increase the performance.....	176
Figure 8-2. A very high-level representation of the workflow of patch-based CNN. After preparing the MDC images we will apply a sliding window to scan the entire image and extract image patches. Next, we will label these patches. Finally, we implement the classification step to predict the labels.	177

Figure 8-3 Feature extraction and classification steps of one of the networks (VVG16). The first block is the feature extraction which consists of consecutive convolution and Max-pooling blocks (purple rectangle in each stack of rectangles is representative of Max-pooling) the second block is a classification module where it is fully connected to the previous layer. The output of this layer is fed to a binary logistic regression for prediction. The number of rectangles shown above is not representative of the actual number of filters in each layer.

..... 179

Figure 8-4 Simple Convolutional Neural Network. This very simple network consists of one convolution and the ReLU layer followed by a Max-pooling layer, a Flatten layer, another dense layer followed by a dropout layer with (20% drop-out) and at the end a there is a SoftMax layer (with only two classes)..... 185

Figure 8-5 This figure shows the flow diagram of the simple model. 186

Figure 8-6 VGG16 Convolutional Neural Network. This relatively deep network consists of few layers of convolution and ReLU layers followed by a Max pooling layers, two last layers are a fully connected layer and a SoftMax layer (with only two classes). 188

Figure 8-7 VGG19 Convolutional Neural Network. This relatively deep network consists of few layers of convolution and ReLU layers followed by a Max pooling layers, two last layers are a fully connected layer and a SoftMax layer (with only two classes). 190

Figure 8-8 One block of ResNet-50. It has two learning paths, the bottom path contains layers of convolution, batch normalization, and ReLU, and the top path has no learning module.	191
Figure 8-9 Accuracy values per epoch for channel 1. The accuracy level fluctuated at the beginning and eventually plateaus around 0.85. Each color represents one epoch.	193
Figure 8-10 Loss values per epoch for channel 1. The loss level fluctuated at the beginning and eventually becomes plateau around zero. Each color represents one epoch.....	194
Figure 8-11 Accuracy values per channel for each epoch. The accuracy level increased at the beginning and became stable at around 0.85. Each color represents one channel. All channels show relatively similar performance.	195
Figure 8-12 Loss values per channel for each epoch. The loss level decreased drastically at the beginning and became stable at around 0.25. Each color represents one channel. All channels show relatively similar performance.....	195
Figure 8-13 Prediction visualization for one whole image input. The prediction per sliding window is calculated and the whole image is reconstructed per channel (for 6 channels). All channels show relatively similar performance.....	196
Figure 8-14 Comparison between normalized input and nonnormalized input. Accuracy values per channel for each epoch is shown for 10 inputs, half normalized and the other half non-normalized. The accuracy level increased at the beginning and became stable at around	

0.85 for normalized channel and around 0.84 for the non-normalized channel. Each color represents one channel.....	197
Figure 8-15 RGB images are constructed from a ratio image.....	198
Figure 8-16 Monitoring the performance for a larger epoch for the normalized input. The accuracy value for channel 1 during 25 epochs are monitored. The accuracy level increased at the beginning and became stable at around 0.848.	199

List of Abbreviations

AS: Active Surveillance
AUC: Area Under the Curve
BG: Blue Excitation Green Emission
BR: Blue Excitation Red Emission
CAD: Computer-Aided Diagnosis
CC: Connected Component
CCD : Charge-Coupled Device
CCP: Chromophobe RCC
CCRCC: Clear Cell RCC
CIN: Cervical Intraepithelial Neoplasia
CNN: Convolutional Neural Network
ConvNets: Convolutional Neural Networks
DCNN: Deep Convolutional Neural Network
DNN: Deep Neural Network
FAD: Flavin adenine dinucleotide
FFT: Fast Fourier Transform
GW: Gradient-Weighted
H&E: Hematoxylin and Eosin
HDI: High Human Development Index
HER2: Human epidermal growth factor receptor type 2
HIS: Hyper-Spectral Imaging
HPV : Human Papillomavirus
IRB: Institutional Review Board
KNN: K-nearest neighbor
LEEP: Loop Electrical Excision Procedure
LSVM : Latent SVM

MDC: Multispectral Digital Colposcopy
ML: Machine Learning
NADH: Nicotinamide adenine dinucleotide
NRBM: Non-Convex Regularized Bundle Methods
PC: Principal Component
PCA: Principal Component Analysis
PCa: Prostate Cancer
PRCC: Papillary RCC
PSA: Prostate-Specific Antigen
PTFE: Polytetrafluoroethylene
QDP: Quantitative Digital Pathology
RCC : Renal Cell Carcinoma
ROC: Receiver Operating Characteristic
ROI: Region of Interest
SD: Standard Deviation
SIL: Squamous Intraepithelial Lesions
SNP: Single-nucleotide polymorphism
SVM: Support Vector Machine
TMA: Tissue Micro-Array
TNR: True Negative Rate
TPR: True Positive Rate
UB: UV Excitation Blue Emission
VB: Violet Excitation Blue Emission
VG: Violet Excitation Green Emission
VGG: Visual Geometry Group
wB: white Illumination Blue Reflectance
wR: white Illumination Red Reflectance

Acknowledgments

The path toward obtaining my PhD degree is almost at its end. Now that I am closing this chapter of my life, I feel obliged to express my gratitude and appreciation of all the amazing people whose support and help made it possible for me to overcome many ups and downs during these years. It has been a marathon and a life changing experience for me.

First of all, I would like to give many thanks to my supervisor Dr. Calum MacAulay who supported me through it all. Not only he patiently motivated and helped me with my thesis but he also supported me during the many transitions that I went through in my academic, career and personal life. He approved my choices but at the same time provided me with critical feedbacks to direct me on the right track. I am forever indebted to you Dr. Calum. MacAulay.

Besides my supervisor, my deep gratitude goes out to my committee members Dr. Pierre Lane, Dr. Mira Keyes and Dr. Haishan Zeng. Your comments and suggestions were always appreciated and made my thesis become unblemished. You made my defense less stressful and more enjoyable. I should specially mention Dr. Mira Keyes here, who was my role model. I learnt a lot about patient care by shadowing her in operation rooms. Her feedbacks and support will never be forgotten.

Many thanks to my university examiner Dr. Stefan A. Reinsberg and Dr. Shahriar Mirabbasi; I am forever grateful to you for accepting to examine my thesis and giving me your feedbacks, which made it feasible for me to defend my thesis.

I should also give special thanks to my family and friends. To my life long and forever cheerleader, my mom. She has always believed in me, giving me encouragement and support whenever I needed it. Lavishing me with her love even though she was far away. To my beloved husband and friend, Amir. Without him finishing this journey was not possible. He was there for me, helped me

to figure out ways to solve my problems or just patiently listening to them. I am thankful to you for not only understanding that I have to spend most of my time on my research but cherishing it and helping me focus on my thesis and not letting me to give up. You walked by me as if it was your own thesis. Amir, you are the best.

Special thanks go to Fatemeh, I was so lucky to meet her at BC Cancer Research Center. She became my dear friend and family. We spent lots of time drinking coffee and studying together. Thanks, Fatemeh for being there for me.

And lastly, many thanks to my best friend Mehdi. Thank you for the support not only on my thesis but my life. I met you ages ago during first year of my B.S and you helped me throughout my past and current degrees and everything else. Words can't express my gratitude Mehdi.

Dedication

To my Mom, Zohre

Chapter 1 Introduction

This dissertation presents techniques for cancer detection/diagnosis/prognosis using Artificial intelligence (AI). The term AI describes the techniques that simulate human behaviour. The focus is on the development of Machine Learning (ML), a subset of AI, techniques and advanced image analysis algorithms that are applied to tissue of the prostate, kidney, and cervix to detect/diagnose/prognose cancer. The methods and approaches however could be easily modified and extended to other sites of cancer.

In the first half of this thesis, we introduce methodologies to assist with the quantification of prostate cancer nuclei in histopathological tissue images. In the second half of this thesis, we present our novel method for the detection and classification of larger-scale structures in tissue for two cancer types, renal carcinoma, and cervical pre-cancers.

In this chapter, we provide background information (cancer statistics, the importance of medical imaging and advanced techniques for image analysis) relevant to the study presented in this thesis. We also summarize some of the current state of the art research relevant to our subjects of interest. Finally, we present an overview of the general framework of this thesis and provide a summary of the dissertation.

1.1 Cancer Statistics

Cancer is a major public health problem worldwide and is the second leading cause of death in the United States.

Despite advances in technology and extensive research in cancer domain, cancer incident rate (during 2005-2014 period) was stable for women and only declined %2 for men, similarly, the

mortality rate only declined about 1.5%/per year for both women and men (during 2006-2015 period). It is estimated in 2018, we will have about 1,735,350 new cancer cases and 609,640 cancer deaths in the United States (Siegel, Miller, et al. 2018).

Lung, prostate, and colorectum are the most common cancer types in men, and in women, however, lung, breast, and colorectum are the most common causes of cancer death. These four cancers account for 45% of all cancer deaths, with one-quarter due to lung cancer (Siegel, Miller, et al. 2018).

Men have a slightly higher chance of being diagnosed with cancer. The reason for higher susceptibility in men could be related to environmental factors, lifestyle, endogenous hormones, and many more other factors (Siegel, Miller, et al. 2018)

These factors highlight the significance of personalized medicine and are the reasons behind the fast-growing trend of the application of machine learning techniques in medicine.

1.2 Personalized Medicine – Oncology

Personalized treatment in oncology has attracted lots of attention, yet it requires more investigations.

A fast, cost-efficient and comprehensive evaluation of the genome, gene sequencing/profiling, has encouraged personalized cancer medicine¹. In the cancer domain, personalized oncology means analysis of an individual patient's tumor which allows the selection of effective drugs to control that tumor and thereby improve survival rates.

¹ Personalized cancer medicine refers to the study of a person's genetic makeup and tumor growth

Examples of such targeted treatments are, Trastuzumab and Vemurafenib which are used for human epidermal growth factor receptor type 2 (HER2)–expressing breast cancer and for melanomas that express mutated *BRAF* respectively (Tannock, Hickman 2016). Although most of these targeted therapies are effective at first, patients usually show resistance and succumb to the disease. This shows we have still a long way to completely understand cancer genomic as well as all the underlying cancer pathways. Furthermore, target therapy is very costly and not available to all patients.

One of the ways to potentially gain an understanding of cancer biology is through quantitatively analyzing different molecules or observing the expression of a certain biological state or condition (e.g. angiogenesis and proliferation), some of which have proven to be useful biomarkers for cancer detection and management. Using advanced imaging systems and incorporating different imaging modalities or techniques with machine learning algorithms has the potential to improve cancer management.

1.3 Cellular and Tissue Level Analysis

Optical technologies generate micro and macro level images and thereby provide clinicians with an improved visualization ability which otherwise would be very difficult if not impossible to observe and monitor by the human eye.

The recent improvement in biomedical optics has made it possible to have a better understanding of the molecular and tissue changes by capturing high spatial resolution images even in-vivo locations. An example of such a system is the device developed at BC Cancer Research Center that acquired macroscopic images in an in vivo fashion using multispectral imaging techniques (Milbourne, Park, et al. 2005). This optical technology could be used for cancer diagnosis as it can

guide gynecologists to select more precise biopsies from the cervix (advanced guided biopsies) or completely remove the need for any biopsy. This could improve patients' quality of life and reduce associated medical costs. Medical tools like this could be particularly beneficial in low resource countries where access to preventative and screening resources is limited.

1.4 Digital Pathology

Digital pathology refers to the creation and analysis of digital representations of tissues (through imaging a fixed, sectioned and stained tissue placed on a glass slide) usually using whole slide digital scanners. The acquired images may be used for targeted analysis which can aid with disease diagnosis and possibly cancer prognosis.

With the advances in technology and medical imaging, it is now possible to use very fast, cost-efficient data acquisition tools with very high resolution (at molecular and cellular level).

Although proliferation of whole slide digital scanners results in improvements in clinical and research studies (by scanning slides and capturing digital images which later could be used for further analysis), tissue analysis still suffers from a dependency on human interaction and guidance as well as other challenges that we discuss in the following sections.

1.5 The Necessity for Development of Automated Diagnostic

Methods

Manual evaluation of biomedical images is very labor-intensive, expensive and subjective. Furthermore, it might not be practical in locations where there is limited access to trained personnel. This emphasizes the need for the development of automated diagnostic and prognostic methods. The goal of this thesis is to introduce machine learning-based automated diagnostic

techniques which could be used as the first line in the detection and diagnosis of cancer and pre-cancerous lesions in developing countries. The approaches proposed in this thesis can provide additional information relevant to diagnosis/detection to oncologists/pathologists and improve our understanding of the biological behavior of lesions at the cellular and tissue level for different cancer types, so we can guide and personalize the treatment approach in individual patients.

1.6 Challenges

While digitizing pathology slides offers many advantages, it comes with some challenges that we discuss below.

1.6.1 Staining and Imaging Issue

Cellular heterogeneity of the lesions/ tumor microenvironment poses a significant difficulty in designing an automated system for cancer analysis. This variation becomes more pronounced with cancer as the tissue becomes a sea of different cell types and loses much of its structure and differentiated morphology. The other difficulties are due to the methods and techniques used for preparing the specimen for imaging. Issues such as staining variation, tissue deformation, sectioning inconsistency, and shrinkage complicates the design of automated systems. Additionally, optical imaging systems can also introduce linear and non-linear inconsistencies.

1.6.2 Lack of Gold Standard

The other equivalently challenging issue is the lack of acceptable ground truth for pathology specimens. The definition of appropriate and widely acceptable gold standard can be debatable even within the pathology community. For example, researchers have reported that manual tracing of regions of interest is not reliable and should not be used as the only source for the gold standard (Meiburger, Molinari, et al. 2016). This shows the need for a second (or more) opinion, which is

usually acquired with the assistance of another human (technician/ expert in the field) or a computer-aided device.

Generally, long term patients' outcomes combined with various diagnostic procedures and other biological factors may be considered as an acceptable gold standard. These processes can be very expensive, time-consuming and in some scenarios may not be available.

1.6.3 Diagnoses of Atypia

The tissue obtained from epithelial cancers such as prostate and breast may also contain atypical glands. These glands can be truly atypical, but may also be suspicious or represent malignant epithelium, so the results obtained from biopsies can be unclear (Kopp, Parsons, et al. 2011). The term cell atypia refers to the borderline entities that lie in the spectrum between benign and malignant. It can be hard to find a consistent definition of atypical cells and even more difficult to define the gold standard. Long term clinical follow-up and larger specimen are needed to improve the identification of these cases. Quantitative studies could potentially improve our understanding of these cases and help to clarify these somewhat challenging cases (Boucheron 2008).

1.7 Computer-Aided Diagnosis (CAD)

Computer-aided detection/diagnosis (CAD) is the procedures which use computing power to assist doctors with a better understanding of the case under investigation/study. This can be achieved by presenting quantitative measures or providing an interpretation of the medical images. An example of such a system is a CAD program that can discriminate between progressive and non-progressive cases using images from a prostate tissue microarray. Such a procedure usually involves three main steps, as explained below.

1.7.1 Localization and Segmentation

The first step is to identify region/structure of interest (such as glands in the prostate or colon tissue or vascular structure in skin or kidney) or objects (such as cell/nuclei) in various tissue types (i.e. lung, oral and prostate).

These regions, structures or objects carry biologically important information regarding the pathological behavior or disease progression that is to be recognized/classified/identified by a CAD system.

1.7.2 Feature Extraction and Classification

The next step after the selection and segmentation of the regions of interest is to extract features/characteristics from these segmented areas. Identifying appropriate and relevant features play a very critical role in the development of a powerful CAD classification system. This process usually requires a close collaboration between engineers and pathologists to encode aspects of their domain knowledge into quantitative measures. The ultimate goal is to design a fast, objective solution to assist clinicians with their decision-making tasks. An example of such a procedure is using CAD to generate a Gleason score from prostate cancer tissue sections (Fehr, Veeraraghavan, et al. 2015).

1.7.3 Disease Diagnosis and Prognosis

The final step is to integrate the multiple forms of information obtained from the last two steps into a fully automated system that generates the desired score, classification, grading, etc. (that can be used for disease diagnosis/prognosis). Such a system can provide valuable information which otherwise may not be available/observable even by expert users.

1.8 Application of Machine Learning (ML) in Medicine

Machine learning started to flourish in the 1990s. Arthur Samuel nicely described it as giving “computers the ability to learn without being explicitly programmed”. Machine learning is not used only in healthcare, but its application is very wide and diverse, examples are self-driving cars, speech recognition, cybersecurity, and many more (Babak Ehteshami Bejnordi 2017).

Figure 1-1 illustrates the histogram of published papers on machine learning over the last 20 years. This histogram is normalized by the total number of publications on ML. This information is collected based on the query search of PubMed using “cancer” and “machine learning” keywords. As we can see the distribution is very skewed to the right, which suggests an increase in published papers over time in this domain.

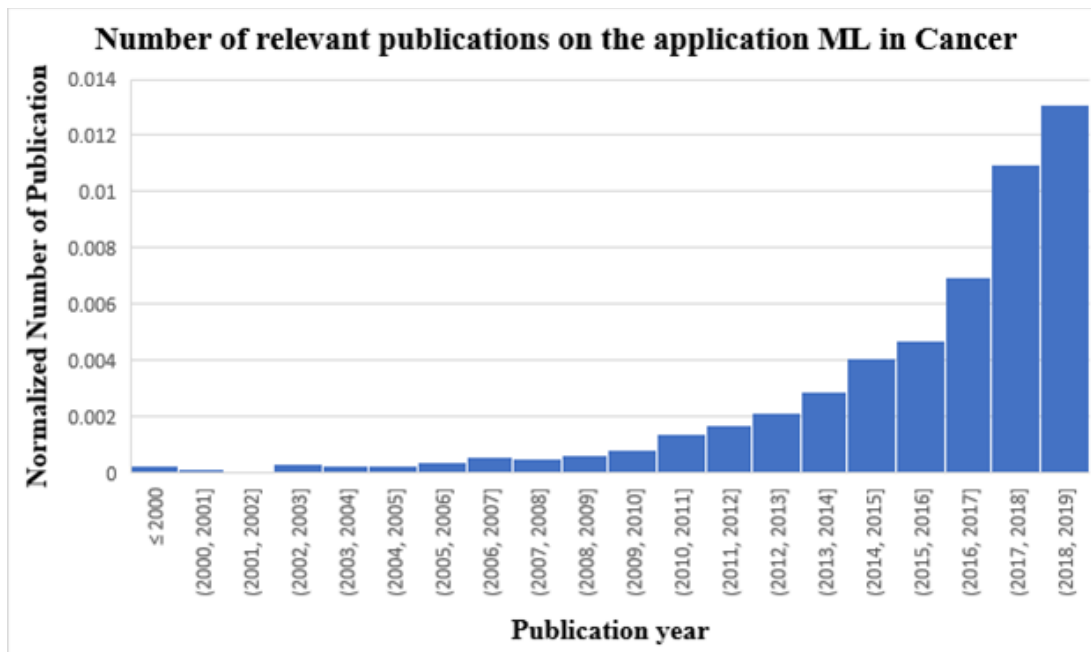


Figure 1-1 This figure shows the distribution of the number of publications related to cancer and machine learning per year. As we can see in this histogram, the number of publications

in this domain increased over the years. This histogram is normalized by the number the total number of publications on Machine Learning.

Depending on the nature of the data and the problem under study, machine learning algorithms can be divided into two main categories: supervised learning, and unsupervised learning.

1.8.1 Supervised Learning

Supervised learning is a data processing task where the training data is labeled. The training dataset consists of a paired input observation (feature vector) and an output (user assigned ground truth) label. Supervised learning methods process the training data and generate an inferred function which later can be used to predict the label of unseen data. In other words, this function will be used to assign the feature vector to the label-mapping of any new dataset. This method can further be divided into two closely related prediction models, regression and classification. When the target property is continuous the task is referred to as regression. However, when the target property is a finite set of values, the task is referred to as classification.

Andriy Burkov in his recent book “The Hundred-Page Machine Learning Book” nicely explained different ML methods. This book provides a very detailed overview of various ML techniques, here we briefly explain the most widely used ML algorithm that we also tried in this thesis.

Support Vector Machines (SVM): This discriminative model separates the training dataset by finding separating hyperplanes that maximize the gap between different classes and partition the feature space into different regions for each class. When the task is a binary classification the hyperplane simply divides the space into two parts where each side belongs to one class. This hyperplane is guaranteed to maximize the distance between the two classes. SVM support different kernel types which means it can use non-linear hyperplane to increase its accuracy. Gaussian

kernel is most frequently used and is shown to improve classification performance over linear kernel in many classification problems. One of the concerns about using SVM is that it tries to focus on the hardpoints(data points closer to decision boundary) which might limit its generalization capabilities.

Decision trees: These predictive models use the observation about an item (as decided by the samples which are falling in the same branch) and try to make a decision about the new target value. Trees are very famous and are widely used as a simple and effective solution for prediction tasks. There are various decision trees such as Random Forest and Boosted trees.

In this thesis, we used **Boosted trees** for cervical cancer classification. Boosted trees are an effective ensemble method that uses more than one decision tree and refine the training weights to improve the prediction decision.

Convolutional Neural Networks (CNN): A very famous and widely used sub-type of Deep Learning is CNN which is mostly used to classify visual imagery, however, they are also used in language processing. These networks are made of neurons that take advantage of hierarchical patterns by applying one filter to different locations in an image. These networks learn the weights during the training process and do not require hand-engineered features. This independence from prior knowledge has made CNN very attractive, however, they require large datasets. We used CNN for cervical cancer classification.

1.8.2 Unsupervised Learning

In contrast with the supervised learning where the labels for the training dataset were known, an unsupervised learning dataset is not labeled. The task, in this case, is to use the structure of the unlabeled data to generate an implicit feature vector to label mapping.

The most common forms of unsupervised learning are clustering algorithms, which automatically partitions the data into groups of similar items called clusters. One of the most well-known clustering methods is called K-means clustering (Wu, SpringerLink ebooks - Computer Science 2012).

K-means clustering: K-means clustering partitions the samples into k clusters which are clustered around k centers. It uses a heuristic algorithm to converge quickly to a local optimum. In each iteration, it tries to minimize the within-cluster variance (by using some measures such as Euclidean distances). We used K-means clustering in the second chapter to help with the prostate gland detection task.

1.8.3 Classical ML vs Deep Learning Methods

Classical machine learning techniques have been used successfully in the medical domain for years; however, their dependency on expert input/evaluation is inevitable. Classical machine learning, in the medical domain, has been used for different tasks such as anomaly detection, disease diagnosis/prognosis, cancer grading, etc. These techniques use the domain-knowledge to find/identify the complex relationships between different patterns and leverage them to accomplish the prediction task.

As discussed earlier, the main limitation of classical machine learning is that it requires expert knowledge and consequently it is very domain-specific. This drawback limits its predictive power and makes it only suitable for specific defined datasets and prevents it from being generalizable to new datasets that are acquired from different tissue types or obtained in a different fashion using a different system or from a different environment.

Deep learning, a subtype of machine learning, in contrast, can automatically discover the representations needed for detection or classification, allowing the system to directly map an input image to an output vector. The main limitation of such a system is that it is very data-driven, and it requires a very large dataset (since it needs to optimize a large set of parameters as a part of its learning process.)

In this thesis, the focus is on both classical machine learning as well as deep learning techniques. We believe both techniques are equivalently powerful/practical depending on the nature of the dataset and the problem under study.

1.9 State of the Art ML for Cancer Diagnosis/ Prognosis

Over the past few decades, machine learning methods have been applied successfully to various fields. However, in the healthcare/medicine domain ML improvements were limited primarily due to the limited size of the available annotated dataset which consequently led to insufficient data acquisition and prevented the design of advanced and complex ML-techniques. Recently, ML has started to flourish in healthcare and medicine. ML techniques and feature selection algorithms have been widely used to assist with disease diagnosis and prognosis.

In the medical domain, classical machine learning is a mean of selecting features, relevant to the task under study, from an exhaustive list of features space. Frequently, most of these features are redundant, irrelevant and not very practical for the topic under study. The selection of relevant features is usually achieved in two main ways: one by asking the clinician/pathologist to identify useful diagnostic features (Canzonieri, Monfardini et al. 1998) and the other approach is to use an automated feature extraction method (Nafe, Yan et al. 2006, Latson, Sebek et al. 2003).

ML techniques can be used for cancer detection and disease diagnosis and also for disease prognosis. The main purpose of prognostic ML tools is to predict (i) individual cancer susceptibility, (ii) recurrence and (iii) survival (Kourou, Exarchos, et al. 2015).

1.9.1 Cancer Detection/ Diagnosis

Machine learning techniques have been used broadly for anomaly detection with the aim of finding cancerous or pre-cancerous lesions.

Machine learning techniques have been used for detection/diagnosis of cancer in many sites such as lung (Cui, Luo et al. 2019a), oral (Kourou, Exarchos, et al. 2015), cervix (Tseng, Lu et al. 2014), ovarian (Paik, Lee et al. 2019), renal (Bektas, Kocak et al. 2019), breast (Kim, Kim et al. 2012), prostate (Vos, Hambrock et al. 2008, Shah, Turkbey et al. 2012), and etc. There are multiple studies that recommended to apply machine learning techniques for the detection of abnormalities in different imaging modalities such as mammogram images of breast cancer (Jalalian, Afsaneh|Mashohor, Syamsiah B.T.|Mahmud, Hajjah Rozi|Saripan, M. Iqbal B.|Ramli, Abdul Rahman B.|Karasfi, Babak 2013).

For example, support vector machines (SVM) are powerful ML technique that has been successfully applied to clinical problems. Over the last 20 years, SVM showed promising results for the classification of many real-life problems, mostly because of its strong generalization ability (Wang, S., Burt, et al. 2014). Multiple research groups such as Shah et al. (Shah, Turkbey, et al. 2012) and Vos et al. (Vos, Hambrock et al. 2008) used machine learning techniques on Multiparametric MRI for prostate cancer detection. These two groups employed SVM as their classifier for localizing prostate cancer and analyzing prostate lesions.

In this thesis, we presented a modified SVM technique for prostate cancer grading. We also used a similar strategy for discriminating progressive from non-progressive prostate cancer cases which we will explain in detail in chapter five of this thesis.

ML techniques will not only improve the accuracy of detection and diagnosis, but it can also improve prediction of cancer recurrence and survival (Wang, Burt, et al. 2014) Wishart et al. in 2006 showed that it is possible to improve predicting biological behavior of cancer, probability of recurrence, and mortality rate by 15–25% by using techniques which took advantage of ML algorithm (Cruz, Wishart 2006).

1.9.2 Prediction of Individual Cancer Susceptibility

Inherited mutations in genes can result in a higher risk of developing cancer in specific tissues. For example, a mutation in tumor suppressor genes, oncogenes, and mutation in genes involved with DNA repair/ cell cycle control might increase individual susceptibility to cancer.

Waddell M et al. used SVM techniques with leave-one-out cross-validation to predict an individual's cancer susceptibility for multiple myeloma. The goal of this study was to answer this question: is it possible to predict individual susceptibility to early onset of myeloma by using ML techniques? Single-nucleotide polymorphism (SNP) profiles were used as the feature vectors for this study. They achieved a prediction accuracy of 71%. Although this seems to be very close to random guessing it is important to point out that in this study only 3000 SNPs were used in profiling, whereas several million SNPs are known which shows a feasible window of possibility for further research and improvement (Waddell, Page, et al. 2005).

1.9.3 Prediction of Cancer Recurrence

Similarly, multiple groups have used SVM to predict cancer recurrence. Groups such as Kim W et al (Kim, Kim, et al. 2012), and Eshlaghy A et al. (Lg, et al. 2013), Tseng C-J et al. (Tseng, Lu et al. 2014). All three groups used SVM techniques for cancer recurrence prediction, the first two groups for breast cancer and the last group for cervical.

Exarchos K et al. presented a very interesting study where they employed Bayesian models for prediction of oral cancer recurrence. They used a systematic and multiparametric approach where features were selected from clinical, imaging and genomics. In this study, different classifiers were used for each feature category and at the last step a consensus classifier was used (in a complementary manner) to increase the accuracy. They reported 100% accuracy on discriminating disease recurrence vs non-recurrence (Kourou, Exarchos, et al. 2015).

1.9.4 Prediction of Cancer Survival

As explained earlier, ML can also predict cancer survival. Similar to cancer diagnosis and cancer recurrence prediction, many groups reported SVM as their choice of classifier for cancer survival prediction. Groups such as Chang S-W et al. (Chang, Abdul-Kareem, et al. 2013) and Xu X et al. (Xu, X., Zhang, et al. 2012) respectively used SVM for oral and breast cancer survival prediction.

1.9.5 Deep Learning and Cancer

So far, we discussed classical machine learning techniques for cancer diagnosis/prognosis. These techniques are very powerful since they use features that are highly dependent on the subject under the study. However, since these features are tailored for specific cancer; it is not possible to generalize the method to other cancer types. There are even cases where it is not possible to apply the same technique to the same tissue image captured from a different imaging system. This

limitation and the dependency on the human expert are among the challenges that deep-learning techniques try to overcome.

Although, deep-learning methods seem to be superior to classical machine learning they are very data-driven. Access to a large dataset especially in the medical domain is usually considered a luxury and most often is impossible to obtain.

In Table 1-1 we summarized a few recent studies which used deep learning techniques for cancer diagnosis/prognosis. In chapter 8 of this thesis, we discuss deep-learning approaches for cervical cancer detection.

Table 1-1 Selected summary of deep-learning publication for cancer detection/diagnosis/prognosis

Publication	Neural Network	Cancer type	No of patients/Images	Type of data	Reported Accuracy or AUC
Saikia et al. (Saikia, Bora et al. 2019)	VGG 16, VGG 19, ResNet 50, GoogLeNet V3	breast	2120 images	FNAC ² images.	96.25%
Alom et al. (Alom, Yakopcic et al. 2019)	RCNN ³ , RRCNN ⁴	retina, lung, skin	40/retina, 900/skin, 534/lung	retina/DRIVE dataset ⁵ , white light lung/CT, skin/ histopat	retina/%99, lung/ %98, skin/%86

² FNAC: Fine needle aspiration cytology

³ RCNN: Recurrent Convolutional Neural Net (Alom, Yakopcic et al. 2019)

⁴ RRCNN: Residual Convolutional Neural Net (Alom, Yakopcic et al. 2019)

⁵ DRIVE dataset: Digital Retinal Images for Vessel Extraction (DRIVE: Digital Retinal Images for Vessel Extraction)

Cai et al.(Cai, Huang et al. 2019)	AlexNet	Breast	990 images	full-field digital mammography	%86.89
Sedghi et al.(Sedghi, Pesteie et al. 2019)	DNM ⁶	Prostate	255/sample	TRUS-guided biopsies	AUC>0.8
Halicek M et al. (Halicek, Little, et al. 2019)	3-layer custom CNN	head and neck	21/patients	Hyper-spectral imaging	92%
Cui S et al. (Cui, Luo et al. 2019b)	classical/deep-learning/joint	lung	106/patient	non-small cell lung cancer	~0.83%
Lee JH et al.(Lee, Ha et al. 2019)	ResNet50, VGG16, VGG19, etc.	cervix	995/image	CT images	AUC>0.95

1.10 Motivation

While much work has been done using ML on image data for cancer localization/detection/grading and prognosis, there is still much more that can be done to harness the potential of ML combined with imaging systems. Automated quantitative image-based methods built around ML predictive power could be beneficial to patients of all cancer types and could significantly improve the cancer screening and detection process.

In this dissertation, we propose and demonstrate across many spatial scales the application of advanced ML techniques to create new, diagnostic, prognostic and prediction tools.

In this thesis, we applied ML technique to image data focusing across various spatial scales. Starting at the microscopic scale (the individual pixel of microscopy image) to collections of pixels representing a biological entity (nuclei) for prostate cancer nuclei detection. The next step up in

⁶ DNM: Deep Neural Maps (Sedghi, Pesteie et al. 2019)

scale was microscopic tissue structures that are collections of cells (microvascular system in renal carcinoma tissues) and finally, we applied ML techniques to the tissue macroscopic scale for localization of cervical pre-cancers.

1.11 Structure of This Thesis

In the second chapter of this dissertation, we introduce new ML techniques for the detection and segmentation of glandular structures in prostate cancer that are applied predominately at the pixel level ($\sim 0.3\mu\text{m}$). These glandular structures are important since they contain clinically relevant information regarding the disease status and Gleason grade of prostate cancer. The manual gland segmentation process is very time consuming and subjective, also existing automated methods are not robust and reliable. In this chapter, we introduce an automated, fast and objective method for glandular structure detection using hyperspectral image pixel characteristics and some pixel-level morphology operations, Figure 1-2.

Recognizing and locating cell nuclei in images of stained tissue sections is a key step in the design and automation of Quantitative Digital Pathology, (QDP). A powerful QDP system can measure the visible structural chromatin changes (within nuclei) associated with neoplasia and enable prognostic and predictive assays based upon imaging of sectioned prostate tissue. However, tissue complexity, overlapping cell clusters, image and tissue inhomogeneities, and other imaging issues introduce uncertainties into this predictive process and are barriers to the development of a fully automated system. In the third and fourth chapters, we present solutions to overcome these challenges.

In the third chapter, we present a new method to automatically segment Thionin-stained prostate cancer nuclei (4 μ m-15 μ m spatial scale), which enable the extraction of various nuclear and cell sociology features.

In the fourth chapter, we present an interactive segmentation algorithm that provides the user the ability to evaluate nuclei segmentation results and possibly add or remove nuclei centers to correct/improve the segmentation. This chapter complements the previous chapter by providing a tool for interactive cell-segmentation correction.

In the fifth chapter, we present a machine learning approach to extract clinically relevant features from segmented nuclei and use these features for two goals 1) prostate cancer grading 2) prostate cancer prognosis. Figure 1-2 summarizes the procedure used in chapter 2 through 5.

The cancer under study in chapter six is renal carcinoma. Despite advances in technology and the introduction of new diagnostics methods, the renal carcinoma mortality rate is still very high and so there is a need for a reliable, robust, cost-effective diagnostic/classification method in this site. In this chapter, we present our approach for renal carcinoma sub-type classification. In this study, we move up the spatial scale (10-200 μ m) and apply ML techniques to identify blood vessels and vascular patterns in renal carcinoma. The required steps are summarized in Figure 1-3.

In chapters seven and eight we present two different classification approaches for cervical cancer detection at the macro scale (~1-4 cm). This is the largest scale we examine in this thesis. Early detection is highly critical for successful cervical cancer screening as this cancer develops over several years without any obvious symptoms. Although due to cervical cancer screening programs, cervical cancer incidence and mortality rates have decreased in developed countries, developing

countries on the other hand still suffer from high mortality rates and a lack of screening/evaluation programs.

In chapter seven, we present a classical machine learning approach based on the analysis of the Multispectral Digital Colposcopy (MDC) image of the cervix.

In chapter eight, we present and compare different deep-learning approaches for classification of normal vs abnormal areas in the MDC images (the same dataset as chapter 7).

Figure 1-4 summarized the procedure used in chapters 7 and 8.

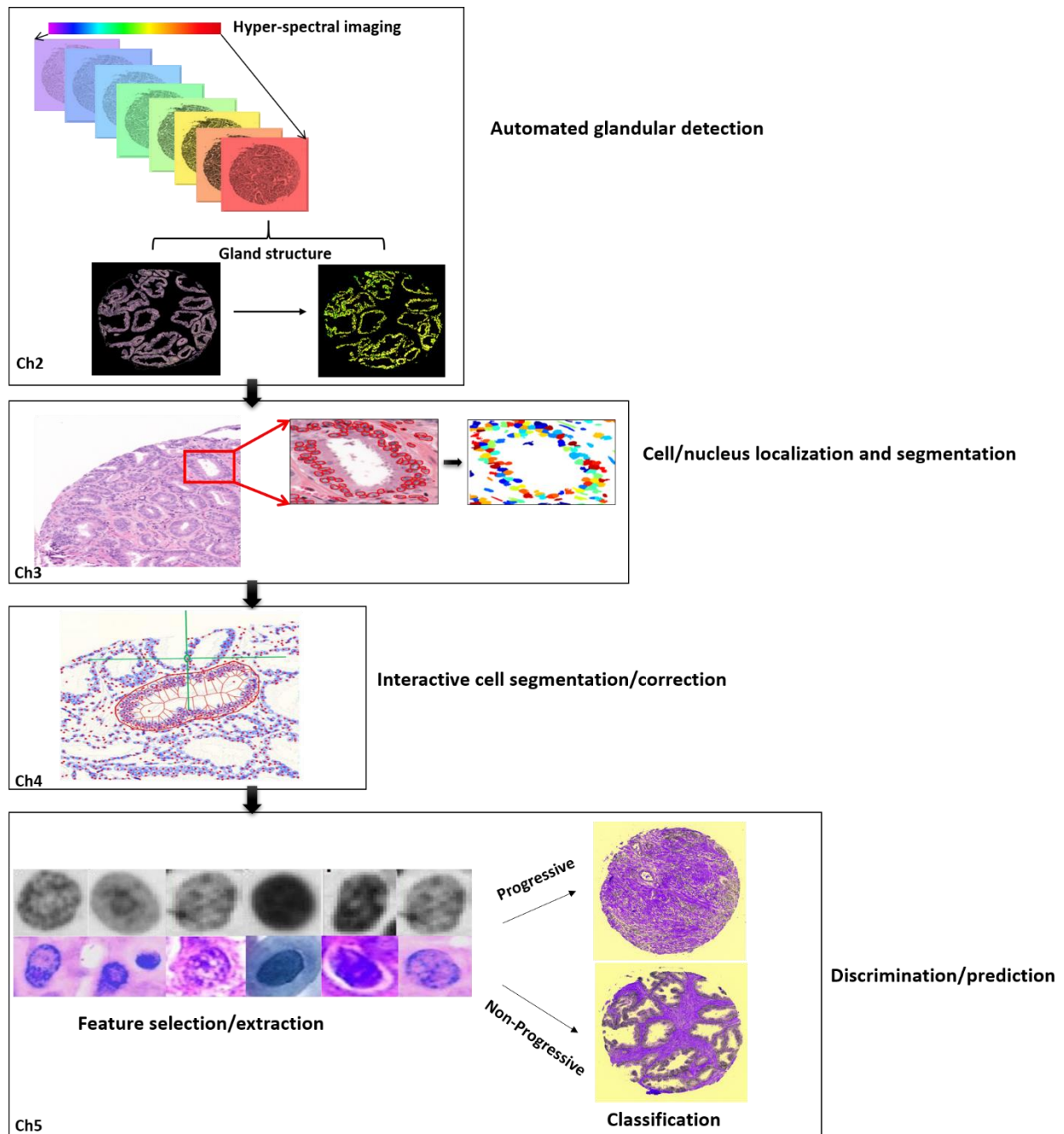


Figure 1-2 Flowchart of dissertation, chapters 2, 3, 4 and 5.

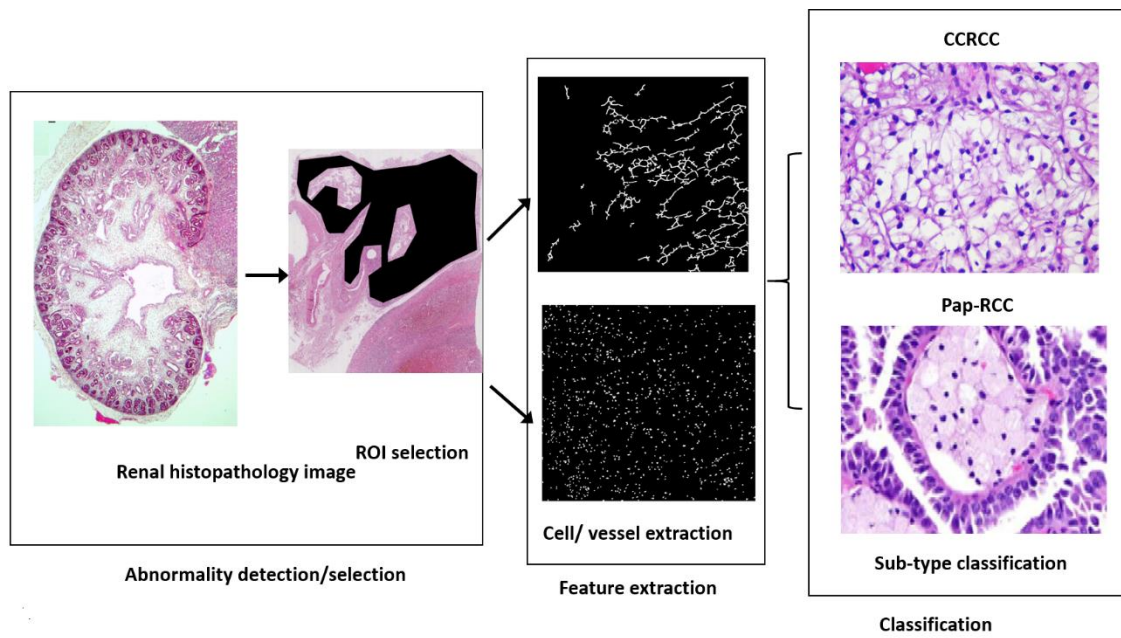
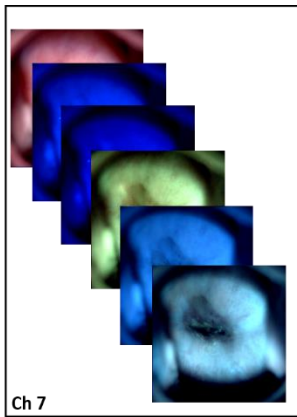


Figure 1-3 Flowchart of dissertation, chapter 6.



Multi-spectral imaging

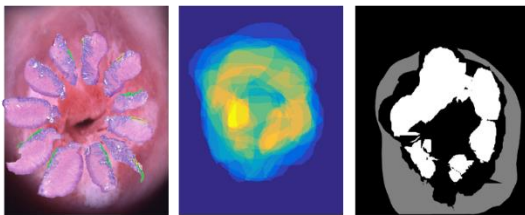
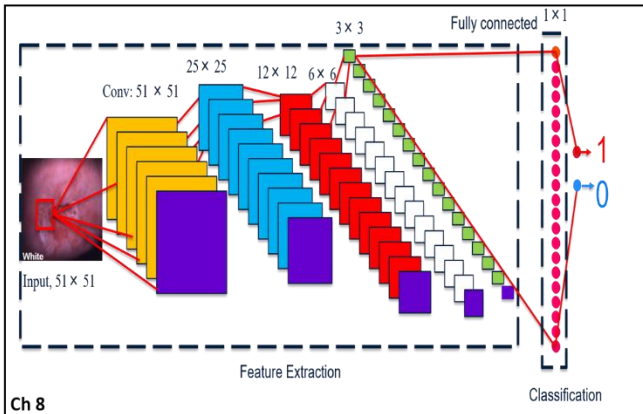
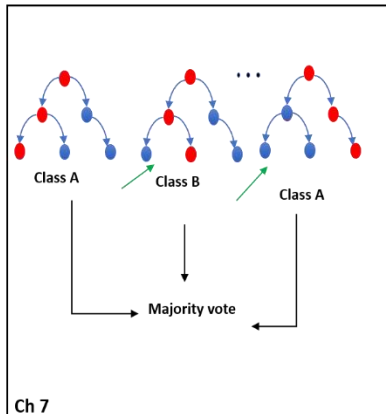


Image annotation and labeling

Conventional ML

Deep Neural Network



Classification

Figure 1-4 Flowchart of dissertation, chapters 7 and 8.

Chapter 2 Automated Prostate Glandular Detection Using Hyper-Spectral Imaging

2.1 Introduction

2.1.1 Prostate Cancer Epidemiology, Screening and Diagnostic Techniques

Prostate Cancer (PCa) is the most frequently diagnosed non-cutaneous malignancy in men. PCa is known as the second and third leading cause of cancer mortality for American and Canadian men, respectively (Prostate Cancer Statistics | CDC2019). More than two million men in the United States have had prostate cancer. Over 15% of men are estimated to be diagnosed with prostate cancer during their lives, in 2019 it is estimated there will be 174,650 new prostate cancer cases and the estimated number of deaths in 2019 is 31,620 (Cancer of the Prostate - Cancer Stat Facts). PCa-related deaths are estimated to increase despite advances in diagnostic/prognostics techniques.

PCa is a disease of disrupted prostate cell genomes and their corresponding proteins. In the process of cancer development, normal prostate epithelium cells will be deformed and disrupted, cells move from androgen-dependent and eventually androgen-resistant (referred to as castrate-resistant carcinoma). This process is very complex and involves processes associated with up/downregulation (malfunction) of different molecular pathways.

With the assistance of Prostate-Specific Antigen (PSA) screening, it is now possible to detect prostate cancer in its early stages; however, this method has caused significant over-diagnosis/-treatment of PCa patients which led to unnecessary morbidity, decrease the quality of life and

significant increase in health care costs. There are recent studies which consider low-grade Gleason pattern prostate lesions as non-malignant lesion as they do not have the characteristic hallmarks of cancer (Hashim Ahmed, Mani Arya, et al. 2012) however this approach has not been clinically adopted. Most men with low-risk disease will be offered Active Surveillance (AS), an appealing approach for men with early PCa stage. In contrast, patients with more aggressive prostate cancer and, in particular, those with high Gleason grade would likely benefit from active treatments such as radiation with or without androgen deprivation therapy and/or brachytherapy or surgery. Needless to say, it is of great interest to design methods that can better discriminate between aggressive/non-aggressive lesions.

As explained earlier, epigenetic and genetic changes in normal epithelium cells can result in cancer progression, it is very important to design techniques to phenotypically monitor the epigenetic and genetic changes in cellular and tissue architecture.

Currently, the diagnosis of prostate cancer is done by visually observing stained sections from prostate biopsies and determining the Gleason grade of these samples (Fenner 2015). This procedure is based on observing the morphological structures and changes in architectural tissue characteristics. Although this method is very effective and has been used for decades, it is very human-dependent/subjective and consequently, it is subject to inter-observer variability. This process is also very expensive and may not be available in countries with low resources. This fact motivated us to take the initiative to tackle this issue and design a fully automated system for prostate cancer grading and aggressive/non-aggressive prediction. The first step toward designing a fully automated system is to detect the region of interest (in our case, glandular structures). We published the processes involved with the design of an automated glandular detection method

(Zarei, Bakhtiari, et al. 2017). In the upcoming chapters, we will present methods to use these structures for our classification purposes.

2.1.2 Automated Gland Detection

Many groups previously tackled the problem of automated gland detection in different tissues, such as colon (Gunduz-Demir, Kandemir et al. 2010, Ozgul, Sokmensuer et al. 2011), breast (Naik, Doyle et al. 2008) and prostate (Nguyen, Sabata et al. 2012, Pääkkönen, Päivinen et al. 2015, Singh, Kalaw et al. 2017, Doyle, S., Hwang et al. 2007, Naik, Doyle et al. 2008) . The focus of most of these groups was on the detection of the empty region inside the gland (lumen) and incorporating the color information (Naik, Doyle, et al. 2008, Nguyen, Sabata, et al. 2012) to find the empty spaces. This step is usually followed by different image processing algorithms such as active contours (Xu, J., Sparks, et al. 2010), level-set (Naik, Doyle, et al. 2008), and etc. to detect and segment gland structures. However, these methods do not seem to be robust with respect to color variation which can be due to the variation of microscopic color calibration and staining variation. Moreover, these methods might fail to detect partially/fully filled or small glands. As such they require human interaction to validate/monitor the process (Xu, Sparks, et al. 2010). In a recent study (Fu, Qiu, et al. 2014), a random field model was used to overcome some of these challenges, however, it failed to detect any glandular structures that did not have star-convexity shapes.

In this chapter, we present a new approach to automatically detect glandular structures in Tissue Micro-Array (TMA) cores and subsequently detect the gland nuclei. The nuclei from these glands can be used for future normal/abnormal and cancer subtype classification.

2.2 Hypothesis

Recently in a pilot study, we obtained promising results on discriminating between aggressive from non-aggressive PCa cases with a 70% correct classification rate using only nuclei of prostate TMA cores. We hypothesize that by incorporating clinically relevant information from a region of interest, it is possible to increase our system performance. Monitoring glandular structures is known to be important since, in the course of cancer development, tissue alternation may result in highly packed, ill-defined, fused and poorly-formed glandular lumina. These architectural changes are important patterns for tissue diagnosis and cancer prognosis. Moreover, the nuclei within each glandular structure could be used for further quantitative analysis. However; manual segmentation of the glandular structures especially in a large-scale study, e.g., thousands of TMA cores or cases is very user intensive and time-consuming. The goal of this chapter is to develop techniques for automated prostate glandular detection to facilitate the diagnosis/prognosis process.

2.3 Objective

Detection and segmentation of glandular structures are important. These structures contain clinically relevant information regarding the disease status and Gleason grade of prostate cancer. The objective of this chapter is to design an automated, fast and objective method for the detection of glandular structures in digitized histopathology images of Tissue MicroArray (TMA).

2.3.1 Specific Objective

- I) Capture prostate TMA images using a Hyper-Spectral Imaging system
- II) Develop methods for glandular structure detection

The workflow for the whole process of our study on prostate cancer is presented in Figure 2-1. The first step is to image the tissue using our hyper-spectral imaging system, followed by

automated detection of the region of interest. The next step is to select and implement ML techniques and the last step is parameter optimization to improve the system performance. The focus of this chapter is the blue box of Figure 2-1. TMA images were used for glandular structure detection.

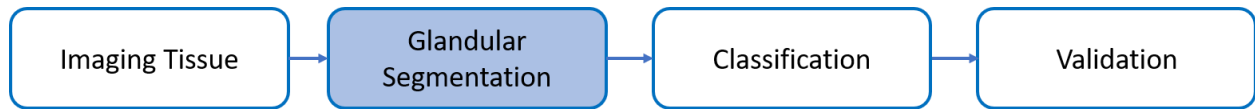


Figure 2-1. This figure shows the pipeline of our study. The focus of this chapter is on the **Glandular Segmentation**. In the later chapters, we will explain the classification and validation processes.

2.4 Materials and Instrumentation

2.4.1 Clinical Data

The TMA slide of prostate tissue was stained with Hematoxylin and Eosin (H&E). The slide consists of 132 cores. This slide was provided by the Vancouver Prostate Center, pathology lab. We scanned this TMA slide with our in-house Hyper-Spectral Imaging (HSI) system, Figure 2-2 (A).

2.4.2 Instrumentation

Here we present the description of our in-house Hyper-Spectral Image (HSI) acquisition system. This imaging system comprises a “NEO” CMOS camera (Andor, Belfast, UK) a CRI varispec liquid crystal tunable filter, a Zeiss Axioskop 2 mot plus, a Marzhauser Wetzlar XY stage, and a standard 100W halogen lamp, all of which are controlled via a custom-designed Matlab GUI.

The TMA slide was imaged with wavelengths of 420nm to 720nm in 20nm steps which results in a total of 16 spectral band images. The transmitted light from the sample is filtered by the tunable liquid crystal filter for the desired wavelength. Objective magnification, numerical aperture, and camera resolution were 20x, 0.75, and 2048x2048 respectively. The field of view was 667 μ m by 667 μ m which resulted in an effective pixel size of 0.106 μ m² (optical spatial resolution of 9.43 pixels/ μ m²). Camera well depth was 30,000 e⁻.

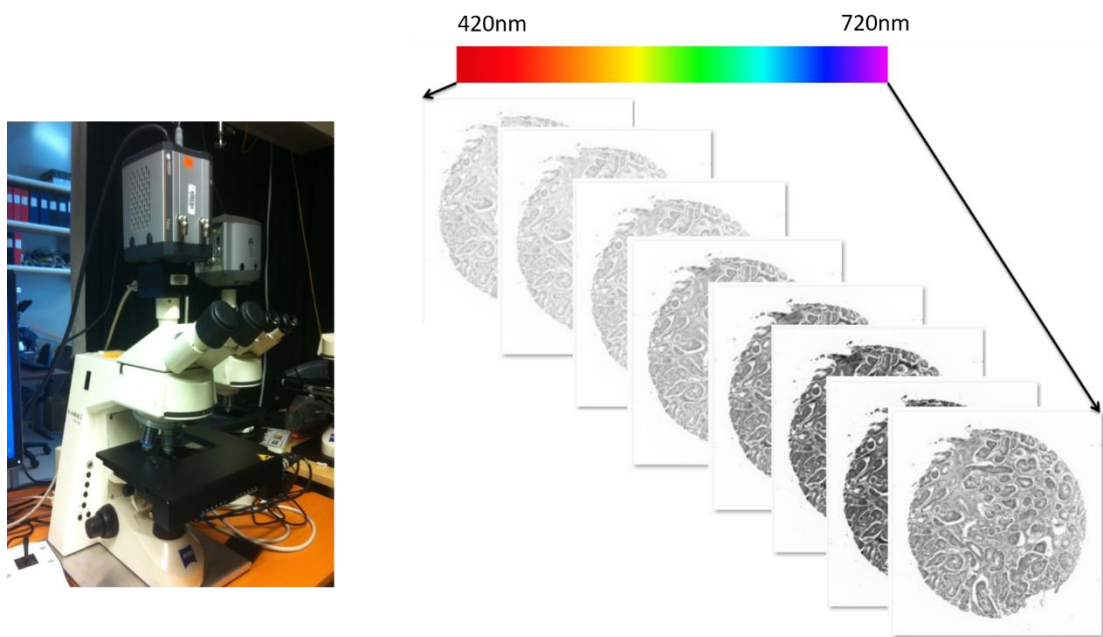


Figure 2-2. This figure shows our hyperspectral image acquisition set-up (on the left) and a representation of hyperspectral image construction that could be acquired by this system (on the right). 16 images with different wavelengths (420-720 nm) were captured with this imaging system (only 8 shown for illustration), the images are then combined and PCA images are constructed

2.4.3 Data Preparation

Prior to acquiring raw images, a blank area of glass is imaged for each wavelength and used as the flat field correction image, equation (2,1).

$$CorrectedImage = \frac{Input\ image}{Flat\ field\ image} \quad (2,1)$$

This image is collected as a measure of the spatial distribution of the incident light on the sample and is used to determine correct exposure time for each raw spectral band image. Each core on the TMA slide required four (2x2) fields of view or tiles.

Every pixel in the raw image is divided by its corresponding flat field pixel value to represent the percentage of light transmission. Pixels were then scaled by $2^{16}-1$ and saved as 16-bit TIFF files.

Each tile is scanned with 50 pixels overlap with adjacent tiles. We then used MATLAB intensity-based image registration function on the tile that is imaged with 540 nm wavelength to register the four tiles. The reason for this selection was that the 540 nm wavelength image showed the highest contrast among the other hyperspectral images. Overlapping strips in this image are aligned, any transitional error is corrected, and the offsets are recorded. Overlapping pixel values in the large stitched image are a weighted sum of the values from each overlapping tile proportional to the pixel's distance from the edge. We performed the same adjustment on the remaining 15 hyperspectral images.

2.5 Methodology

2.5.1 Principal Component Analysis

Principal Component Analysis (PCA) is a common transformation used on multidimensional (hyperspectral) data to reduce data dimension and remove (spectral) correlation. It is a linear

algebra basis change from orthogonal basis vectors defined by the sensor to basis vectors with no spectral correlation. This step is done with MATLAB PCA function, equation (2,2).

$$PCA_{coeff} = pca(X, Name, Value) \quad (2,2)$$

where X is all 16 bands input images (each band is unrolled to a one-dimension array, we put all 16 images in 16 individual columns to form a 2D matrix). Name is “algorithm” and Value is “SVD” (Singular Value Decomposition) of X (Principal component analysis of raw data).

The data represented on the new axis are the principal component (PC) images each with decreased data variance or image contrast. Each PCA image is a weighted sum of the input images. Typically, a significant portion of the dataset variance is contained within the first 3 PC and they can be combined into a single color image with components 1, 2 and 3 in the red, green and blue channels respectively (with the first three PCA we captured almost 98% of the information). It is important to point out that, running PCA on datasets independently will result in PC RGB composite images where identical materials can appear as different colors due to the fact that basis vectors are calculated for each dataset separately; however, using consistent weights results in consistent PCA images. To compare RGB images, datasets are transformed into a common basis vector set and scaled so that color intensities can be compared between RGB images. Figure 2-3 shows the steps for constructing hyperspectral images.

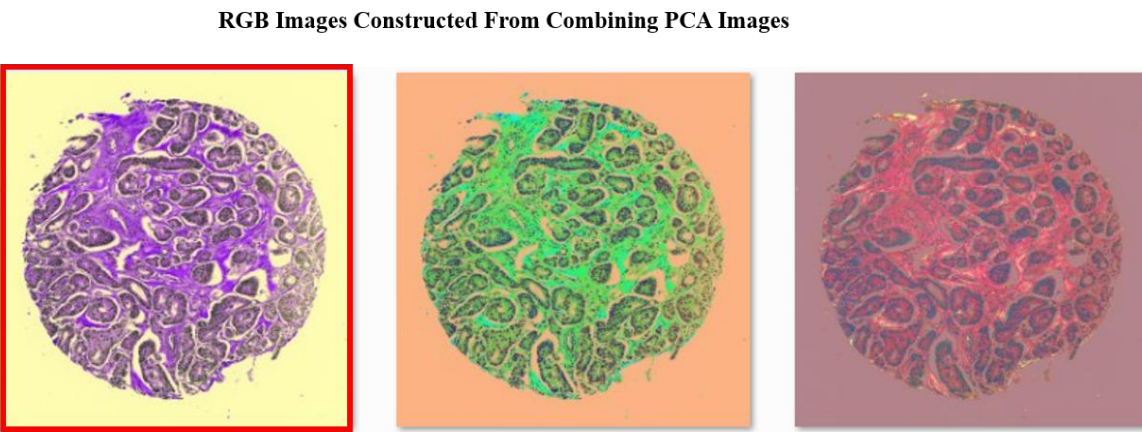
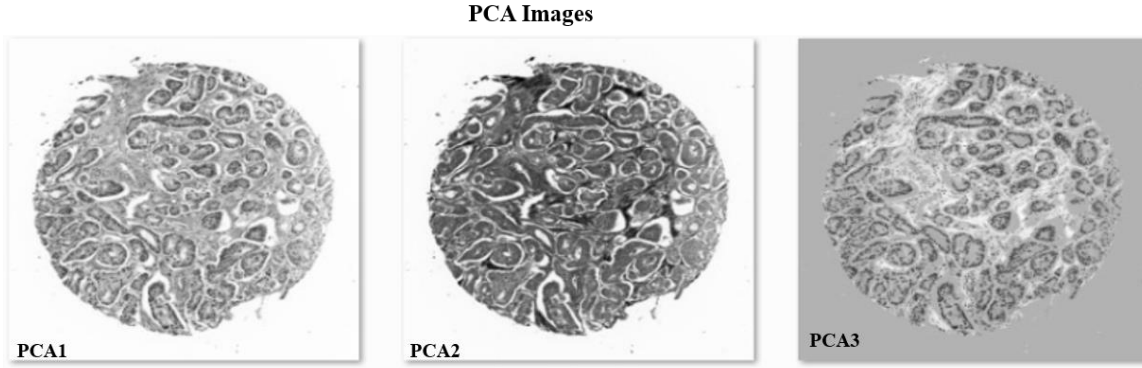


Figure 2-3. Process of constructing the RGB images from PCA. The image in the red-box is the first three PCA components and is used for further analysis. The second image (with orange background) is second, third and fourth PCA components and the last image is the third fourth and fifth PCA components. No satisfactory result is obtained using the last two images. They are presented here only for illustration.

2.5.2 K-means Clustering and Morphological Cleaning

The PC RGB images are now ready for further analysis. We first transformed them into LAB color space. We used LAB color transformation because Euclidian distance between colors does not correspond to the color perception in the RGB model (Color difference2019). LAB colour space is a three dimensional (L,a,b) colour/ intensity coding. The “L” axis, extends from black to white,

the “a” axis extends from green (-a) to red (+a) and the “b” axis extends from blue (-b) to yellow (+b). We then used K-means clustering in two steps:

Step 1: At this step, we separate out three materials, glass, stroma and glands. So, we set the number of clusters, K, equal to 3 and we applied K-means clustering on all images. We obtained three classes as shown in Figure 2-4 (B). Here we are interested in only the gland class; however, the gland class is not clean and contains some residual stroma as shown in the first image of Figure 2-4 (B). So, it was necessary to apply some image processing techniques to clean out the glands and remove the artifacts.

In order to clean up the input image we used the following morphological processes:

1. Removed all the connected components with an area of less than 30 pixels.
2. Applied morphological image opening with a disk structure element of size 5 to the remaining set (obtained from above) to remove the small pieces and connect the components that are close together and assumed to be related.
3. At the last step we removed all objects smaller than 1000 pixels, in this stage we expect the glands to be well-formed and connected so the objects smaller than 1000 pixels in size are considered noise/unwanted structures.

The result is well-segmented glands as shown in Figure 2-4 (C).

So far, we were successful in detecting glands, however, as we know glands contain cells with nuclei and cytoplasm, but we are only interested in the segmentation of the nuclei within each gland. In order to accomplish this goal, we need to continue with the second step as described below:

Step 2: In this step, we use another K-means clustering technique to segment nuclei inside the detected gland cells (nuclei of stroma are removed from the image in step 1). Here again, we set the number of clusters, K , equal to 3 since we have 3 clusters to classify (background, cytoplasm, and nuclei). Similar to step 1, we performed morphological cleaning (connected components smaller than 30 pixels are removed), Figure 2-4 (D).

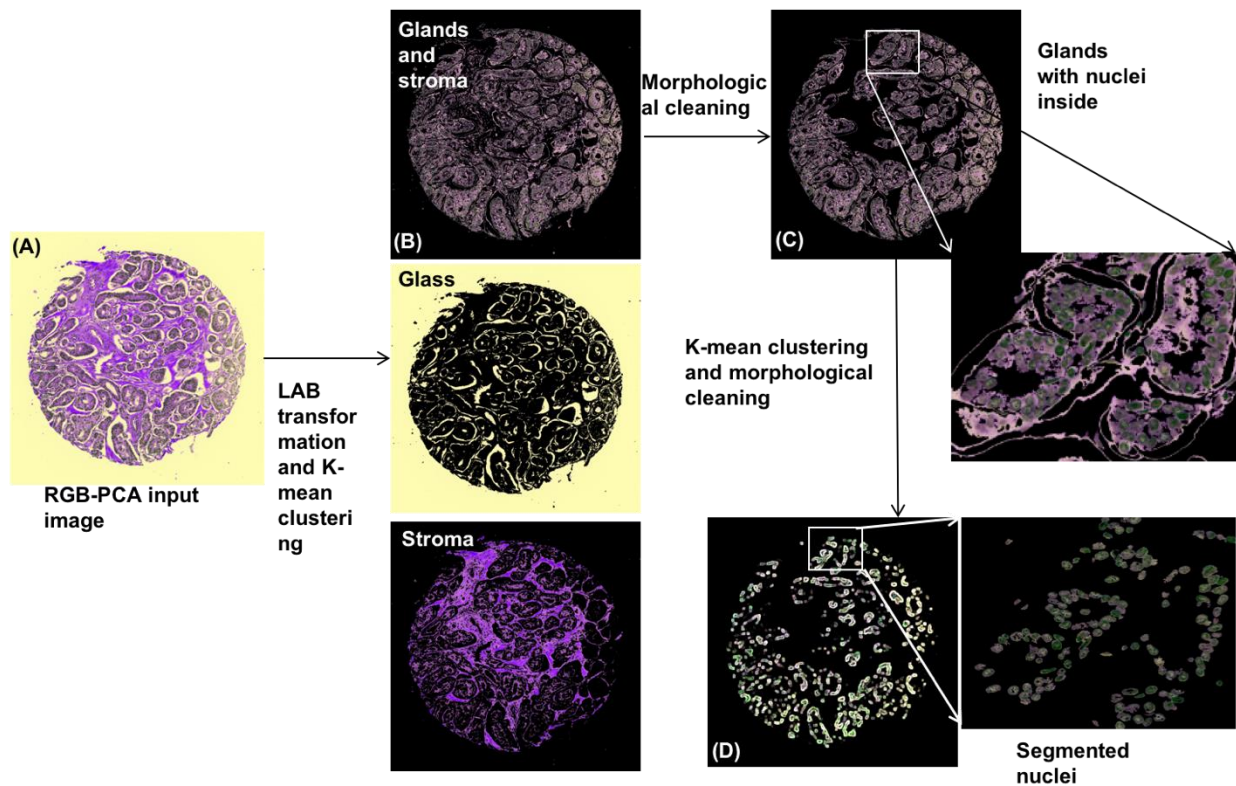


Figure 2-4 This figure shows the steps required for prostate glandular detection as well as detection of its nuclei. First, the PC RGB image transferred into LAB space and then using K-means clustering, gland/stroma, stroma, and glass are segmented. Morphological cleaning is used to remove undesired stroma in the gland/stroma image. At the last step, another K-means clustering sequence is applied to detect nuclei within each detected gland.

2.6 Results

We applied our automated method to all 132 TMA images. On average, it took two minutes for the algorithm to complete the glandular structures detection process and one minute to separate out the nuclei. To assess the performance of our algorithm we used a human observer. If more than 80% of the glands per TMA core were detected successfully, we denoted it as “good” detection. If 50% - 80% of the glands were detected, we denoted it as “moderate detection” and “bad detection” otherwise. Figure 2-5 shows an example of “good” gland detection. In the last column, we showed that our algorithm successfully segmented other tissue structures such as neural fibers (this structure is categorized as stroma).

In Figure 2-6 and Figure 2-7 we show examples of “moderate” gland detection and lastly, in Figure 2-8 and Figure 2-9 we show an example of “bad” detection.

Sixty-one percent of the cores had “good” detection, 15% had “moderate” detection and 24% of the cases had “bad” detection. About 50% of cases with “moderate” detection had uneven color intensity distribution, likely, due to various tissue processing/preparation protocol issues, (e.g. under-fixed or under-stained structures) and more than 60% of the cases in the “bad” detection category had gland infiltration filled with inflammatory cells and/or they had very muscular/hyalinized stroma. These are some of the characteristics that our algorithm was not tuned to detect and consequently reduced its detection power.

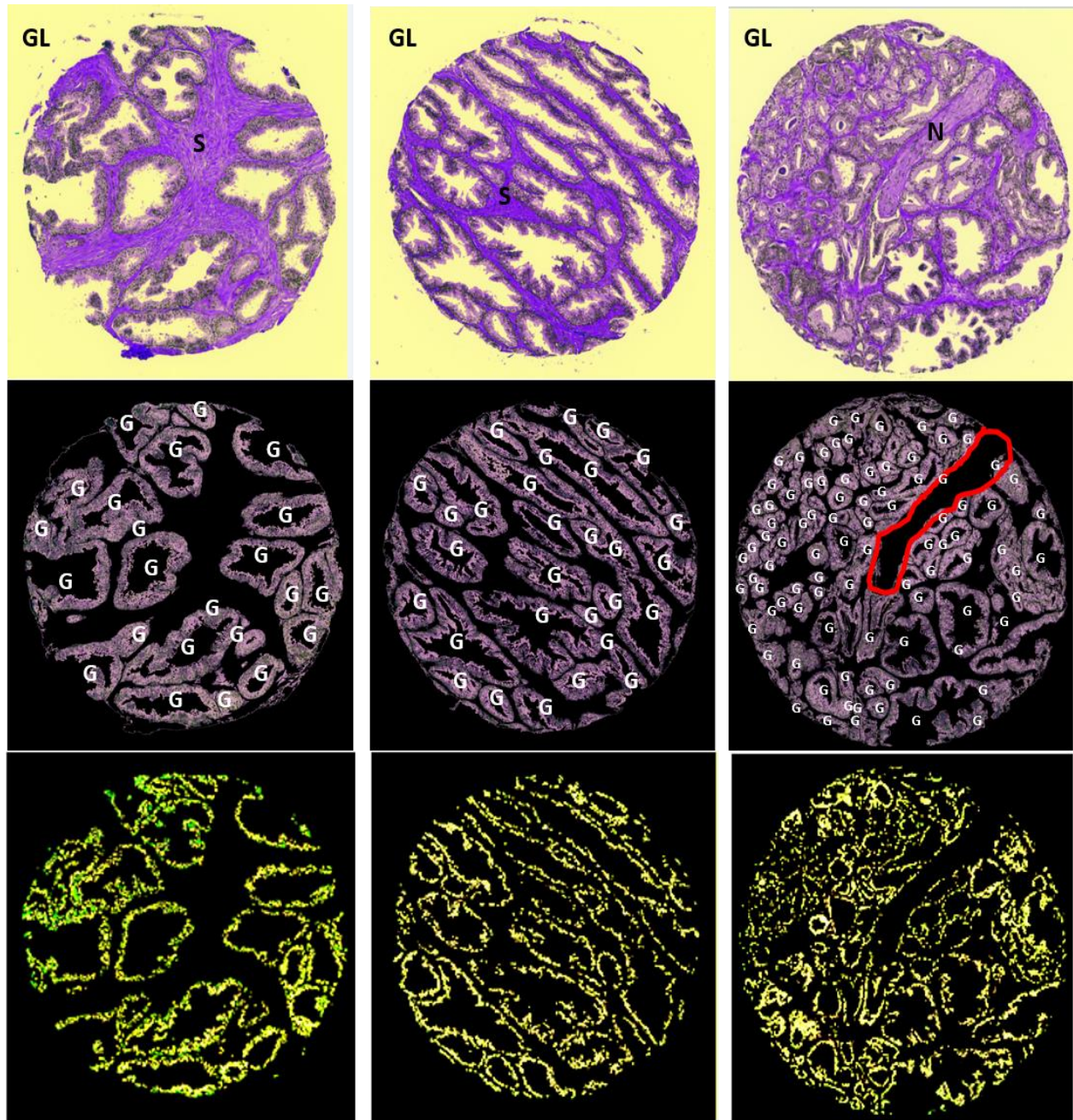


Figure 2-5 This figure shows examples of “good” glandular structure detection, the top row shows the PC RGB images. GL stands for glass, S stands for stroma and N stands for nerve trunk. The second row shows the gland detection results where G stands for gland; we showed an example of nerve trunk detection and removal in the third image. The third row shows the nuclei segmentation results.

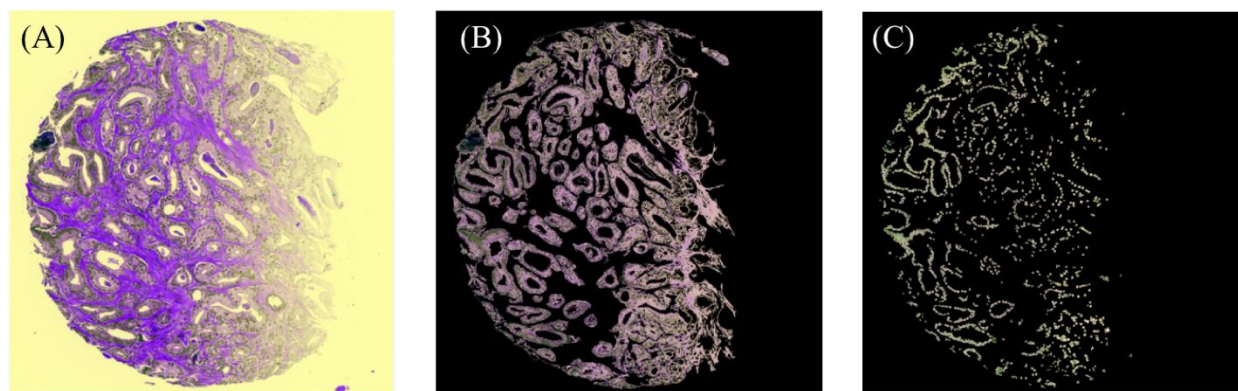


Figure 2-6 This figure shows an example of “moderate” glandular detection. Some glands are missed due to poor sectioning/scanning/imaging processes.

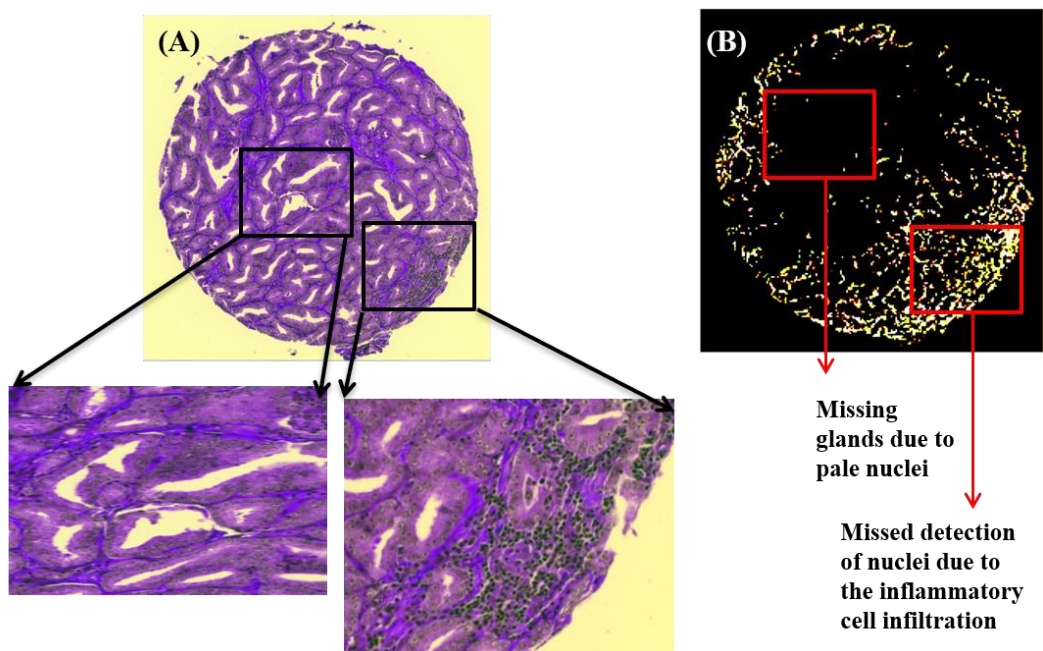


Figure 2-7 This figure shows an example of “moderate” glandular detection. This is an example of a core with densely packed tumor glands and small luminal spaces. In this example, the tumor has glandular structures with slit-like luminal spaces. The gland image is not shown here.

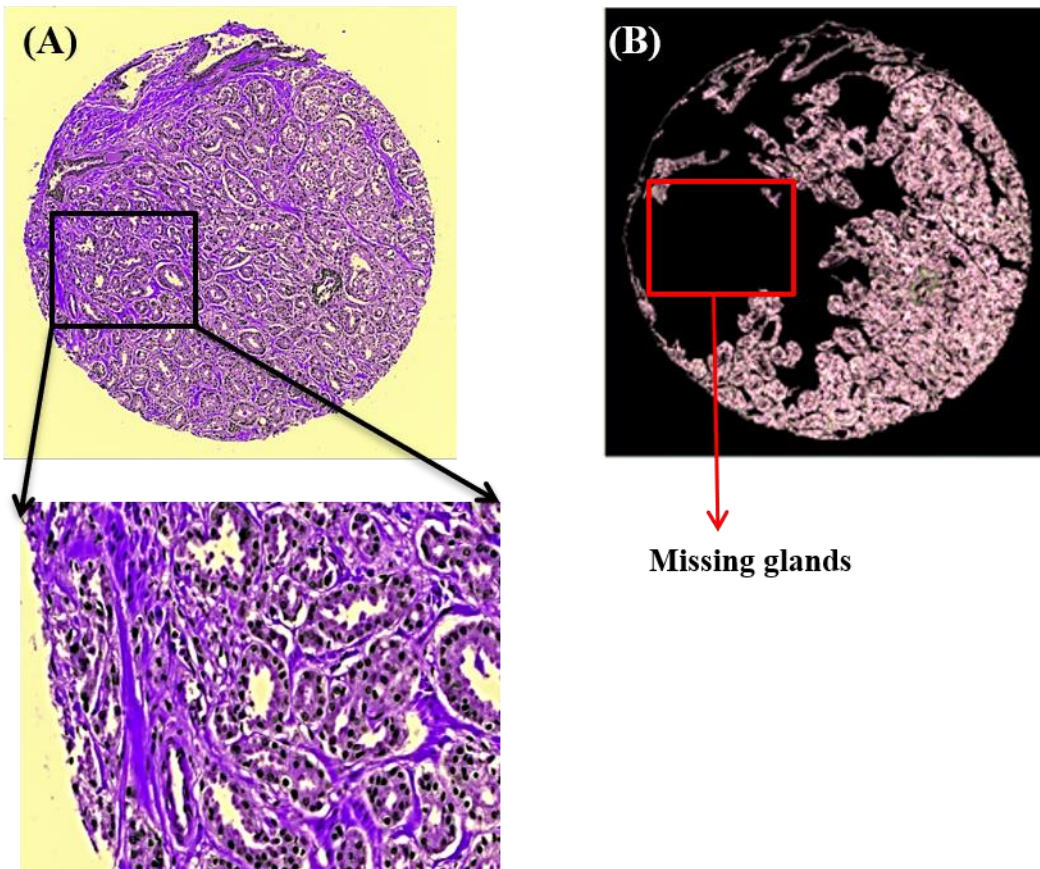


Figure 2-8 This figure shows examples of “bad” glandular detection in which the stroma is infiltrated by inflammatory cells. (A) the PCA image with a zoom-out of a section of the core with glands. (B) shows the failure of the gland detection. In the red box, multiple glands were mislabeled and were not detected. The nuclei image is not shown here.

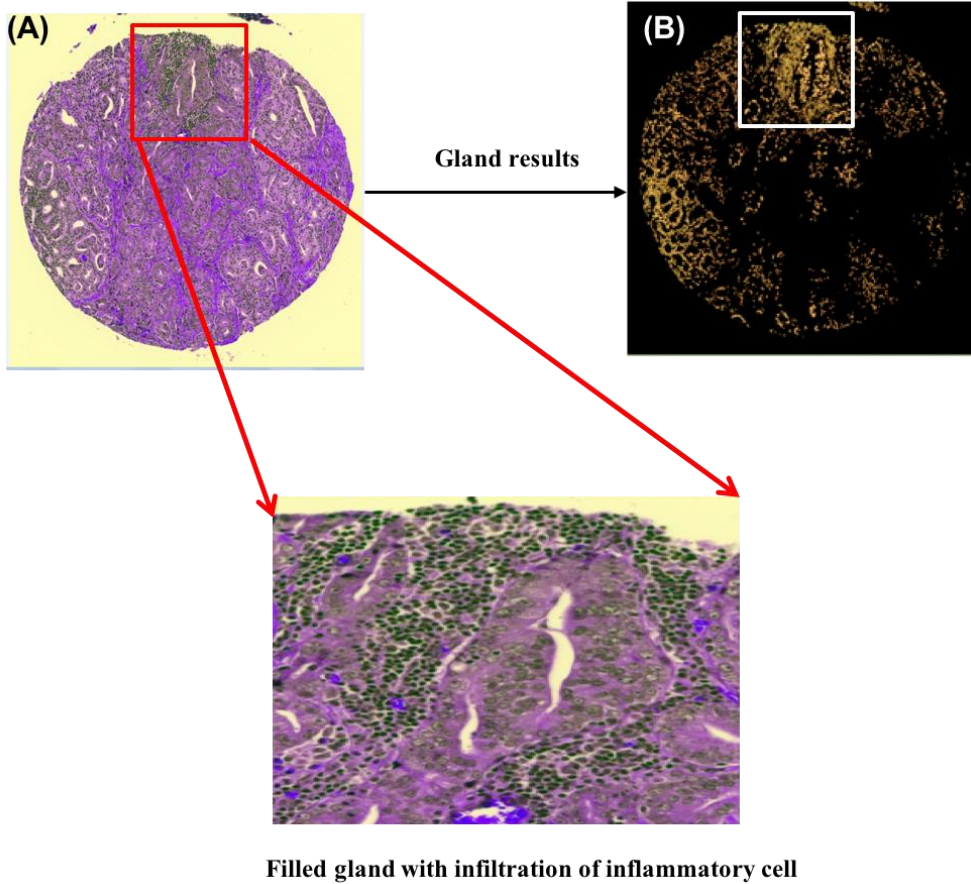


Figure 2-9 This figure shows examples of “bad” glandular detection, image (A) and image (B) show examples of cores in which the stroma is infiltrated by inflammatory cells.

2.7 Discussion

We compared our results with the state of art presented in (Alexe, Deselaers et al. June 2010). We directly used their publicly available code and tuned the parameters for our specific gland detection purposes (since the algorithm is dependent on the input image). The result is shown in Figure 2-10. This figure shows a very high probability of gland locations in the middle. Although our method works significantly better on this dataset, this comparison might not be completely fair since the method presented in (Alexe, Deselaers et al. June 2010) is designed for generic objects detection and is not designed specifically to detect glands in prostate tissue. From this comparison, we can

draw this conclusion that our method performs better for TMA prostate gland detection which is basically the purpose of this study.

One of the weaknesses of our algorithm is its sensitivity to color intensity. This limitation is due to the PCA transformation and K-means clustering steps. We believe uneven color intensities, especially at the edges, is one of the reasons for our method misclassification. In this study, we applied our method on 132 prostate TMA cores and obtained promising segmentation results. It would be beneficial to investigate the performance of this algorithm on a larger dataset as well as observing its performance on different stains.

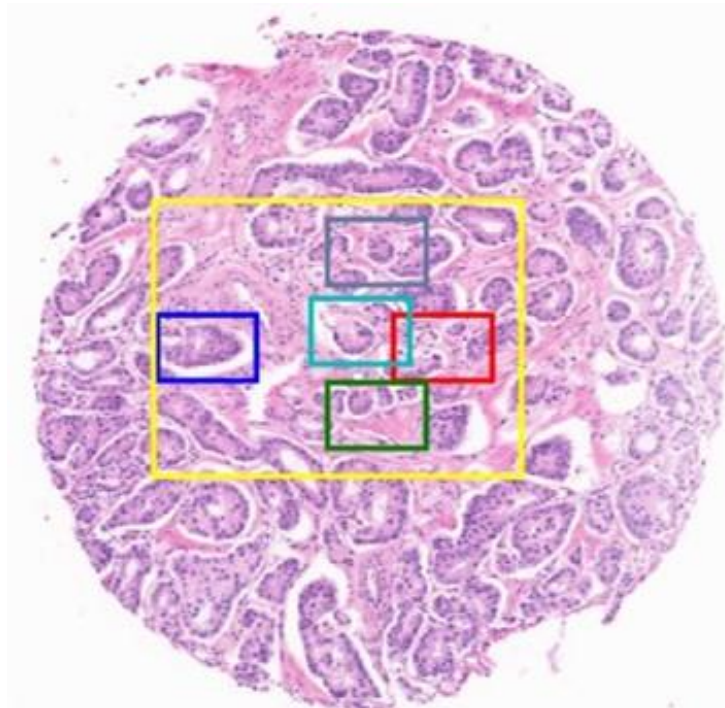


Figure 2-10 Example of gland detection result using the code presented by (Alexe, Deselaers et al. June 2010). The input image is an RGB image from the same core as shown in image 2 before the PCA transformation. The superimposed rectangles show the locations with a high probability of the presence of glands (mostly in the center.)

2.8 Conclusion

Morphological structures and architectural tissue characteristics are important for the diagnosis and prognosis of prostate cancer by determining the Gleason score/grade of prostate cancer tissues. Glandular structures are the regions of interest and they provide clinically relevant information about the tissue. Although the visual inspection of tissues has been around for years as an effective primary procedure for diagnostic purposes, the inter-intraobserver variability, among other challenges limits this procedure.

In this chapter, we presented our study results on the development of a new automated technique for glandular detection. This is the first step toward the design of a fully automated QDP system. We believe one of the significant aspects of our study is incorporating different techniques that result in a fully automated and robust detection method. We integrated hyperspectral image capabilities along with advanced image processing techniques to develop a fully automated algorithm for prostate gland detection.

Chapter 3 Design of an Automated Technique for Digital Pathology

Cell Segmentation

3.1 Introduction

In the process of cancer development, normal cells deform into cancerous cells. This complex process involves genetic and epigenetic disruption of the normal DNA organization, which results in phenotypic changes in nucleus/cellular/tissue structure. Therefore, to understand the altered functionality of cells and their interactions within the tissue, it is necessary to quantitatively study individual cells in the context of the whole tissue. With the flourish of light microscopy in 1670, study of tissue and cellular structures became popular (Madabhushi, Lee 2016, Meijering 2012) and continued with other imaging modalities such as fluorescence labeling of tissue samples (Irshad, Veillard et al. 2014, Doyle, Scott 2011) and digital microscopy (Alomari, Allen et al. 2009, Langer, Binenbaum et al. 2015). Quantitative assessment of nuclear structure using optical microscopy is now possible due to advances in image acquisition hardware and image processing algorithms. With the data obtained from Quantitative Digital Pathology (QDP), it is possible to measure the structural chromatin changes and provide prognostic and predictive outcomes. Image segmentation is the core of QDP. A very good segmentation algorithm greatly enhances the effectiveness of the QDP process and facilitates the decision-making in clinical pathology studies (Alomari, Allen, et al. 2009). With the assistance of medical image segmentation methods, it is possible to detect and segment cells and extract the relevant features. In the following section, we will present two main segmentation categories (Pixel-wise methods and Region-based methods). We have implemented similar techniques to design our minimally interactive cell segmentation technique.

3.1.1 Medical Image Segmentation Methods

Partitioning an image into the desired non-overlapping regions is called image segmentation. There are various segmentation methods which can be divided into two major categories:

- I) Pixel-wise methods
- II) Region-based methods

3.1.1.1 Pixel-wise Methods

This method is based on a simple yet important assumption: Pixels of the same segmented region should have similar characteristics. This reasonable hypothesis helped the flourish of many segmentation algorithms (Mao, Xie 2013, Wang, X., Zhang, et al. 2012). This method is further divided into two main categories, Pixel-wise thresholding, and Pixel-wise clustering.

(a) Pixel-wise Thresholding

One of the basic methods used in pixel-wise segmentation is Thresholding. This method usually requires some pre-knowledge of the input image and the intensity distribution of the Region of Interest (ROI). This requirement has been thoroughly investigated for one threshold, but it is very challenging for regions that require multi thresholds (Gonzalez, Wood, et al. 2004). Furthermore, thresholding is sensitive to noise, particularly salt & pepper noise (Gonzalez, Woods, et al. 2004). The other problem associated with this method is its ignorance of the spatial relationships of the ROI. Spatial information can be used to further improve the segmentation performance.

Despite these and many more drawbacks, this method is still being used widely due to many advantages that this method has such as being fast, simple and efficient. These flexibilities/advantages have made this method very desirable for segmentation purposes (Manikandan, Ramar, et al. 2014, Ye, Hu et al. 2012, Cheriet, Said et al. 1998, Dong, Li et al.

2014). We have used a modified version of this method in our segmentation algorithm, which we will discuss later in this chapter.

(b) Pixel-wise Clustering

Clustering algorithms are among unsupervised methods. These methods do not require labeled datasets (predefined input-output mapping). The main idea behind this method is to segment the input image into different partitions according to some criteria (such as distance, color, etc) related to the ROI and then check the quality of the segmentation (partitions). The last quality-check estimation can be done by using either an expert feedback or an automated (estimation segmentation) method. If the result is not satisfactory, the algorithm can segment the image again using different initial points and clustering parameters. This iterative process can be repeated multiple times until the desired result is obtained.

This method is sensitive to initialization and may result in unsatisfactory segmentation performance (Zahara, Fan, et al. 2005). Although this method is very popular in the literature and many researchers have used this method or a modified version of it for their segmentation task (Coleman, Andrews 1979, Pappas 1992, Uslan, Bucak 2010, Thilagamani 2011, SrinivasaRao, Suresh, et al. 2015), similar to pixel-wise approach, that we discussed earlier, it doesn't take spatial properties of the pixels into consideration (Zahara, Fan et al. 2005).

To improve our segmentation quality, we used a modified version of this method which incorporates the spatial information as well. Later in this chapter, we will fully discuss our approach.

3.1.1.2 Region-based Methods

Many region-based segmentation methods exist and are used in the literature (R, K et al. 2014, Slabaugh, Unal, et al. 2009, Lalaoui, Mohamadi 2013), in this section we will discuss the most relevant ones to the work presented in this thesis.

a) Deformable Models

Deformable models are a model-based method which require a parametric representation model of segmented objects to produce good results. This method segments the images using closed parametric curves.

Although this approach has many advantages, it is sensitive to initialization and can have poor convergence (Kass, Witkin, et al. 1988).

We incorporated deformable models into our segmentation algorithm. In our approach, we modeled nuclei as ovals using a simple deformable model. Furthermore, we utilized information from another part of our algorithm to overcome the challenges we discussed earlier (such as sensitivity to initialization and poor convergence). We discuss our approach later in this chapter.

b) Region Growing

Region-growing models require initial points to start the process. This method starts from the provided initial points (seeds) and grows the ROI outward from these points. The criteria for this growth can be adjacency/or similarity in the region. The adjacency criterion considers the spatial relation between pixels and improves the performance of segmentation; while the similarity approach uses a similarity metric between two points, or a region and a point and decides whether to grow or not. Region growing methods usually need a manual initialization and similar to thresholding methods are sensitive to noise (Adams, Bischof 1994).

We used a region growing method (Adams, Bischof 1994) as a part of our segmentation algorithm and combined it with other segmentation methods to overcome some of the discussed limitations. We describe our approach later in this chapter.

There are a variety of other segmentation approaches in the literature which apply advanced and complex methods to achieve cell segmentation. Examples include, dynamic programming (which uses gradient vector and boundary information of the objects) (McCullough, Gudla et al. 2008) , graph-cut (which uses fast min-cut approach to form the segmentation mask) (Bensch, Ronneberger April 2015), active mask (uses different combination of active contour methods) (Srinivasa, Fickus et al. 2009). Recently, multiple groups have used machine learning algorithms for their segmentation tasks. Different methods such as Support Vector Machine (SVM) and more advanced methods such as Deep Learning methods are used to achieve better segmentation results. Machine learning approaches require a large data set of images for training purposes (Ronneberger, Fischer, et al. 2015, de Brebisson, Montana 2015, Long, Shelhamer et al. June 2015). It is very hard (if not impossible), expensive and time-consuming to construct a large dataset especially in the medical domain.

In this chapter, we will present our segmentation approach in which we combined different segmentation methods to improve segmentation performance. Furthermore, in this chapter, we will introduce a new method to post-process the segmentation outcome with minimal user interaction.

3.2 Hypothesis

Our hypothesis in this chapter is very similar to the previous chapter. We believe digital pathology has high prognostic potential, however, it is mostly being used for preliminary diagnosis purposes. We hypothesize that an analytical approach based on the combined measurements of nuclei and

cell sociology features (Enfield, Martin, et al. 2019) can better classify aggressive and non-aggressive tissues.

Our goal is to develop simple, objective, automated, and cost-effective tools with the ability to produce reliable prognostic information for differentiating between indolent and aggressive prostate cancers. The first step to achieve this goal requires one to be able to accurately recognize, localize and segment cell nuclei. The segmented nuclei can then be used/accessed as a part of an automated pathology system to monitor and predict prostate cancer progress.

3.3 Objective

Our goal in this chapter is to develop an automated cell-/nuclei segmentation. We will also present the process to design an interactive tool for segmentation correction to improve the performance of the segmentation task.

3.3.1 Specific Objective

In this and next chapter, we have the following specific tasks:

- 1- Prepare the TMA images: capture and store images
- 2- Implement different segmentation algorithms to accomplish an automated cell-segmentation
- 3- Perform validation on the segmented cells (compare with human annotator)
- 4- Design an interactive method to enhance the segmentation quality (next chapter)
- 5- Compare the performance of the algorithm on various tissue type (next chapter)

The workflow for the whole process of our study on prostate cancer is presented in Figure 3-1.

The focus of this chapter, as shown in the blue box of Figure 3-1 is on cell/nuclei segmentation.

In this step, we will first scan the TMA slide with a whole slide scanner, and then prepare the image for the next step (image processing step).

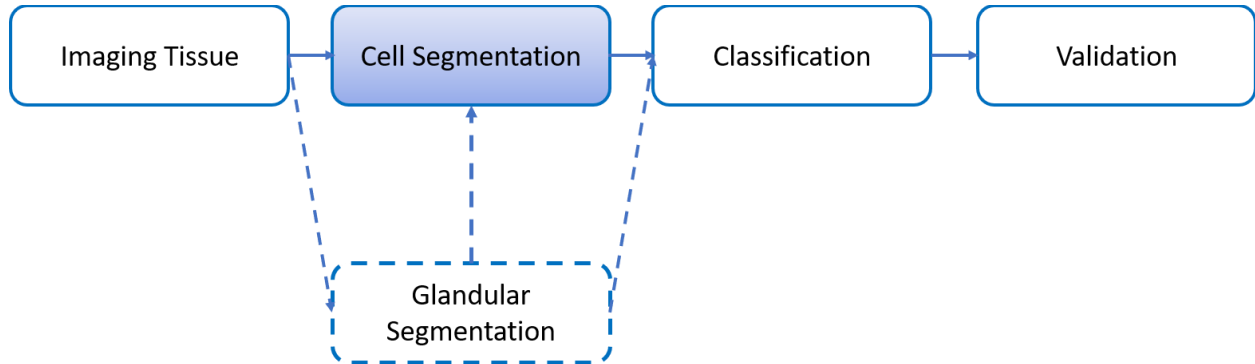


Figure 3-1 This figure shows the pipeline of our study. In the previous chapter, we discussed our algorithm for glandular segmentation. The focus of this chapter is on the cell segmentation. In the later chapters, we will describe the classification and validation processes.

3.4 Materials and Data Preparation

Surgical biopsy specimens from 33 prostate cancer patients were collected and a Tissue Microarray (TMA) was constructed from these specimens. The TMA is stained with Feulgen-Thionin stain. The reason for selecting this stain is because Feulgen-Thionin stain binds in a stoichiometric fashion with DNA and since light absorption per pixel is directly proportional to DNA amount, it will increase the contrast of nuclei in the image and enable the quantitative measurement of the DNA organization within the nucleus.

The TMA slides were scanned using a whole slide image tiling scanner (Pannoramic MIDI scanner, 3DHisttech, Budapest, Hungary, with $\times 40$ objective) where each tile image was 1024 by 1024 pixels and the pixel spacing was $0.3\mu\text{m}$ in both x and y directions.

3.5 Methodology

3.5.1 Automated Prostate Cancer Nuclei Segmentation

We used MATLAB as our prototyping tool for software development, primary because it has a very powerful image processing and machine learning toolboxes. (Once an optimized approach is identified it is possible to implement our methods in JAVA/C++ for even faster processing performance). Furthermore, object-oriented languages such as JAVA provides a better platform for further changes and modification (such as the addition of new features) which might be necessary in the course of software development.

Figure 3-2 shows an image of Thionin-stained prostate core from a TMA. This image is constructed from 16 sub-images (tiles) which are captured from the Pannoramic scanner (Figure 3-2 (a)). To reduce memory usage, we processed each tile separately and then stitch them together to represent the whole core.

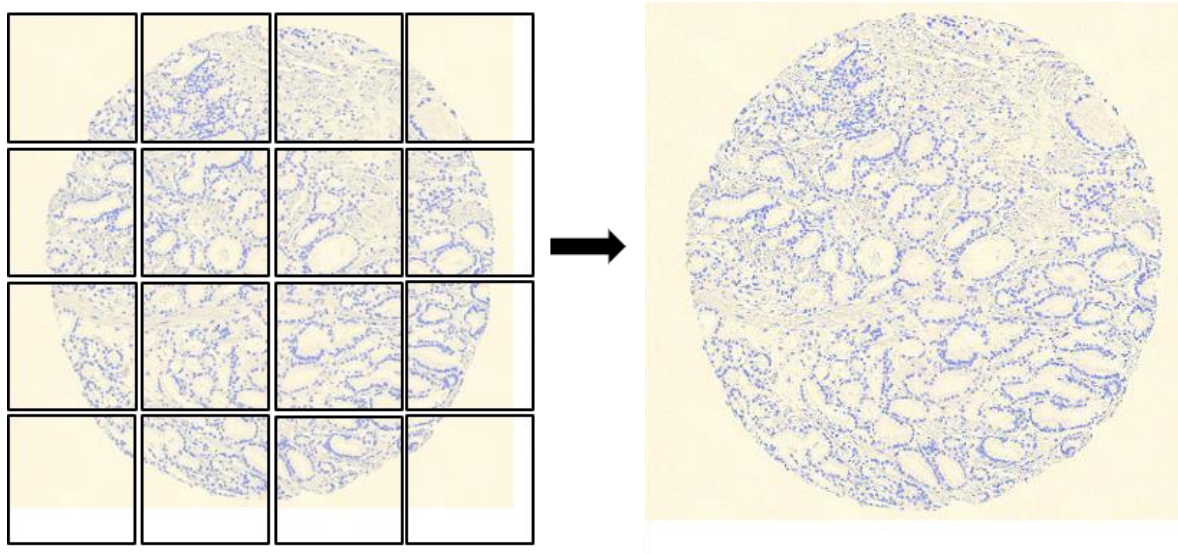


Figure 3-2 Thionin-stained prostate cancer core from a tissue microarray, a) 16 sub-images. b) the whole prostate core image obtained from (a).

Image pre-processing: To reduce photonic and systematic noise, we applied various filters. The trade-off of applying such filters is, losing the image quality and consequently reducing the image segmentation performance.

In this study, we applied a simple Gaussian filter (Gonzalez, Woods 2008). Gaussian filters are among smoothing filters whose weight is chosen according to the degree of smoothing required. This filter is known as a relatively good filter for reducing white noise (Dong, Li et al. 2014). A two-dimensional Gaussian filter in the frequency domain is formulated as shown in equation (3,1). Where D_0 is the filter cut-off frequency (Figure 3-3). We tried different values for D_0 and we chose $D_0 = 25$ based on the best visual result.

$$H(u, v) = e^{-D^2(u, v)/2D_0^2} \quad (3,1)$$

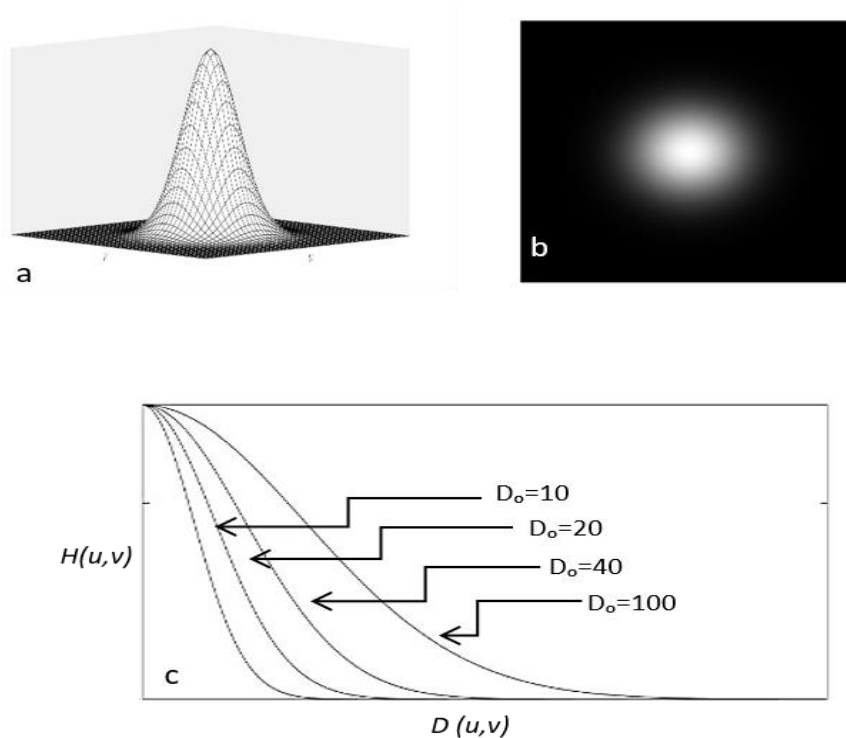


Figure 3-3 Gaussian filter. a) Perspective plot of Gaussian filter b) filter display as an image c) filter structure for different values of D_0 .

Next, we converted the RGB color image into a greyscale image. Since the nuclei are stained blue, we constructed a grey image by averaging the Red and Green channels (Since the intensity of blue in the image is very high, including blue channel in the averaging would skew the image to the blue color).

To construct a binary image from the gray image, two different methods were considered. The first approach was to use the Gradient-Weighted (GW) thresholding method (Gonzalez, Woods, et al. 2004). This method is known as an effective method and is applied to many medical images (Gonzalez, Woods, et al. 2004). In this method, the gradient image is calculated by using a Sobel filter, equation (3, 2). The assumption is that the gradient image will contain higher values for edges and lower values for the rest of the image.

$$\frac{\partial f(x,y)}{\partial x} = (f(x+1, y-1) - f(x-1, y-1) + 2 \times f(x+1, y) - 2 \times f(x-1, y) + f(x+1, y+1) - f(x-1, y+1)) \quad (3,2)$$

$$\frac{\partial f(x,y)}{\partial y} = (f(x-1, y+1) - f(x-1, y-1) + 2 \times f(x, y+1) - 2 \times f(x, y-1) + f(x+1, y+1) - f(x+1, y-1)) \quad (3,3)$$

We define $f(x, y)$ to represent the greyscale image. w and h represent the width and height of the image respectively, the GW threshold is calculated as follows:

$$T = \frac{\sum_{x,y} f(x,y)g(x,y)}{\sum_{x,y} g(x,y)} \quad x = 1, \dots, w \text{ and } y = 1, \dots, h \quad (3,4)$$

where $g(x, y)$ is the Sobel gradient magnitude image.

GW thresholding approach didn't provide a perfect result due to intensity inhomogeneity between and within nuclei, as shown in Figure 3-4(b). To overcome this challenge, we implemented a

second thresholding method, Mean Gradient (MG). This method is reported by Watanabe for the thresholding of cytology images (Ye, Hu et al. 2012). MG is a variant of the GW method. MG tries to reduce the effect of isolated high values of the gradient magnitude by dividing the sum of the gradient of image intensity by the number of pixels of that intensity. The threshold calculated from MG histogram is given by equation (3,5):

$$T = \frac{\operatorname{argmax}_i [\sum_{f(x,y)=i \in (x,y)} g(x,y)]}{n_i} \quad x = 1, \dots, w \text{ and } y = 1, \dots, h \quad (3,5)$$

where n_i is the number of pixels with the intensity of i .

We applied MG thresholding to our images; Figure 3-4 (c). In Figure 3-4 we show the comparison between GW and MG thresholding. From this image, we can observe that MG thresholding produced a better result compared to GW.

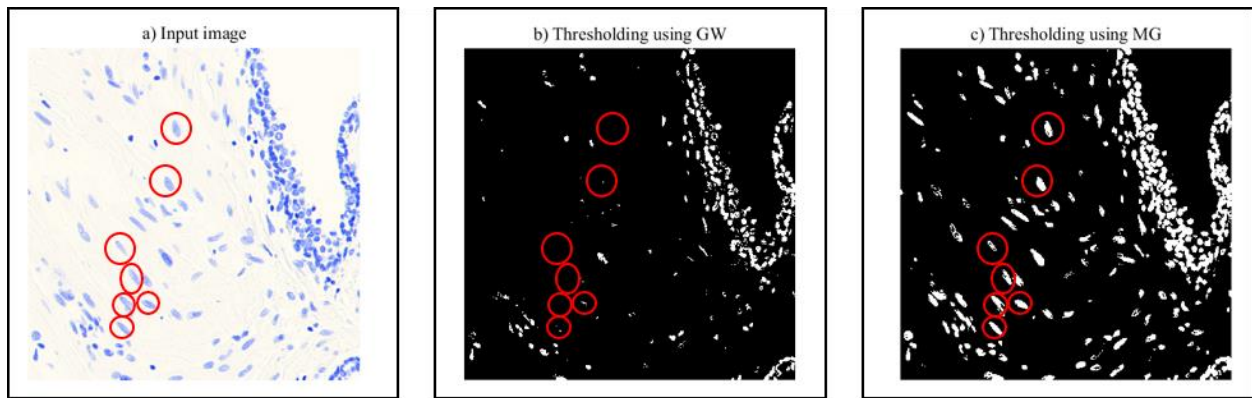


Figure 3-4 Process of image thresholding by Mean Gradient histogram and Gradient weight.

an) input image b) binary image obtained from GW, c) binary image obtained from MG.

The red circles enclose cells where the difference between the two thresholding methods is obvious.

Post-processing: The next step is to post-process the images and remove the artifacts. We removed small debris from the image by detecting and removing objects smaller than 10 pixels. We then labeled the remaining connected components as shown in Figure 3-5.

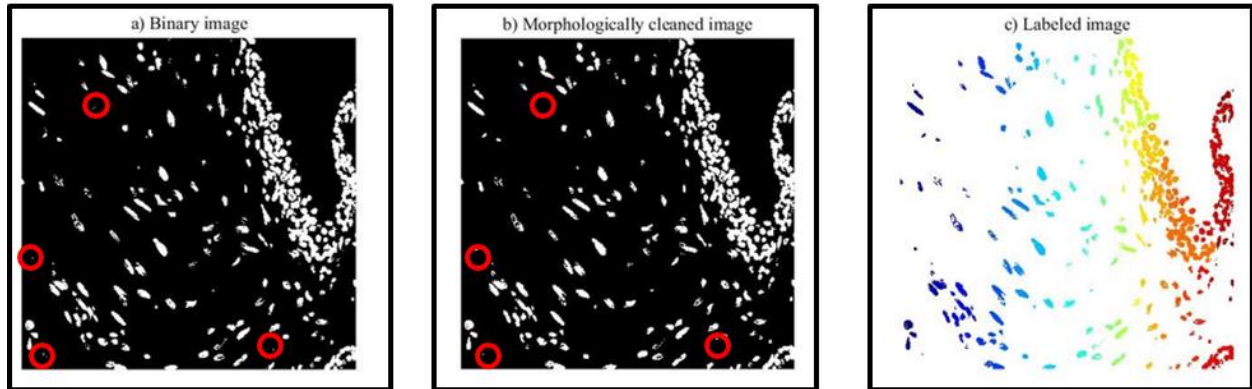


Figure 3-5 Morphological image cleaning and labeling. a) binary image b) cleaned image c) labeled image, each color represents a connected component. The red circles enclose areas where debris existed and where debris removed upon applying the proposed algorithm.

Now, that we have divided each image into multiple regions (connected labeled objects) it is possible to analyze each region independently. In Figure 3-6 we compared the histogram of the whole image and a selected connected component object, more variation could be observed in the histogram of the selected region versus the whole image.

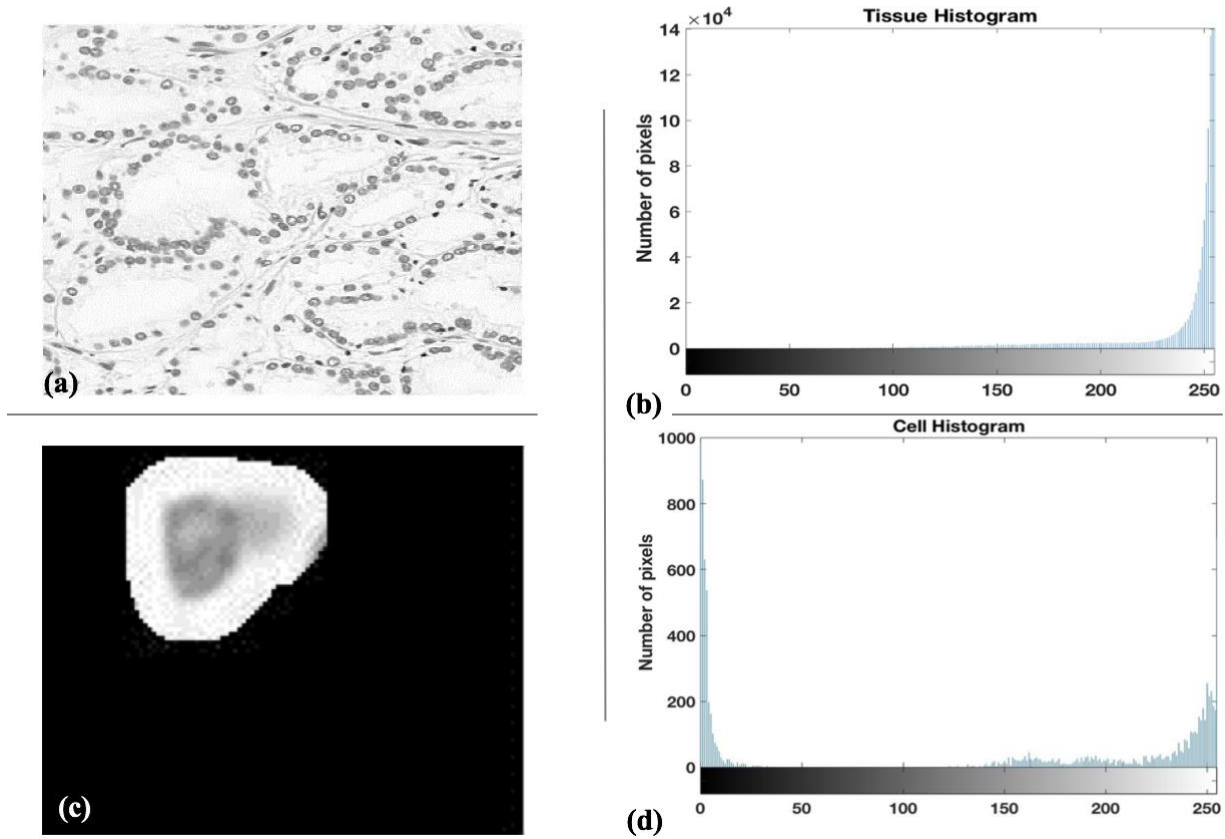


Figure 3-6 Global thresholding versus local thresholding (a) whole image (b) histogram of the whole image (c) region of interest (d) histogram of (c).

To further process each labeled object, and to optimally threshold each individual localized sub-image, we performed another thresholding method referred to as Otsu thresholding (Gonzalez, Woods et al. 2004). The Otsu method minimizes the overall within-class variance by minimizing the sum of class variances (Gonzalez, Woods, et al. 2004).

In this method, the normalized histogram of the image is calculated (each gray level has the probability of $p_i = \frac{n_i}{MN}$, where $M \times N$ is the image size and n_i denotes the number of pixels with the intensity of i).

Then the cumulative sum and cumulative means are calculated using the formula presented in (3,6) and (3,7) respectively. The next steps are computing the mean intensity with (3,8), and then

calculating the between-class variance with (3,9), and finally finding the value of k , which maximize the equation presented in (3,9) and setting this value as the image threshold.

In these equations L is intensity levels $\{0, 1, 2, \dots, L-1\}$ and k is the intensity level between 0 and L which sections the whole image into two segments:

$$P_1(k) = \sum_{i=0}^{t-1} p_i \quad (3,6)$$

$$P_2(k) = \sum_{i=t}^{L-1} p_i \quad (3,7)$$

$$m_1(k) = \frac{1}{P_1(k)} \sum_{i=0}^{t-1} i \quad (3,8)$$

$$m_2(k) = \frac{1}{P_2(k)} \sum_{i=t}^{L-1} i \quad (3,9)$$

$$m_G = \sum_{i=0}^{L-1} i p_i \quad (3,10)$$

$$\sigma_B^2 = P_1(k)(m_1(k) - m_G)^2 + P_2(k)(m_2(k) - m_G)^2 \quad (3,11)$$

We then applied the Otsu thresholding algorithm to the region of interest (localized objects) and subsequently extracted only the objects which contain the original seed point.

This combined thresholding algorithm worked perfectly for the detection of well-separated nuclei, however, it had some difficulty in detecting overlapping nuclei in a cluster, as shown in Figure 3-7.

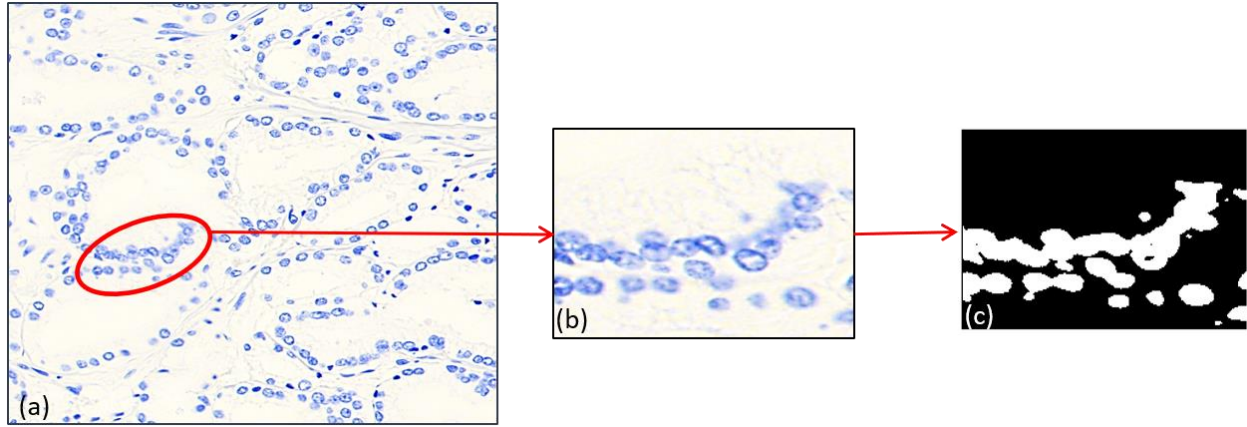


Figure 3-7 An example of segmentation failure with simple thresholding. Failure is due to clumped nuclei. (a) whole image (b) cluster of nuclei, (c) segmentation mask from (b).

Improving the performance of our algorithm for nuclei in clusters: In order to improve the performance of our detection method especially for overlapping nuclei, we computed the image skeleton (Morphology - Skeletonization/Medial Axis Transform) and chose each endpoint of the skeleton as the potential location for nuclei centers. Next, we fit an oval to the edge of the detected area. This oval has two constraints: 1) its center is near the initial guess of the object center and 2) its area (number of pixels) is < 800 pixels (within the average nuclei area range). With these constraints, we can make sure that the oval would only fit one nucleus. In the next step, we remove the detected object from the cluster and continued with the previous step on the remaining cluster's objects Figure 3-8.



Figure 3-8 Localization procedure for clustered nuclei. This method finds the skeleton of a cluster and uses the endpoints as a possible location for the cell center. As we can see in the rightmost figure, the initial four nuclei resulted in six potential nuclei centers, we will explain how to find the four nuclei out of these six spots.

Figure 3-9 demonstrates the performance of the algorithm in the segmentation of different cases (Isolated nucleus, a single nucleus with an overlapping neighbor, and a cluster of nuclei).

Figure 3-9 illustrates the result of our algorithm on three different object types:

- I) Easy case: Isolated nucleus (top row): Perfect detection for isolated nuclei. (This is an example of an easy type where the segmentation is done perfectly).
- II) Moderate case: Single nucleus with a small overlapping neighbor (middle row): Single nucleus is detected successfully and the small (and pale) overlapping nucleus is removed.
- III) Hard case: Cluster of nuclei (bottom row): The processes explained in the previous section is used to separate out these four nuclei (in this image, the result for one nucleus is shown)

The resulted image (right-most image of Figure 3-9) shows the estimated nuclei centers after applying our method.

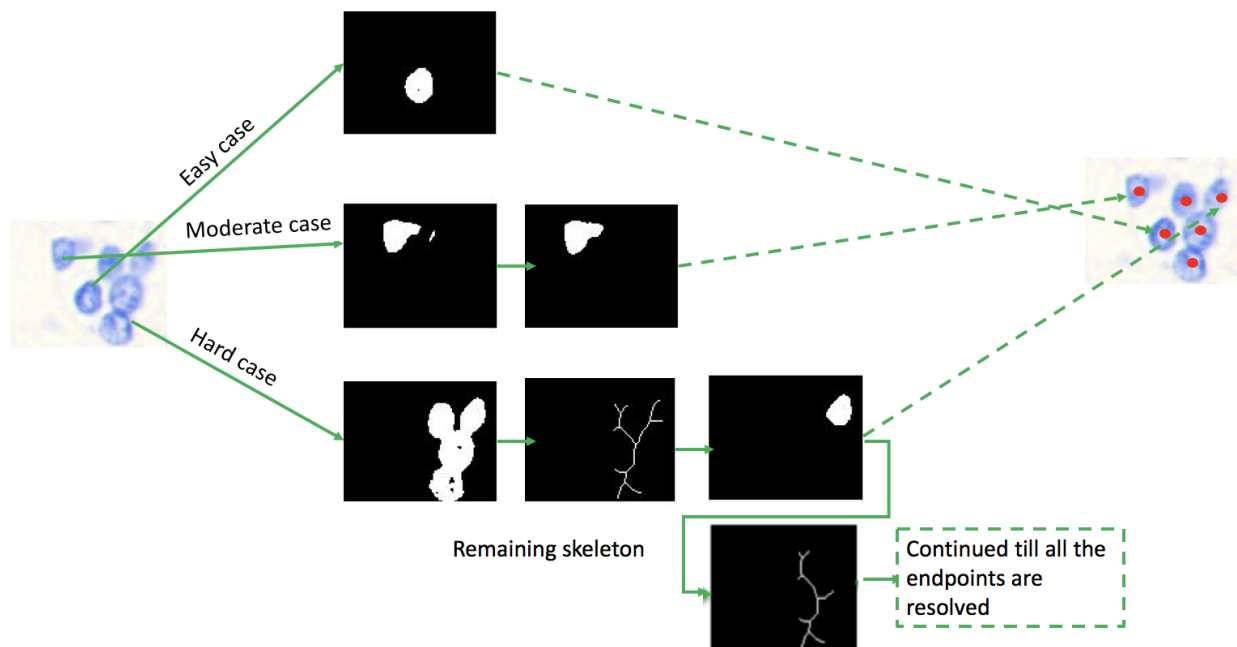


Figure 3-9 Nuclei segmentation for three different object types (isolated nucleus, a single nucleus with overlapping neighbor, and a cluster of nuclei).

Figure 3-10 illustrates the performance of our algorithm on a larger portion of an image. Our method was successful in localizing both isolated nucleus and nuclei within clusters.

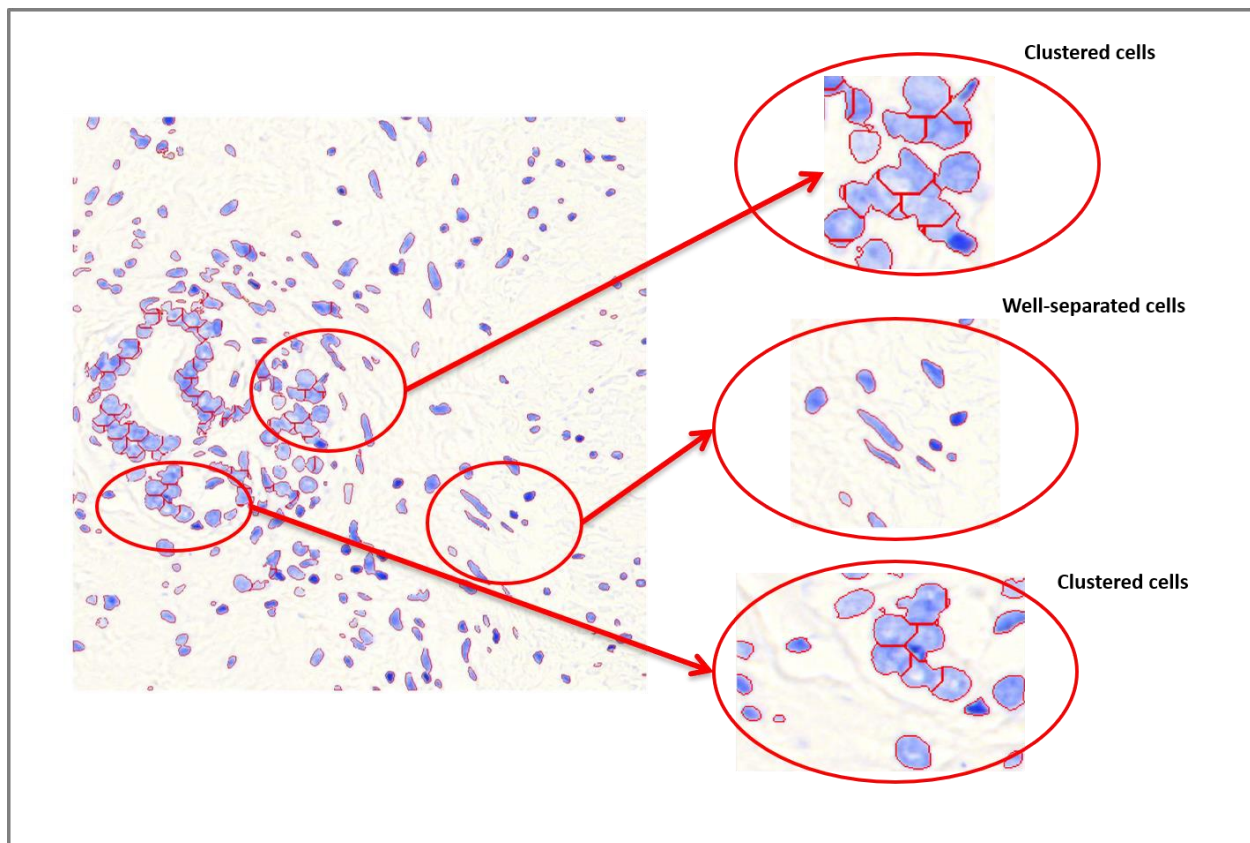


Figure 3-10 Nuclei segmentation results for different nucleus arrangement

We applied our algorithm on the RGB image of prostate TMA cores. The algorithm locates the estimated location of nuclei centers as shown in Figure 3-11. In this image, red dots represent the center of nuclei.

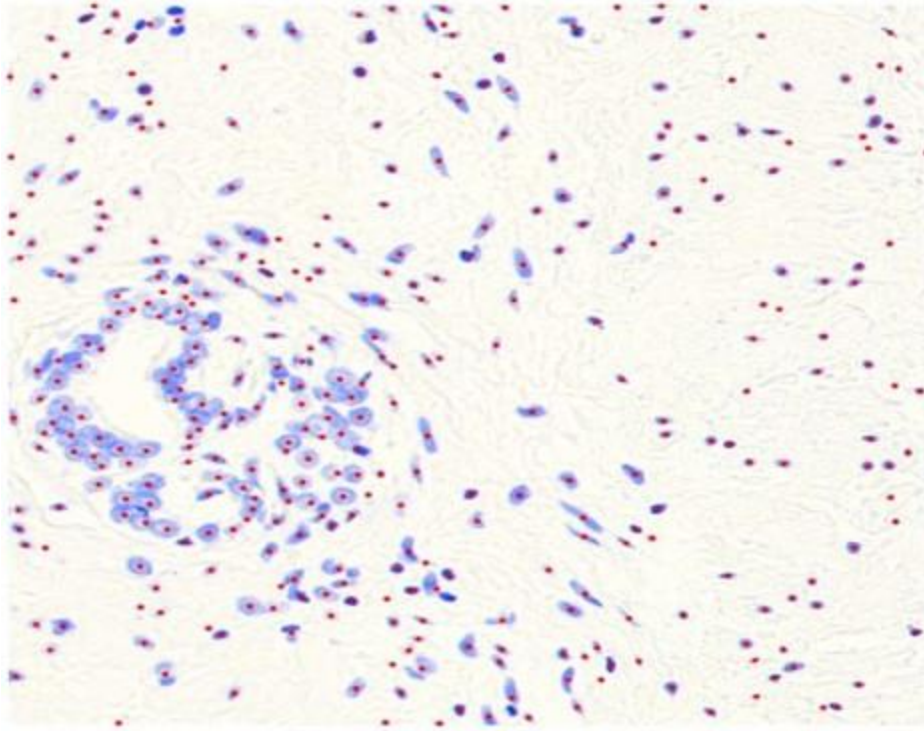


Figure 3-11 Nuclei detection result for prostate TMA tissue. Red dots represent detected nuclei

For a summary of our algorithm that we presented in this chapter, please refer to Table 3-1.

Table 3-1 Automated PCa nuclei segmentation algorithm

Automated Nuclei Segmentation Algorithm

Input data: Scanned RGB image

Output data: Segmented nuclei

Algorithm:

1. Convert RGB image to a greyscale image (using MATLAB `rgb2gray`⁷ function)
2. Generate a binary image using global thresholding methods
3. Apply morphological image cleaning to remove the small objects and remove the artifacts
4. Label connected components in the binary image
5. Dilate the result from above
4. Apply Otsu thresholding
5. Select the connected component (CC)
6.
 - If $CC_{size} \approx Ave(nuclei_{size})$:**
 - Add ROI as a single nucleus to final results and remove it from image
 - else:
 - Generate a skeleton of the mask (referred to as original mask):
 - While there is an end-point in the skeleton of the original mask:
 - i. randomly select one of the endpoints and use it as the initial point.
 - ii. generate a rough mask around the initial point (by fitting an oval with the constraints explained in the text).
 - iii. multiply this mask (referred to as nuclei mask) with the original mask to generate an ROI mask.
 - iv. use an ROI mask to segment the nuclei.
 - v. multiply the inversed of nuclei mask with the original mask and called it original mask.
original mask = original mask * inversed (nuclei mask)
 - vi. Generate a skeleton of the original mask
 - vii. go the beginning of the while loop and check the condition.

The last step after segmentation is saving each nucleus for further analysis

⁷ `rgb2gray`: this is a MATLAB function which Converts RGB image to grayscale

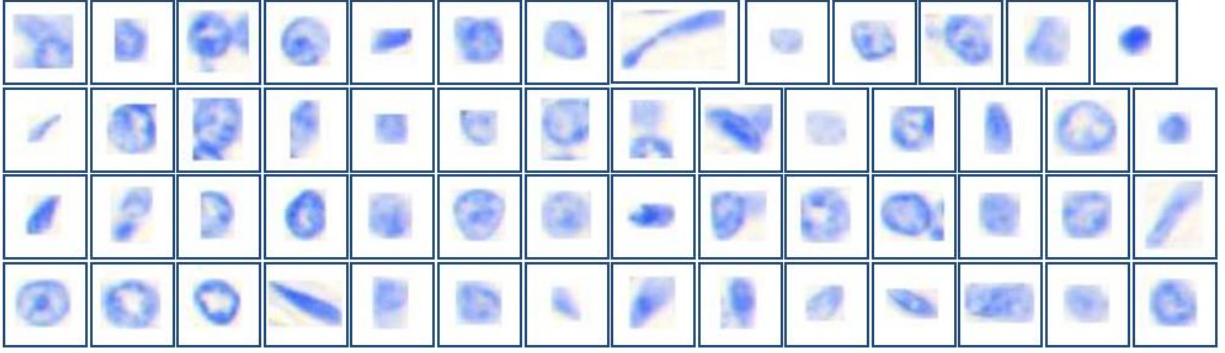


Figure 3-12 Gallery of segmented nuclei using our segmentation algorithm

A summary of our whole algorithm design is illustrated in Figure 3-13.

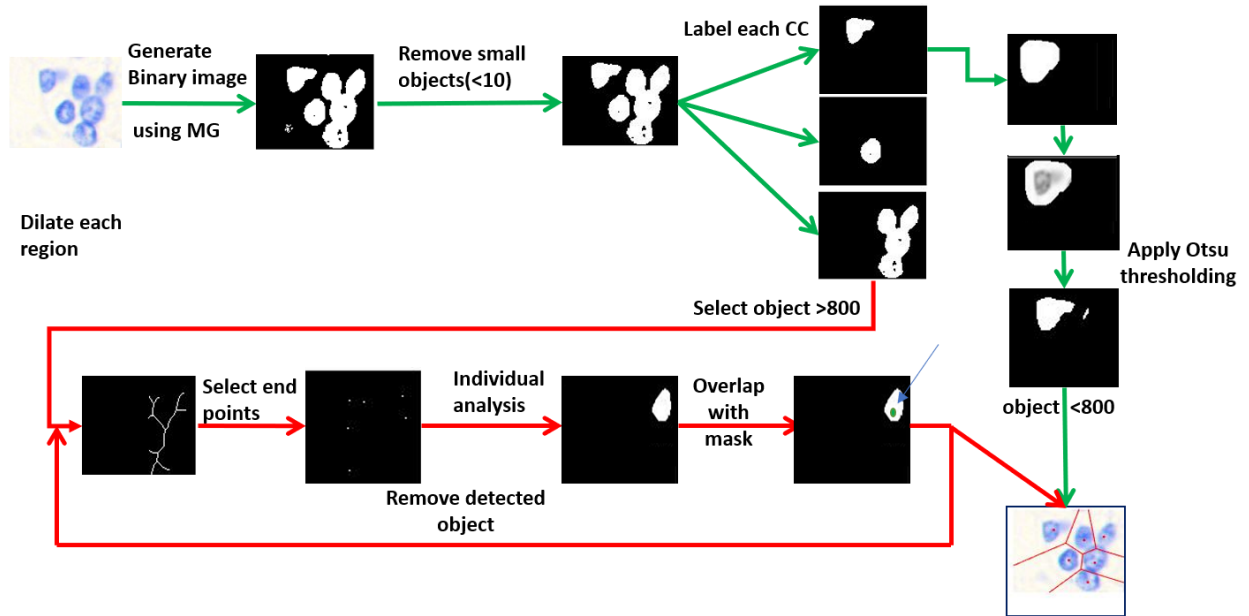


Figure 3-13 A cartoon that represents the whole processes used in our algorithm for the detection of prostate cancer nuclei in TMA slides.

3.5.2 Algorithm Performance and Experiment Results

We ran our algorithm on Thionin stained prostate cancer dataset containing 596 images (each image was 1024 by 1024 pixels). We randomly selected 20 images from a total of 596 images and compared the results of our algorithm with the results of a human annotator (human annotator

identified 13820 nuclei in the 20 randomly selected images). The comparison result is shown in Table 3-2.

Table 3-2 Processed nuclei segmentation results with our proposed algorithm.

Total Processed Nuclei	Number of False Positive Nuclei	Number of Missed Detection Nuclei	Average False Positive Percentage	Average Missed Nucleus Percentage
13820	470	774	3.4 %	5.6 %

According to Table 3-2, our proposed method missed around 5.6% of nuclei and had a 3.4% false nuclei detection rate.

On average, using our algorithm we can process each image in about 2 minutes on a 3 GHz machine with 4 Gigabytes of RAM while the same process done by a human expert requires 45 minutes.

3.6 Discussion

Although our algorithm was very successful in the detection of nuclei, it had some challenges in detecting nuclei in a few situations. One of the most common difficult situations is when the nucleus is lightly stained and has a pale color. In this case, the intensity of the nucleus is very similar to the background and may result in poor detection/segmentation.

These limitations motivated us to design a method which can post-process the image (in an interactive fashion if required). We implemented such a system and we will explain our approach in the next chapter.

Chapter 4 Interactive Cell Segmentation for Digital Pathology

4.1 Introduction

This chapter is the continuation of the previous chapter, as a result, the introduction, hypothesis, and objectives are very similar to the prior chapter. The extra information that we will present in this chapter is around the improvement of the technique that was presented in the previous chapter. Furthermore, in this chapter, we will discuss how our method can be generalized and applied to other cancer sites.

The method we presented in the previous chapter had some difficulties in the segmentation of a few cases, particularly in the segmentation of nuclei within clusters. To accomplish a complete nuclei segmentation, we have designed an interactive segmentation methodology. Our algorithm utilizes user input to re-segment locations where the user recognizes that the automated method failed to localize the cell nuclei. This method gives the user the ability to select/unselect miss-segmentation and re-segment areas with minimum interaction.

4.2 Motivation

As explained earlier, the technique presented in the previous chapter had a very good performance on localized cells, however, the performance was relatively worse in challenging cases such as lightly stained nucleus, (Figure 4-1). In this case, for example, the intensity of the nucleus is very similar to the background resulting in a poor segmentation.

Another challenging situation happens when we have a cluster of nuclei or overlapping nuclei. In these cases, our technique might miss a few nuclei or over-segment the cluster, Figure 4-2.

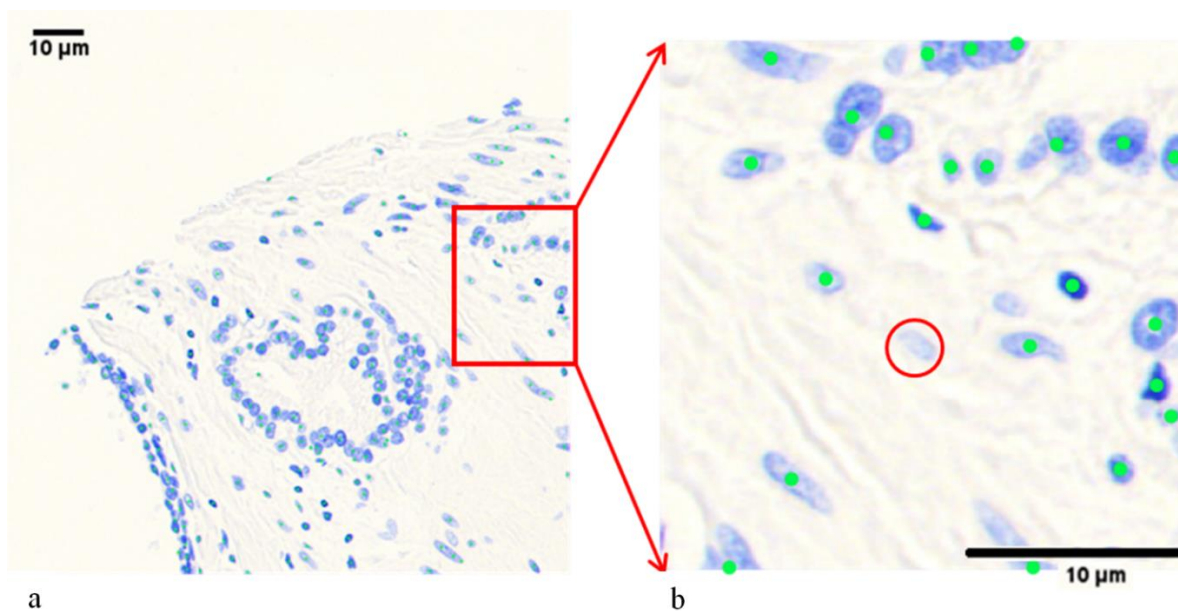


Figure 4-1 Faint nuclei resulted in segmentation failure. a) a section of prostate tissue (Thionin stained), blue objects are cell nuclei, light green dots (not very visible) are cell locations predicted by our algorithm. **b)** zoomed-in section of image(a) red circle circumscribes a pale nucleus that our algorithm missed.

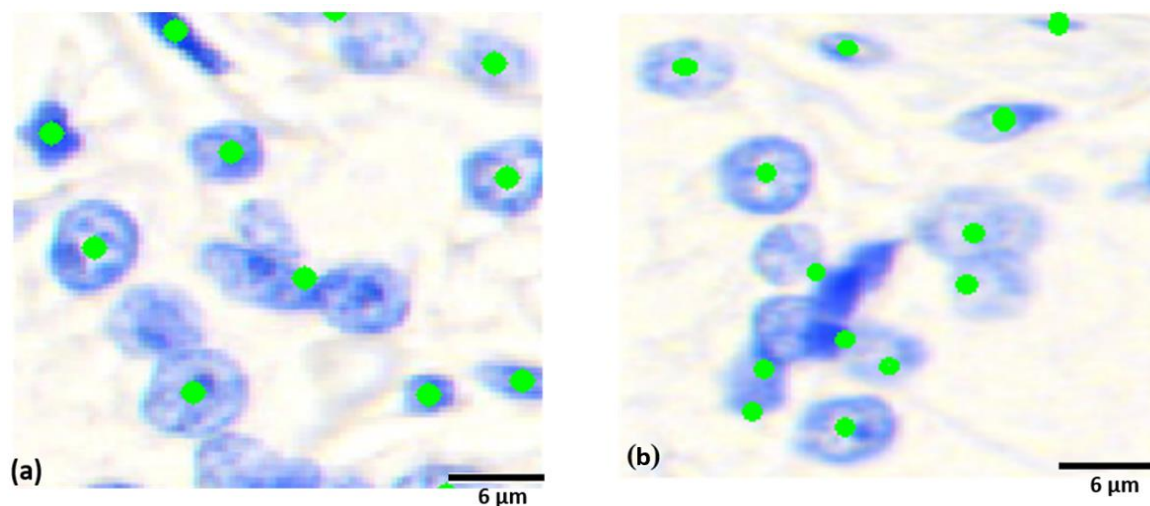


Figure 4-2 Failure of segmentation method on a cluster of nuclei. a) blue objects are cell nuclei and the light green is the algorithm prediction result for the location of nuclei (missing nuclei). **b)** incorrect localization of cell nuclei by our algorithm (over-segmentation) for a cluster of nuclei.

4.3 Objective/Hypothesis

The goal of an automated segmentation technique is to detect/localize/segment the objects of interest for a rapid interpretation. Thus, asking the user to draw nucleus boundaries (Ning, Zhang, et al. 2010) is counterintuitive and should be avoided in any tool that is to be used clinically.

We hypothesize that it is possible to improve the performance of our technique (presented at the last chapter) by designing a powerful segmentation algorithm that can segment nuclei by utilizing the additional high-level information about the nuclei center locations that the user has provided (add/removed nuclei centers). To be useful in clinical settings, this process should require minimal interaction from users

Our goal is to design a fast and accurate post-process method to be combined with an automated algorithm (presented in the previous chapter) to improve segmentation performance.

In the next section, we will present our algorithm that uses the information provided by the user to improve the segmentation accuracy.

4.4 Methodology

In this section, we will discuss the steps required to accomplish the design of an integrative segmentation method which requires minimal input from the users. If no input is provided, the algorithm will still complete the segmentation procedure using the process explained in the previous chapter.

Consider a situation where the nucleus has poor contrast and the user provides a rough estimation of the nuclei center, our algorithm will take the following steps to complete the segmentation. First, it uses the provided point as an initial guess and dilates that point with a disk of radius 50. This radius is selected by heuristically experimenting with different sizes from 5-60 pixels (each pixel spacing is $0.3\ \mu\text{m}$ and 50 pixels is equal to $15\ \mu\text{m}$ which is near nuclei size $7\text{-}20\ \mu\text{m}$). It uses

this mask to extract a portion of the image that should contain the intended nucleus as shown in Figure 4-3.

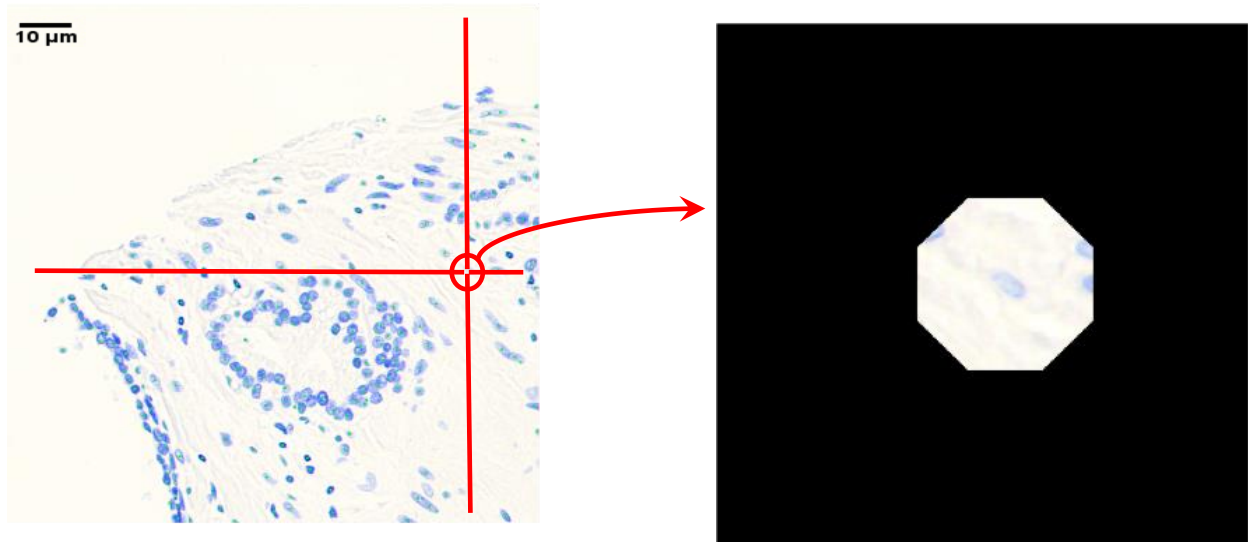


Figure 4-3 Interactive correction method to add/remove the nuclei. a) original image, the red cross represents the interactive method on the location where we plan to apply some correction. b) extracted the nucleus and surrounding neighbors.

The next step is to find the edges in the extracted area. This edge detection step is done using the Canny edge detection method (Canny 1986). Next, we dilated the extracted edges with a square structure element of size 3 (large enough to cover some part of nuclei as well as the background). We used this mask as our new ROI mask. Our assumption here is that the area under this new mask contains approximately the same amount of background and nucleus. Therefore, we can use a simple Otsu thresholding method to determine an optimal threshold for extracting the nucleus Figure 4-4.

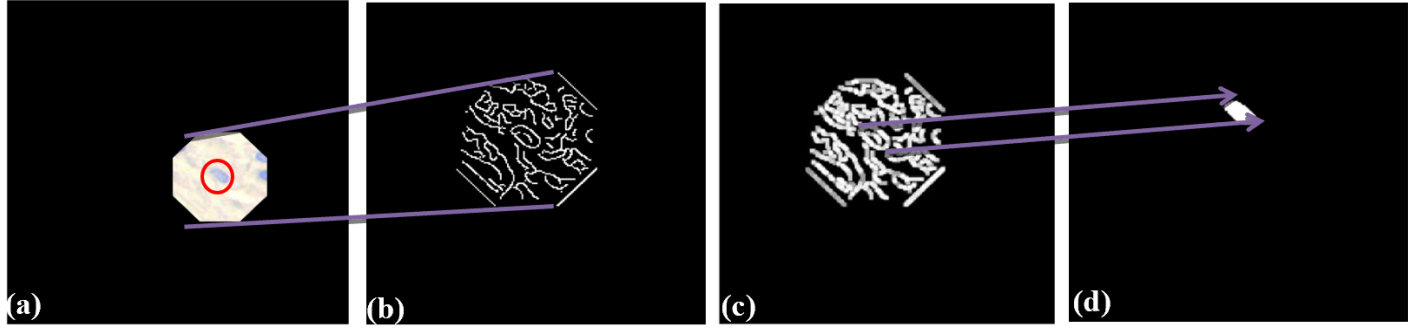


Figure 4-4 Thresholding based on local information a) masked image. b) extracted edges in the area c) dilated edges using Canny d) extracted mask using Otsu thresholding.

Although Otsu thresholding usually gives us reasonable results, it fails to extract the whole nucleus in some cases. To overcome this challenge and improve our result, we have applied a region growing algorithm combined with a clustering method. The details of our implementation are as follows:

We implemented a modified region-growing algorithm technique. This method is based on the intensity values of the object of interest. In addition, we added a hysteresis mechanism to make the region-growing method robust with respect to noise and variations in the staining intensity. We have formulated the hysteresis step similar to the double thresholding method presented by Canny, (Canny 1986) with a couple of modifications to make it suitable for our study. We incorporated the pixels in the extracted mask obtained from Otsu thresholding as the region-growing initial seed points. Next, we inverted the image so that the background becomes dark and cells become white. This makes it possible to apply the region-growing by starting from the white regions. It is important to highlight that, by implementing these steps we eliminated the need for human initialization or random initialization which is frequently required for the region growing algorithms.

Hysteresis region-growing: In our model, the initial seeds are usually a cluster of pixels instead of a single value, in order to decide whether to add a new point (pixel) into the region we used some statistical values of the cluster as the algorithm constraint, details of this implementation is provided below:

Our model uses two thresholds ($T1$ & $T2$ ($T1 > T2$)). The average intensity of the initial seeds was used as $T1$, with this threshold we generated a binary image referred to as $G1$. For the second threshold ($T2$) we used the standard deviation (SD) of the intensities of the initial seeds and formulated $T2$ as $T2 = T1 - SD$. With the $T2$ threshold, we generated a binary image referred to as $G2$. $G1$ gave us strong candidates whereas $G2 = G2 - G1$ provided us with the weak candidates. We start with a pixel in $G1$ since we are confident that it belongs to the nuclei, we then choose the corresponding pixel in $G2$ and check for eight connectivity to decide on region growth.

During the region-growing procedure, two or more different regions might merge together. In these situations, we would calculate new threshold values for $T1$ and $T2$ from the initial seed points of these regions.

Figure 4-5, shows an example of cell localization after applying the above technique.

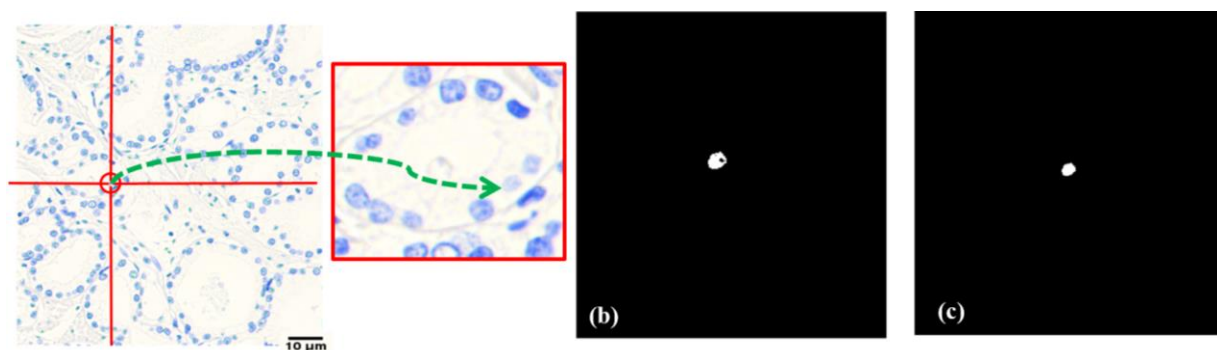


Figure 4-5 Applying region-growing methods a) original image, selected nucleus and zoomed nucleus. b) original region growing method c) region growing result using hysteresis.

Hybrid clustering: To help our algorithm robustness to noise we integrated a hysteresis approach in our proposed region growing method. However, this technique still has some difficulties with intensity variations in the objects and may not connect all parts of a single nucleus together. Therefore, we employed a novel hybrid clustering method.

This method combines two well-known clustering methods (K-means clustering and Density-based clustering). K-means clustering tends to return convex clusters as it usually does not use spatial information (Likas, Vlassis, et al. 2003). On the other hand, Density-based clustering uses spatial information but there is no guarantee of returning a convex shape (Gan, Tao 2017). In this study, we combined K-means clustering and density-based clustering technique to incorporate the strength of both methods.

K-means clustering: In the K-means clustering method we set the number of clusters, K , equal to the number of distinct objects in the area of interest (ROI: is obtained from the previous region-growing step).

Density-based clustering: To calculate the density-based clustering we first need to define a few parameters, one of these parameters is the distance function presented in equation (4-1).

$$f(p_x, p_y) = d(Xp_a, Xp_b)^\alpha \times d(Lp_a, Lp_b)^\beta \quad (4-1)$$

where Xp_a and Xp_b represents the location of a and b respectively. Lp_a and Lp_b represents color in L.a.b space for a and b respectively, and d is the function that finds the Euclidian distance between these points.

The proposed distance function in equation (4-1) incorporates both the intensity level of a pixel as well as its location. This function has two tuning parameters (α and β) which allow us to make a trade-off between the importance of pixels location and their intensity values.

Since the Euclidian distance between colors does not correspond to the color perception in the RGB model we decided to use the Lab color map.

Lab Color Space is a three dimensional (L,a,b) color/ intensity coding. The L axis extends from black to white, the “ a ” axis extends from green ($-a$) to red ($+a$) and the “ b ” axis extends from blue ($-b$) to yellow ($+b$).

The initial points can be selected randomly. To find the clusters, we set α to 0.5 and β to 0.8 in equation (4-1). The other parameter is the maximum distance between two adjacent points in the cluster, we chose this value in equation 4-1 to be equal to 4 and finally the last parameter which is the minimum number of the pixels to form a cluster is selected to be equal to 10 pixels. All these parameters were selected heuristically by experimenting with various values and selecting those with the best visual outcome (This evaluation was done on a subset of images).

Hybrid clustering workflow: As discussed earlier, we used two clustering techniques (K-means clustering with K detected centroids as the initial cluster centers and density-based clustering) for segmentation purposes.

We first apply these two clustering methods and use their segmentation agreement as our initial segmentation result (in other words, those pixels that obtained the same labels from these two clustering methods, are considered as correctly clustered pixels.)

In the next step, we use density-based clustering to make the final decision on the remaining pixels (that are not selected by both clustering methods from the previous step). We use the size of the resulted cluster from this step as the factor for making the decision. If the cluster obtained from this step is large enough (>10 pixels) we consider it as a nucleus otherwise we will discard it to avoid debris in our final results. An overview of the whole clustering step is depicted in Figure 4-6.

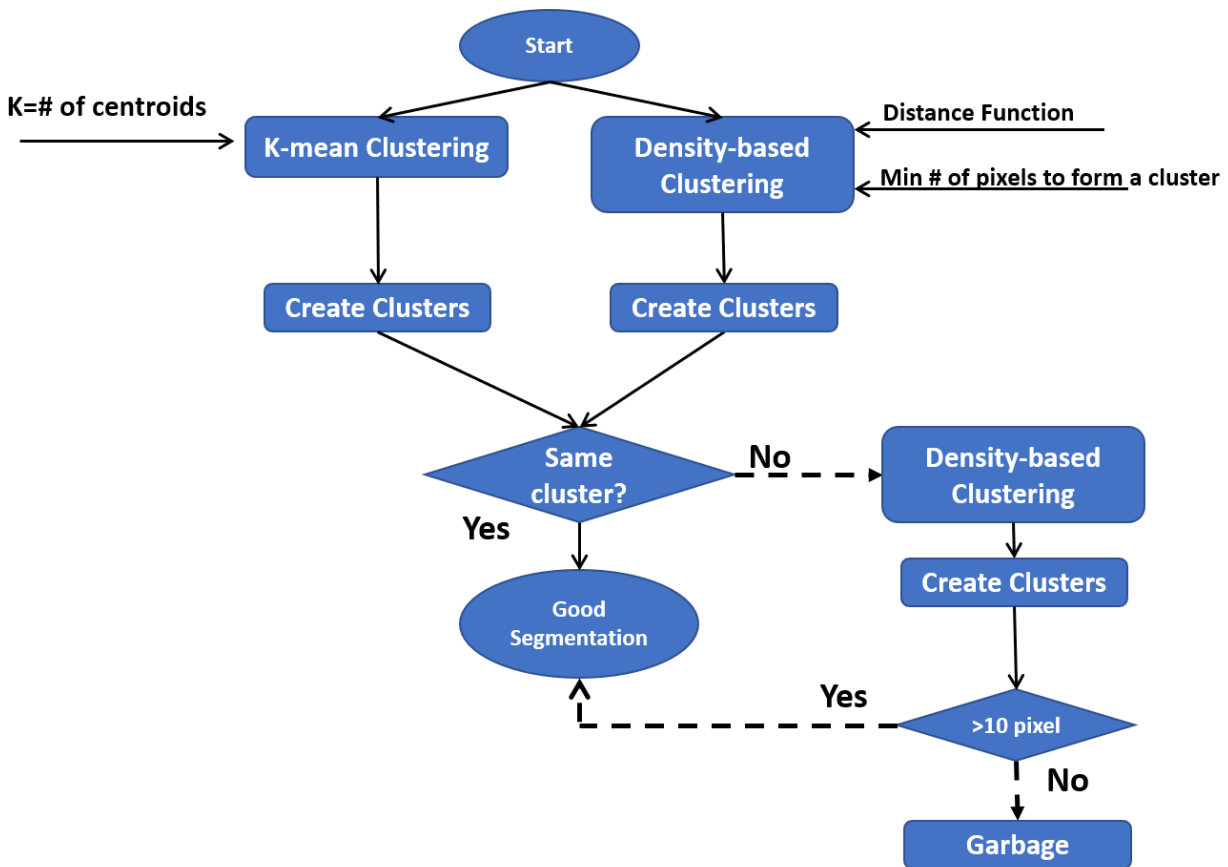


Figure 4-6 Clustering flowchart, shows the steps necessary to select relevant clusters. This algorithm consists of two clustering methods (K-means clustering and density-based clustering) in the first run, pixels are selected if they belong to the same cluster (generated from two clustering methods), on the second run we feed the remaining pixels to the density-based clustering and choose those with large area.

Figure 4-7 shows the result of comparing different clustering algorithms. In this figure, we show one nuclei cluster cropped from the original image as an example.

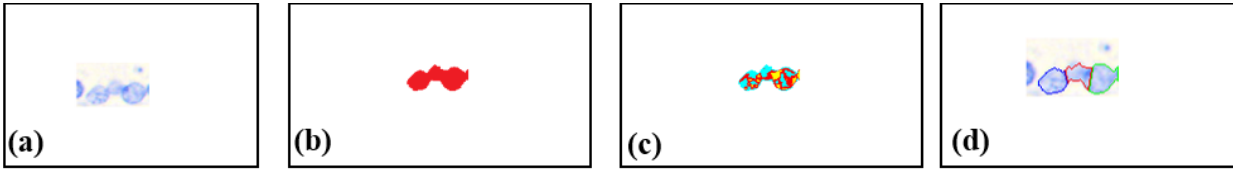


Figure 4-7 Comparison between clustering methods (Density-based clustering, K-means clustering and hybrid method) a) input image. b) clustering result of Density-based clustering. c) clustering result of K-means clustering. d) results of our hybrid clustering method.

In Table 4-1 we summarized the whole algorithm design that we introduced in this chapter.

Table 4-1 Interactive nuclei segmentation algorithm steps

Interactive Nuclei Segmentation Algorithm
Input data: Scanned RGB image Output data: Segmented nuclei
Algorithm: <ol style="list-style-type: none"> 1. Apply automated segmentation method (presented in the previous chapter) 2. Ask the user to evaluate the outcome (nuclei centers) 3. Based on the outcome, the user should make a decision (add/remove nuclei or do nothing) <ul style="list-style-type: none"> if the user removed a center from a cluster <ol style="list-style-type: none"> a. Find the labeled objects that the removed center belonged to b. Assign pixels of the removed nucleus to their nearest neighbor object in the cluster else if the user added a center <ul style="list-style-type: none"> Dilate selected point with a disk with a radius of 50 as the ROI Extract edges in the ROI and dilate them with square of width 3 Use dilated edges to create a new ROI mask and apply Otsu to the pixels within this. (Use Otsu results to binarize the initial ROI) Find the connected components that contain nucleus marked by the user and automatically detect nuclei center and use them as seed points for region growing algorithm Apply the hybrid clustering method on the region results of above step to find cell Use the final mask to extract the nucleus

4.5 Results

We applied our localization method presented in the previous chapter on Thionin stained prostate cancer TMA dataset from 33 patients.

Our dataset contains more than 170,000 nuclei. We compared our segmentation results with the results obtained from the human annotator (which here is considered as our gold standard). We obtained about 12% false-positive cell segmentation and about 3.4% missed cell detection. To improve the segmentation accuracy, we used the interactive algorithm presented in this chapter.

Direction to use the algorithm: the user loads the image and runs the original automated segmentation algorithm (presented in the previous chapter). Then the user uses our interactive method to add or remove nuclei centers. Adding and removing nuclei is a real-time process (means the process is very fast and the user can see the result right away). This interactive method can continuously show the results of any user interaction in a real-time fashion and gives the user the ability to observe the result of his/her actions. In general, human experts require 45 minutes of processing time to select nuclei of each core while our automated method combined with this interactive method takes about only 6 minutes, more importantly, our algorithm has the same accuracy as the manual process. This means our algorithm is fast, efficient, and accurate which makes this method suitable to be used in the clinical settings.

For illustration purposes, we selected an image, Figure 4-8, and ran our segmentation method with/without our interactive procedure, the result is presented in Figure 4-9 and Figure 4-10 respectively.

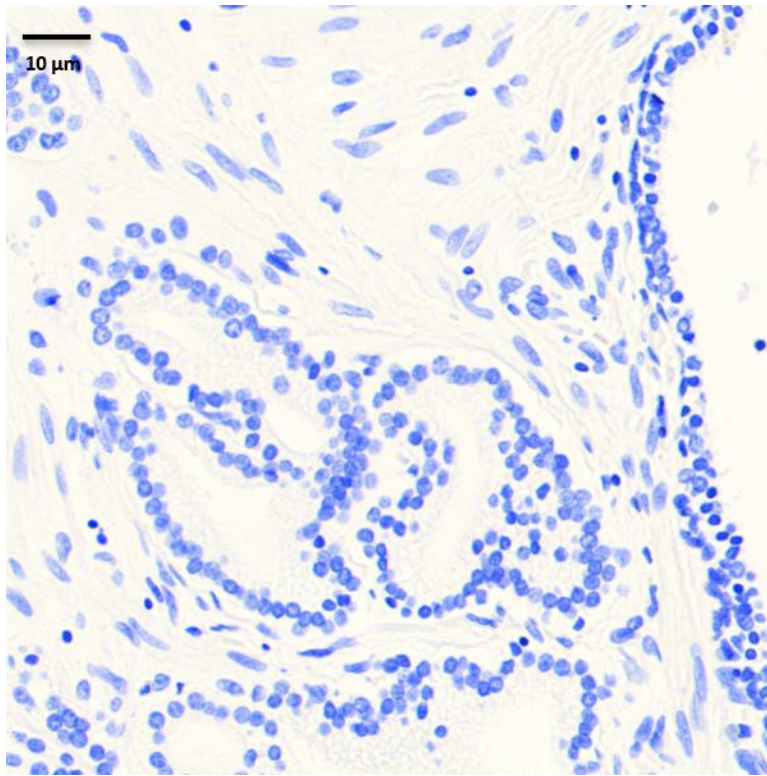


Figure 4-8 Section of Prostate TMA core, selected for comparing the segmentation results with/without our interactive method.

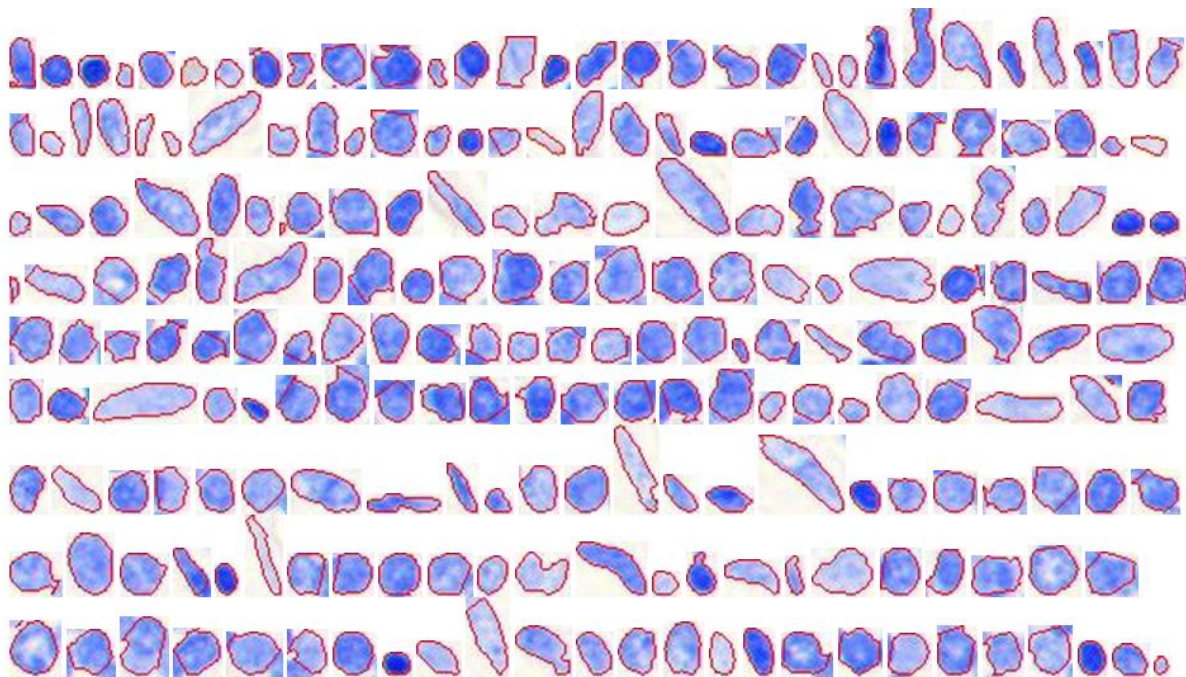


Figure 4-9 Automated segmentation result with our interactive method.

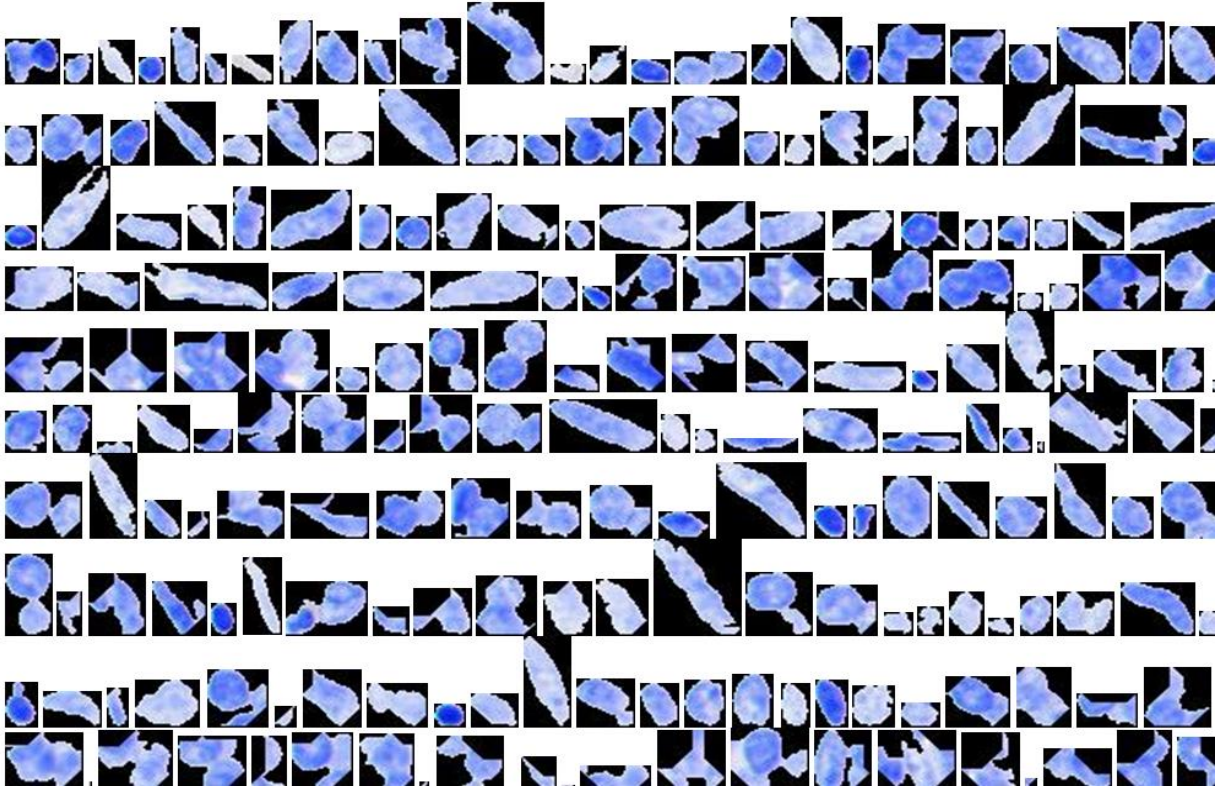


Figure 4-10 Automated segmentation result without our interactive method.

4.6 Discussion

4.6.1 Generalization to Other Cancer Sites and Comparison with ILASTIK

In this section, we discuss the result of comparing our method with an open-source segmentation software, referred to as ILASTIK (Sommer, Straehle, et al. 2011).

Although ILASTIK provides a user-friendly environment for interactive image segmentation, it is designed for expert-domain users (requires that the users have some image processing knowledge).

ILASTIK uses a random forest algorithm that learns from labels provided by the user, each pixel's neighborhood is characterized by a set of nonlinear features. Although ILASTIK is very powerful,

it requires an experienced user to continuously provide labels and refine/correct the labels, this drawback makes Ilastik very time-consuming, user-intensive and subject to inter-intra variability.

We compared our segmentation results with the result obtained from Ilastik on the same sample of Thionin stained prostate tissue. We tuned Ilastik parameters such as cell size and thresholds to obtain the best possible segmentation result. Figure 4-11, shows the outcome of this comparison.

We chose two labels (colors), red for the cells, and, and green for the background (Figure 4-11 (a) and Figure 4-11(b)). Part of the segmentation results from Ilastik is shown in Figure 4-11(c). Manual detection resolved 449 cells in the original image, our algorithm detected/segmented 450 cells (after two correction process), however, Ilastik detected only 206 objects (nuclei, cluster of nuclei). We can see in Figure 11(c) that Ilastik failed to segment cells within clusters. One of the reasons for this failure could be due to its algorithm implementation which uses local features (cues, brightness, color, and texture) and is not designed to capture spatial configurations (Sommer, Straehle, et al. 2011). On the other hand, our algorithm uses global features to successfully detect and segment overlapping nuclei. Another Ilastik drawback is the necessity for user input, the user has to manually draw lines to separate cells from the background. This process is very time consuming and is not practical in clinical use, where the user has to screen more than 10,000 cells per slide. Our model uses a fast-automated segmentation method to select cells and then the user has the chance to post-process incorrect/missed cells by removing/adding cell centers.

Prostate Tissue Stained with Thionin

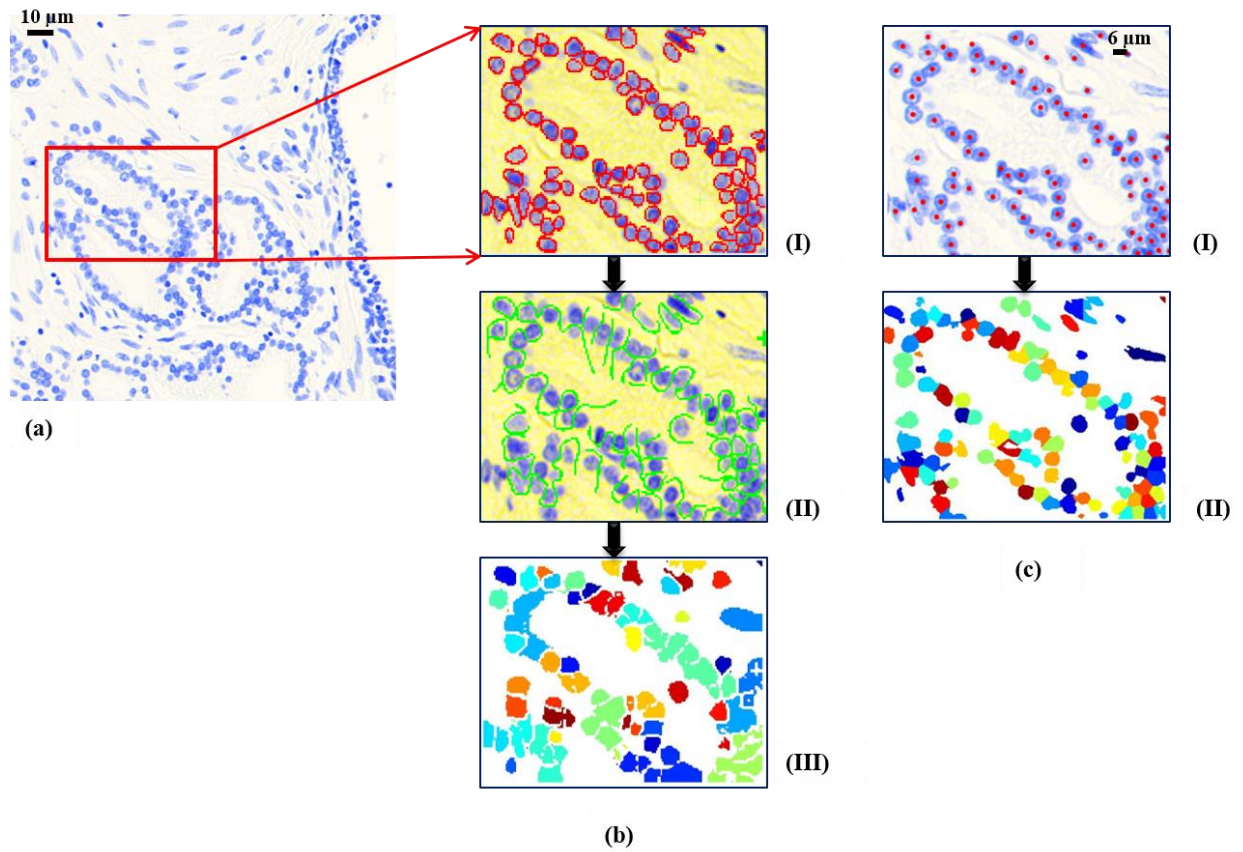


Figure 4-11 Comparison between Ilastik nuclei segmentation and our algorithm. (a) a selected image from prostate TMA which is Thionin stained (column b) Ilastik nuclei segmentation required steps (manually selection of first label, nuclei, (I) manual selection of background, stroma, (II) final results is presented in (III) each color in (III) represents an object) (column c) our algorithm nuclei segmentation steps (automated localization of nuclei with minor adjustments(I), final results presented in (II), again each color in (II) represents an object)

In this study, we tuned our algorithm for only Thionin stain. As explained before, one of the advantages of using Thionin stain for nuclei segmentation is that it binds in a stoichiometric fashion with DNA which results in high contrast images, which makes segmentation much easier.

Here, as a proof of concept, we showed and compared two methods performances (our algorithm and Ilastik). In this experiment, we didn't adjust for the new color since the purpose of this study is to use this algorithm specifically on Thionin stained tissues, however, we still want to investigate how our algorithm performs on other stains.

To investigate the performance of our algorithm on other stains, we decided to apply our method to prostate tissue stained with Hematoxylin and Eosin (H&E). In this experiment, we used the same dataset of TMA prostate core biopsies; however, in this experiment the TMA cores were stained with H&E. We did not change any of the algorithm parameters to adjust for the new stain. We randomly selected 10 samples from the whole dataset and only select $\frac{1}{4}$ of each image in our experiment. The human impression of correct segmentation is used as the gold standard for the comparison. The result is presented in Table 4-2.

The result of the comparison between our algorithm performance and Ilastik is shown in Figure 4-1. We again chose two labels (colors), red for the cells, and green for the background (Figure 4-12(b-I) and Figure 4-12(b-II)). By comparing Figure 4-12 (b-III) and (c-II) we can appreciate that our algorithm performed much better than Ilastik. Ilastik had poor performance for detecting cells in clusters. Although our algorithm was relatively better, it was not perfect since the method was not adjusted for the new stain.

Prostate Tissue Stained with H&E

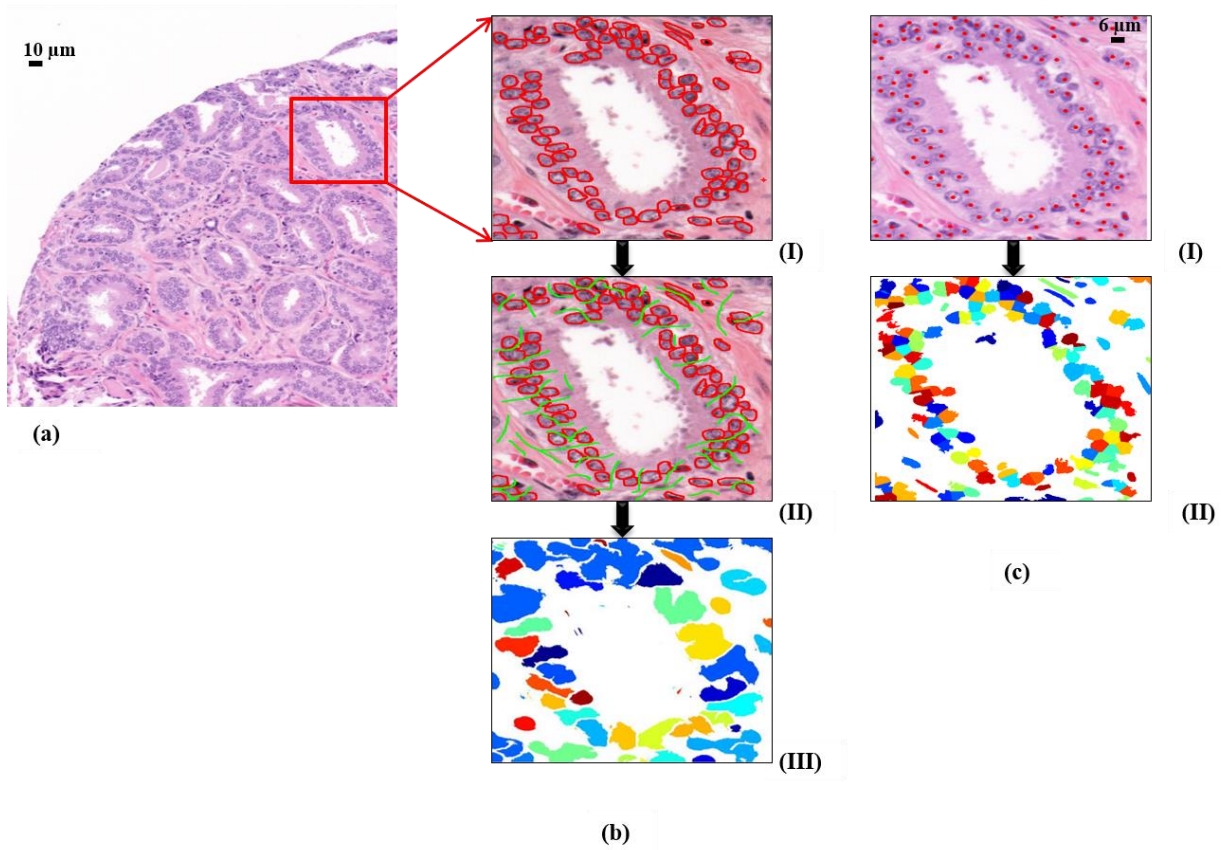


Figure 4-12 This figure shows the comparison between ILASTIK nuclei segmentation and our algorithm. (a) a selected image from prostate TMA, Thionin stained (column b) ILASTIK nuclei segmentation steps (manually selection of first label, nuclei, (I) manual selection of background, stroma, (II) and final results presented in (III) each color in (III) represents an object) (column c) our algorithm nuclei segmentation steps (automated localization of nuclei with minor adjustments(I), final results presented in (II), again each color in (II) represents an object)

Table 4-2 Prostate tissue (H&E) segmentation results for 10 randomly selected images.

PROSTATE TISSUE (H&E) NO.	FALSE NEGATIVE (%)	FALSE POSITIVE (%)	TOTAL OBJECTS
1	1.49	1.12	801
2	1.13	0.20	965
3	2.35	0	85
4	5.12	5.38	390
5	0.52	6.25	384
6	1.10	4.15	361
7	0	5.55	36
8	0.28	2.11	708
9	0.13	0.97	718
10	0	3.22	186
TOTAL	1.16	2.17	4634

We also compared our segmentation method with Ilastik on other tissue types. We explored oral, cervix and lung tissues, all tissues were stained with Thionin.

Oral Tissue

In Figure 4-13 we show the comparison between the performance of our segmentation algorithm and Ilastik on oral tissue. Figure 4-13-a is the original image, the segmentation results and a gallery of selected segmented cell are shown in Figure 4-13 (b) and Figure 4-13 (c) respectively. By comparing Figure 4-13 (b) and Figure 4-13 (e), we can see that our algorithm outperformed Ilastik. In this segmentation, we did not adjust any part of the code based on the size and color of oral tissue, however, to increase the performance of Ilastik we manually drew a boundary for each individual cell. As mentioned earlier one of the main drawbacks of using Ilastik is the

necessity of extensive user intercalation to select all the objects in an image. This process might require hours for large samples.

We compared our segmentation performance with the human impression of correct segmentation.

For the outcome of our segmentation on 10 randomly selected oral tissue please refer to Table 4-3.

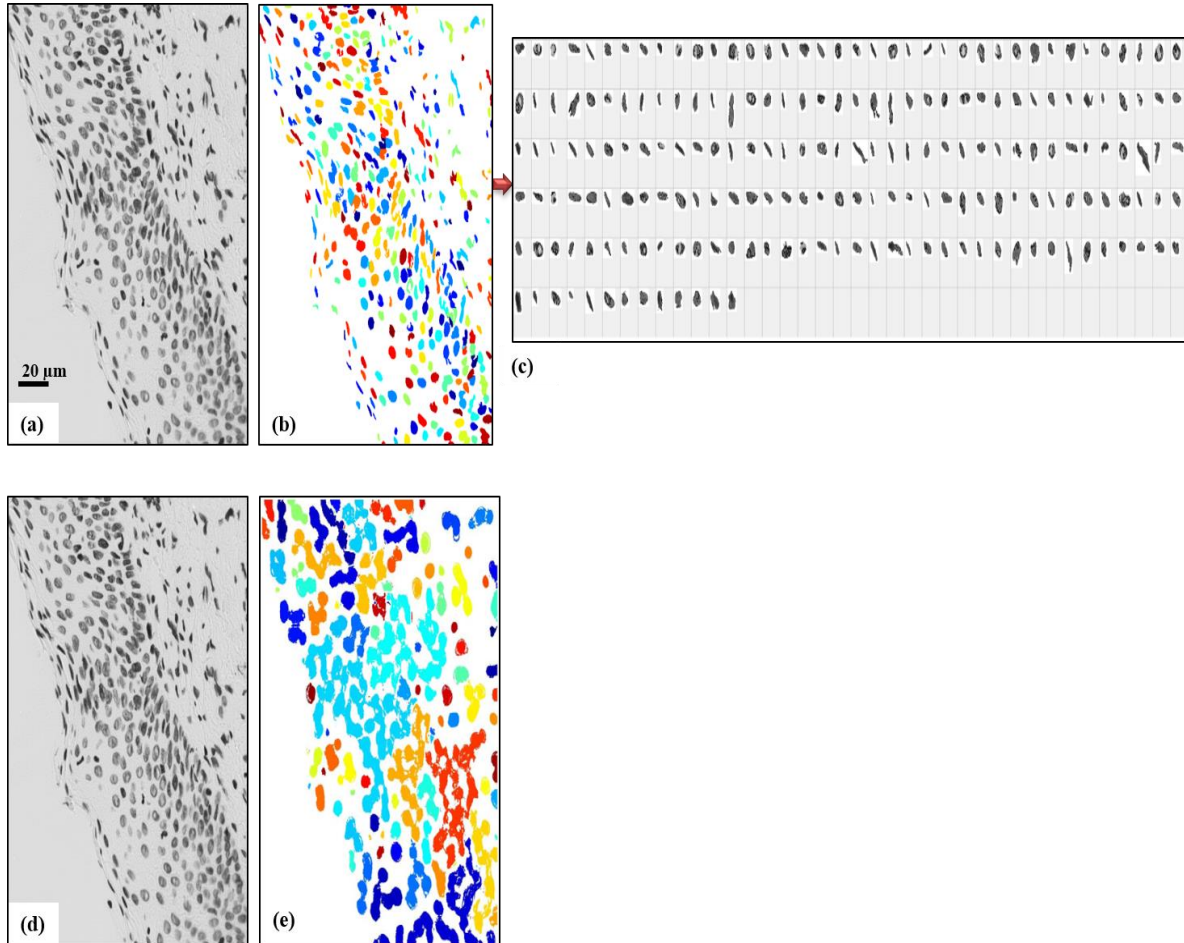


Figure 4-13 This figure shows the comparisons between ILASTIK nuclei segmentation and our algorithm performance on oral tissue. (a) a selected image from oral tissue, which is Thionin stained, (b) our segmentation results where each color represents an object. (c) gallery of selected nuclei from our segmentation method. (d) original oral image (the same

as (a)), and segmentation result from ILASTIK nuclei segmentation is presented in (e) where each color represents an object.

Table 4-3 Oral tissue (Thionin) segmentation results for 10 randomly selected images.

ORAL TISSUE NO.	FALSE NEGATIVE (%)	FALSE POSITIVE (%)	TOTAL OBJECTS
1	3.69	1.58	379
2	3.32	1.10	271
3	5.74	1.20	331
4	9.21	5.26	152
5	8.28	3.18	157
6	0	7.40	54
7	0.81	4.32	370
8	0.14	1.40	711
9	0.35	7.16	279
10	4.50	1.80	111
TOTAL	2.80	2.7	2815

Cervical Tissue

Figure 4-14 shows the results of our method compared with ILASTIK for the segmentation of the epithelial layer of Thionin stained cervix tissue. Similar to oral tissue, we visually investigated the performance of our method on 10 randomly selected cervix tissues. The segmentation result of our algorithm on 10 randomly selected cervix tissues is presented in Table 4-4.

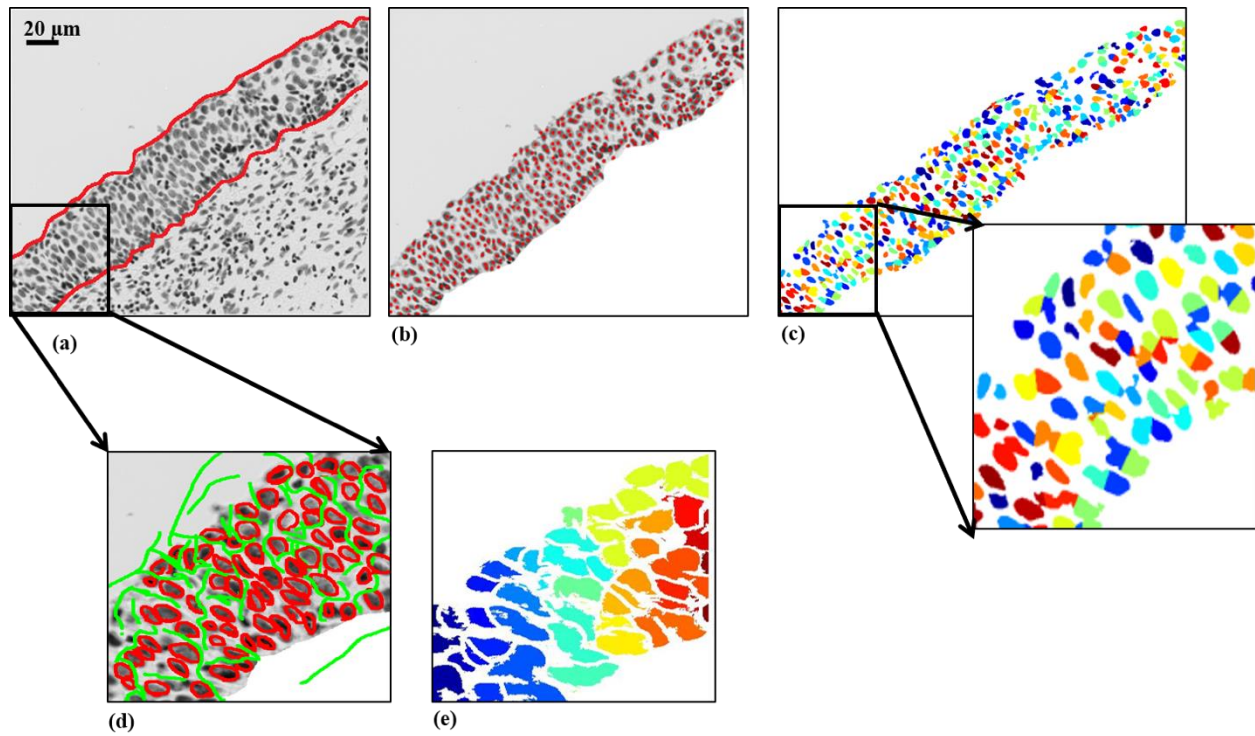


Figure 4-14 Comparison between Ilastik nuclei segmentation and our algorithm for cervix tissue. (a) a selected image from cervix tissue, Thionin stain, we manually select the epithelial layer for segmentation purposes. (b) Automated nuclei localization with our algorithm, (c) the segmentation results from our algorithm where each color represents an individual object. (d) the first and second steps of Ilastik segmentation, manual selections of two labels (red for cells and green for background/stroma). (e) segmentation results from Ilastik where each color represents an object.

Table 4-4 Cervix tissue (Thionin) segmentation results for 10 randomly selected images.

CERVIX TISSUE NO.	FALSE NEGATIVE (%)	FALSE POSITIVE (%)	TOTAL OBJECTS
1	0.71	0.28	2106
2	0.74	0.22	1346
3	0.26	0.11	3399
4	0.53	0.42	1861
5	0.28	0.28	1764
6	0.84	0.42	944
7	1.56	3.13	510
8	3.60	1.71	582
9	2.36	2.48	805
10	1.24	0.24	803
TOTAL	0.81	0.55	14120

Lung tissue

We also experimented on a sample dataset of lung tissue, stained with Thionin. We compared our algorithm performance with ILASTIK segmentation performance. For the result of this comparison please refer to Figure 4-15. Again, we used the standard human impression of correct segmentation as our gold standard. For the outcome of our segmentation result on 10 randomly selected lung tissue samples, please refer to Table 4-5.

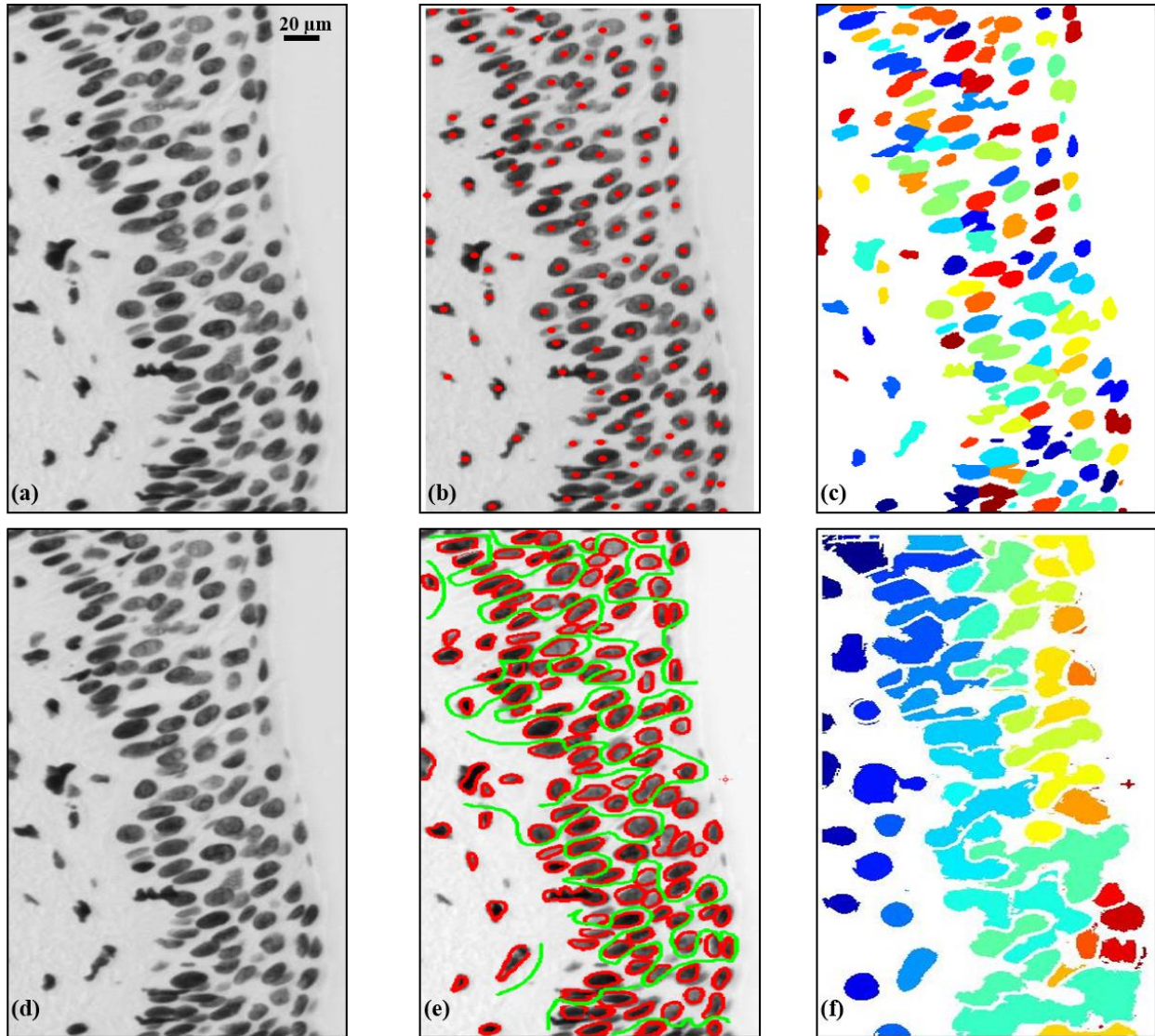


Figure 4-15 Comparison between Ilastik nuclei segmentation and our algorithm for lung tissue. (a) a selected image from lung tissue, Thionin stain, (b) Automated nuclei localization with our algorithm, (c) the segmentation results from our algorithm where each color represents an individual object. (d) original image presented in (a). (e) the first and second steps of Ilastik segmentation, manual selections of two labels (red for cells and green for background/stroma). (f) segmentation results from Ilastik where each color represents an object.

Table 4-5 Lung tissue (Thionin) segmentation results for 10 randomly selected images.

LUNG TISSUE NO.	FALSE NEGATIVE (%)	FALSE POSITIVE (%)	TOTAL OBJECTS
1	0.04	0.04	2106
2	1.04	0.14	1346
3	0.50	0.05	3399
4	0.48	0.21	1861
5	0.28	0.17	1764
6	2.11	0.10	944
7	0.98	0.58	510
8	0.51	0	582
9	0.86	0.37	805
10	1.61	0	803
TOTAL	0.66	0.13	14120

4.6.2 Reproducibility Analysis

To investigate the intra- and inter-observer variability of our method, we asked a user to select nuclei centers of H&E stained prostate tissue (blue dots in Figure 4-16). To simulate another observer's possible selections, we moved the originally selected cell center (blue dot) by ± 5 pixels in four directions (top, down, left and right) and then we ran our algorithm for all five different cases. After the above process was completed we calculated relevant statistics/features (such as area, perimeter, major and minor axes) from the segmented objects.

We then calculated the difference between the center features and the displaced feature and normalized it by the value of the center feature, equation (4-2).

$$f_{normalized} = \frac{f_{displaced} - f_{center}}{f_{center}} \quad (4-2)$$

where f stands for calculated features depending on the subscript either from the center or the displaced center.

The result of this experiment is shown in Figure 4-17. As we can see in this figure the observed difference is very small (approximately less than 0.01).

With this experiment, we demonstrated that our algorithm is robust and reproducible with respect to the selection of initial points.

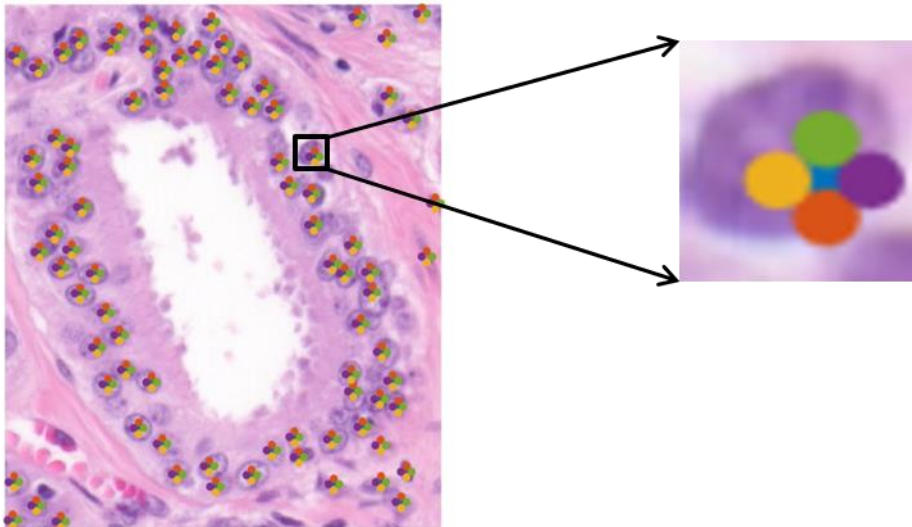


Figure 4-16 Reproducibility analysis (segmentation results and the dependency of the algorithm to user input/selection). The blue dot represents the original location selected by the segmentation algorithm and other dots are potential locations that the user could possibly select.

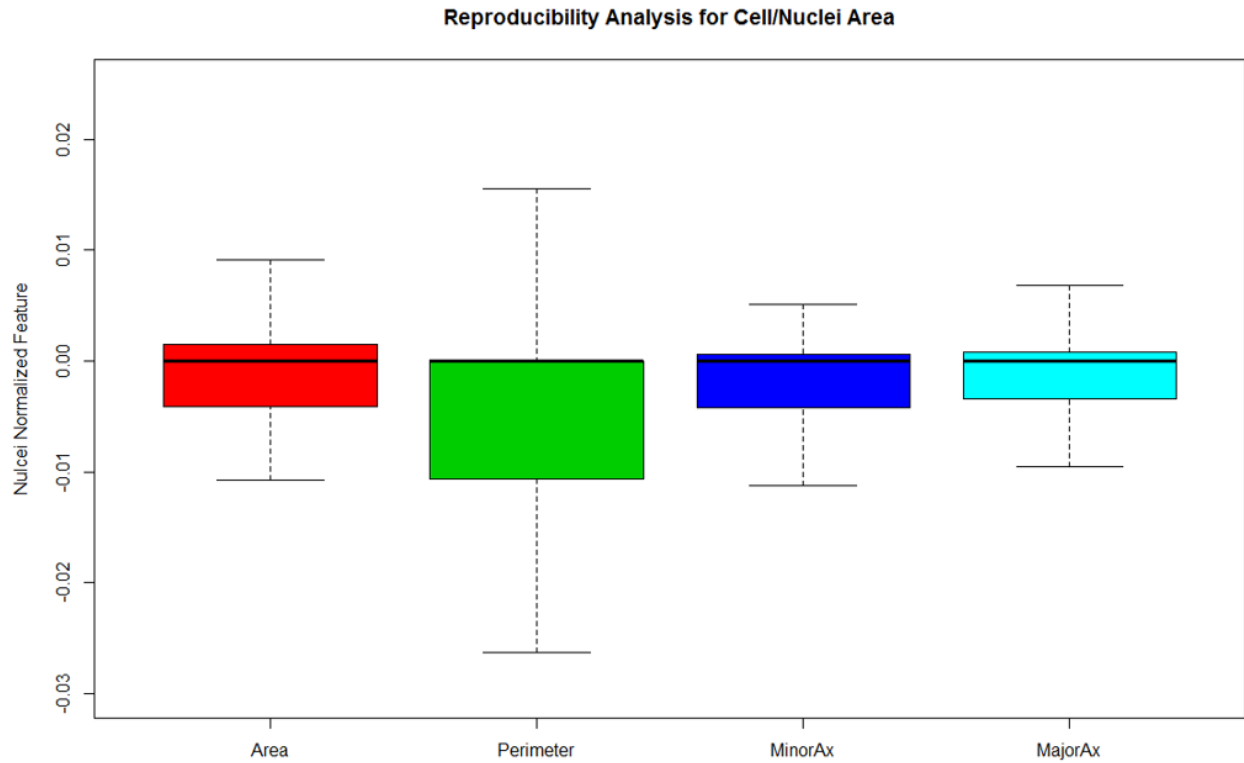


Figure 4-17 Reproducibility analysis of five cores. Statistic results on segmented cells obtained from cells/nuclei features.

As discussed earlier, the analyses on the other tissue sites and other colors were on a very small dataset. It would be interesting to expand this work and continue the experiment on a larger sample size. Another limitation of this study is the dependency to criteria-based algorithms such as K-means clustering and density-based clustering.

4.7 Conclusion

In this chapter, we presented an interactive segmentation method to improve the extraction of nuclei from the image of prostate TMA cores. The nuclei were stained with Feulgen-Thionin, and the cores were selected from 33 patients.

Furthermore, we investigated the performance of our method on other tissue types (oral, cervix and lung) and other stains (H&E) as a proof of concept (the datasets for these experiments were very limited).

Although we tuned our methods for prostate cancer cells, we showed that it is possible to use this method to segment other cell types. This generalization capability made our interactive method an essential core for developing an automated cell segmentation tool for different tissue types.

Chapter 5 Design of Machine Learning Methods as a Predictive Tool for Prostate Cancer

5.1 Introduction

Classical Machine learning for automated image processing is a topic of great interest, especially in medical imaging. The majority of these analyses are based on classical supervised learning where domain knowledge plays a critical role in the design and handcrafted features. In cancer research, usually, the diagnostic hallmarks are defined and selected by pathologists/clinicians. Scientist incorporates these features with other statistical information to create a comprehensive or sometimes rather comprehensive feature set.

The difficulty of classification arises from the fact that the vast majority of relevant information such as underlying class probabilities, class-conditional probabilities, relevant features, and etc. are unknown. Therefore, the introduction of exhaustive feature space, although not very desirable and computationally expensive, is inevitable. This might result in redundant and irrelevant features that later with proper feature reduction methods or a smart classifier design could be addressed.

The focus of this chapter is to select and extract clinically relevant features from segmented nuclei and use these features for prostate cancer grading.

5.2 Hypothesis

It is our hypothesis that using nuclei features from segmented cells (as explained in the previous chapters) and a sophisticated classification method (as we introduce later in this chapter) makes it possible to classify prostate cancer grades. To test our hypothesis, we selected cell-level features

from prostate cancer TMA cores and applied a modified SVM algorithm to classify prostate cancer grade.

5.3 Objective

In the previous chapters, we discussed the importance of the segmentation step for designing a fully automated method for prostate cancer classification. The next step is to use these well-segmented cells and extract relevant information from them. This information is called features. Later we use these features as the input of our classifier. Our objective for this chapter is to introduce the whole orchestration of feature selection and design/implementation of classifiers.

5.3.1 Specific Objective

- III) Selection and implementation of relevant features.
- IV) Development of a machine learning technique to use the above feature to classify prostate cancer grades.

The workflow for this chapter is presented in Figure 5-1. The first step as always is imaging the tissue, followed by image/ label preparation. The next step is feature selection and the design of the ML technique for classification and finally, parameter optimization to improve the performance.

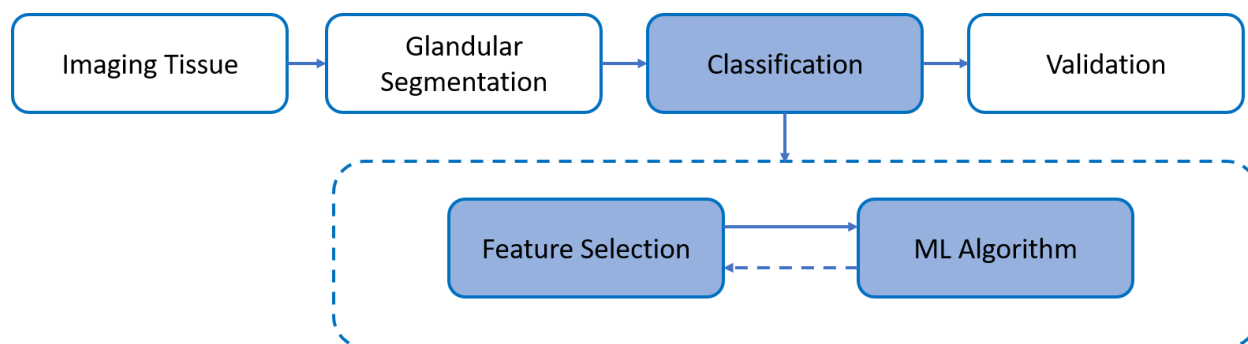


Figure 5-1 This figure shows the pipeline of our study. The focus of this chapter is on the classification step.

5.4 Materials and Instrumentation

5.4.1 Clinical data

A Tissue Microarray (TMA) was obtained from the University of British Columbia and the Vancouver Prostate Center. The surgical specimen tissue cores were obtained from 32 prostate cancer patients treated with brachytherapy with a median follow-up of 8.9 years. The specimens were placed into a TMA and a section from the TMA stained with Thionin. The patients were divided into two groups based on whether they relapsed (Failure group) or not (Non-failures group), Table 5-1.

Table 5-1 Number of patients with different Gleason scores in failure and no- failure groups

GLEASON SCORE	FAILURE (NO. PATIENTS)	NON-FAILURES(NO. PATIENTS)
6	0	8
7	9	7
8	4	1
9	3	0

In the failure group, the Gleason score was seven in nine patients, eight in four patients, and nine in three patients. In the non-failures group, the Gleason score was six in eight patients, seven in seven patients, and eight in one patient.

The slides were scanned digitally with the whole slide Panoramic MIDI scanner (3DHISTECH) using a 20x magnification and 0.8 NA objective.

We used our segmentation algorithm as explained in the earlier chapter to construct a gallery of cells that will be used to select features from them, Figure 5-2.

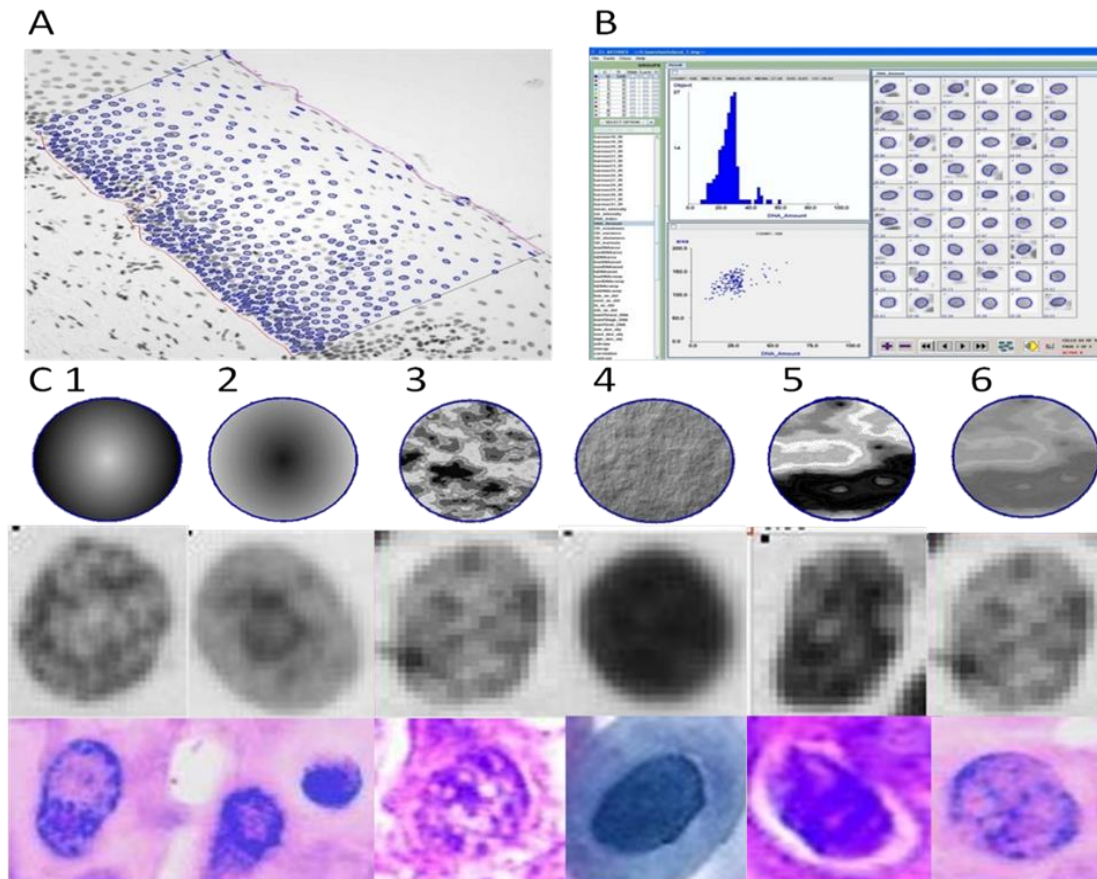


Figure 5-2 This figure shows the steps required from segmentation, (A) shows a part of an image used for segmentation (B) shows segmentation results and some statistics (area of cells) on the left panel and (C) shows some relevant features, such as area, nucleus density, etc.

5.5 Methodology

5.5.1 Feature Selection

5.5.1.1 Feature Selection-Size and Other Related Properties

Here we present the details regarding the features that we selected for our classification purposes.

The first feature set is related to the selected object size.

- **Area** is defined as the total number of pixels in an object (MacAulay, Keyes, et al. 2017).

$$Area = \sum_n \sum_m Img_{mask}(n, m) \quad (5,1)$$

where $Img_{mask}(n, m)$ is the mask of nuclei (the object).

- **Perimeter** connected boundary of an object, defined as below (MacAulay, Keyes, et al. 2017)

$$Perimeter = \sum_n \sqrt{(x(n+1) - x(n))^2 + (y(n+1) - y(n))^2} \quad (5,2)$$

- **Compactness** is a measure of the roundness of an object

$$Compactness = \frac{4 \cdot \pi \cdot Area}{Perimeter^2} \quad (5,3)$$

where x and y are the x-coordinate and y-coordinate of the N boundary pixels.

- **Radii** distance between the boundary pixels and the object centroid (MacAulay, Keyes, et al. 2017).

$$Radii = \sqrt{((x - x_0)^2 + (y - y_0)^2)} \quad (5,4)$$

x_0 and y_0 are the x and y centroids of the object as defined below:

$$x_0 = \frac{\sum_n \sum_m n \cdot Img_{mask}(n, m)}{Area} \quad (5,5)$$

$$y_0 = \frac{\sum_n \sum_m m \cdot Img_{mask}(n, m)}{Area}$$

- **Sphericity** is a measure of the largest circle fully inscribed in the object to that of the smallest circle which fully circumscribes the object.

$$Sphericity = \frac{\min(Radii)}{\max(Radii)} \quad (5,6)$$

- **Inertial Shape** similar to compactness is another measure of roundness:

$$InertialShape = \frac{2\pi \sum_n \sum_m Radii(n, m)^2}{Area^2} \quad (5,7)$$

- **Center of Mass** is defined as below (MacAulay, Keyes, et al. 2017) :

$$CenterMass = \frac{\sqrt{(x_{CM}-x_0)^2 + (y_{CM}-y_0)^2}}{2m_{xy}} \quad (5,8)$$

where

$$x_{CM} = \frac{1}{N} \sum_n \sum_m n \cdot Img_{gary}(n, m) \quad (5,9)$$

$$y_{CM} = \frac{1}{N} \sum_n \sum_m m \cdot Img_{gary}(n, m)$$

- **Equivalence Diameter** is defined below:

$$Equiv = 2 \sqrt{\frac{\pi}{Area}} \quad (5,10)$$

- **Major Axis Length** and **Minor Axis Length** is defined as below (MacAulay, Keyes, et al. 2017).

$$Maj_Axis_Len = 2\sqrt{2} \cdot \sqrt{m_{xx} + m_{yy} + \sqrt{(m_{xx} - m_{yy})^2 + 4m_{xy}^2}} \quad (5,11)$$

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

where

$$m_{xx} = \frac{1}{M} \sum_{i=1}^M (x_i - x_o)^2 + \frac{1}{12} \quad (5,12)$$

$$m_{yy} = \frac{1}{M} \sum_{i=1}^M (y_i - y_o)^2 + \frac{1}{12}$$

$$m_{xy} = \frac{1}{M} \sum_{i=1}^M (x_i - x_o)(y_i - y_o)$$

where M is the total number of pixels in an object.

- **Eccentricity** measures the deviation from the circular (a circle has an eccentricity of zero and a line has an eccentricity of one) (*Eccentricity (mathematics)*. 2019).

$$Eccentricity = \frac{2\sqrt{\frac{Maj_Axis_Len^2}{4} - \frac{Min_Axis_Len^2}{4}}}{Min_Axis_Len} \quad (5,13)$$

Orientation is defined as the angle between the major axis and the x-axis of the image (*System and method for automatically detecting malignant cells and cells having malignancy-associated changes.*).

$$\theta_o = \begin{cases} \tan^{-1} \frac{m_{xx} - m_{yy} + \sqrt{(m_{xx} - m_{yy})^2 + 4m_{xy}^2}}{2m_{xy}}, & m_{yy} < m_{xx} \\ \tan^{-1} \frac{2m_{xy}}{m_{xx} - m_{yy} + \sqrt{(m_{xx} - m_{yy})^2 + 4m_{xy}^2}}, & else \end{cases} \quad (5,14)$$

Elliptical Deviation which is a measure of deviation from ellipse and considers the double-side Minus operation (considers pixels that are presented in the image and not present in the ellipse and vice versa), the result is also normalized by the area (Bengtsson, Rodenacker 2003).

$$x_0 = \frac{\sum_{(n,m) \in Img \& \nexists Ellipse} Img(n, m) + \sum_{(n,m) \in Ellipse \& \nexists Img} Ellipse(n, m)}{Area} \quad (5,15)$$

where Ellipse (n, m) is the ellipse mask.

- **Convex Area** which is a measure of the convex hall of the object mask (Bengtsson, Rodenacker 2003).

$$ConvArea = \sum_n \sum_m Convex(Img_{mask}(n, m)) \quad (5,16)$$

- **Convex Deficiency** as defined below calculates the normalized difference between the convex hull area and the actual area (Bengtsson, Rodenacker 2003).

$$ConvexDefi = \frac{ConvArea - Area}{Area} \quad (5,17)$$

- **Solidity** as defined below (Bengtsson, Rodenacker 2003) :

$$Solidity = \frac{Area}{ConvArea} \quad (5,18)$$

- **Filled Area** which considers only the areas of the image that are filled and disregard the holes (not-filled areas).

$$FilledArea = \sum_n \sum_m fill(Img_{mask}(n, m)) \quad (5,19)$$

- **Euler Number** which calculates the difference between the total number of regions vs total number of holes (*Euler's number (mathematical constant)*. 2019).

$$EulerNumber = N_{regions} - N_{holes} \quad (5,20)$$

- **Extent** which is defined as below:

$$Extent = \frac{Area}{W \cdot H} \quad (5,21)$$

where W and H are width and height respectively.

- **Aspect Ratio** which is defined as below:

$$AspectRatio = \frac{H}{W} \quad (5,22)$$

where W and H are width and height respectively.

- **Perimeter FFT**: uses the Fast Fourier Transform (FFT) to measure energies at various frequencies to characterize the variations in the object boundary (MacAulay, Keyes, et al. 2017).

$$Perimeter\ FFT = \sum_{n=0}^{N-1} x(n) \frac{-j2\pi nk}{N}, k = 0:N-1 \quad (5,23)$$

We used $N = 1024$ and if $x(n)$ is smaller than N , it is zero-padded.

- **Perimeter FFT boundaries** are defined as below (MacAulay, Keyes, et al. 2017) :

$$PerimeterFFTLow = \sum_{n=0}^{340} |PerimeterFFT(n)|^2 \quad (5,24)$$

- **Perimeter FFT Medium** describes as below:

$$PerimeterFFTMed = \sum_{n=0}^{683} |PerimeterFFT(n)|^2 \quad (5,25)$$

- **Perimeter FFT High** describes as below:

$$PerimeterFFTHi = \sum_{n=0}^{1023} |PerimeterFFT(n)|^2 \quad (5,26)$$

- **Perimeter FFT high** describes as below:

$$PerimeterFFTh = \sum_{n=32h}^{32.(h+1)-1} |PerimeterFFT(n)|^2, h = 0:32 \quad (5,27)$$

- **Perimeter Curvature** is defined as below (Bengtsson, Rodenacker 2003) :

$$\theta_{curve} = \tan^{-1} \frac{y(n+1) - y(n)}{x(n+1) - x(n)}, n = 1:N \quad (5,28)$$

- **Bending Energy** is measured as the sum of the difference between the curvature of two boundary points, as defined below (Bengtsson, Rodenacker 2003) :

$$BendingEnergy = \sum_{n=1}^N (\theta_{curve}(n+1) - \theta_{curve}(n)), n = 0:N \quad (5,29)$$

5.5.1.2 Feature Selection - Size and Other Related Properties

These features are calculated using the Gray Level Co-occurrence Matrix (GLCM) which calculates the occurrence of the specific grey value in a specific location (pairs of value and spatial location) as defined below.

- **GLCM Inertia** is defined as below (Bengtsson, Rodenacker 2003, System, and method for automatically detecting malignant cells and cells having malignancy-associated changes.):

$$GLCMIntertia = \sum_n \sum_m (n - m)^2 GLCM(n, m) \quad (5,30)$$

- **GLCM Energy** is defined as below (Bengtsson, Rodenacker 2003)

$$GLCMEnergy = \sum_n \sum_m GLCM^2(n, m) \quad (5,31)$$

- **GLCM Entropy** is defined as below (Bengtsson, Rodenacker 2003):

$$GLCMEntropy = \sum_n \sum_m GLCM(n, m) \log_2 (GLCM(n, m)) \quad (5,32)$$

- **GLCM Homogeneity** is defined as below (Bengtsson, Rodenacker 2003):

$$GLCMHomogeneity = \sum_n \sum_m \frac{GLCM(n, m)}{1 + (n - m)^2} \quad (5,33)$$

- **GLCM Max Prob** is defined as below (Weyn, van de Wouwer et al. 1998) :

$$GLCMMaxProx = \max_{n, m} GLCM(n, m)$$

- **GLCM Cluster Shade** is defined as below (Bengtsson, Rodenacker 2003):

$$GLCMClusterShade = \sum_n \sum_m (k - M_x + m - M_y)^3 GLCM(n, m) \quad (5,34)$$

where

$$GLCMCorrelation = \frac{\sum_n \sum_m (n - M_x)(m - M_y) GLCM(n, m)}{S_x \cdot S_y} \quad (5,35)$$

$$M_x = \sum_n \sum_m n GLCM(n, m)$$

$$M_y = \sum_n \sum_m m GLCM(n, m)$$

- **GLCM Correlation** is defined as below (Bengtsson, Rodenacker 2003):

where

$$S_x = \sqrt{\sum_n (n - M_x)^2 GLCM(n, m)}$$

$$S_y = \sqrt{\sum_m (m - M_y)^2 GLCM(n, m)}$$

Here we discuss properties related to the optical properties of the selected object.

- **Image Band** which measures image bands for each object as defined below:

$$ImageBand = [A(n, m, 1), \dots, A(n, m, B)], B = \text{number of bands} \quad (5,36)$$

- **Intensity** is defined as below (Bengtsson, Rodenacker 2003):

$$Intensity(k) = \frac{1}{B} \sum_{b=1}^B ImageBands(k, B) \quad (5,37)$$

- **Values of Optical Densities** is calculated as below (Bengtsson, Rodenacker 2003):

$$OpticalDensities = [OD_1, OD_2, \dots, OD_B, OD_I] \quad (5,38)$$

where

$$OD_b = -\log_{10} \left(\frac{ImageBands(k, b)}{A_{mx}} \right), \quad k = 1:Area, b = 1:B \quad (5,39)$$

$$OD_I = -\log_{10} \left(\frac{Intensity(k)}{A_{mx}} \right), \quad k = 1:Area \quad (5,40)$$

where A_{mx} is the maximum transmission value for a blank slide, which is 255 for an 8-bit image.

- **Integrated Optical Density** is defined as below (Bengtsson, Rodenacker 2003):

$$IntegratedOpticalDensities(b) \quad (5,41)$$

$$= \sum_{k=1}^{Area} Optical\ Densities(k, b), b = 1:B + 1$$

MOD or the integrated optical density is defined as below (MacAulay, Keyes, et al. 2017)

$$MOD = \frac{IntegratedOpticalDensities(b)}{Area} \quad (5,42)$$

$$Hue = \begin{cases} \theta(k), & B < G \\ 2\pi - \theta(k), & else \end{cases} \quad (5,43)$$

$$\theta = \arccos\left(\frac{1}{2} \cdot \frac{2R - G - B}{\sqrt{(R - G)^2 + (R - B)(G - B)}}\right) \quad (5,44)$$

5.5.1.3 Feature Selection-Texture Features

These features are calculated using the gray level run length (GLRL) matrix given θ .

- **Run Length** features are defined as below (MacAulay, Keyes, et al. 2017)
 - **Short Run Emphasis** is defined as below:

$$ShortRunEmphasis = \frac{1}{S} \cdot \sum_m \frac{r(m|\theta)}{m^2} \quad (5,45)$$

where $\sum_i \sum_j G(i, j|\theta) = \sum_i g(i|\theta) = \sum_j r(j|\theta)$, $g(i|\theta) = \sum_j G(i, j|\theta)$,
 $r(j|\theta) = \sum_i G(i, j|\theta)$,

- **Long Run Emphasis** is defined as below:

$$LongRunsEmphasis = \frac{1}{S} \cdot \sum_m r(m|\theta) \cdot m^2 \quad (5,46)$$

- **Gray Level Non-uniformity** is defined as below:

$$grayLevelNonUniformity = \frac{1}{S} \cdot \sum_n g(n|\theta)^2 \quad (5,47)$$

- **Run Length Non-uniformity** is defined as below:

$$RunLength nonUniformity = \frac{1}{S} \cdot \sum_m r(n|\theta)^2 \quad (5,48)$$

- **Run Length Percentage** is defined as below:

$$RunsPercentage = \frac{1}{Area} \cdot \sum_m r(m|\theta) \quad (5,49)$$

- **Low Gray Level Run Emphasis** is defined as below:

$$LowGrayLevelRunsEmphasis = \frac{1}{S} \cdot \sum_n \frac{g(n|\theta)}{n^2} \quad (5,50)$$

High Gray Level Run Emphasis is defined as below:

$$HighGrayLevelRunsEmphasis = \frac{1}{S} \cdot \sum_n g(n|\theta) \cdot n^2 \quad (5,51)$$

5.5.1.4 Feature Selection-Chromatin Information

Here we presented features that are related to the chromatin within nuclei of the cell.

- **Area:** this measure is defined earlier, here we calculate this value for different regions.
 - **Low DNA Area** is defined as below (MacAulay, Keyes, et al. 2017) :

$$LowDNAArea = \frac{\sum_n \sum_m OD(n, m)_L}{Area} \quad (5,52)$$

- **Medium DNA Area** is defined as below (MacAulay, Keyes, et al. 2017) :

$$MedDNAArea = \frac{\sum_n \sum_m OD(n, m)_M}{Area} \quad (5,53)$$

- **High DNA Area** is defined as below (MacAulay, Keyes, et al. 2017) :

$$HighDNAArea = \frac{\sum_n \sum_m OD(n, m)_H}{Area} \quad (5,54)$$

- **Low DNA Integral Optical Density** is defined as below (MacAulay, Keyes, et al. 2017) :

$$LowDNAIntOpticalDensity = \frac{\sum_n \sum_m OD(n, m) \cdot OD(n, m)_L}{IOD} \quad (5,55)$$

- **Medium DNA Integral Optical Density** is defined as below (MacAulay, Keyes, et al. 2017) :

$$MedNAIntOpticalDensity = \frac{\sum_n \sum_m OD(n, m) \cdot OD(n, m)_M}{IOD} \quad (5,56)$$

- **High DNA Integral Optical Density** is defined as below (MacAulay, Keyes, et al. 2017) :

$$HighDNAIntOpticalDensity = \frac{\sum_n \sum_m OD(n, m) \cdot OD(n, m)_H}{IOD} \quad (5,57)$$

- **Low DNA Mean Optical Density** is defined as below (MacAulay, Keyes, et al. 2017) :

$$LowDNAMeanOpticalDensity = \frac{\sum_n \sum_m OD(n, m) \cdot OD(n, m)_M}{\sum_n \sum_m OD(n, m) \cdot OD(n, m)_L} \quad (5,58)$$

- **Medium DNA Mean Optical Density** is defined as below (MacAulay, Keyes, et al. 2017)

$$MedDNAMeanOpticalDensity = \frac{\sum_n \sum_m OD(n, m) \cdot OD(n, m)_H}{\sum_n \sum_m OD(n, m) \cdot OD(n, m)_L} \quad (5,59)$$

- **High DNA Mean Optical Density** is defined as below (MacAulay, Keyes, et al. 2017) :

$$HighDNAMeanOpicalDensity = \frac{\sum_n \sum_m OD(n, m) \cdot OD(n, m)_{MH}}{\sum_n \sum_m OD(n, m) \cdot OD(n, m)_L} \quad (5,60)$$

- **Number of Regions** measures the number of discrete regions based on the defined connectivity, in this study we considered 4- and 8-connectivity.
- **Compactness**: this measure is defined earlier; here we calculate this value for different regions.

- **Low Compactness** is defined as below (MacAulay, Keyes, et al. 2017) :

$$LowCompactness = \frac{4\pi LowDNAArea \cdot Area}{P_L^2} \quad (5,61)$$

- **Medium Compactness** is defined as below (MacAulay, Keyes, et al. 2017) :

$$MedCompactness = \frac{4\pi MediumDNAArea \cdot Area}{P_M^2} \quad (5,62)$$

- **High Compactness** is defined as below (MacAulay, Keyes, et al. 2017) :

$$HighCompactness = \frac{4\pi HighDNAArea \cdot Area}{P_H^2} \quad (5,63)$$

- **Medium-High Compactness** is defined as below (MacAulay, Keyes, et al. 2017) :

$$MedHighCompactness = \frac{4\pi MediumHighDNAArea \cdot Area}{P_{MH}^2} \quad (5,64)$$

Where P_* is the perimeter of the region of interest.

- **Distance**: is calculated as the average distance between all pixels of the regions of interest and the centroid.

- **Low Distance** is defined as below (MacAulay, Keyes, et al. 2017) :

$$LowDistance = \frac{\sum_{(n,m) \in OD_L} \sqrt{(n - x_0)^2 + (m - y_0)^2}}{LDNAArea \cdot Area \cdot mean(Radii)} \quad (5,65)$$

- **Medium Distance** is defined as below (MacAulay, Keyes, et al. 2017) :

$$MedDistance = \frac{\sum_{(n,m) \in OD_M} \sqrt{(n - x_0)^2 + (m - y_0)^2}}{MDNAArea \cdot Area \cdot mean(Radii)} \quad (5,66)$$

- **High Distance** is defined as below (MacAulay, Keyes, et al. 2017) :

$$HighDistance = \frac{\sum_{(n,m) \in OD_H} \sqrt{(n - x_0)^2 + (m - y_0)^2}}{HDNAArea \cdot Area \cdot mean(Radii)} \quad (5,67)$$

- **Medium-High Distance** is defined as below (MacAulay, Keyes, et al. 2017) :

$$MedHighDistance = \frac{\sum_{(n,m) \in OD_H} \sqrt{(n - x_0)^2 + (m - y_0)^2}}{HDNAArea \cdot Area \cdot mean(Radii)} \quad (5,68)$$

- **Center of the Mass** is calculated by finding the distance between the center of the mass and the object centroid (MacAulay, Keyes, et al. 2017).

- **Low Center Mass** is defined as below:

$$LowCntMass = \frac{\sqrt{(x_{CML} - x_0)^2 + (y_{CML} - y_0)^2}}{LDNAArea \cdot Area \cdot mean(Radii)} \quad (5,69)$$

- **Medium Center Mass** is defined as below:

$$MedCntMass = \frac{\sqrt{(x_{CML} - x_0)^2 + (y_{CML} - y_0)^2}}{MDNAArea \cdot Area \cdot mean(Radii)} \quad (5,70)$$

- **High Center Mass** is defined as below:

$$HighCntMass = \frac{\sqrt{(x_{CML} - x_0)^2 + (y_{CML} - y_0)^2}}{HDNAArea \cdot Area \cdot mean(Radii)} \quad (5,71)$$

- **Medium High Center Mass** is defined as below:

$$HighCntMass = \frac{\sqrt{(x_{CML} - x_0)^2 + (y_{CML} - y_0)^2}}{MHDNAArea \cdot Area \cdot mean(Radii)} \quad (5,72)$$

5.5.1.5 Feature Selection- Cell Sociology

- **Voronoi Tessellation or Dirichlet** finds polygonal cells around the nodes (*Voronoi diagram*. 2019). In this region, all pixels are closer to the nuclei than any other nuclei outside the polygonal region, Figure 5-3.
- **Delaunay Triangulation** is a complement (dual graph) to Voronoi Tessellation.
- **Minimum Spanning Tree** is the tree that, as the name suggests, connects all the cells for which the sum of the edge lengths is minimum.
- **Number of Nodes** is the total number of nodes in a graph (N_e).

Number of Edges is calculated as below (5,73)

$$N_e = \frac{1}{2} \sum_{i,j} a_{i,j}$$

- **Edge Length** is calculated by finding the Euclidean distance between each two neighboring nuclei.

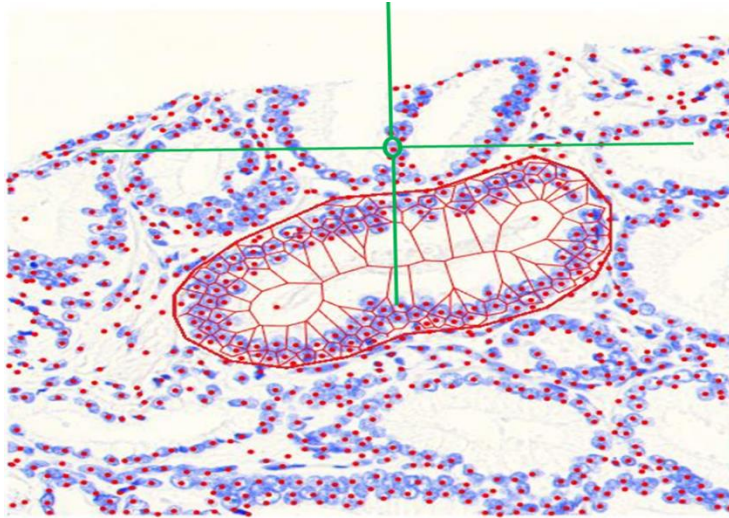


Figure 5-3 An image of the Voronoi Tessellation feature on a part of TMA tissue.

5.5.1.6 Latent Support Vector Machine Algorithm for Classification of Prostate Cancer

Grading

A common way to classify tissues and determine whether it is progressive or non-progressive is to classify each nucleus within tissue as progressive-non-progressive and then aggregate these labels and compute a score for the whole tissue and then determine how likely it is for the tissue to become progressive. It is shown that the accuracy of this method relies on many parameters such as the anatomical location of the nucleus (whether it is part of a clustered group of nuclei) the imaging properties (such as image quality, out of focus issue, contrast, etc.) and many other parameters. There are two major ways to handle these problems. In the first method, the nuclei which have high quality and are part of anatomically relevant areas are selected manually. The main drawback of this approach is that it requires extensive human expert interaction which makes the process expensive, time consuming and biased. The second approach is to implement an algorithm to first predict relevant areas/features and then use another classifier to predict the labels.

Although the second approach is relatively faster it needs labeled datasets for training the classifiers and these classifiers usually do not have perfect accuracy. These inaccuracies may also propagate to the later stages of the prediction pipeline.

Here we introduce an elegant method to overcome the aforementioned difficulties. The intuition behind our proposed method comes from the fact that the goal of all the predictive systems is to distinguish between progressive and non-progressive parameters and whatever comes in between is considered as unnecessary information.

Here we present a predictive method using the Latent Support Vector Machine (LSVM) framework. By using this method, we can consider all the intermediate labels as latent variables and then classify the nuclei. Using the methods expressed in the previous chapters, we have constructed a collection of segmented nuclei from Thionin stained tissue micro-array of prostate cancer patients with a Gleason score of 6-9. Our goal is to introduce a powerful method to classify these nuclei based on the Gleason score and the progressive/non-progressive outcome.

We formulated our latent SVM model using the following scoring function:

$$\begin{aligned}
 f_w(x, h) &= w^T \Psi(h, x) \\
 &= w_0^T \varphi_0(x_0) + \sum_{j \in V} w_1^T \varphi_1(x_j, h_j) + \sum_{j \in V} w_2^T \varphi_2(h_j)
 \end{aligned} \tag{5,74}$$

where $f_w(x, h)$ calculates a predictive scoring value for a nucleus. This value is correlated to the likeliness of cancer progression, h_i is a set of latent variables representing binary values of different parameters/features such as belonging to epithelial layers, image quality, being a cluster member and others. We will explain each part of this equation in the following section.

For an overview of our system please refer to Figure 5-4.

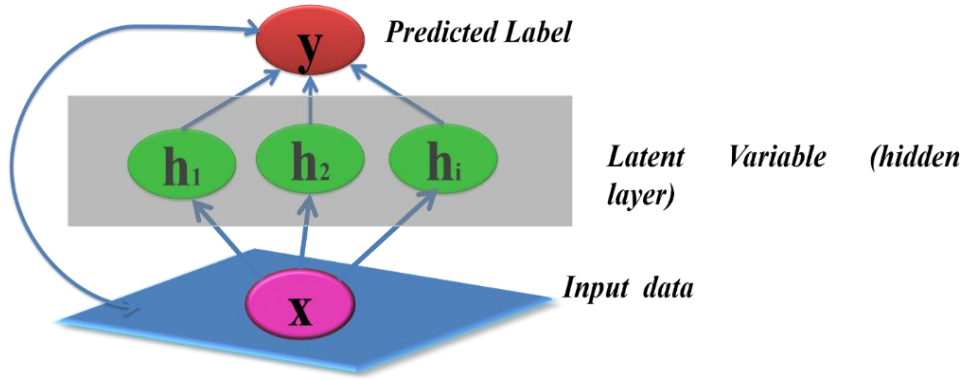


Figure 5-4 A Simplified cartoon representing an overview of our LSVM algorithm.

Nucleus progressive\non-progressive potential $w_0^T \varphi_0(x_0)$: This potential function is a global function that models the relationship between the nucleus shape information and prediction. x_0 represents a set of features describing the shape of the nucleus, and w_0 represents the weight vector for these features. These terms guarantee that our model at the worst case would perform like any other classification method that uses nuclei information for prediction of cancer progression.

Image-Nucleus Property potential $w_1^T \varphi_1(x_j, h_j)$: Models the compatibility between image features x_j and the predicted nucleus property h_j . We used the following formula for this potential function.

$$w_1^T \varphi_1(x_j, h_j) = \sum_{b \in H} w_{1,b}^T 1(h_j == b) \cdot x_j \quad (5,75)$$

where function $1(.)$ is an indicator function. This function is equal to one if the inner term is true and is zero otherwise. To calculate x_j we used some information from cell sociology features for latent variables related to anatomical location on nucleus (as presented in Feature Selection- Cell Sociology section), blur (Narayanan Ramakrishnan 2010) and contrast (Peli 1990) information

for the quality of the nucleus and adjacency information for determining whether it belongs to a cluster or not. H represents the set of values for the latent variables which is usually 0 or 1 (for example whether it belongs to the epithelial layer or not). For the quality measures latent variable, we have a set of (0,1,2,3) values where 0 is the highest out of focus and low quality and 3 is the highest quality.

Nucleus Property-progressive\non-progressive potential $w_2^T \varphi_2(h_j)$: This potential function models the relation between nucleus properties and cancer progression, we modeled this term using the following equation:

$$w_2^T \varphi_2(h_j) = \sum_{b \in H} w_{2,b}^T 1(h_j == b) \quad (5,76)$$

Inference and Learning: We found the values for the latent values as well as the output in the inference part of the algorithm. Since the search space for our problem is relatively small we used the brute force method to find latent values.

The most challenging part of our approach is to find the weights in our formula. Since our algorithm is nonconvex it is not possible to use gradient descent optimization method as our solver. To solve this problem, we used Non-Convex Regularized Bundle Methods (NRBM) package (Do, Artières 2009) This package is guaranteed to converge to local optima. Furthermore, the kernel used in the scoring function is linear and the NRBM package works efficiently to find the weights. From our experiments, the estimated local optima appear to be good enough and they can outperform the existing prediction methods such as regular SVM.

We used $y^* = \operatorname{argmax}_y E(x, y; w)$ as the predicted label of x , where the value of y can be ± 1 . Given $((x^1, y^1), (x^2, y^2), \dots, (x^n, y^n))$, the set of training data, our aim is to find parameters such that the sign of score function x^i represents the class of the X^i .

We formulated the training criteria in the Max-Margin framework (Felzenszwalb, Girshick, et al. 2010). We set w by using the following formula:

$$\min_{w, \xi^i} \frac{\lambda}{2} ||w||^2 + \sum_{i=1}^N \xi^i \quad (5,77)$$

where $\xi^i = \max(0, 1 - E(x^i; w))$

$$s. t \ y_i(E(x^i; w) + b) \geq 1 - \xi^i, \xi^i \geq \max(0, 1 - y_i(E(x^i; w) + b)) \quad (5,78)$$

where λ represents the trade-off constant and is determined using cross-validation in different experiments. These constraints force the score of progressive samples to be positive and greater than 1 while forces the score for non-progressive samples to be negative and less than -1.

5.6 Experiment and Result

In this section, we present the results of our classifier and investigate the performance of our classification approach.

We used Matlab 2014 to implement the features presented in the previous sections. Then we used Python 2.7 to implement classifiers and apply models to the data.

In this study, we investigated two important questions: I) the accuracy of our algorithm for Gleason score grading, II) Prostate cancer prognosis (progressive and non-progressive prostate cancer case

for patients with moderate Gleason score). In both these two experiments, we applied LSVM to all the features presented earlier. Also in both studies, 80% of the cases (cells) are randomly selected for training and the other remaining 20% is reserved for testing.

5.6.1 ML for Prostate Gleason Score Grading

In this experiment, the goal is to predict the prostate cancer Gleason score. In this study, we used images of all 32 patients with the descriptions mentioned in the “Material” section. Each TMA core was imaged and saved as 16 smaller tile sub-images. Randomly one sub-image per patient is selected for further processing. The total number of cells under this study was 84,527 cells.

In the first step, we tried to distinguish between high Gleason scores (8 and above) and low Gleason scores (7 and below) and then within these categories we tried to do a fine level classification to predict the accurate Gleason score. For this experiment, we used the features presented in the previous sections and applied different models to our data. The latent SVM classifier performed the best compared with the other classifiers that we tried. The results of this experiment are presented in Table 5-2f and Figure 5-5.

As shown in Figure 5 5 Latent SVM method performed very well for distinguishing between high Gleason grades and low Gleason grades, however, the accuracy dropped during fine level classification (still higher than the baseline classifier, regular SVM).

5.6.2 ML for Prostate Prognoses

In the second experiment, our goal is to predict whether cancer progression would happen for patients with a Gleason score of 7.

Among 32 patients we had 9 patients with a Gleason score of 7 in the failure group and 7 patients with a Gleason score of 7 in the non-failure group. The total number of patients for this experiment was 16. Similar to the first experiment, we randomly selected one sub-image per patient for further analyses. The total number of cells under this study was 46,464 cells.

For this experiment, we randomly divided the input dataset to test and train datasets. For the test set, we took 20% of the data and used the rest as our training set. The results of this experiment can be seen in Figure 5-6. Our proposed Latent SVM method significantly outperformed the regular SVM and has an accuracy of 92% while the regular SVM only achieves 84% accuracy. Some Statistics that are described in equations (5,80)-(5, 85) are calculated and summarizes in Table 5-3.

Table 5-2 The classification statistics for classifying aggressive vs non-aggressive on the test dataset.

True Positive	False Negative	Accuracy	Precision
7898	2471	0.925	0.887
False Positive	True Negative	Recall	F1 Score
1005	35090	0.761	0.819

$$Accuracy = \frac{\Sigma TP + \Sigma TN}{\Sigma total\ population} \quad (5,79)$$

$$Precision = \frac{\Sigma TP}{\Sigma TP + \Sigma FP} \quad (5,80)$$

$$Recall = \frac{\Sigma TP}{\Sigma TP + \Sigma FN} \quad (5,81)$$

$$F1 = \frac{2 * Precision * Recall}{Precision + Recall} \quad (5,82)$$

$$Sensitivity(TPR) = \frac{\Sigma TP}{\Sigma TP + \Sigma FN} \quad (5,83)$$

$$Sensitivity(TPR) = \frac{\Sigma TN}{\Sigma TN + \Sigma FP} \quad (5,84)$$

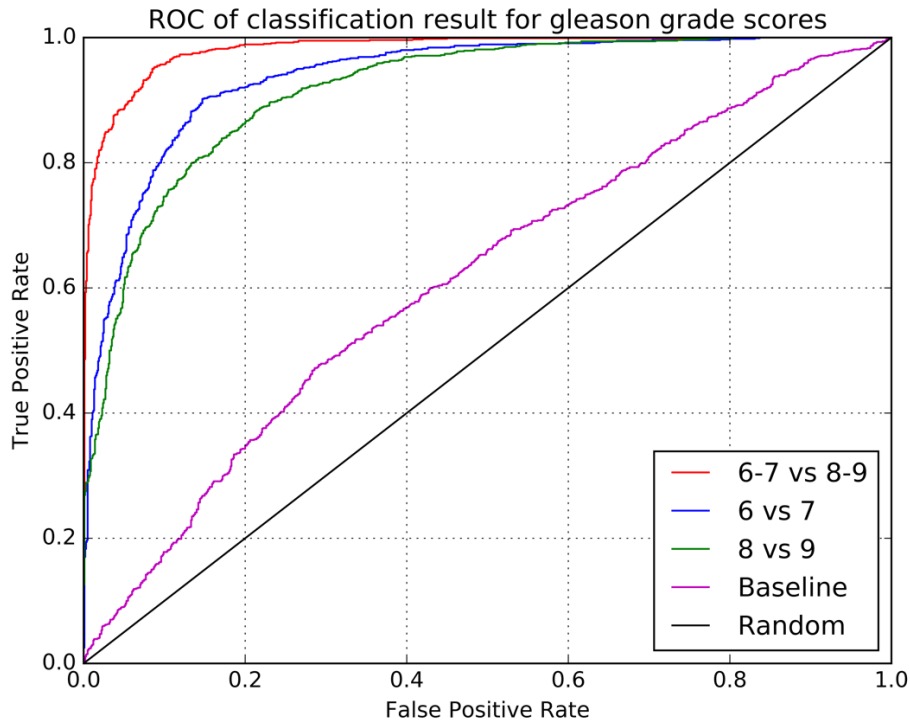


Figure 5-5 In this figure we present the ROC result of Glean grade classifications between different scores. The black line represents a random line. The purple line is the baseline using SVM. LSVM is used to classify between 6-7 vs 8-9 red line, 6 vs 7 shown with the blue line and 8 vs 9 shown with the green line.

Table 5-3 The classification statistics for classifying 6-7 vs 8-9 Gleason score cases on the test set.

True Positive	False Negative	Accuracy	Precision
14327	3231	0.876	0.665
False Positive	True Negative	Recall	F1 Score
7233	59736	0.816	0.732

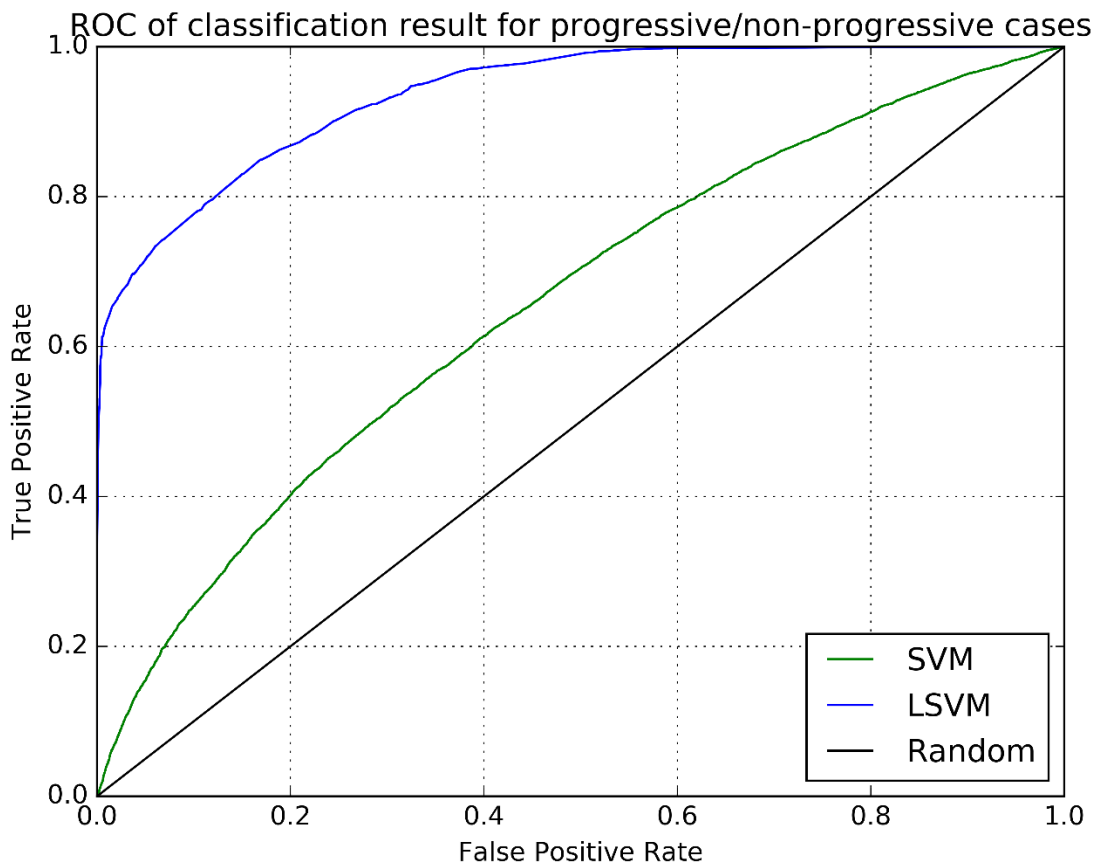


Figure 5-6 In this figure we present the ROC result of aggressive vs non-aggressive classification for patients with Gleason score 7. We used both SVM and LSVM as our classification method, the results are shown with green and blue lines respectively.

5.7 Conclusion

We implemented our algorithm and ran our proposed method on our dataset and achieved 95% and 87% accuracy for Gleason score prediction on the training-set and test-set respectively. Also, we achieved 98% and 92% for aggressive vs non-aggressive prediction on the training-set and test-set respectively. The differences in the accuracies obtained from the test-set and training set could be due to over-fitting.

One of the advantages of our method is that it requires much less human interaction which achieved the purpose of having an automated Quantitative Digital Pathology (QDP) system.

LSVM algorithm could serve as a predictive tool for prostate cancer behavior prediction. This technique can be combined with a high-resolution imaging system to reduce over-diagnosis/treatment.

One of the limitations of the method presented in this chapter is the necessity to design hand-crafted (hand-engineered) features. These features served as an excellent solution for the problem under the study however they limit the predictive power of classical ML to the topic under the study.

Although this method still needs extensive validation to be clinically useful, future application of this method can change clinical pathology by identifying aggressive lesions that require up-front treatments vs. non-aggressive lesions that need to be under active surveillance.

Chapter 6 Machine Learning for Renal Cell Carcinoma (RCC)

Detection

This study is based on the collaboration between our group at BC Cancer Research Center, Canada and Morpheme group at Inria (institution for research in digital science and technology), France. The goal is to use explicit measures of tissue structures for tissue type prediction.

6.1 Background-Renal Carcinoma

Kidney cancer is one of the most common malignancies worldwide. More than 338,000 new kidney cancer cases diagnosed in 2012 (Kidney cancer statistics 2018). Only in the United States, it is estimated that about 62,700 patients will be diagnosed with kidney cancer in 2016 and ¼ of them succumb to death (Key Statistics About Kidney Cancer). More than 85-90% of kidney cancers arise in the renal parenchyma, and many of the cases are renal cell carcinomas (RCCs) (Campbell, Rini, et al. 2012). WHO in 2004 classified RCC into 10 groups primarily based on RCC morphology. Clear Cell RCC (CCRCC) with 60-70% prevalence and Papillary RCC (PRCC) with 10–15% prevalence account for first and second most common RCC respectively (Campbell, Rini, et al. 2012). The focus of this chapter is on these two prevalent renal carcinomas.

Kidney cancer incidence is estimated to increase due to the aging of the population, obesity, diet, lifestyle, etc. (Campbell, Rini, et al. 2012). Despite advances in technology and the introduction of new diagnostics methods, the RCC mortality rate is still high and so a reliable, robust, cost-effective diagnostic method is needed. We are interested to develop such a system with high sensitivity and specificity to classify normal versus RCC, CCRCC versus PRCC.

The study of the cell out of the context of its microenvironment might not be a very reliable approach since the microenvironment orchestrates the cancer development and cells' spatial orientation and relative locations toward each other are influenced by cancer heterogeneity. As a result, hallmarks such as stromal vascularity (Iakovlev, Gabril, et al. 2012), stromal hyalinization and others are clinically important features. The other important characteristic of CCRCCs is their architectural patterns, the common ones are compact-alveolar (nested), tubular (acinar), and microcystic. Figure 6-1 (b) shows an architectural pattern of CCRCC cells.

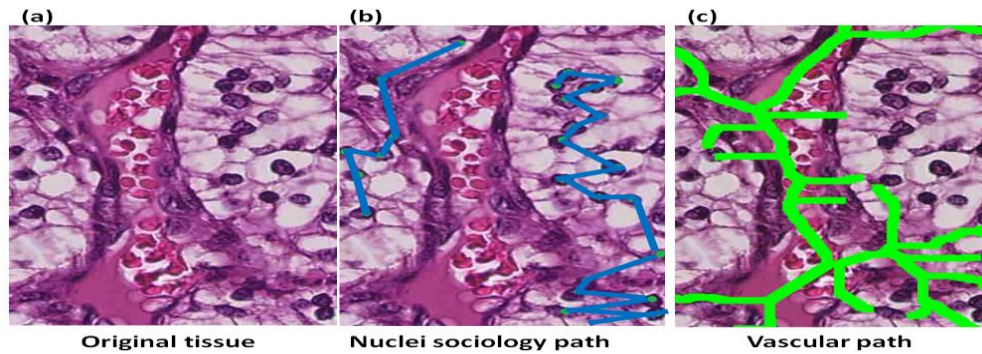


Figure 6-1 Microenvironment features. (a) shows a clinical image of renal carcinoma where lots of large cells with clear cytoplasm and a very large vessel in the tissue are present. (b) shows the cell/nuclei sociology/ tree-pattern. (c) shows the vessel and detected branches which is an indication of angiogenesis.

6.2 Hypothesis

Digital pathology using advanced image analysis technologies has revolutionized histopathology by providing an objective assessment of cellular morphology, aiding with cancer grading and obtaining clinically relevant quantitative data from pathological samples (Heindl, Nawaz, et al. 2015). Since the development of the cancerous cell is a complex process which results in different nucleus, cellular and architectural morphology, we hypothesize that a discriminative algorithm

which uses information from cancer hallmarks such as blood vessel development associated with angiogenesis can effectively assist with normal and RCC classification. Figure 6-2 shows some pathological features of renal carcinoma that might be helpful to consider.

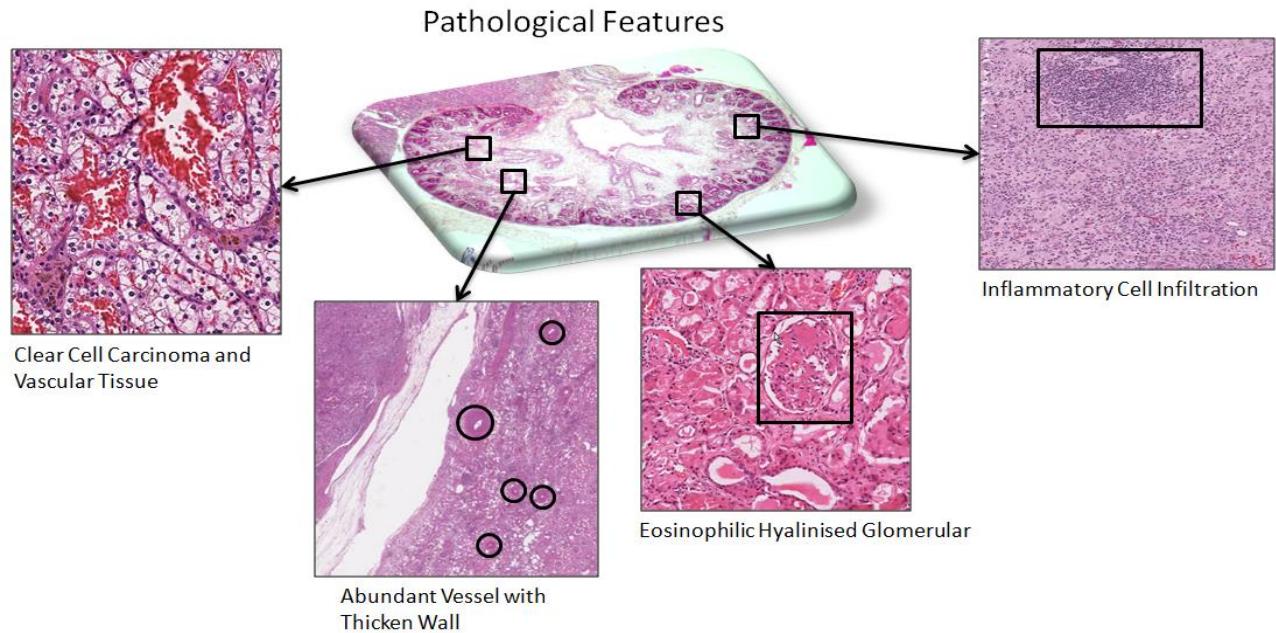


Figure 6-2 Renal tissue, pathological features. In this image, many pathological features are illustrated, although, in this study, we are interested in angiogenesis hallmark, other pathological indications such as infiltration of inflammatory cells are worth mentioning. Other features such as abundant clear cell carcinoma, although is not discussed in this chapter, were the topic of previous chapters where we showed the importance of nuclei features and how they can be used for classification purposes.

6.3 Objective

The objective of this study is to design a reliable and robust machine learning classifier to discriminate between different renal carcinoma subtypes.

6.3.1 Specific Objective

- Collecting the image, selecting the region of interest, labeling the images.
- Using cancer hallmark features for the input of the classifier.
- Implementing machine learning algorithms to evaluate the renal images.

The workflow for this chapter is presented in Figure 6-3.

The first step, like most of the biomedical studies, was to prepare the clinical data. The next step was to image the collected data/sample/specimen, followed by consecutive image processing techniques such as image enhancement, image segmentation, Region of Interest (ROI) selection, etc. The next step was to use more advanced image processing techniques to select and extract features that are going to be used in the next step for classification. The last step as explained and shown in Figure 6-3 is the implementation of machine learning techniques to use the extracted features from the previous step to discriminate between normal and abnormal cases. This step consists of feature, parameter optimization, selection of important features, implementation of classification methods (test, train, and validation) and finally parameter optimization to improve the performance.



Figure 6-3 This figure shows the pipeline of our renal carcinoma study. The processes consist of imaging and pre-processing of the captured images. The next step is post-processing and extracting/segmenting clinical important features and the last step is ML techniques implementation for classification of normal and abnormal cases.

6.1 Material

Tissues from 50 consecutive patients with renal cell carcinoma were collected by Pasteur Hospital, Nice, France. These tissues were annotated by the study pathologist into CCRCC (60% of the cases), PRCC (60% of the cases) and mixed CCRCC/PRCC (10% of the cases).

All tissues are sliced into 3 μ thick slices and stained with H&E. This stain has two components Eosin and Hematoxylin. Eosin stains the cytoplasm pink and Hematoxylin stains the nuclei blue.

A Leica SCN400 was used to scan the samples. The captured images have a resolution of 0.25 μ m and they are around 100,000 \times 100,000 pixels in size. The Images were manually annotated, as shown in black areas in Figure 6-4 (b). Since the images were very large, to reduce the computation costs we implemented a sliding window, part (c) in Figure 6-4. The sliding window was 50 \times 50 μ m and the overlapping area was 5 μ m.

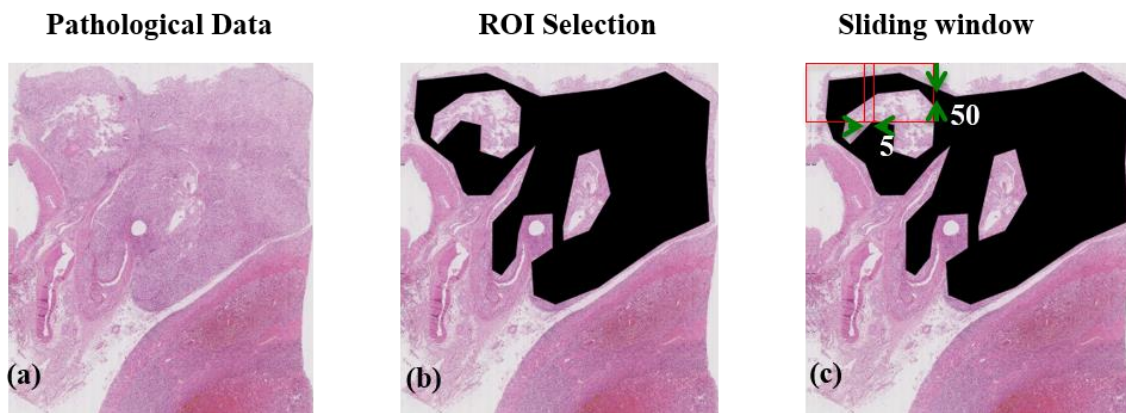


Figure 6-4 This image shows the Sliding window implementation. This image illustrates the process required to slice the large image to smaller pieces so that our computing instrument could handle the pre/post-processing step (required image processing). (a) an image of renal carcinoma tumor, (b) manually annotated image where the black area

represents the tumor and other areas represent the healthy tissue. (c) a sliding window is shown on the image.

6.2 Methodology

In this section, we explain the step and processes required to extract the clinically important features, we describe the required image processing steps and in the end, we will describe the classification step. One of the features that is clinically important is the vascular structures. As explained earlier, angiogenesis is one of the hallmarks of cancer and we believe it could be a discriminative feature for our classification goal.

The whole workflow for extracting the vessels is shown in Figure 6-5.

To enhance the presence of vascular structures, a color-deconvolution method is used. In this step, new channels are defined: 1) vessel channel (V) and 2) nuclei channel (H), calculated by equations (6,1) and (6,2) respectively.

$$V = 0.5 \frac{R+B}{\sqrt{R^2+G^2+B^2}} \quad (6,1)$$

where R , B , and G represent red, blue and green pixel intensities in the RGB image.

$$H = \frac{B}{\arctan\left(\frac{B}{\max(R,G)}\right)} \quad (6,2)$$

In the following step, we applied a modified Gabor filter (Zubiolo, Debreuve, et al. 2016), equation (6,3) followed by Harr filter (Haar 1-D wavelet transform - MATLAB haar).

$$G^+(x, y) = \begin{cases} G(x, y) & \text{if } x' \geq 0 \\ 0 & \text{otherwise} \end{cases} \quad (6,3)$$

$$G^-(x, y) = \begin{cases} G(x, y) & \text{if } x' \leq 0 \\ 0 & \text{otherwise} \end{cases} \quad (6,4)$$

where:

$$G(x, y) = \cos\left(\frac{x'\pi}{\sqrt{2}\sigma}\right) \exp\left(-\frac{x'^2 + y'^2}{2\sigma^2}\right) \quad (6,5)$$

where:

$$x' = x \cos\theta + y \sin\theta \quad (6,6)$$

$$y' = x \sin\theta - y \cos\theta \quad (6,7)$$

The next step is to convert the image to a binary image, for this step we used the Otsu algorithm to make the binary image. Then we cleaned the binary image using morphological cleaning (removing objects smaller than 5 pixels and opening with a disk structure of size 5 pixel). Once we have the cleaned image, we applied hit or missed based skeleton extraction method (Soille, SpringerLINK eBooks - English/International Collection (Archive) et al. 2004) to extract the skeleton. Figure 6-5 represent the whole pipeline of this process. This image, at a high level, explains the steps taken from an RGB image (input) to the image of the vessel skeleton (output).

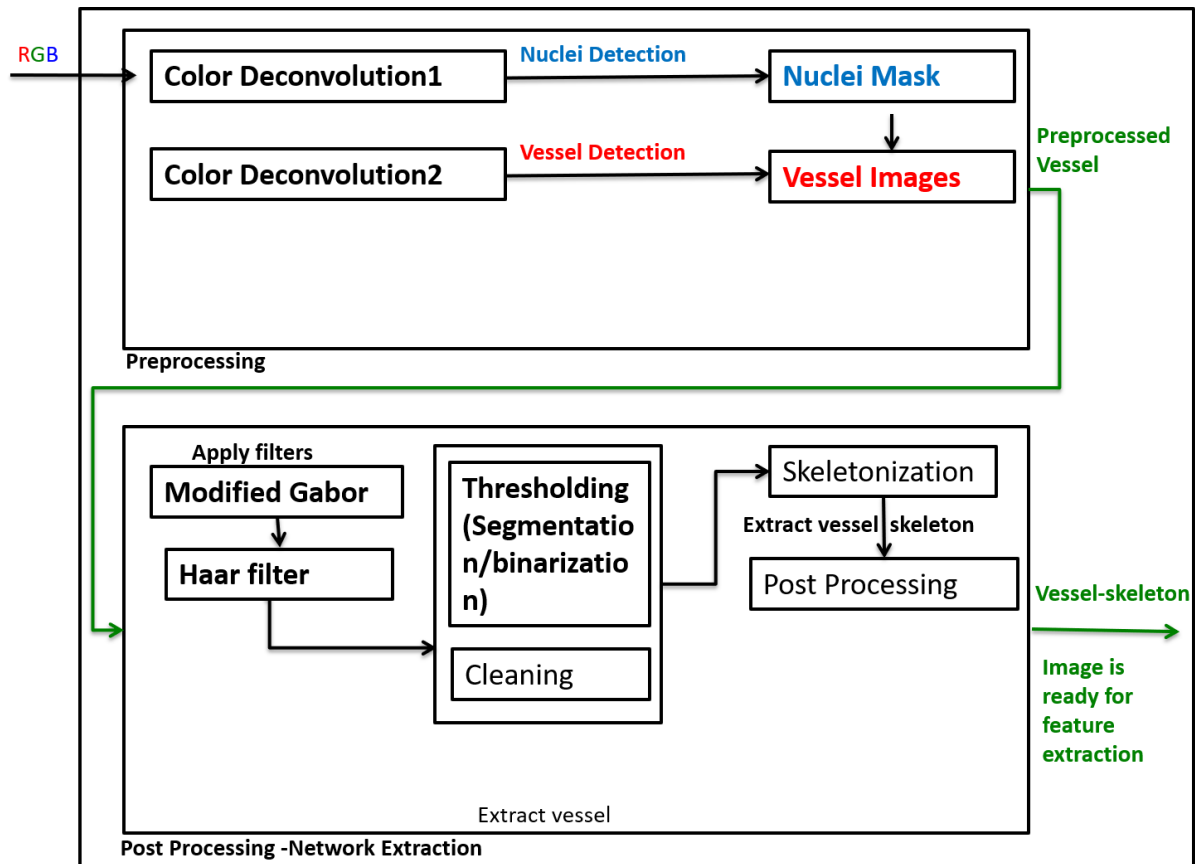


Figure 6-5 Pipeline of the pre/post-processing. This workflow starts with an RGB image, pathological image (input) and the resulted output is the skeleton of the vessels where all the nuclei and other structures are removed.

In Figure 6-6, we showed the whole procedures that we explained so far using a part of our renal sample. The steps showed in Figure 6-6 aligns with the process explained in the diagram of Figure 6-5.

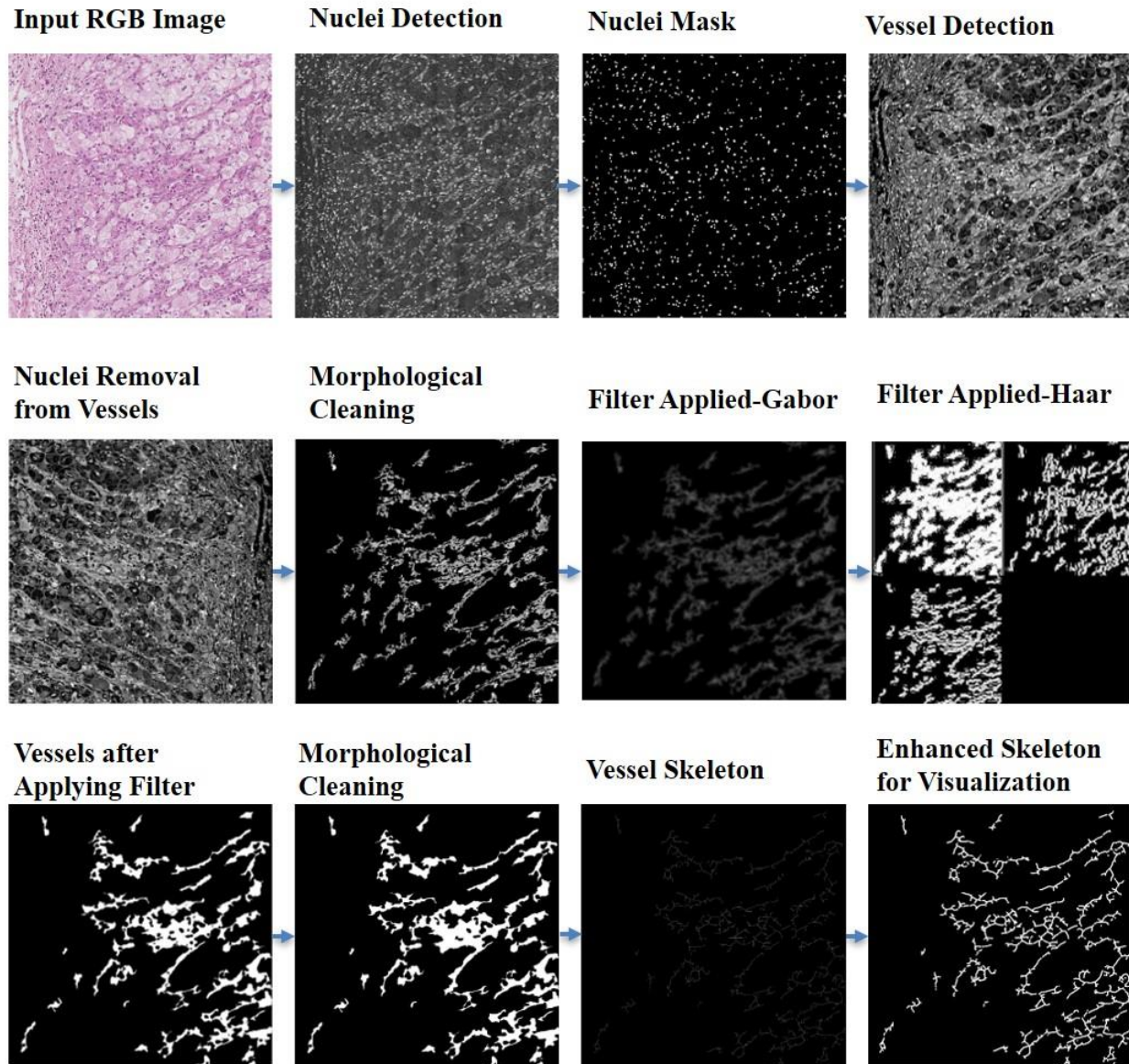


Figure 6-6 The whole process of the pre-post processing step. This workflow starts with an RGB image, the pathological image of renal carcinoma (input) and ends with the output which is the skeleton of the vessels where all the nuclei and other structures are removed.

6.3 Feature Extraction and ML Algorithm Design

The next step in our workflow is developing an ML method to classify different RCC subtypes.

This step consists of two main tasks:

- 1- Implementing /selecting the features;
- 2- Developing different machine learning classifiers using these features.

6.3.1 Feature Extraction

A 50×50 sliding window scans the whole image, per each sliding window an RGB image will be generated. As explained above and shown in Figure 6-5 and Figure 6-6 an image of the vessel skeleton is generated from each input RGB image. The next step is to extract features for these vessels.

Here we define features and extract them from the skeleton. (these features are selected as they are more relevant to tree's (here vessel structure) pattern).

The skeleton features that I decided to calculate are listed as below:

- 1- Number of branches,
- 2-Number of the actual junctions (close junctions, merged together),
- 3- Number of endpoint voxels (with less than 2 neighbour) ,
- 4- Number of junction voxels (with more than 2 neighbour),
- 5-Number of slap voxels (if a pixel has exactly 2 neighbors),
- 6-Average branch length,
- 7-Number of triple points,
- 8- Number of quadruple points,
- 9- Maximum branch length,

10-Largest-shortest path (Floyd–Warshall algorithm),

$$11 - \frac{\text{Number of End point voxels}}{\text{Number of Junctions}}$$

6.3.2 Design/Select the Classifier

At this point, we prepared a dataset in which each row represents a sliding window and each column represents the feature vector for that window. The next step is to select a classifier, we experimented with different classifiers and selected an ensemble of five classifiers with the best outcome for our study, as explained below.

One of the classifiers that we used is “Discriminant Analysis”, this classifier uses a linear decision boundary (Linear discriminant analysis. 2019). This boundary is generated by fitting class conditional densities to the data using Bayes’ rule, equation (6,8).

Like PCA, this model can also be used to reduced class dimensionality by projecting the dataset into the most discriminative directions.

$$P(y = k|X) = \frac{P(X|y = k)P(y=k)}{P(X)} = \frac{P(X|y = k)P(y=k)}{\sum_l P(X|y = l) \cdot P(y=l)} \quad (6,8)$$

where y , X and k are the label, sample, and class respectively.

The model fits a Gaussian density to each class assuming all classes share the same covariance matrix, equation (6,9).

More specifically, for linear and quadratic discriminant analysis, $P(X|y)$ is modeled as a multivariate Gaussian distribution with density:

$$P(X|y = k) = \frac{1}{(2\pi)^{d/2} |\Sigma_k|^{1/2}} \exp \left(-\frac{1}{2} (X - \mu_k)^T \Sigma_k^{-1} (X - \mu_k) \right) \quad (6,9)$$

where d is the number of features. In this model, we need to estimate class priors $P(y = k)$, class means μ_k and covariance matrices.

We can assume $\Sigma_k = \Sigma$ and this will lead to linear decision surfaces, equation (6,10).

$$\begin{aligned} \log \left(\frac{P(X|y = k)P(y=k)}{\sum_l P(X|y = l) \cdot P(y=l)} \right) = 0 &\Leftrightarrow (\mu_k - \mu_l)^t \Sigma^{-1} X \\ &= \frac{1}{2} (\mu_k^t \Sigma^{-1} \mu_k - \mu_l^t \Sigma^{-1} \mu_l) - \log \frac{P(y = k)}{P(y = l)} \end{aligned} \quad (6,10)$$

The next classifier we used was logistic regression, also known in the literature as logit, maximum-entropy classification or log-linear classifier (Generalized Linear Models — scikit-learn 0.20.2 documentation.). In this model, the probabilities describing the possible outcomes of a single trial are modeled using a logistic function, a loss function using maximum likelihood is defined and the optimization goal is to minimize this loss function, equation (6,12).

The cost function that we tried to minimize is shown in equation (6,11).

$$J(\theta) = - \frac{1}{m} \sum_{i=1}^m y^{(i)} \log((h_w(x^i)) + (1 - y^{(i)}) \log(1 - h_w(x^i))) \quad (6,11)$$

where

$$h_w = \frac{1}{1 + \exp(-W^T X)} \quad (6,12)$$

where W^T , is the transposed matrix of weights.

We also used tree-based classifiers such as decision tree which is a decision support tool that uses a tree-like model of decisions, as well as Random Forest and boosted tree algorithms (Introduction to Boosted Trees).

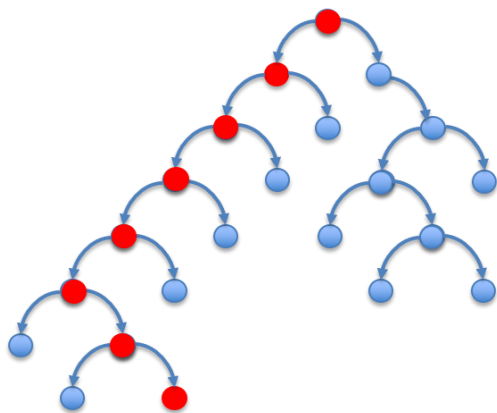


Figure 6-7 Diagram of a simple decision tree. In each node, a selection is made as shown in red.

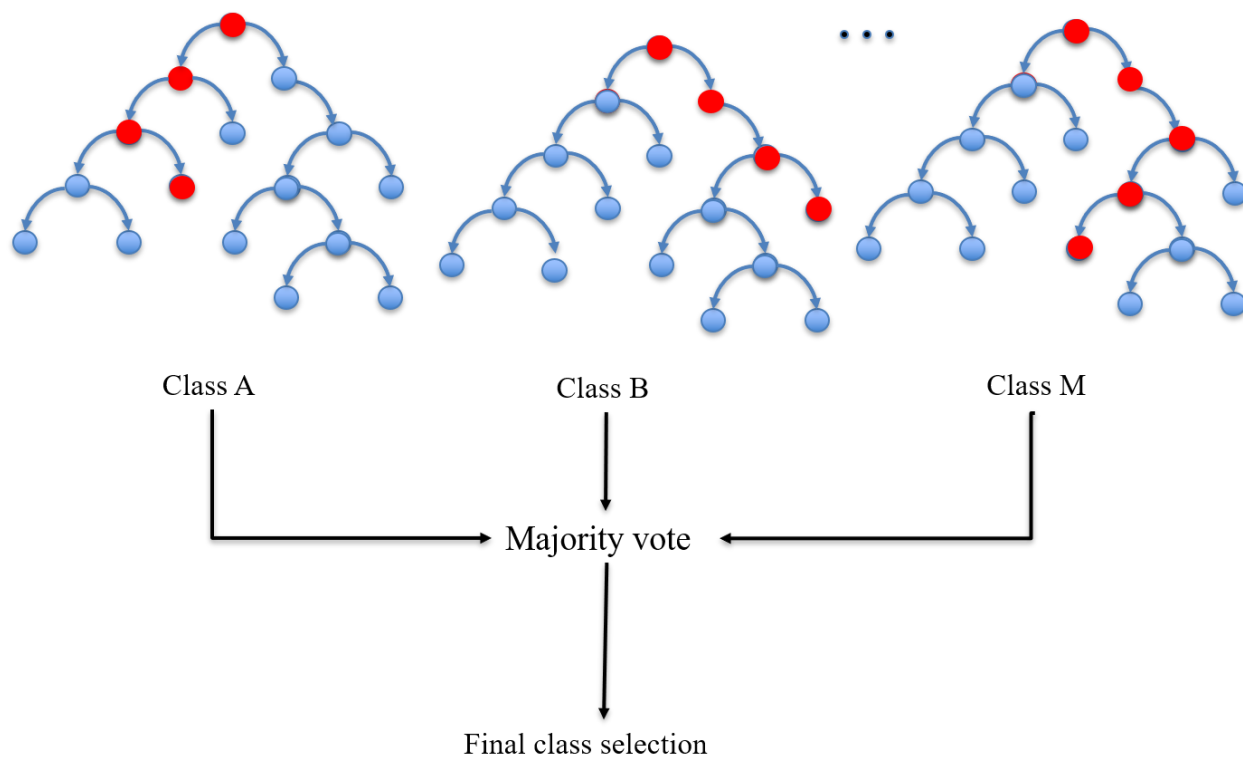


Figure 6-8 Diagram of the Random Forest algorithm, this algorithm is based on an ensemble decision making where each tree can have many branches. The final class selection would be based on the majority votes of all trees.

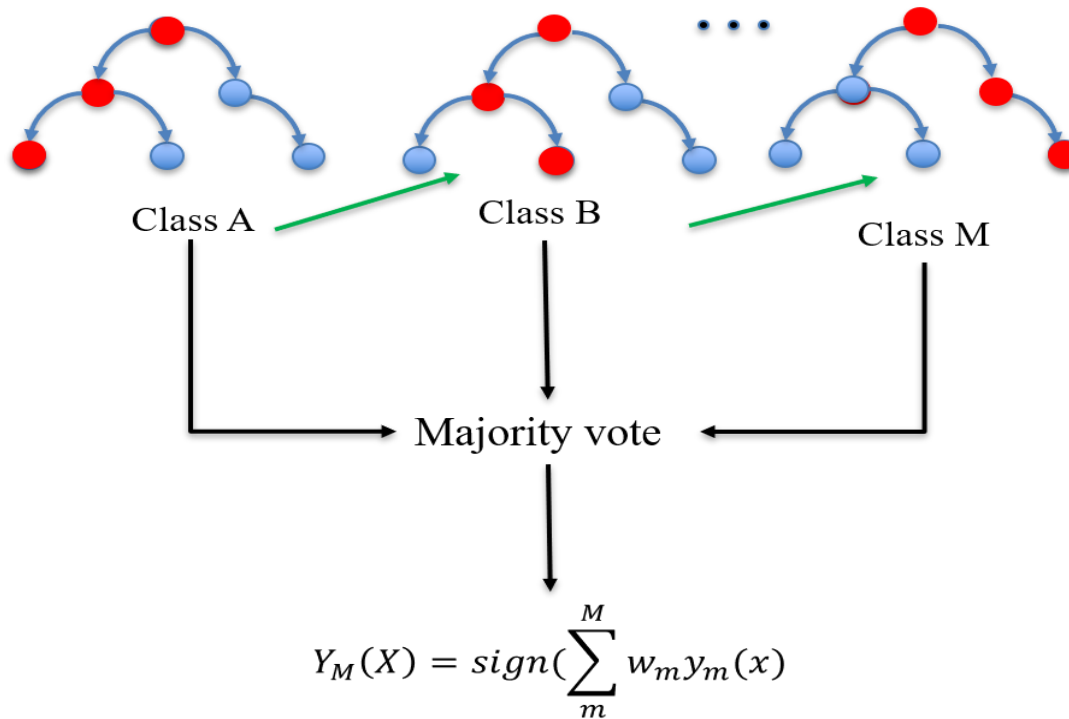


Figure 6-9 Diagram of Boosted Tree algorithm. Boosted tree is an ensemble algorithm that uses the majority votes of some stump trees, in each step the weights for the next decision making would be updated as shown with the green arrows above.

6.4 Results

For the first experiment, we decided to classify CCRCC vs other tissues. We applied different classifiers (linear discriminate, logistic regression, decision tree, random forest, boosted tree classifier) to our dataset. Figure 6-10 shows the ROCs of all the classifiers we used, we obtained 95% training accuracy (using 80% of the dataset). Table 6-1 summarizes the classification results on the training -set.

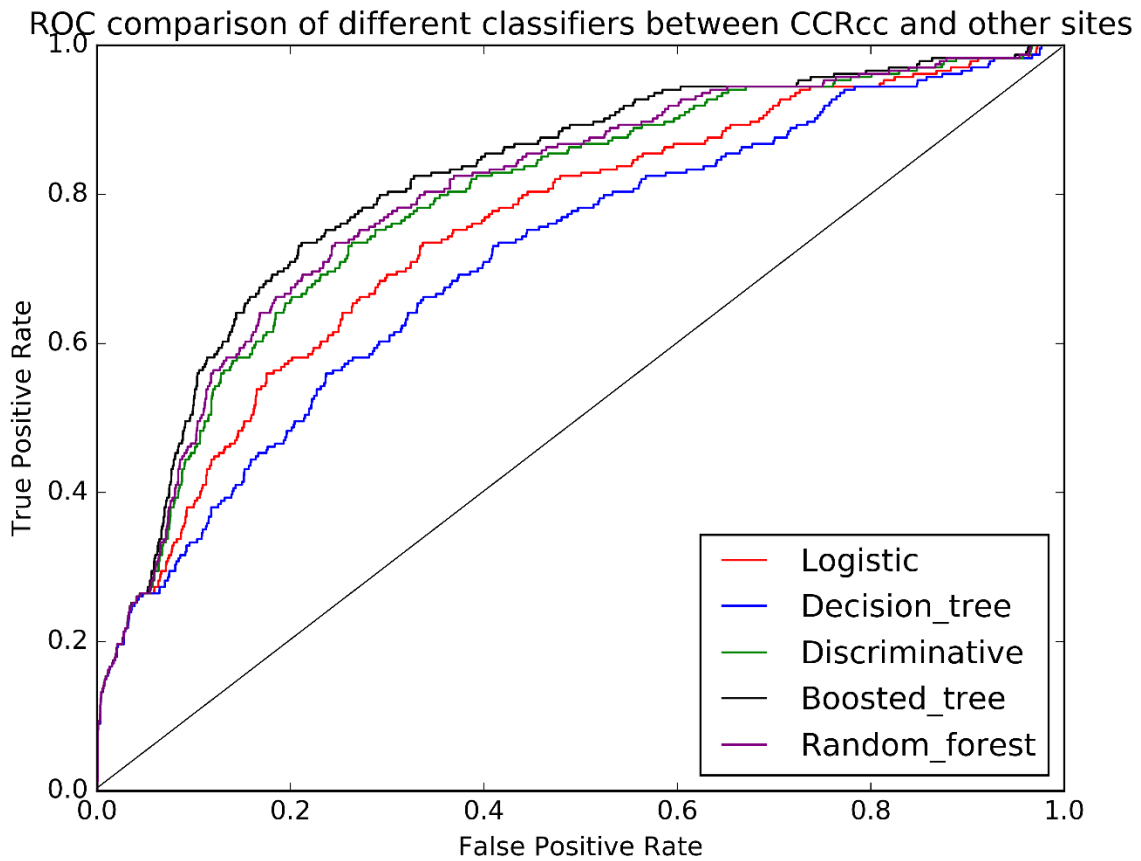


Figure 6-10 This figure represents the ROC of all 5 classifiers (linear discriminate, logistic regression, decision tree, random forest, boosted tree classifier) for classification between CCRCC and other tissue sites.

Table 6-1 classification results for CCRCC and other tissue sites using a boosted tree classifier on the training set

Iteration	Elapsed Time	Training-accuracy	Training-log loss
1	0.009	0.950	0.484
2	0.018	0.951	0.370
3	0.028	0.951	0.300

4	0.037	0.952	0.254
5	0.049	0.952	0.225
6	0.064	0.952	0.204
Number of examples: 14683			
Number of classes: 2			
Number of feature columns:11			

Next, we decided to normalize the features, using the Z-score normalization method (Feature scaling) and observe the performance of our method. The accuracy didn't improve very much and stayed relatively the same. Figure 6-11 shows the ROCs of all the classifiers that we picked for this study. Comparing this figure with Figure 6-10, we can observe similar behavior.

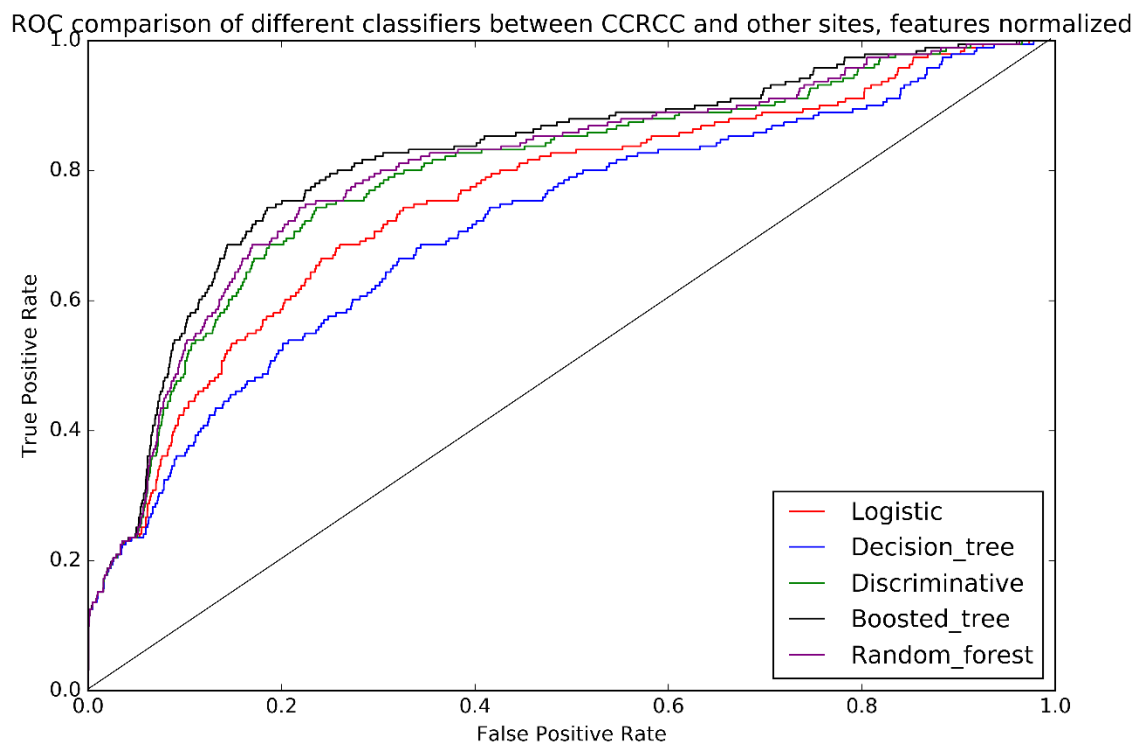


Figure 6-11 This figure represents the ROC of all 5 classifiers (linear discriminate, logistic regression, decision tree, random forest, boosted tree classifier) for classification between CCRCC and other tissue sites when the features are normalized.

We also classified (Chromophobe RCC) CCP and other tissue sites, Figure 6-12 as well as PCC and other tissue sites Figure 6-13. We did not obtain a promising result for classifying these two cancer sites. In the future section of this chapter, we discuss the possible reasons for this poor result, we also discuss methods/avenue to explore for possible better classification results.

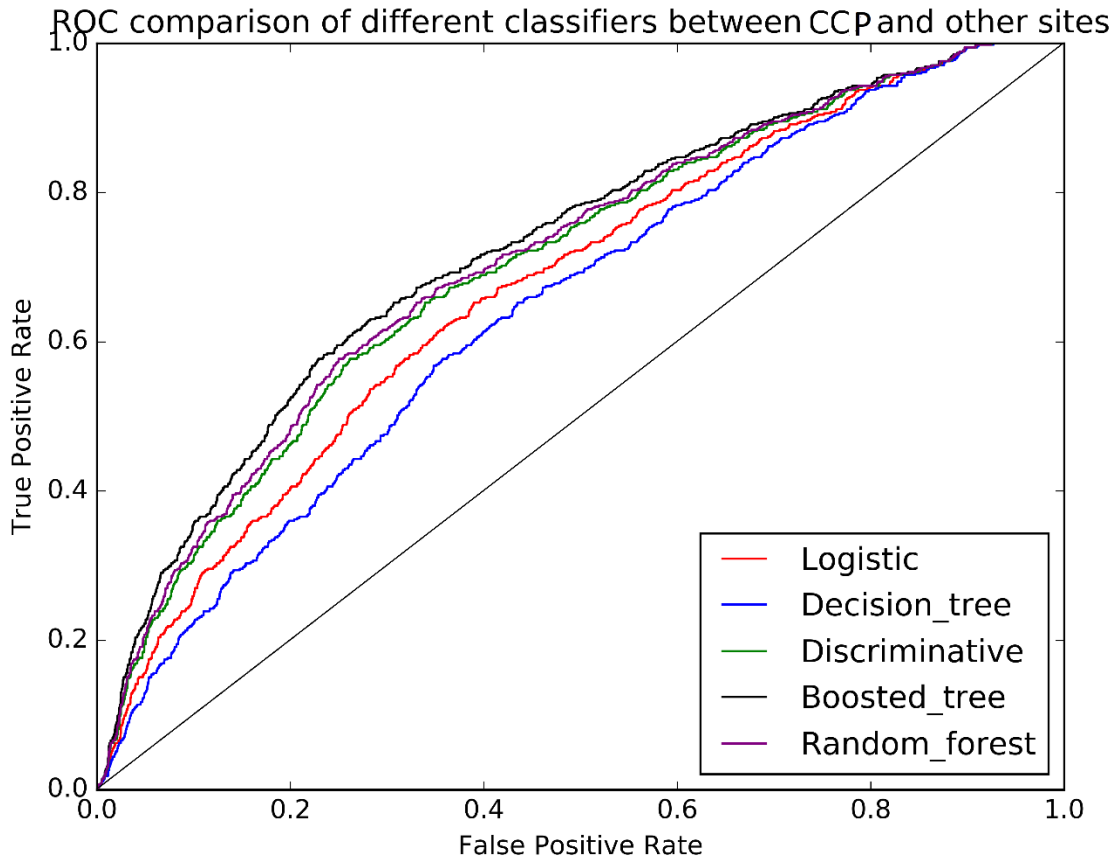


Figure 6-12 This figure represents the ROC of all 5 classifiers (logistic regression, decision tree, random forest, boosted tree classifier) for classification between CCP and other tissue sites.

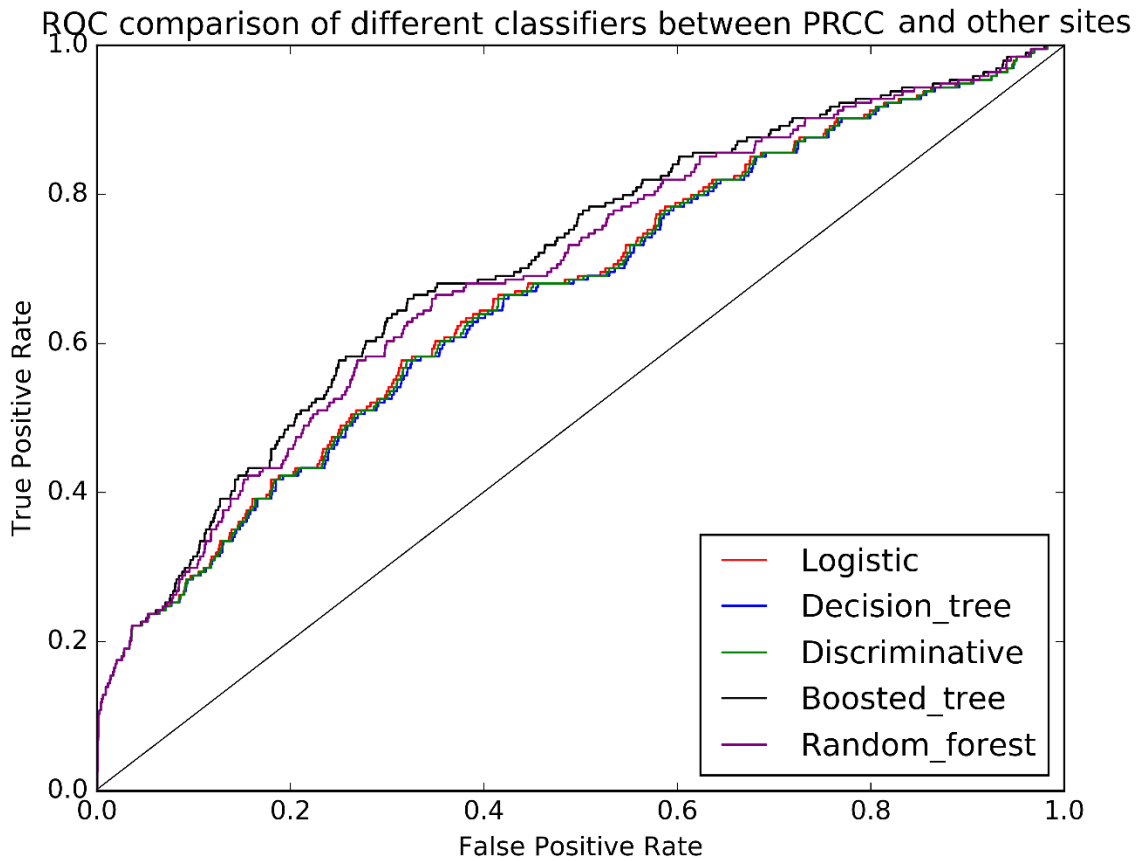


Figure 6-13 This figure represents the ROC of all 5 classifiers (logistic regression, decision tree, random forest, boosted tree classifier) for classification between PRCC and other tissue sites.

6.5 Discussion and Future Directions

There are a few avenues to investigate to improve the performance of our algorithm, for example:

- 1) Extracting other clinically relevant features such as:
 - (a) cell/nuclei feature: There are atypical features indicating of CCRCC, for example, epithelial cells with clear cytoplasm, prominent nuclei, and distinctive cell membrane.
- In the previous chapters, we showed a custom-designed automated segmentation

- algorithm for prostate tissue, this method can also be tuned for renal carcinoma to locate nuclei and later extract nuclei features.
- (b) Population-based features such as stroma and cell sociology feature: Stroma structures and features, as well as the formation and orientation of the cells, could be among the features to investigate.
 - (c) Other statistical quantities such as mean, median, entropy, size, shape, compactness, major/minor axes could be used to increase the accuracy.
- 2) Using other stains such as Thionin: Nuclei/ blood specific stains could be used to specially stain the desired structure and improve the segmentation process.
 - 3) Applying different color de-convolution method: More sophisticated color maps such as LAB or clustering methods such as K-means clustering or other methods such as principal component analysis (PCA) could be used to improve the performance.
 - 4) Collecting more data: in general, in the ML domain, more data usually would result in more accurate outcomes. The size of the data that we had for this study was very limited and wasn't a good representation of the whole population. In the future, we plan to collect more data to better generalize our algorithm.

The most important limitation of this study was related to the small number of images, resulting in imbalanced dataset (the number of normal cases was very small). To overcome this issue we could potentially use cost-sensitive metrics or to apply up-sampling (for the minority class) and down-sampling for the (majority class).

Table 6-2 potential feature list for renal carcinoma study

<i>Image Representation</i>	<i>Quantitative Histomorphometric Features</i>
<i>Individual Cells</i>	Area, Mean, Standard Deviation, Elliptical Features: Minimum-Maximum axes, length Orientation Convex Hull: Convex areas Boundary Features: perimeter, radii, FFT, etc.
<i>Cell Sociology (e.g. using minimum Spanning Tree)</i>	Edge Length: Average, Standard Deviation, Minimum-Maximum Ratio, Disorder, etc
<i>Vascular Pattern (E.G. using graph-based models)</i>	Total polygon Area, Average polygon Area, Number of terminal (or end) branches, Average length of the terminal branches, number of junctions, the average length of the non-terminal branches, and different ratios, etc.
<i>Radiometric and Densitometry</i>	Image Bands, Intensity Optical density, integrated optical density, Mean optical density, etc.
<i>Texture</i>	Hue, co-occurrence matrix features (gray and optical density), Inertial, energy, entropy, homogeneity, maximum probability, cluster shade, cluster prominence, correlation, etc.
<i>Image Representation</i>	Quantitative Histomorphometric Features

Chapter 7 Cervical Cancer Epidemiology, Screening and Diagnostic Techniques

Cervical cancer is the fourth common diagnosed cancer in women worldwide (*Cervical cancer statistics*. 2018). Although it is a preventable disease, approximately 500,000 patients with cervical cancer are diagnosed worldwide every year and about half of these people may succumb to the disease (*Cervical cancer statistics*. 2018). Despite the introduction of the advanced diagnostic and prognostic methods/tools, it's been predicted that the occurrence of cervical cancer will increase due to lifestyle, diet, smoking, and aging of the population. In developing countries, cervical cancer is often the most common cancer in women and may constitute up to 25% of all female cancers. Countries with High Human Development Index (HDI) have a three times higher 5-year survival rate comparing to countries with low HDI (David Forman, Catherine de Martel, Charles J Lacey 2012). Cervical Intraepithelial Neoplasia (CIN) or Squamous Intraepithelial Lesions (SIL) is a cervical cancer precursor which can take 3 to 20 years to progress into cancer. It is very important to diagnose cervical cancer at an early stage when the treatment is effective. Unfortunately, in developing countries, cervical cancer is frequently diagnosed at a late stage which results in a high mortality rate (Botha 2015).

Cervical cancer incidence rates have dramatically decreased since the introduction of population-based organized cytological screening programs, using the Papanicolaou (Pap) smear, in western countries. A Pap-smear screening program detects potential precancerous changes in the cervical tissue which may be caused by certain types of human papillomavirus (HPV)⁸. In the presence of abnormalities in the Pap smear, patients are referred for examination with colposcopy followed by

⁸ An HPV infection is caused by human papillomavirus, which is a DNA virus from the papillomavirus family. Over 170 types of HPV are known most which are harmless and resolved eventually, there are certain types, however, which may persist and cause precancerous lesions (*Human papillomavirus infection*. 2018)

possible biopsies and lesion treatments if necessary. This procedure can decrease the cervical cancer mortality rate; however, this labor-intensive process might not be accessible in developing countries (Chang, Follen et al. 2002a).

Despite advanced screening tests, cervical cancer is still the leading cause of cancer-related death in women in developing countries (Origoni, Prendiville, et al. 2015), which could be due to limited screening resources. Although the need to develop a more advanced detection method is more viable in developing countries, such a method can also be beneficial to developed countries as an adjunct to current screening procedures to increase the overall reliability and improve the detection outcomes.

Optical technologies such as fluorescence spectroscopy have been extensively investigated as an effective and non-invasive method for cancer screening. It has been shown that morphological, cytological and histopathological changes can be quantified using fluorescence and reflectance imaging (Milbourne, Park et al. 2005a). We are interested in developing optical technologies for cervical cancer screening for the developing world.

Fluorescence spectroscopy has showed promising results for the detection of precancerous lesions. For example, with the application of fluorescence imaging, it is possible to clinically visualize invisible cancer/pre-cancer locations in the oral cavity (Lane, Pierre, Poh, et al. 2011, Poh, Ng et al. 2007, Lane, Pierre M., Gilhuly, et al. 2006). Similar results have been obtained from the lung (Lam, S., MacAulay, et al. 1993) and gastrointestinal tract (Zeng, Haishan, Weiss, et al. 1998). More importantly, several clinical trials have used the same approach for the real-time detection of CIN (Ferris, Lawhead, et al. 2001, Ramanujam, Mitchell, et al. 1994). Another important technique is point spectroscopy (which uses a fiber optic probe). This technique showed 86% and 74% sensitivity and specificity respectively (Ramanujam, Mitchell, Mahadevan-Jansen, et al.

1996, Ramanujam, Mitchell, Mahadevan, et al. 1996, Ramanujam, Mitchell, et al. 1994) ,however, as larger field of view is very critical in cancer imaging and can help with cancer diagnosis, field spectroscopy is being conducted in the lung, bladder, skin, and cervix (Weingandt, Stepp et al. 2002, Nordstrom, Burke et al. 2001, Zeng, H., MacAulay et al. 1998, Lam, Stephen, MacAulay et al. 2000, Mourant, Bigio et al. 1995). Our group at BC Cancer Research Center has developed a Multispectral Digital Colposcopy (MDC) technique to visualize the whole cervix. We used different light wavelengths chosen to have the most diagnostic outcome based on a large point spectroscopy trial (Benavides, Chang et al. 2003a). This device can also be used adjunct to current colposcopy device as a guide for better, more accurate biopsies in the developed countries with more robust screening infrastructure.

7.1 Objective

The objective of this chapter is to design an automated classification method using classical machine learning techniques that use the reflectance and fluorescence images of the cervix and localize the at-risk cervical cancer tissue within these images. In this chapter, we discuss the design process to implement this analytic tool which could be beneficial for cervical cancer screening and diagnostic procedures.

The technique presented here incorporates MDC images and apply ML techniques to eliminate the need for further biopsies or to improve current colposcopy devices by making the detection process faster and more accurate.

MDC images which are captured in different bandwidth along with image processing and advanced machine learning techniques are blended together to provide a more powerful detection tool. This new analytics tool would help with the current standard care system, improve the diagnostic process, and increase the life quality of patients.

7.1.1 Specific Objective

Annotating the MDC image data based on the level of agreement of a group of expert reviewers who were asked to assess the histopathology tissue maps and trace the locations of abnormal tissue in the images.

Defining and selecting features with the most discrimination capabilities.

Developing advanced machine learning algorithms such as boosted tree and convolution neural network to evaluate the MDC images.

The workflow for this chapter is presented in Figure 7-1. The first step is to image the tissue, followed by image/ label preparation. The next step is to select features and implement ML technique (training and testing the dataset) and finally, parameter optimization to improve the performance.

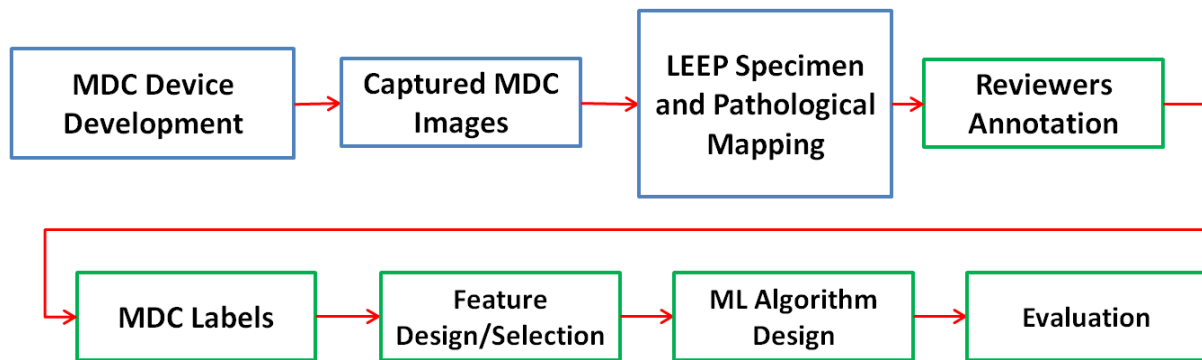


Figure 7-1. This figure shows the pipeline of our study. The processes mentioned in the blue boxes are completed and fully developed by our group. The processes illustrated in green boxes are the focus of this chapter.

7.2 Materials and Methods

7.2.1 Instrumentation

The MDC is a device that combines the image of the whole cervix and a point probe imaging. The analyses presented in this chapter refers only to the whole cervix images. This device uses cross-polarized white-light illumination, reflectance, and fluorescence imaging. The system consists of a colposcope, and an RGB color charge-coupled device (CCD) (MicroPublisher 3.3, QImaging) to capture and filter fluorescent and reflectance light and a light source, which is a Xenon arc lamp in the base of the device. The lamp provides monochromatic and broadband illumination. The fluorescence excitation light is produced using band-pass filters enclosed in a motorized filter wheel.

Figure 7-2 shows the diagram of the system, a photograph of the system and the specifications of the reflectance and fluorescence light used in this device. These excitation and emission wavelengths were obtained from a previously reported study (Chang, Follen et al. 2002b). Previous versions of the device generated automated diagnoses of 80% sensitivity and 80% specificity for the diagnosis of high-grade SIL (Windeatt, Ardeshir 2002, Milbourne, Park et al. 2005b, Park, Collier et al. 2005, Benavides, Chang, et al. 2003b).

For this study, the MDC devices were calibrated before each patient measurement.

A diffuse reflectance standard made from sintered PTFE (Polytetrafluoroethylene) was used to white balance and calibrate the reflectance pathways of the system. Similarly, a fluorescence standard, composed of the same material but doped with two inorganic fluorophores was used to calibrate the fluorescence pathways, Table 7-1.

Fluorescence imaging detects metabolic changes in tissue including changes in NADH and FAD as well as alteration in tissue structure such as collagen and elastin content. Blue excitation

wavelength is used to monitor FAD and collagen, UV excitation is used to monitor NADH. Reflectance imaging can show information regarding the presence and saturation of hemoglobin, as well as the stroma vascularity (angiogenesis) (Freeberg et al. 2007).

Table 7-1 Tiles used for MDC device calibration

<i>Tile</i>	<i>Reflectance</i>	<i>Fluorescence</i>
<i>Measurement</i>		
<i>Reflectance</i>	Positive Control	NA
<i>Fluorescence</i>	Negative Control	Positive Control

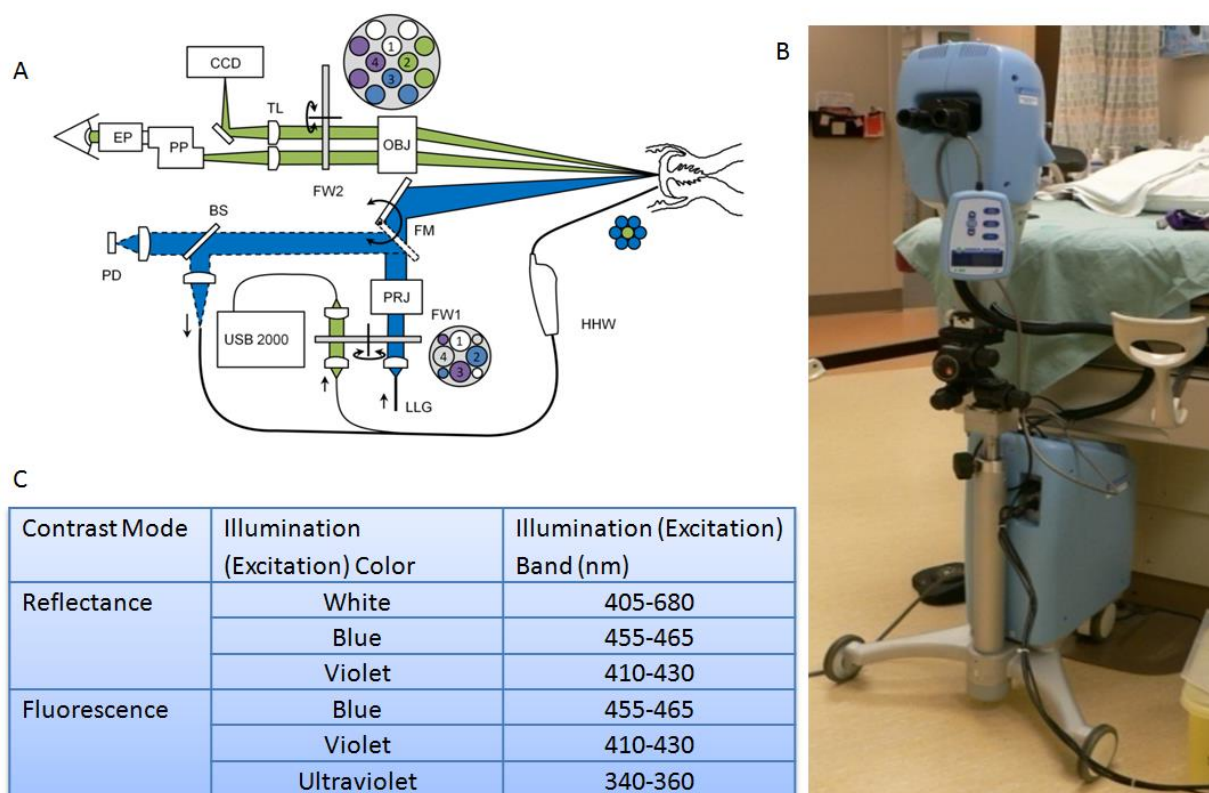


Figure 7-2 This figure shows **A)** a schematic diagram of the Multispectral Digital Colposcope (MDC) device, **B)** a photo of the “in house” device, and **C)** the table of the illumination color and excitation wavelengths.

7.2.2 Clinical Data

49 patients were recruited for this study. Each patient was to undergo standard of care Loop Electrical Excision Procedure (LEEP). Six images were taken using the MDC; three reflectance images and three fluorescence images.

7.3 Methodology-Conventional ML from Data Preparation to Data Analysis

7.3.1 Data Preparation

The study design is shown in Figure 7-3. The study protocol was reviewed and approved by the Institutional Review Board (IRB) at the British Columbia Cancer Agency.

During the colposcopic examination of each patient, acetic acid (6%) was applied to the cervix for 2 minutes. Six images were taken using the MDC; three reflectance images and three fluorescence images.

The colposcopy physician was asked to denote in the white light image 1) margins/edges of the LEEP specimen 2) areas which were thought to be most abnormal.

The removed specimen (via LEEP) was oriented by the surgeon. The LEEP pieces were sectioned and reviewed by a clinical pathologist for final diagnosis and annotated by the study pathologist. Each section was carefully marked identifying areas of low-grade SIL, high-grade SIL, and cancer by an expert cervical cancer pathologist, the description of patients' CIN level is presented in Table 7-2.

Table 7-2 CIN level of patients for this study.

<i>patients</i>	<i>CIN1</i>	<i>CIN2</i>	<i>CIN3</i>	<i>Invasive & CIN3</i>
49	7	16	26	2

Preserving the orientation of each section from each piece helped us to reconstruct the histopathologic map which served as the gold standard for the rest of this study.

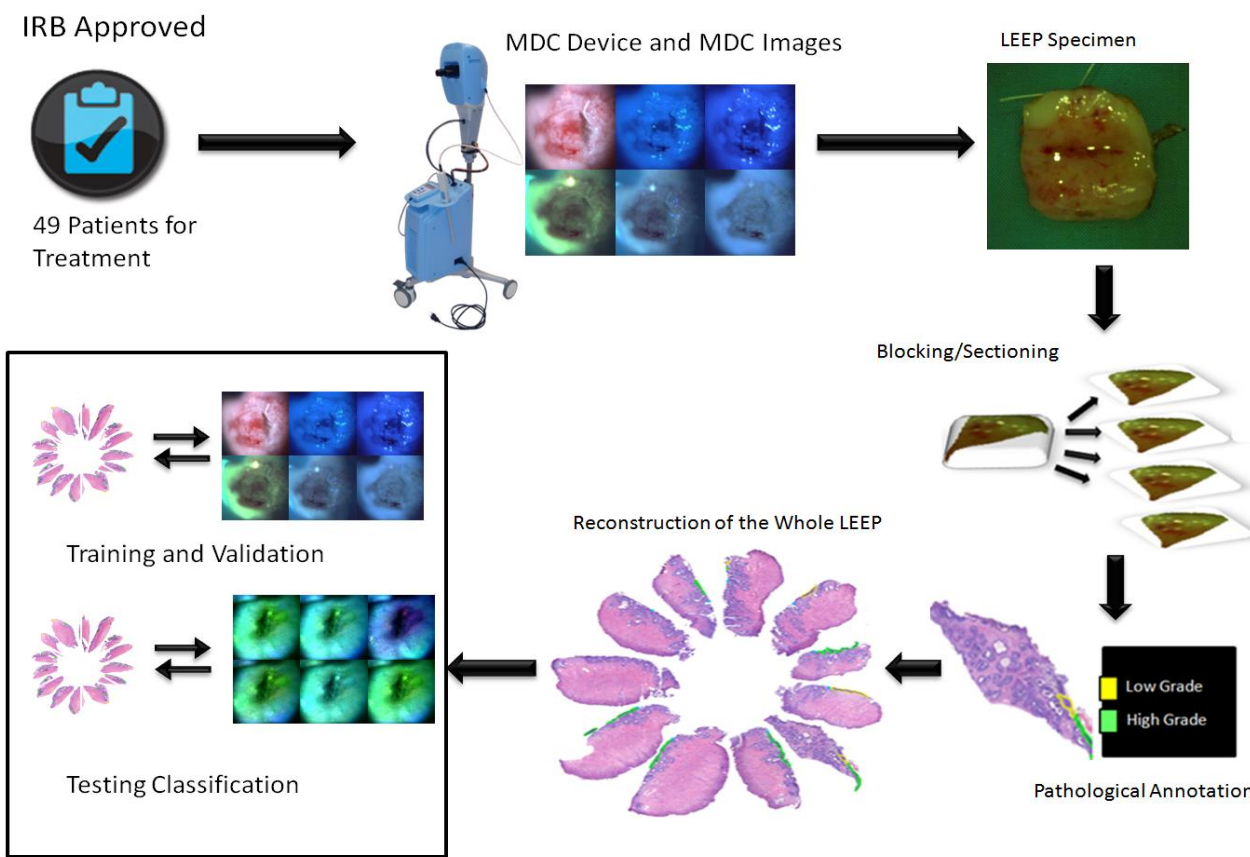


Figure 7-3. This figure shows the study design. Patients were consented for an IRB-approved study, images were acquired, the Loop Electrical Excision Procedure (LEEP) was carried out, the specimen was processed and annotated, and the dataset was subjected to analysis for training and testing classification

7.3.2 MDC Images and Pathological Map

A group of 11 experts were tasked to trace the areas (Region of Interest, ROI) where they believed were associated with high-grade disease based upon only the reconstruction pathology map, Figure 7-4 (a). These 11 traces were superimposed on top of each other, Figure 7-4 (b) and regions of most agreement are shown in the heat-map image of Figure 7-4 (c). Hot regions (glowing yellow) indicate the highest level of agreement on the presence of abnormalities (these are the regions that we are interested to detect via an automated tool). From this image, we generated a mask (description comes later) which serves as the pixel-wise label for each image.

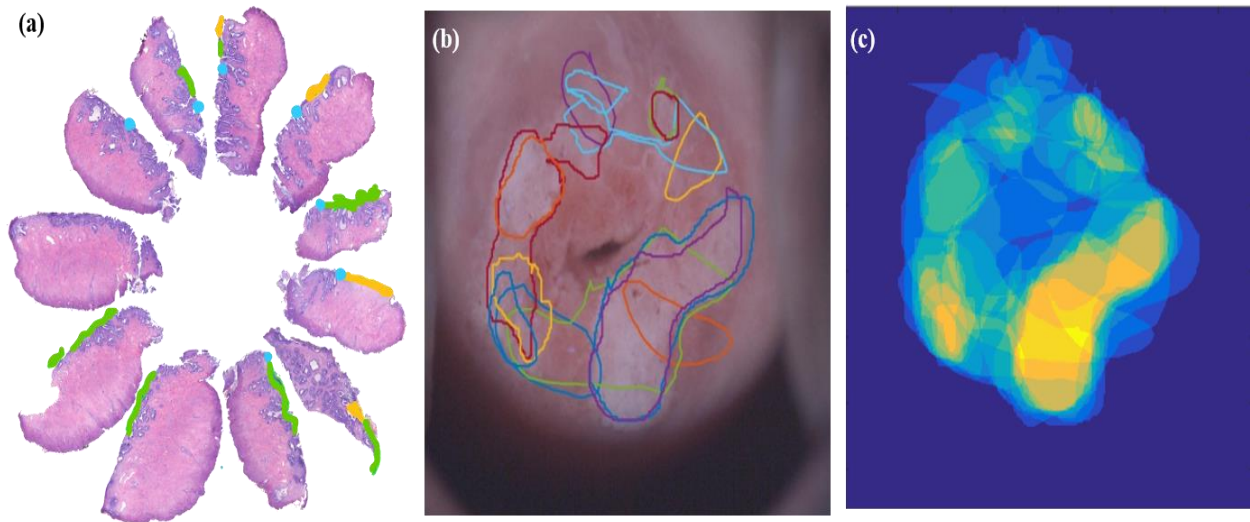


Figure 7-4. The protocol to select regions of interest. (a) pathology map, where the green yellow and blue annotations represent high grade, low grade, and squamous columnar junction respectively. (b) reviewers selected the region of interest (each color boundaries represents a reviewer's drawing) on the RGB image of MDC using the image provided in (a). (c) shows the heatmaps where the hot or glowing areas represent the most agreement between reviewers' selections of ROIs.

For comparison, in the next image (Figure 7 5) we selected four patients and we show the clinical impression of the cervix (first row), cervical pathological annotation, marker by the expert cervical pathologist (second row), traces of abnormalities drawn by our reviewers (third row) and generated heatmaps and mask (fourth and fifth row respectively) constructed from the reviewers' selections. The mask, referred to as the agreement mask, is generated from the level of agreement between the reviewers. Less than 20% agreement on the abnormalities is considered as normal and shown as black areas. Between 20% and 60% agreement is considered as "don't care" areas and shown in gray, and finally above 60% agreement is considered as abnormalities and shown in white.

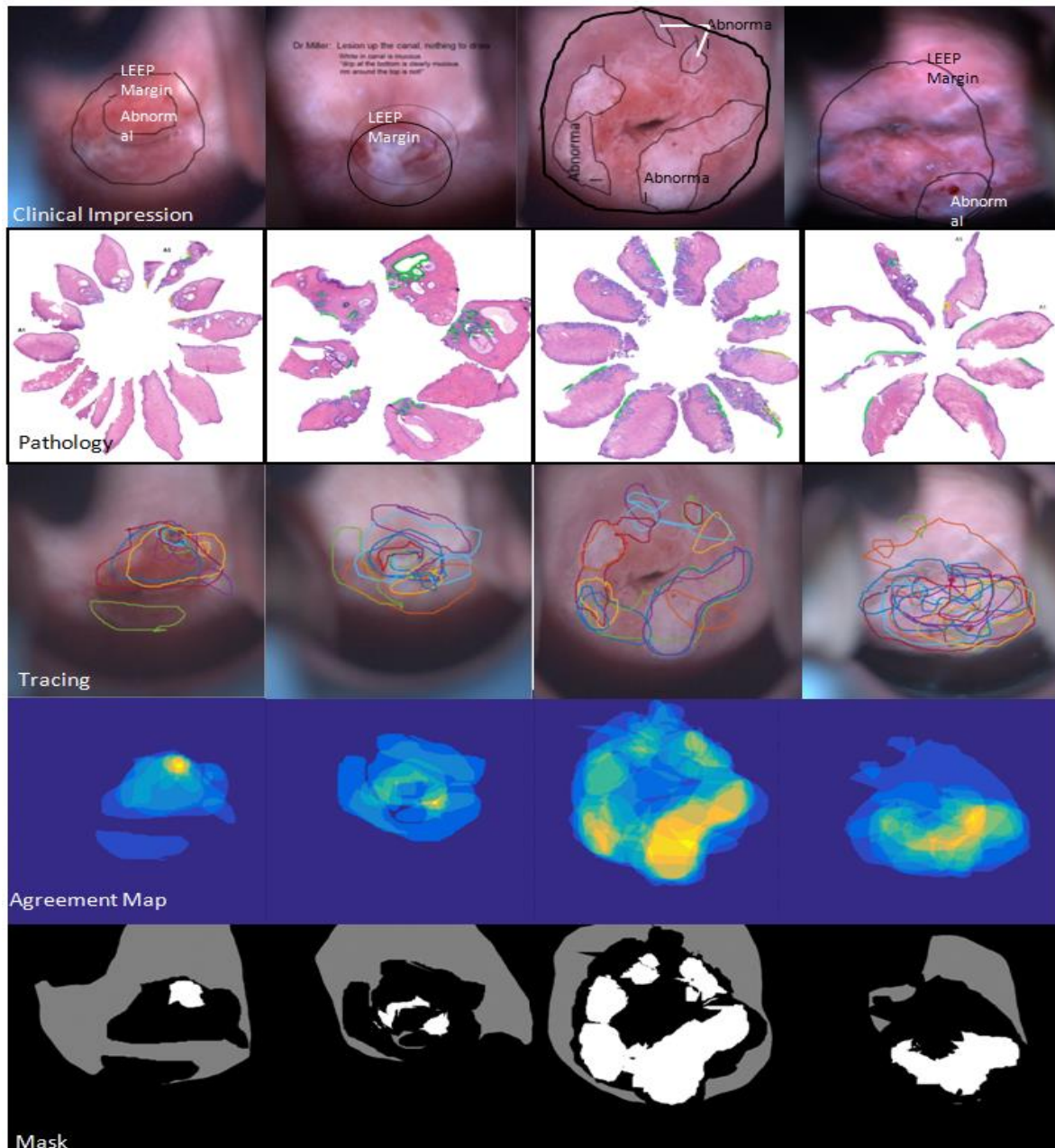


Figure 7-5 This figure shows four patients' results. In each column, we presented the clinical impression of the colposcopy physician, the histopathologic map, the tracing by the 11 reviewers, the agreement map, and the mask generated by the calculation of 60% agreement.

7.3.3 Feature Extraction and ML Algorithm Design

7.3.3.1 Feature Extraction

The next step in our pipeline is developing a classifier to differentiate between normal and low-grade CIN vs. high-grade CIN. This step consists of two main tasks:

- 3- Implementing /selecting the features;
- 4- Developing different machine learning methods using the features selected from the previous step.

7.3.3.2 Design/Select the Features

We designed a 10× 10 sliding window to scan the whole image and used the histogram intensities of the image in each sliding window as our features (20 bins are used to generate the histogram intensities for each image). Histograms are a good choice for the feature because they would remove the spatial sensitivity of the data and only uses the statistical characteristics of image intensity distribution such as mean and variance.

We divided our dataset into two sets: 80% of patients (39 patients) were used for training-validation and 20% (10 patients) were used as the test set. We used different classifiers to identify the normal vs. abnormal regions. Using a boosted-tree we obtained the best results (shown in Figure 7-6). We constructed the feature space using equation (7,1).

$$F(x) = [F_1(x) + F_2(x) + \dots + F_C(x)] \quad (7,1)$$

where C is the number of channels (in our study we had 18 channels). Each channel as defined below is an ensemble of weak classifiers with a fixed depth (d) (in our study we used d equal to 6). The input of our classifier is the features defined as above (histogram of intensities for all MDC images) represented by x . The output vector is defined as y :

$$Y_k(x) = \text{sign} \sum_{n=1}^M w_n h_n(x), n = 1, 2, \dots, M \quad (7,2)$$

where h stands for the estimated probability and M is the number of week classifiers.

Based on a voting strategy, each classifier predicts a label (Windeatt, Ardeshtir 2002), in other words, a label is predicated for each pixel based on the majority votes of the classifiers. We also compared our results with the ensemble classifier that consisted of four classifiers, a linear classifier with Euclidian distance, a linear classifier with Mahalanobis distance, a K -nearest neighbor (KNN) classifier with eight neighbors, and a Support Vector Machine (SVM) with a linear kernel (Young Park, Follen et al. 2008).

Figure 7-6 shows the strategy for our analyses. As shown in this figure, a probability map and a Receiver Operating Characteristic (ROC) curve can be generated for pixels of each image.

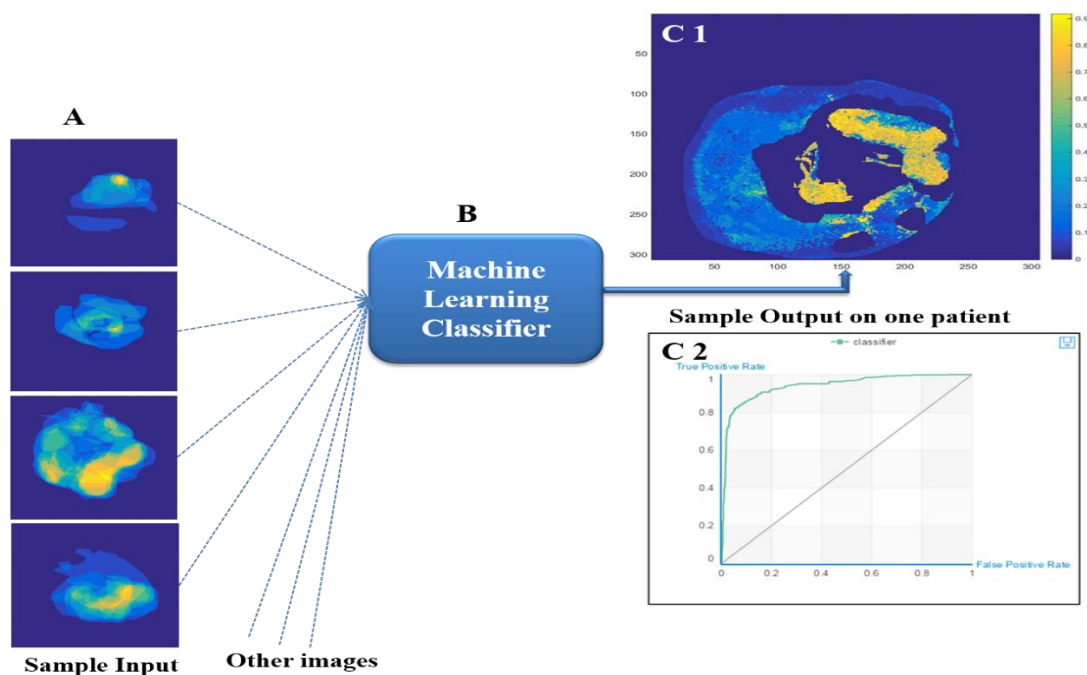


Figure 7-6 This figure shows the strategy for the analysis. Eighty percent of the data was used to train and validate a classifier, shown in A. The classifier was tested on the remaining data (B). Output from a sample patient is shown in C1 and C2. The probability of disease is shown in C1 and the Receiver Operating Characteristic (ROC) curve is shown in C2.

7.4 Experiment and Result

The classification results on both test and training sets are summarized in Figure 7-7 and Figure 7-8 respectively. We obtained 93% and 91% accuracy when we applied a boosted-tree classifier to the training and test set respectively. Area Under the Curve (AUC) was 0.99 and 0.86 for training and test set respectively. Details of the statistics calculated on both test and training datasets is given in Table 7-3 and Table 7-4 respectively.

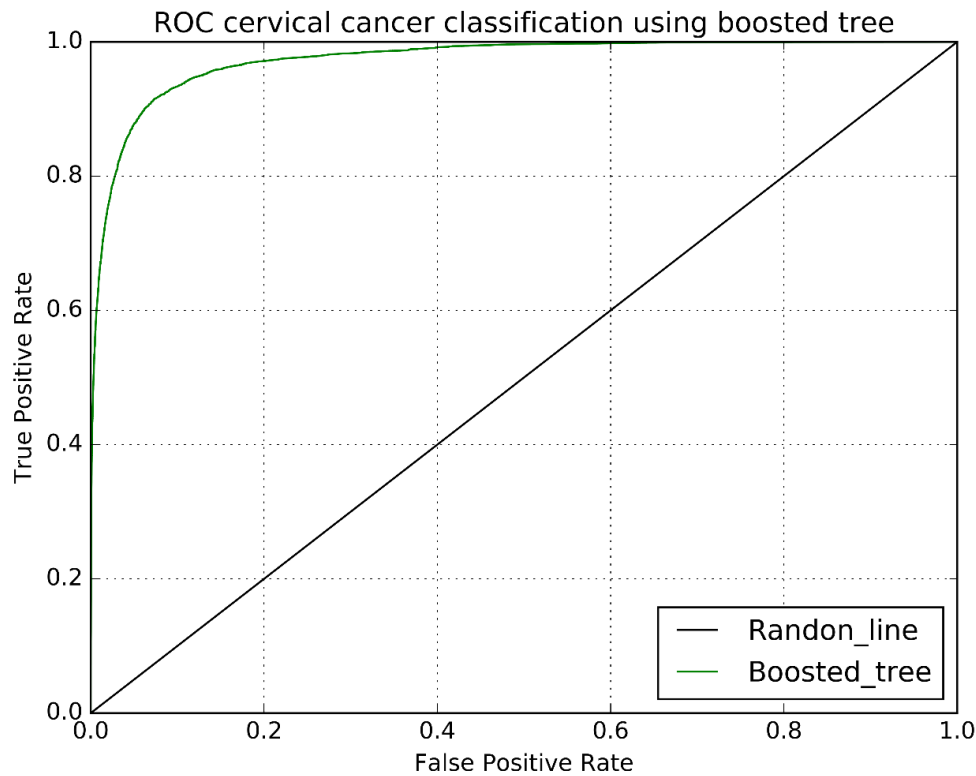


Figure 7-7 This figure represents the ROC of the boosted tree classifier on the training set (using 80% of the data (for 39 patients)) and using histogram intensities as the features.

Table 7-3 The statistics results using a boosted tree classifier on the training set, using histogram intensities as the features.

True Positive	False Negative	Accuracy	Precision	Sensitivity (TPR)
30,992	7,128	0.93	0.89	0.81
False Positive	True Negative	Recall	F1 Score	Specificity (TNR)
3,711	116,002	0.81	0.85	0.97

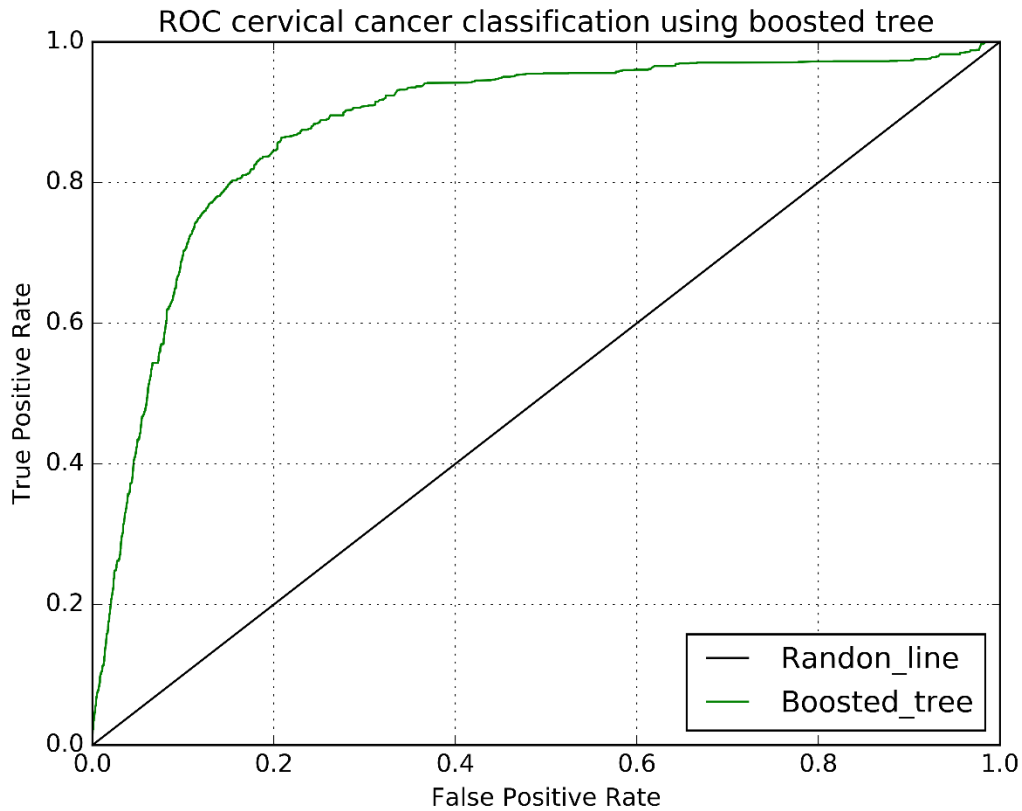


Figure 7-8 This figure represents the ROC of the boosted tree classifier on the test-set (using 20% of the data (for 10 patients)) using histogram intensities as the features.

Table 7-4 The statistics results using a boosted tree classifier on the test set, using histogram intensities as the features.

True Positive	False Negative	Accuracy	Precision	Sensitivity (TPR)
1,070	2,360	0.91	0.47	0.31
False Positive	True Negative	Recall	F1 Score	Specificity (TNR)
1,225	34,803	0.31	0.37	0.97

To observe the performance of the ML algorithm per image, we generated the heatmaps of the probability of each pixel within each image. We can see in Figure 7-9 that our method is very sensitive on the edges where we can see fuzzy predictions, Figure 7-9 (second row). Locations close to the edges are usually harder to predict since even the reviewers were not sure of their selections.

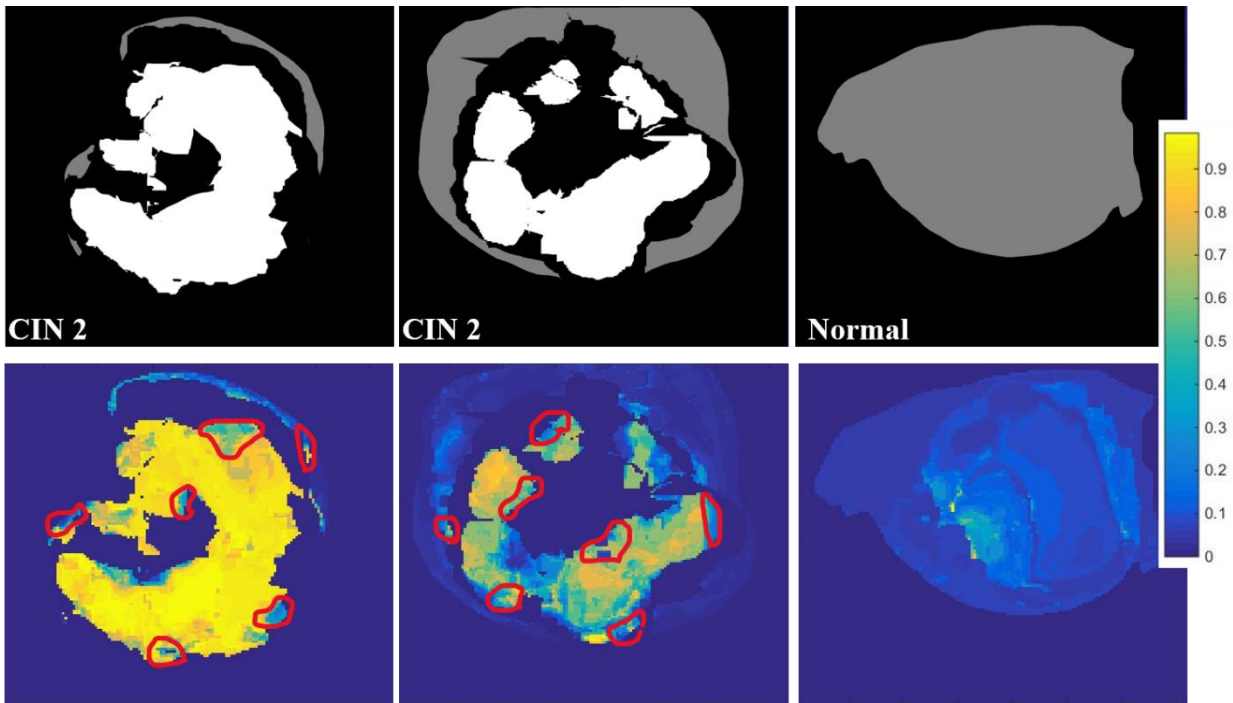


Figure 7-9 The algorithm performance per image. The first row is the label, where white, gray and back represents abnormal, normal and don't-care regions. The second row represents the probability prediction for each pixel within an image. Red drawing illustrates the fuzzy regions such as edges where even the reviewers might be unsure of the selection.

7.4.1 Improving the Performance

7.4.1.1 Define and Select Features with the Most Discrimination Capabilities.

Based on the findings of previous studies (Milbourne, Park et al. 2005) we identified a few ratios that have higher discriminative capabilities. To improve the performance of our algorithm we decided to first investigate these ratios and based on their performance decided whether or not to integrate these features with the features that we designed before (histogram of the intensities within each sliding window).

As we explained earlier, the 18 channels construct six RGB images as shown in Table 7-5, from these images we calculate the new features based on the ratios presented in equation (7,3) to (7,13).

Table 7-5 Table of MDC channel names and acronyms.

<i>Channel No.</i>	<i>Channel Name</i>	<i>Acronym</i>
1	RGB while image	White
2	Blue Reflectance	BRF
3	Violet Reflectance	VRF
4	Blue Excitation	BE
5	Violet Excitation	VE
6	UV Excitation	UVE

$$BEG_WRFR = \frac{BE_{green}}{White_{red}} \quad (7,3)$$

$$VEG_WRFR = \frac{VE_{green}}{White_{red}} \quad (7,4)$$

$$VEG_BER = \frac{VE_{green}}{BE_{red}} \quad (7,5)$$

$$UVEB_VEB = \frac{UVE_{blue}}{VE_{blue}} \quad (7,6)$$

$$VER_BER = \frac{VE_{red}}{BE_{red}} \quad (7,7)$$

$$VEB_VRFB = \frac{VE_{blue}}{VRF_{blue}} \quad (7,7)$$

$$VEG_VRFB = \frac{VE_{green}}{VRF_{blue}} \quad (7,8)$$

$$BEG_VEG = \frac{BE_{green}}{White_{red}} \quad (7,9)$$

$$UVEG_WRFR = \frac{UVE_{green}}{White_{red}} \quad (7,10)$$

$$UVEG_VRFB = \frac{UVE_{green}}{VRF_{blue}} \quad (7,11)$$

$$UVEB_VRFB = \frac{UVE_{blue}}{White_{blue}} \quad (7,12)$$

$$VEG_WRFG = \frac{VE_{green}}{White_{green}} \quad (7,13)$$

Using only the above 12 features and gradient boosting classifier, we constructed the feature importance image (Figure 7-10). From this image we can see that VEB_VRFB (violet excitation blue channel fluorescence emission pixel-wise normalized by violet illumination detected by the camera's blue channel) is the most important feature selected by the algorithm, next important one is VEG_WRFR (violet excitation green channel fluorescence emission pixel-wise normalized by white illumination detected by the camera's red channel). The ROC result of boosted tree classifier (using only these ratio features) on the training and test-set are depicted in Figure 7-11 and Figure 7-12 respectively.

The statistics results on the training set and the test set are shown in Table 7-6 and Table 7-7 respectively. The accuracy dropped on both training and test-set by using these features (comparing with the boosted-tree classifier that uses histogram intensities as the features). To investigate further and observe the effectiveness of ratio feature along with histogram intensities features we combined all features together (the result is presented later in this chapter),

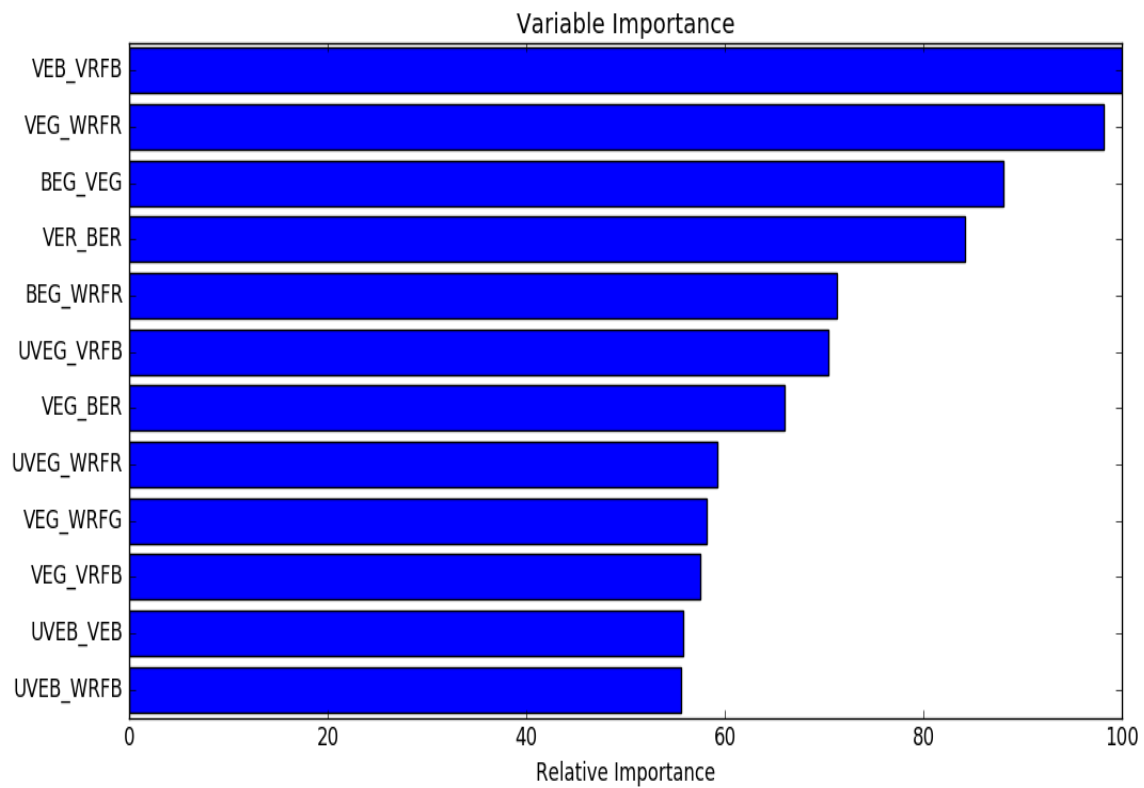


Figure 7-10 This image shows the importance of each feature using our algorithm when only ratio features are used.

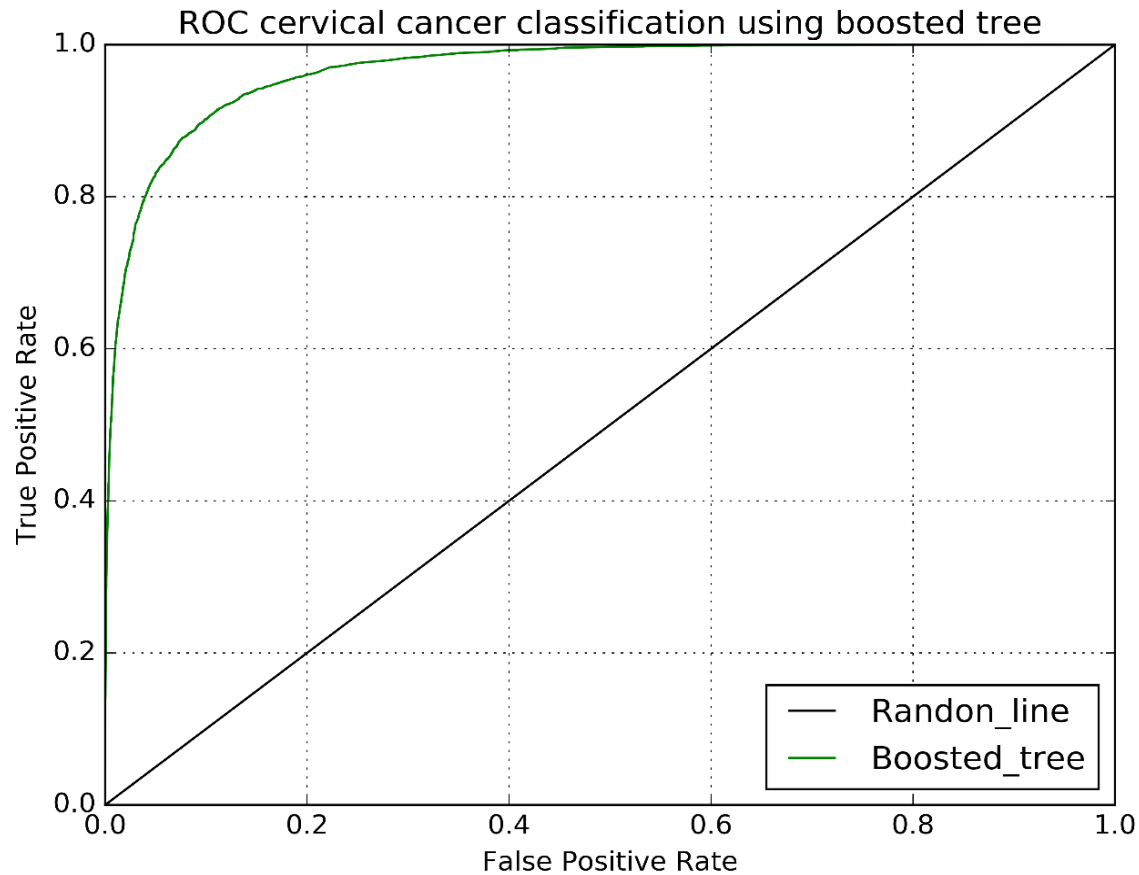


Figure 7-11 This figure represents the ROC of the boosted tree classifier on the training set and (using 80% of the data (for 39 patients)), only ratio features are used.

Table 7-6 The statistics results using a boosted tree classifier on the training set.

True Positive	False Negative	Accuracy	Precision	Sensitivity (TPR)
29619	8501	0.92	0.879	0.777
False Positive	True Negative	Recall	F1 Score	Specificity (TNR)
4070	115643	0.777	0.825	0.97

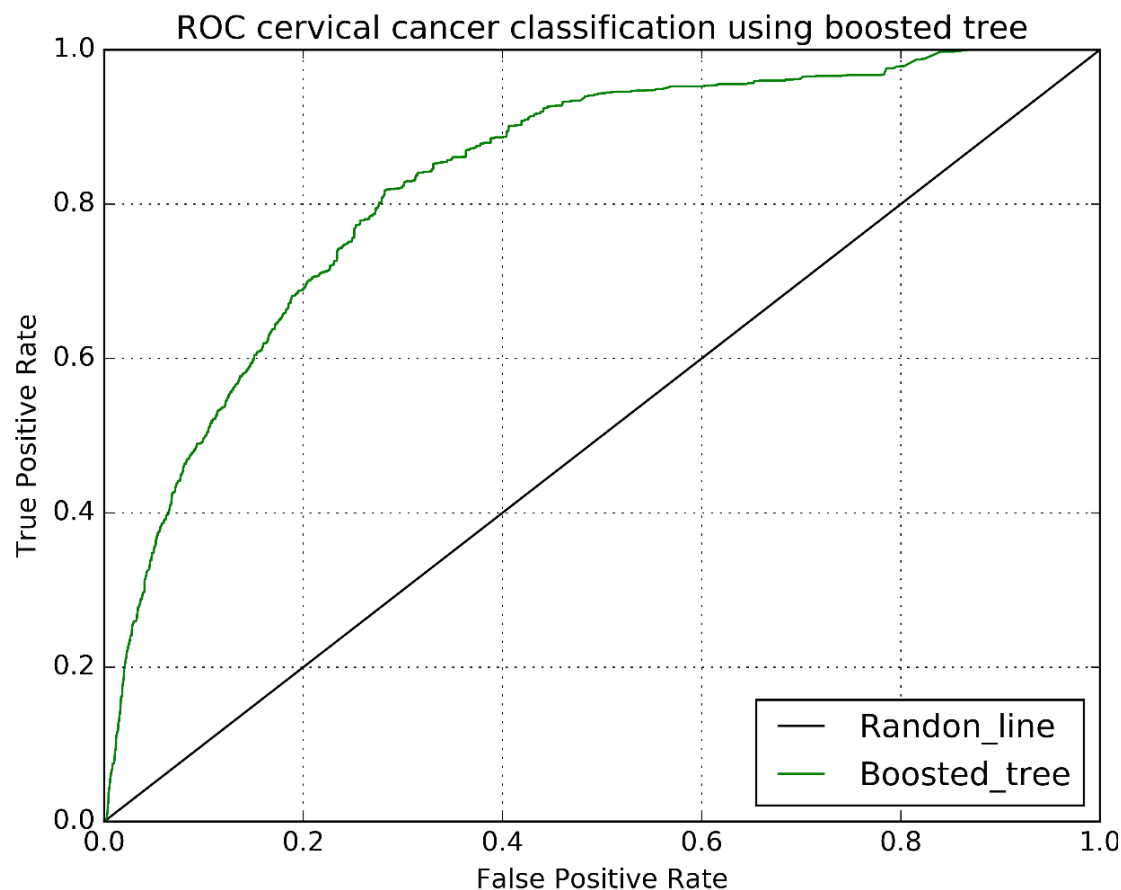


Figure 7-12 This figure represents the ROC of the boosted tree classifier on the test-set (using 20% of the data (for 10 patients))), only ratio features are used.

Table 7-7 The statistics results using a boosted tree classifier on the test set.

True Positive	False Negative	Accuracy	Precision	Sensitivity (TPR)
1499	1931	0.885	0.366	0.437
False Positive	True Negative	Recall	F1 Score	Specificity (TNR)
2594	33434	0.437	0.399	0.93

Next, we combined all features together to form the feature-set and used the same classifier. Figure 7-13 shows the selection of important features. The top five selected features are, VEG_WRFR,

VEB_VRFB, BEG_WRFR, UVEB_WRFB, and 7 (indicates bin 7 of histogram intensities capturing from RGB image).

The ROC result of using all features (ratio features and histogram intensities features) on the training and test-set please are given in Figure 7-14 and Figure 7-15 respectively.

The statistics results on the training set and the test set are shown in Table 7-9 and Table 7-8. We obtained 94% accuracy and 91% accuracy on training and test-set respectively, which is an improvement compared with using only ratio or histogram intensities as the features.

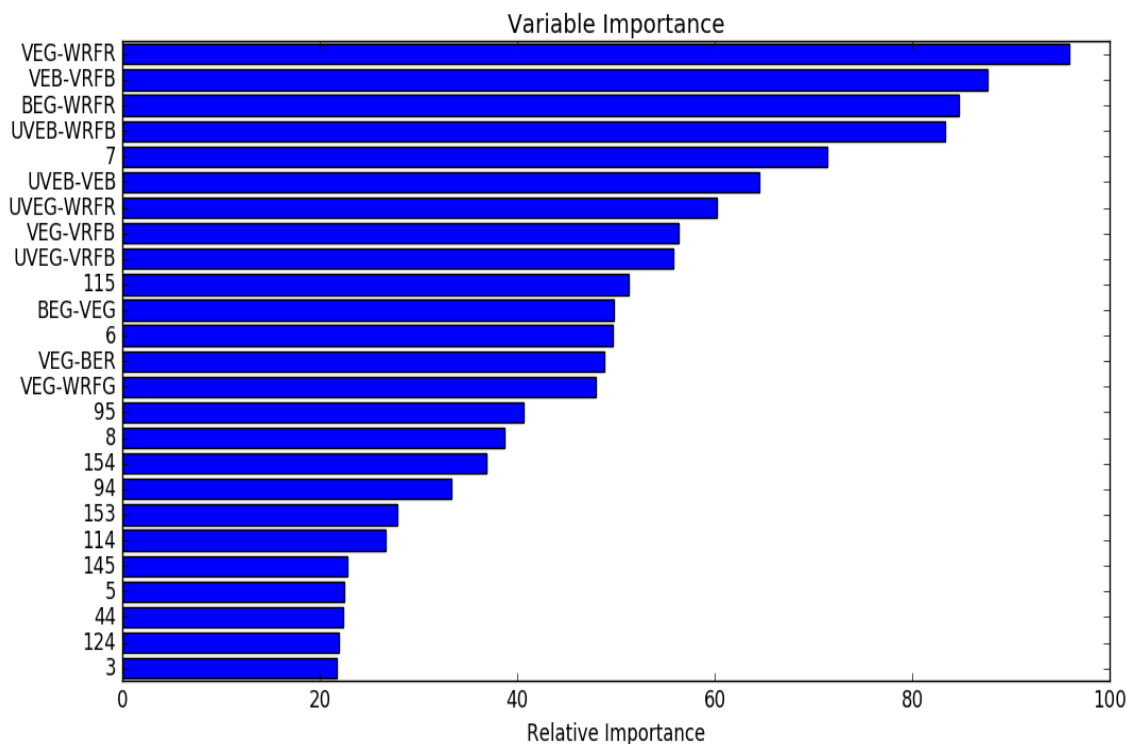


Figure 7-13 This image shows the importance of each feature using our algorithm when all features are used.

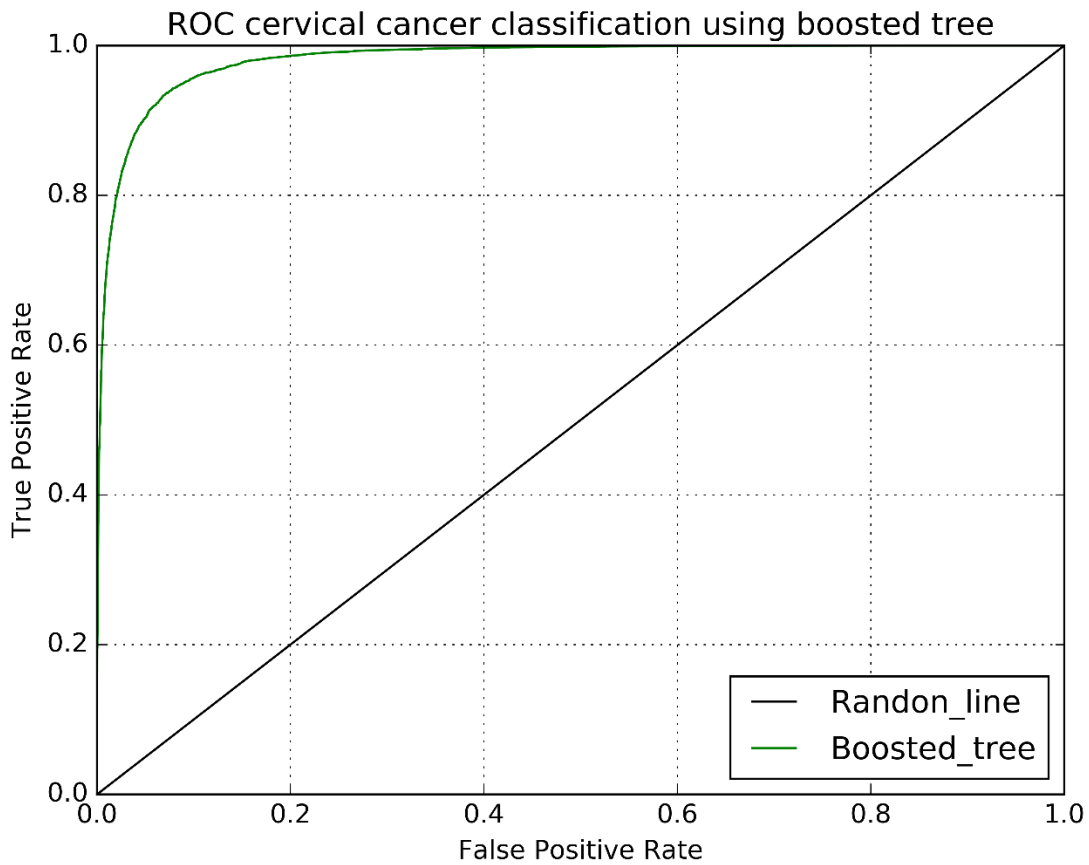


Figure 7-14 This figure represents the ROC of the boosted classifier on the training set (using 80% of the data (for 39 patients)), using all features(ratio and histogram intensities).

Table 7-8 The statistics results using a boosted tree classifier on the training set.

True Positive	False Negative	Accuracy	Precision	Sensitivity (TPR)
32211	5909	0.941	0.903	0.845
False Positive	True Negative	Recall	F1 Score	Specificity (TNR)
3472	116241	0.845	0.873	0.97

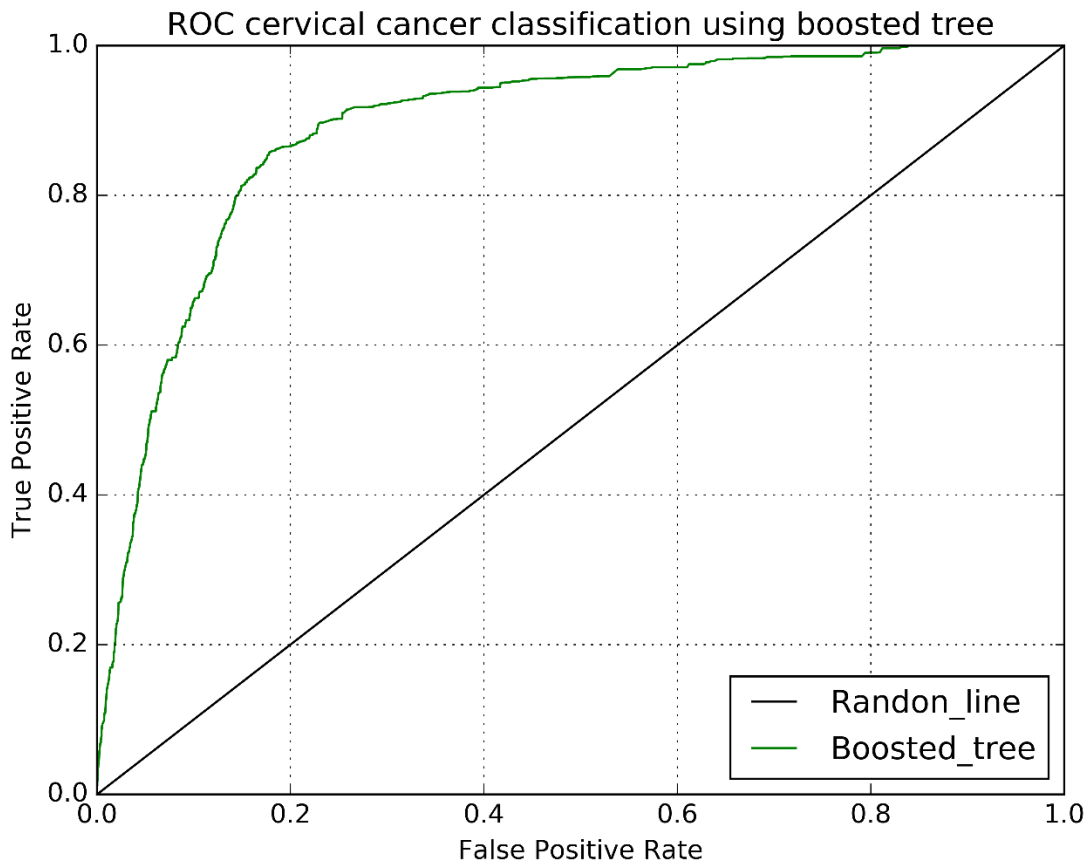


Figure 7-15 This figure represents the ROC of the boosted tree classifier on the test-set (using 20% of the data (for 10 patients)), using all features(ratio and histogram intensities).

Table 7-9 The statistics results using a boosted tree classifier on the test set.

True Positive	False Negative	Accuracy	Precision	Sensitivity (TPR)
1303	2127	0.909	0.469	0.38
False Positive	True Negative	Recall	F1 Score	Specificity (TNR)
1477	34551	0.38	0.42	0.96

7.5 Discussion and Future Directions

Tree classifier results were promising. We compared our method with the existing method presented by (Young Park, Follen, et al. 2008). The method presented in (Young Park, Follen, et al. 2008) only used information from the white light image (red dashed line in Figure 7-16). When we added information from other MDC images into their proposed method (Ensemble classifier) the classifier performance improved (blue line in Figure 7-16), more importantly, our boosted tree classifier (green line in Figure 7-16) outperformed their method.

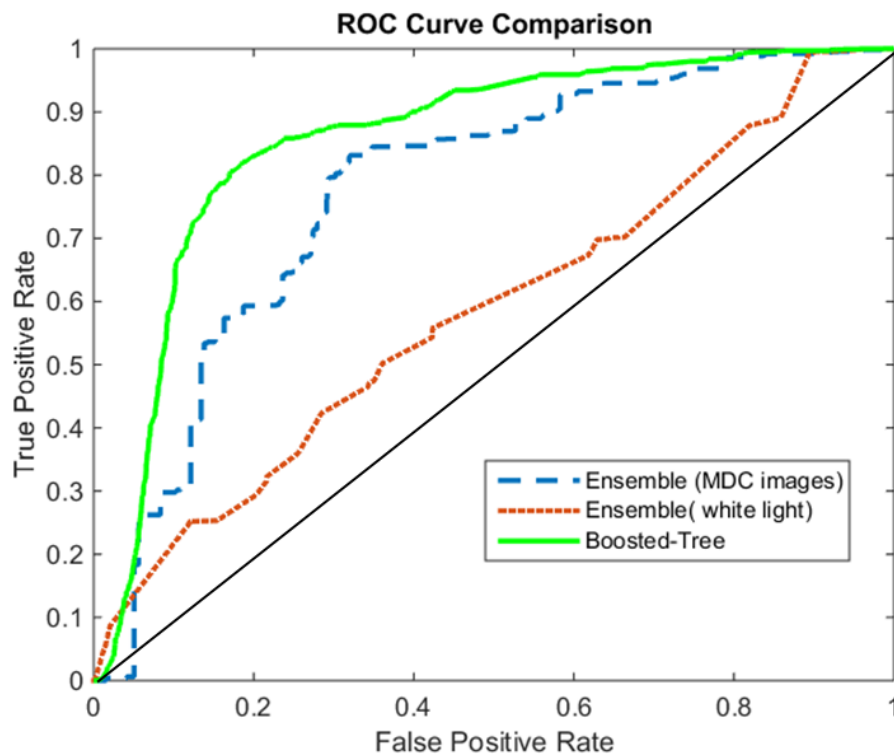


Figure 7-16 This figure shows the results of the Ensemble (MDC images and only white light) and the Boosted-Tree classifier being applied on the test set.

The strengths of this study are found in the histological review and mapping of the specimens. This careful mapping allowed the classification of the five remaining image's pixels. The large number of pixels provides a relatively large data set for analysis.

One of the limitations of this study originate from the histopathologic map not covering every square millimeter of the sample. Thus, assumptions were made about the tissue in-between the images. Another potential weakness of the study is how the reviewers interpreted the instructions for the selection of the abnormalities and accurately defining the lesions.

7.5.1 Future Direction and More Evaluation of the Study Procedure

In this study, we have evaluated the performance of our method (as a proof of concept) on a small dataset of 49 patients. We obtained 41 more patients (description of CIN level of each patient is given in Table 7-10) and extended our method to a larger dataset of 89 patients (which contains old and new images). These data are collected at a different time point (over 10-12 months), with three different devices.

Table 7-10 CIN level of the second half of the patients recruited for further studies.

<i>Patients</i>	<i>CIN1</i>	<i>CIN2</i>	<i>CIN3</i>	<i>Invasive & CIN3</i>
41	1	15	20	5

Using the histogram intensities as the features and boosted tree classifier on the second half of newly acquired images we obtained 92 % accuracy on the training set and 83% accuracy on the test set. We can see that the accuracies slightly dropped on this new dataset.

The ROC results on the training and test-set are shown Figure 7-17 and Figure 7-18 respectively. The statistic results on the training and test set are shown in Table 7-11 and Table 7-12 respectively.

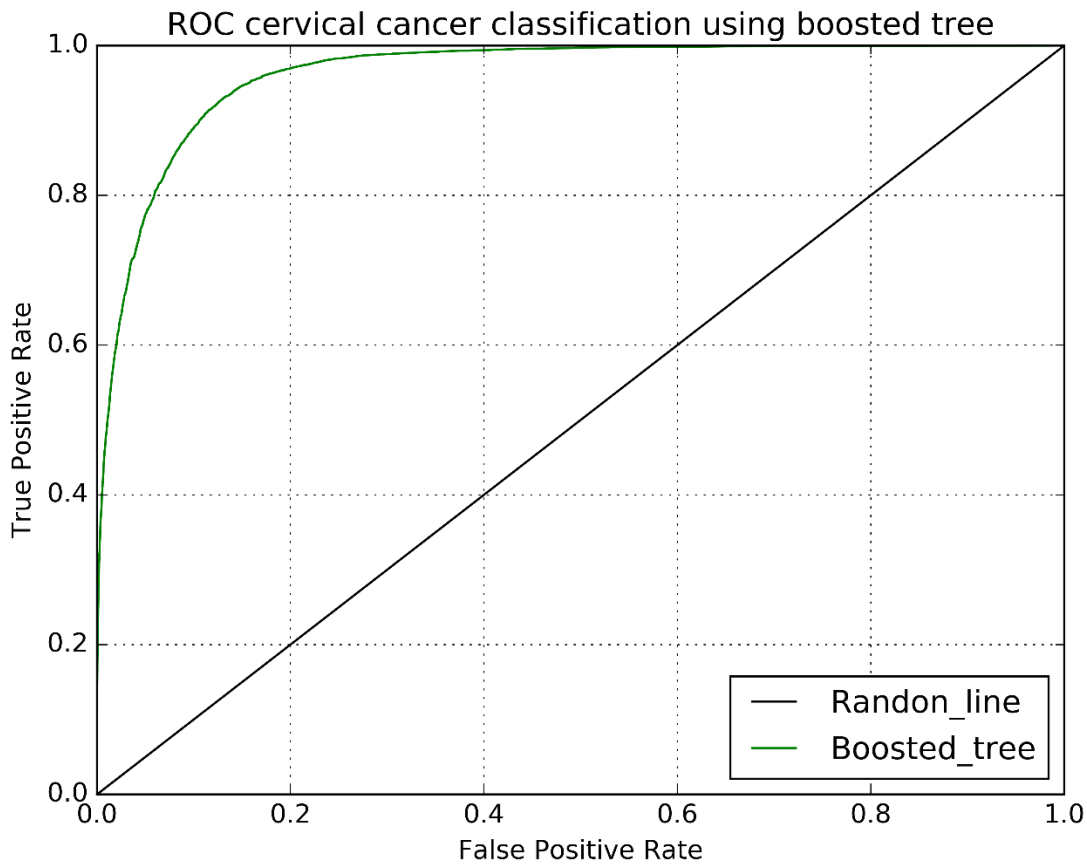


Figure 7-17 This figure represents the ROC of the boosted tree classifier on the training set on the second half of LEEP patients (using 80% of the data (for 32 patients)), using all features (ratio and histogram intensities).

Table 7-11 The statistics results using a boosted tree classifier on the training set.

True Positive	False Negative	Accuracy	Precision	Sensitivity (TPR)
20085	11200	0.923	0.825	0.642
False Positive	True Negative	Recall	F1 Score	Specificity (TNR)
4256	165977	0.642	0.722	0.97

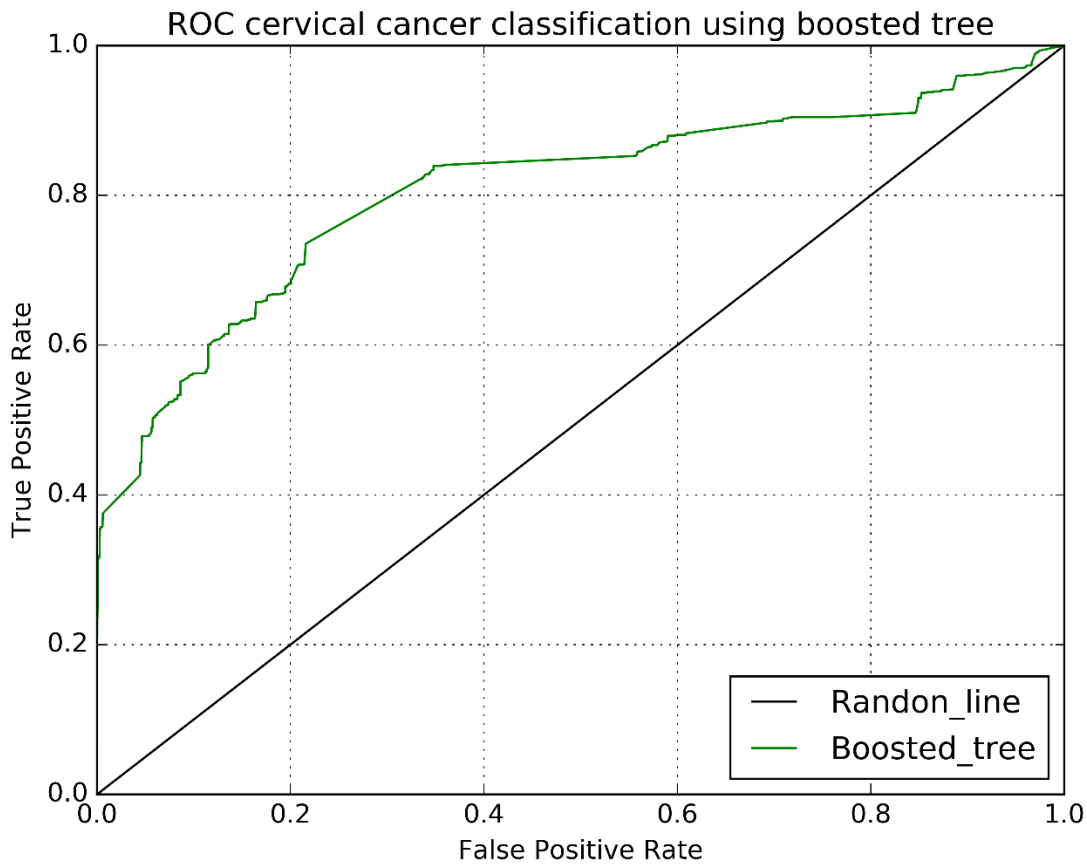


Figure 7-18 This figure represents the ROC of the boosted tree classifier on the test-set on the second half of LEEP patients (using 20% of the data (for 9 patients)), using all features(ratio and histogram intensities).

Table 7-12 The statistics results using a boosted tree classifier on the test set.

True Positive	False Negative	Accuracy	Precision	Sensitivity (TPR)
6248	5070	0.832	0.648	0.552
False Positive	True Negative	Recall	F1 Score	Specificity (TNR)
3398	35663	0.552	0.596	0.91

Then we investigated the performance of our classifier trained on the old dataset and tested on the new data-set. The ROC result is presented in Figure 7-19 and the statistics are presented in Table 7-13.

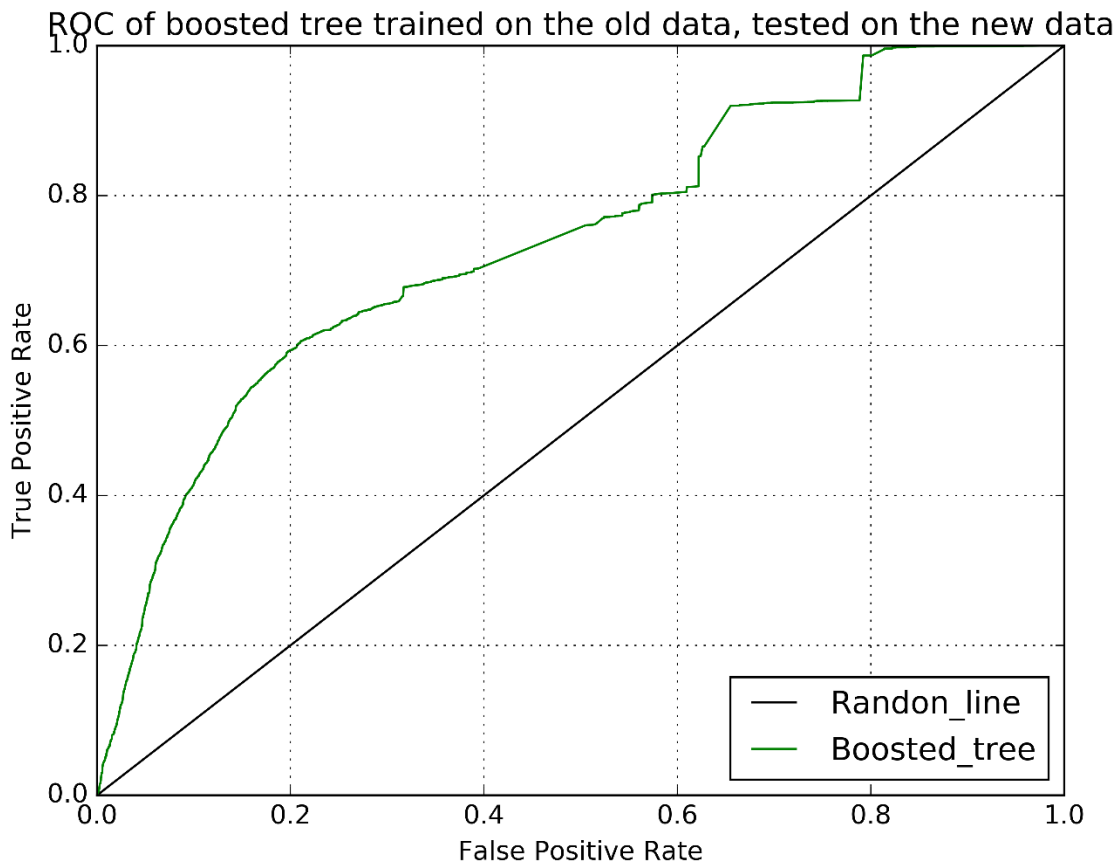


Figure 7-19 This figure represents the ROC of the boosted tree classifier on the new dataset (the second half of LEEP patients), all 50 images are used as the input. The classifier is trained on the first half of the images (49 patients).

Table 7-13 The statistics results using a boosted tree classifier trained on the old-data and tested on the new data.

True Positive	False Negative	Accuracy	Precision	Sensitivity (TPR)
14400	28203	0.83	0.496	0.34
False Positive	True Negative	Recall	F1 Score	Specificity (TNR)
14651	194644	0.34	0.402	0.92

In a pilot study, we also investigated the performance of our algorithm when all 90 images are using all features (ratio and histogram intensities features). Using 10-fold cross-validation we obtained about 79% accuracy, the ROC result is presented in Figure 7-20.

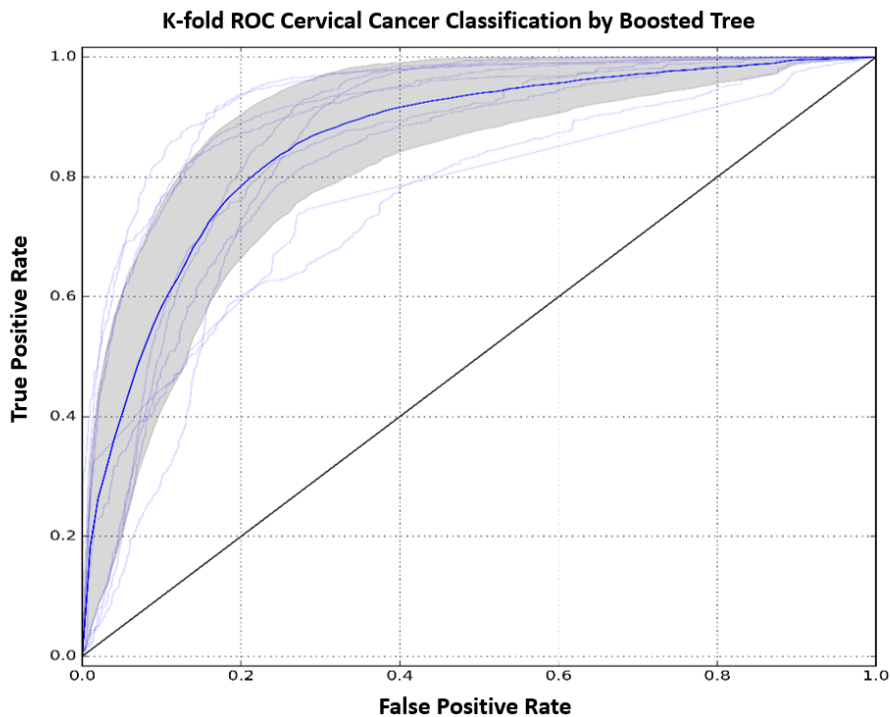


Figure 7-20 ROC, 10-fold cross-validation, using an image of 90 patients (old and new dataset).

As expected, we observed some variations due to changes in the device performances. Moreover, some of the recruited reviewers for the second part of the study are different from the first part of the study which might have also resulted in some intra variation for labeling the dataset based on the difference between reviewers' expertise. More evaluation and investigation of the new dataset are required.

7.6 Conclusion

In this study, we implemented a boosted tree classifier using the histogram intensity as our features. Later we integrated ratio features to improve the performance. Our algorithm performance on a subset of the data (49 patients) was promising (result summary is given in Figure 7-16), however, the classifier performance on the new dataset was not as good as Figure 7-20. Some of the reasons for misclassification could be 1) variation in the performance of different MDC devices, 2) variation in device performance over time 3) variation in the labels resulted from different MDC traces due to different levels of expert's expertise. For future studies, these challenges should be completely investigated. Furthermore, using classical machine learning, it is always necessary to design handcrafted features. In addition the limitation for generalization always exist. As a result, we decided to explore more robust methods such as deep convolutional neural networks (for the details of this study please refer to the next chapter).

Chapter 8 Deep Learning for Cervical Cancer Detection

8.1 Introduction

Deep Neural Networks (DNN) have shown promising results by automatically learning important features through multiple forward and backward propagations (LeCun, Bengio, et al. 2015). The power of DNNs is stemmed from the fact that the model is trained in an end-to-end fashion which has helped to deliver similar or even better results than conventional systems where they rely on hand-crafted features. It is important to keep in mind that although DNN is very data-driven (data-dependent) and could possibly learn/identify robust representations of the data, it can be very expensive (cost and time) and require a complex model. The other drawback of DNN could be the loss of feature interpretability. Moreover, because of performing in a black-box fashion, it gives rise to skepticism by the medical community where the feature interpretability is an important factor (Guo, Liu, et al. 2016)

Transfer learning ability by Deep Convolutional Neural Networks (DCNNs) has significantly improved the performance of image processing, image classification, object detection and computer vision in general.

In this chapter, we propose a deep convolutional learning method to detect and locate abnormality (CIN2 or worse areas) within the cervix (for details of abnormalities refer to the previous chapter). The deep learning network that we propose herein takes MDC images as input and detects and localizes the regions where abnormality exists in the image (for the information regarding the MDC images refer to the previous chapter). The neural networks presented in this chapter are an extension of convolutional neural networks which is used by other groups such as Visual Geometry Group (VGG) at Oxford University. We modified and fine-tuned these networks and used them for our classification task.

In the previous chapter, we explained how the data is collected, labeled and prepared. In this chapter, we first explain the fundamentals of DCNNs and how this method applied to our question. Then we explain how we formulated the image analysis question and handled our data to fit the requirements of the network. We continue by showing the experimental design and conclude the chapter by showing our results and validation process.

8.2 Objective

The objective of this chapter is to investigate the performance of different convolutional neural networks using reflectance and fluorescence images from the cervix (macroscopic MDC images). Our goal is to design an automated technique to localize at-risk tissue (CIN2 above) within the cervix.

8.2.1 Specific Objectives

- I) Investigating and comparing the performance of different Convolutional Neural Networks
- II) Preparing the dataset (MDC images) to fit the selected CNN
- III) Applying different CNN to our dataset and compare the results.

The workflow for this chapter is presented in Figure 8-1. The first step, as always is imaging the tissue, followed by image/ label preparation. The next step is the selection and implementation of the CNNs (using training and test datasets) and finally, the last step is parameter optimization to improve the performance.

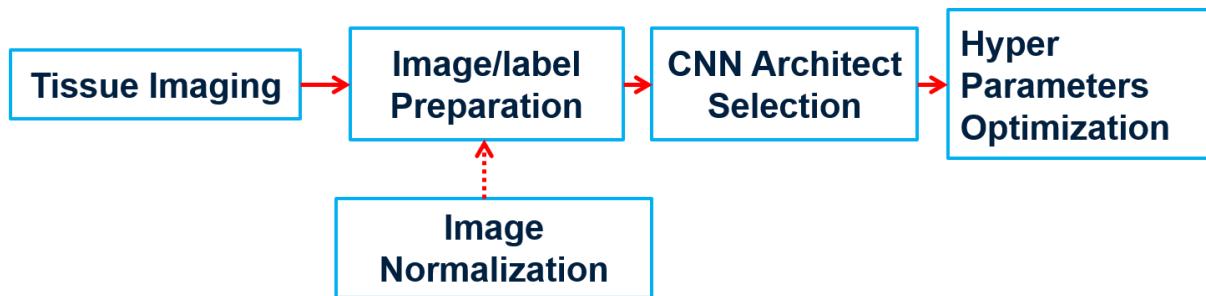


Figure 8-1. Workflow of our process from tissue imaging to hyperparameter optimization. The first step is imaging the tissue with MDC technology, next, we will prepare the data and labels according to the type of CNN that we plan to use (we might also need to center the data, by using normalization procedures) and finally, we will optimize some parameters to increase the performance.

8.3 Methodology-Convolutional Neural Network Architecture

8.3.1 Patch-based CNNs

We proposed a Patch-based CNN algorithm. In this method, the training step is divided into two main sections as illustrated in Figure 8-2. We selected this method due to some limitations that we were facing such as 1) limited number of images and 2) minimal availability of normal-looking cervical MDC images. Although these limitations constrained us from implementing a (whole) image-level classifier, the patch-based CNN reduced the computational cost of handling the high-resolution MDC images. The challenges that we are facing are 1) how to design such a system to derive the best discriminative result and 2) how to aggregate the processed patches to represent the whole image.

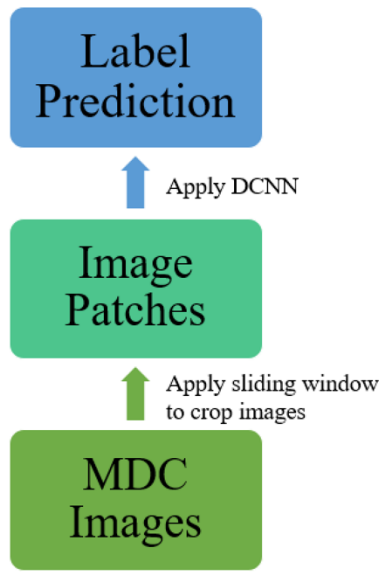


Figure 8-2. A very high-level representation of the workflow of patch-based CNN. After preparing the MDC images we will apply a sliding window to scan the entire image and extract image patches. Next, we will label these patches. Finally, we implement the classification step to predict the labels.

8.3.1.1 Image-patch Generation

We defined various window sizes and strides based on the selected CNN architecture. A sliding window of size $s_x \times s_x$ scans through the entire image using a stride of d . At each location (i, j) , the image within the window is cropped and the image beneath is saved. Setting s_x too small would not capture enough spatial context for object extraction while setting s_x too large would result in the retrieved context being too complicated for the extraction and increase the amount of computation required in the prediction step of a deep CNN. After experimenting with different values for s_x we concluded that 51 would provide the best result.

We set the stride d to be 25 pixels. Although a smaller d could increase the accuracy, it would also increase the amount of overall computation. Through extensive experimentation, 25 pixels was found to achieve a good balance between these two issues.

8.3.1.2 Deep CNN Prediction

We used the images obtained from the sliding window patches and used them to train a deep CNN (DCNN). The goal of this CNN was to predict the label of the pixel located exactly in the center of each patch. Since we are down-sampling the image by generating patches and assigning the label in the middle as the label of the whole patch, we do not need to deal with generating label-mapping and applying the threshold ruling (e.g. calculating the average) for over-lapping areas (areas where two sliding windows have overlaps).

8.3.2 Deep CNNs on Patches

As shown in Figure 8-3 our deep CNNs take an RGB image patch N from a channel as an input, apply a sequence of linear/non-linear transformations to extract features from the patch in a layer-by-layer fashion, and map the extracted features to probabilities of being a label patch M (certain object as the output).

The architecture of deep CNNs is inspired by the system presented by (LeCun, 1989). The convolution and pooling operations enable the networks to learn features that are shift-invariant and robust to small image distortions. These features have been proven to be extremely useful for image-based applications (LeCun et al., 1998; Krizhevsky et al., 2012; Mnih 2013; Girshick et al. 2014).

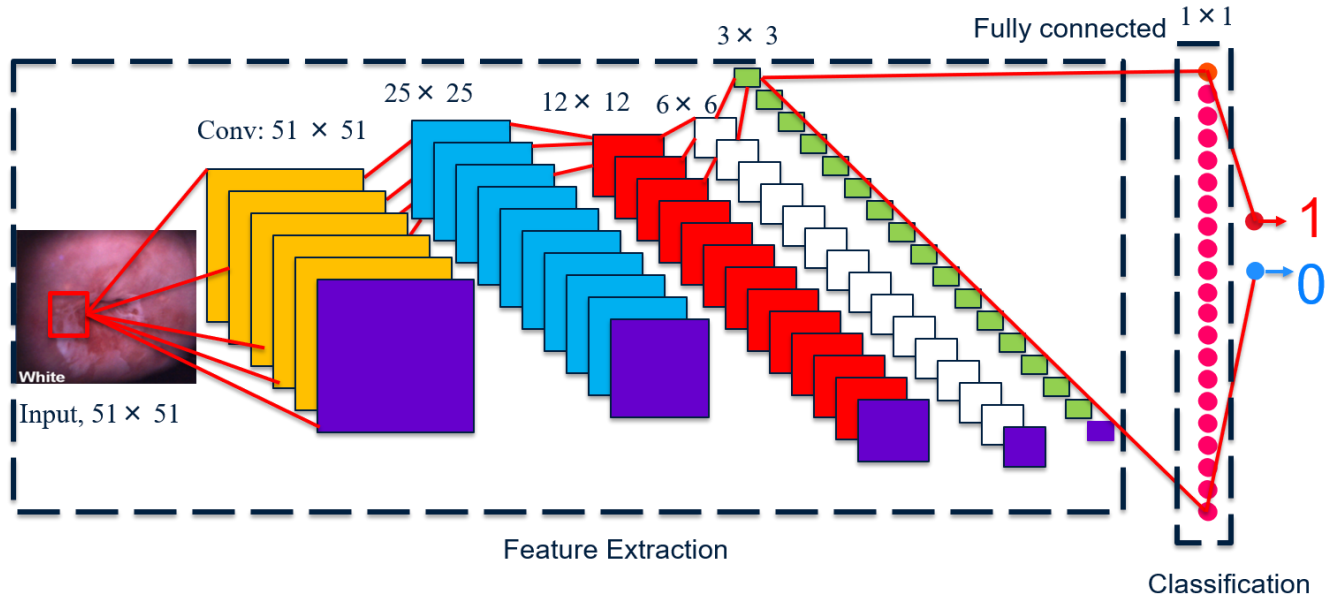


Figure 8-3 Feature extraction and classification steps of one of the networks (VVG16). The first block is the feature extraction which consists of consecutive convolution and Max-pooling blocks (purple rectangle in each stack of rectangles is representative of Max-pooling) the second block is a classification module where it is fully connected to the previous layer. The output of this layer is fed to a binary logistic regression for prediction. The number of rectangles shown above is not representative of the actual number of filters in each layer.

The details of the mathematical model of using patch-based CNNs for object extraction, the architecture of DCNNs, and the approach of the learning parameters of DCNNs, are presented in the following sections:

8.3.3 Problem Formulation

Here we formulate the proposed method. As mentioned above, the goal is to predict the label given an image:

$$P(M|N) \quad (8.1)$$

Here N denotes a three-dimensional patch image. Information regarding the combination of images used to construct the input 3D image can be found in the previous chapter. These images have the size of $s_x \times s_x \times c$ where s_x is the spatial dimension and c is the channel dimension (in our algorithm c is equal 3 but it is possible to have between 1 to n channels). M is the predicted label, which takes the values of a set of possible object labels $\{0, 1, \dots, K\}$, in our case 0-1 since we are implementing a binary classifier.

If we assume the labels to be independent, we can simplify equation 8.1 to sum of all predictions as shown in equation 8.2:

$$P(M|N) = \prod_{i=1}^n P(M_i|N) \quad (8.2)$$

We can use a logistic output, (8.3) and (8.4) unit is a binary classifier. In the case of more classes, it is possible to use other classifier methods such as SoftMax unit (, Softmax function2019).

$$f_i = \sigma(a_i(N)) \quad (8.3)$$

$$\sigma(x) = \frac{1}{1+\exp^{-x}} \quad (8.4)$$

where f_i is the value of the i^{th} output unit; a_i is the total input to the i^{th} output unit, and x is the input images.

8.3.3.1 I-Convolutional Layer

The convolution layers consist of three modules, convolution, non-linear transformation and spatial pooling (all these modules can be eliminated or bypassed).

The reason for using the convolution unit is to enable weight sharing across the input. This feature will make it possible to learn shift-invariant features and eventually help the detection/identification of certain features.

Convolution Layer:

The use of a convolution operator leads to weight sharing across the input (due to the fact that weights W_{ij} of each filter are replicated across the input field). Such weight sharing enables CNNs to learn shift-invariant features, which have proven to be very useful for solving visual problems. Here are some notations:

X : input to the convolution layer

$s_x \times s_x \times c_x$: dimension of the input image (s_x and c_x denote the spatial dimension and number of channels respectively)

$s_y \times s_y \times c_y$: dimension of the output image (s_y and c_y denote the spatial dimension and number of channels respectively)

$s_w \times s_w$: matrix of the weights defined by the filter dimension

The output of the convolution layer can be expressed as shown in (8.5):

$$Y_j = \text{pool} \left(g \left(b_j + \sum_{i=1}^{c_x} W_{ij} * X_i \right) \right) \quad (8.5)$$

where Y_j denotes the array in the j^{th} channel of output Y ; X_i denotes the input in the i^{th} channel of input X ; W_{ij} denotes the weights of the filter connecting input X_i to output Y_j ; b_j denotes a vector of biases; $*$ denotes a two-dimensional convolution operator. g is a point-wise non-linear activation function.

The last function on the convolution unit is a pooling unit. This can apply different pooling strategies such as max, mean or median pooling. In all these methods, the defined statistic will be calculated by considering the neighboring cells. The pooling unit will increase the robustness to

shift and distortion, as a result, CNN can benefit from better generalization. The equation for max-pooling is shown in equation (8.6).

$$g_{(x)} = \text{maxPooling}(x, 0) \quad (8.6)$$

pool is a function that applies spatial pooling to activations (the result of the activation function).

In the case where a max-pooling operator is used, it considers a neighborhood of activations and produces one pooling per neighborhood by taking the maximum activation within the neighborhood. Pooling over a small neighborhood provides additional variance resistance to small input shifts and distortions, which help CNNs achieve better generalization on object recognition.

Fully-Connected layer:

The fully connected layer has a mapping between all nodes with a weight of W . The output of the layer Y can be expressed as,

$$Y = g(b + WX) \quad (8.7)$$

where b denotes a vector of biases; $g(x)$ is a nonlinear activation function.

8.3.3.2 II-Training Step

We have a total of 885,600 data points and we reserved 797,040 for training (90% of the whole dataset) and the rest for testing. The total trained parameters for VGG16 is 1,026 which includes weights W and biases b in each layer. The reason for not training the whole parameter is to take advantage of transfer learning, where a model designed and trained for a specific task is reused for another task (Lisa Torrey 2009). We used ImageNet weights which are already trained on a very large image-set. Using these weights, we can reduce the training time significantly and the model only trains the last few layers (the other non-trainable weights remain constant during forward and backward propagations).

The loss function that we used is the cross-entropy. It learns the best set of parameters by minimizing the total cross-entropy (equation (8.8)) between the ground truth and predicted labels.

$$L(W, b) = -\frac{1}{m} \sum_{i=1}^m (y_i \ln(f_i(N, (W, b))) + (1 - y_i) \ln(1 - f_i(N, (W, b)))) \quad (8.8)$$

where y_i represent the actual label and $f_i(N, (W, b))$ represents the formula that was being used to predict the label. In our study, we used the ReLU function as shown in equation (8.4) (Rectifier ,neural networks,2019).

We used stochastic gradient descent with mini-batches and used the Adam optimization algorithm as our optimizer method. The list of used hyperparameters is given in Table 8-1. These hyperparameters are proven in the literature (Kingma, Ba 2015) to produce the best results.

$$V_{dw} = \beta_1 V_{dw} + (1 - \beta_1) dw \quad (8.9)$$

$$V_{db} = \beta_1 V_{db} + (1 - \beta_1) db \quad (8.10)$$

$$S_{dw} = \beta_2 S_{dw} + (1 - \beta_2) dw^2 \quad (8.11)$$

$$S_{db} = \beta_2 S_{db} + (1 - \beta_2) db^2 \quad (8.12)$$

$$V_{dw}^c = \frac{V_{dw}}{1 - \beta_1^t} \quad (8.13)$$

$$V_{db}^c = \frac{V_{db}}{1 - \beta_1^t} \quad (8.14)$$

$$S_{dw}^c = \frac{S_{dw}}{1 - \beta_2^t} \quad (8.15)$$

$$S_{db}^c = \frac{S_{db}}{1 - \beta_2^t} \quad (8.16)$$

$$W := W - \alpha \frac{V_{dw}^c}{\sqrt{S_{dw}^c + \epsilon}} \quad (8.17)$$

$$b := b - \alpha \frac{V_{db}^c}{\sqrt{S_{db}^c + \epsilon}} \quad (8.18)$$

Where V_{dw} , V_{db} , S_{dw} and S_{db} are set to zero at the initial step.

Table 8-1 Hyper-parameter table for Adam optimizer

<i>Hyper parameter name</i>	<i>Value</i>
α	0.001
β_1	0.9
β_2	0.999
ϵ	0.0001

8.3.4 Data Preparation

We designed a sliding window of size $s_x \times s_x$ and step size of d . The sliding window scans the whole image and save the image underneath. We used this sliding window for all six channels (three in reflectance mode and three in fluorescence mode). The label for each sliding window is the label of the pixel at location $s_x/2 \times s_x/2$ (middle of the window). We experimented with different values for s_x and d . 51 and 25 have given us the optimum results for s_x and d respectively.

We divided our dataset into two sets: 90% were used for training-validation and 10% were used as the test set.

To help the optimizer to converge faster and possibly increase the performance instead of random initialization of the weights, we used the model weights of the ImageNet pre-trained⁹ networks and adapt them to our task of binary classification. We replaced the 1000-way ImageNet classifier in the last layer with a binary SoftMax. All positions and labels are equally weighted in the overall

⁹ ImageNet is an image database organized according to the WordNet hierarchy (currently only the nouns), in which each node of the hierarchy is depicted by hundreds and thousands of images. Recent success in Convolutional networks (ConvNets) is due to the large public image repositories, such as ImageNet (Deng, Dong et al. 2009).

loss function (except for unlabeled pixels which are ignored). Our targets are the ground truth labels (subsamped by the sliding windows of different sizes).

In the next section, we will explain the experiment setup and results.

8.4 Experiment and Results

We used four architectures of CNNs, I) simple architecture II) VGG16, III) VGG19, IV) ResNet50 (Saikia, Bora et al. 2019). A diagram representative of each network is shown in Figure 8-4- Figure 8-8.

We started with a very simple model, as shown in Figure 8-4 and continued with deeper, more complex networks. The list of parameters (learned and not-learned for this model) can be found in Table 8-2.

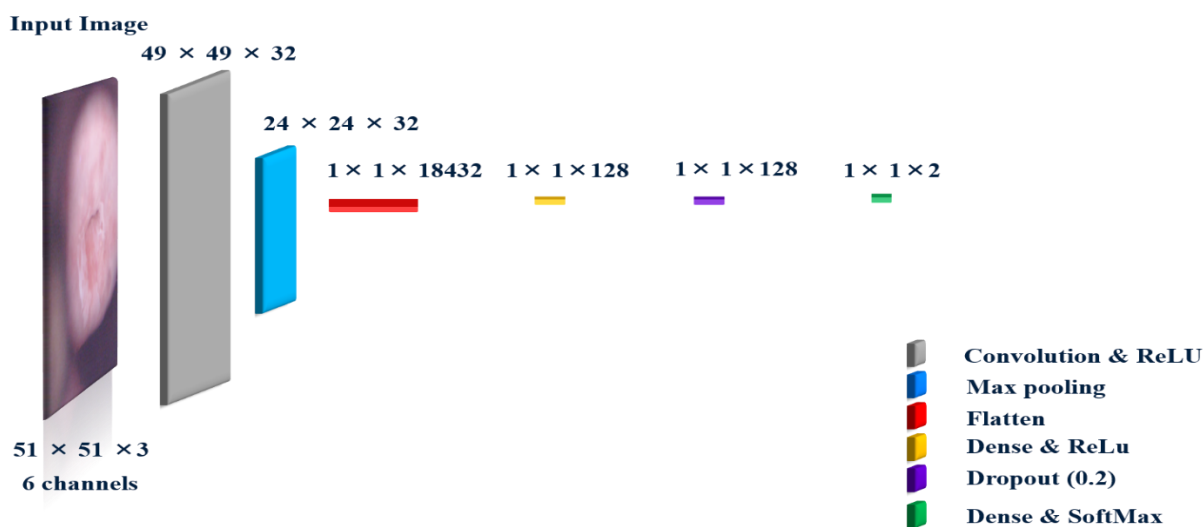


Figure 8-4 Simple Convolutional Neural Network. This very simple network consists of one convolution and the ReLU layer followed by a Max-pooling layer, a Flatten layer, another dense layer followed by a dropout layer with (20% drop-out) and at the end a SoftMax layer (with only two classes).

Using the network presented in Figure 8-4 (referred to as Simple Network) and only 3 channels we obtained around 77% accuracy. This indicated that the network is not deep enough. To increase the accuracy we decided to train deeper models.

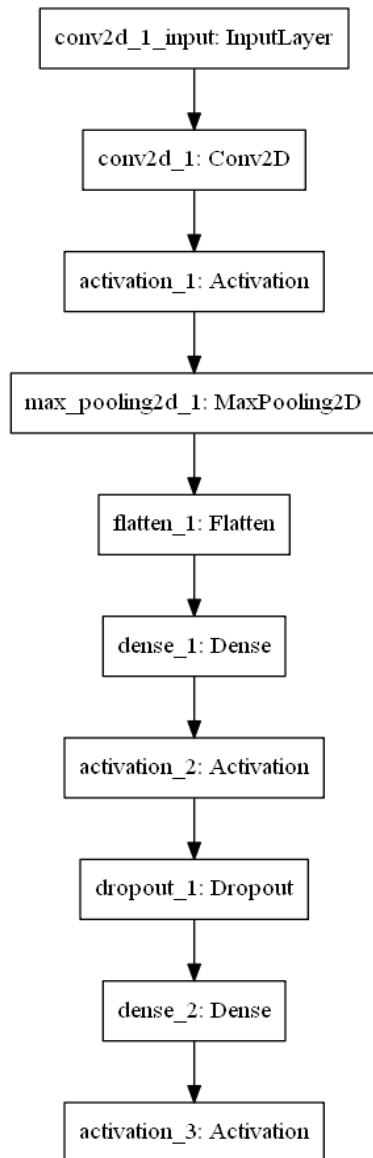


Figure 8-5 This figure shows the flow diagram of the simple model.

Table 8-2 Simple model summary table, shows input to each layer and number of parameters

<i>Layer (type)</i>	<i>Output Shape</i>	<i>Param No.</i>
<i>conv2d_1 (Conv2D)</i>	<i>(None, 49, 49, 32)</i>	<i>896</i>
<i>activation_1 (Activation)</i>	<i>(None, 49, 49, 32)</i>	<i>0</i>
<i>max_pooling2d_1 (MaxPooling2)</i>	<i>(None, 24, 24, 32)</i>	<i>0</i>
<i>flatten_1 (Flatten)</i>	<i>(None, 18432)</i>	<i>0</i>
<i>dense_1 (Dense)</i>	<i>(None, 128)</i>	<i>2359424</i>
<i>activation_2 (Activation)</i>	<i>(None, 128)</i>	<i>0</i>
<i>dropout_1 (Dropout)</i>	<i>(None, 128)</i>	<i>0</i>
<i>dense_2 (Dense)</i>	<i>(None, 2)</i>	<i>258</i>
<i>activation_3 (Activation)</i>	<i>(None, 2)</i>	<i>0</i>
=====		
<i>Total parameters: 2,360,578</i>		
<i>Trainable params: 2,360,578</i>		
<i>Non-trainable params: 0</i>		

To increase the accuracy we decided to use deeper models. One of the networks that we investigated was VGG16. VGG is a convolutional neural network model proposed by VGG (Visual Geometry Group) (Simonyan, Zisserman 2014) from the University of Oxford. This model achieves 92.7% top-5 test accuracy on ImageNet data. They won the ImageNet ILSVRC-2014 competition and secured the first and the second places in the localization and classification respectively. Later they modified and added few layers to their current VGG16 and invented a deeper model called VGG19 (later, in this chapter, we also show how we leveraged this network for our prediction purposes).

We used VGG16 as our CNN model and used our MDC image patches as the input (please refer to the data preparation section for the details on data preparation). The architecture of VGG16 is shown in Figure 8-6 . Using this network, we obtained about 0.846 accuracy which is an improvement over using a simple model. The number of learned and non-learned parameters is given in Table 8-4. Although the training time was longer (14 hrs. to train the VGG16 using MDC patches vs 3 hrs. using the simple model), this is a cost we are usually ready to pay when training on deep neural networks.

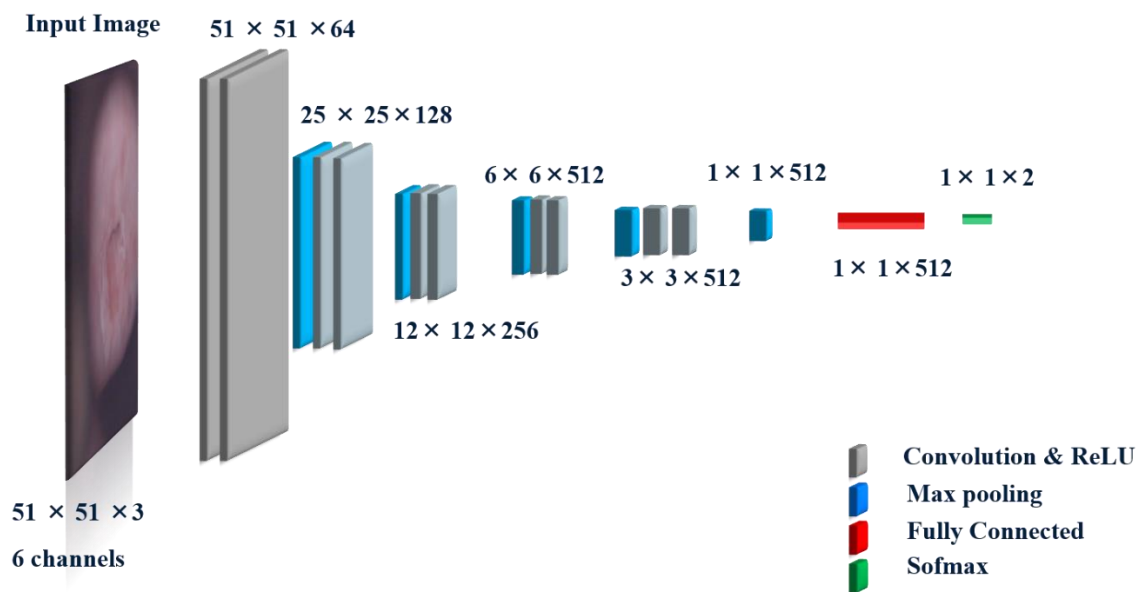


Figure 8-6 VGG16 Convolutional Neural Network. This relatively deep network consists of few layers of convolution and ReLU layers followed by a Max pooling layers, two last layers are a fully connected layer and a SoftMax layer (with only two classes).

Table 8-3 VGG16 summary table, shows input to each layer and number of parameters

<i>Layer (type)</i>	<i>Output Shape</i>	<i>Param No.</i>
<i>input_1 (InputLayer)</i>	<i>(None, 51, 51, 3)</i>	<i>0</i>
<i>block1_conv1 (Conv2D)</i>	<i>(None, 51, 51, 64)</i>	<i>1792</i>
<i>block1_conv2 (Conv2D)</i>	<i>(None, 51, 51, 64)</i>	<i>36928</i>
<i>block1_pool (MaxPooling2D)</i>	<i>(None, 25, 25, 64)</i>	<i>0</i>
<i>Block2-4</i>	<i>continued</i>	<i>...</i>
<i>block5_conv1 (Conv2D)</i>	<i>(None, 3, 3, 512)</i>	<i>2359808</i>
<i>block5_conv2 (Conv2D)</i>	<i>(None, 3, 3, 512)</i>	<i>2359808</i>
<i>block5_conv3 (Conv2D)</i>	<i>(None, 3, 3, 512)</i>	<i>2359808</i>
<i>block5_pool (MaxPooling2D)</i>	<i>(None, 1, 1, 512)</i>	<i>0</i>
<i>flatten_1 (Flatten)</i>	<i>(None, 512)</i>	<i>0</i>
<i>dense_1 (Dense)</i>	<i>(None, 2)</i>	<i>1026</i>
=====		
<i>Total parameters: 14,715,714</i>		
<i>Trainable parameters: 1,026</i>		
<i>Non-trainable parameters: 14,714,688</i>		

To see whether we can increase the accuracy of the prediction using an even a deeper model, we decided to use VGG19. The architecture of VGG19 (shown in Figure 8-7) is very similar to VGG16 (shown in Figure 8-6), however, due to deeper structure, there is a possibility of exploding/vanishing gradient (Hanin 2018). Figure 8-7 shows a carton of the network, and Table 8-4 gives the number of learnt and non-learnt parameters.

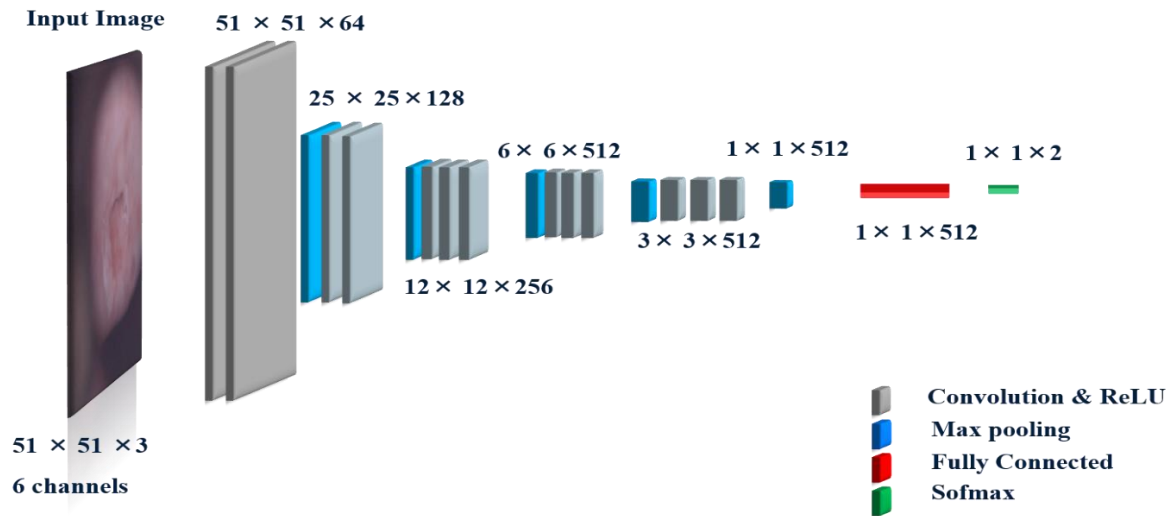


Figure 8-7 VGG19 Convolutional Neural Network. This relatively deep network consists of few layers of convolution and ReLU layers followed by a Max pooling layers, two last layers are a fully connected layer and a SoftMax layer (with only two classes).

Table 8-4 VGG16 summary table, shows input to each layer and number of parameters

<i>Layer (type)</i>	<i>Output Shape</i>	<i>Param No.</i>
<i>input_1 (InputLayer)</i>	<i>(None, 51, 51, 3)</i>	<i>0</i>
<i>block1_conv1 (Conv2D)</i>	<i>(None, 51, 51, 64)</i>	<i>1792</i>
<i>block1_conv2 (Conv2D)</i>	<i>(None, 51, 51, 64)</i>	<i>36928</i>
<i>block1_pool (MaxPooling2D)</i>	<i>(None, 25, 25, 64)</i>	<i>0</i>
<i>Block2-4</i>	<i>continued</i>	<i>...</i>
<i>block5_conv2 (Conv2D)</i>	<i>(None, 3, 3, 512)</i>	<i>2359808</i>
<i>block5_conv3 (Conv2D)</i>	<i>(None, 3, 3, 512)</i>	<i>2359808</i>
<i>block5_conv4 (Conv2D)</i>	<i>(None, 3, 3, 512)</i>	<i>2359808</i>
<i>block5_pool (MaxPooling2D)</i>	<i>(None, 1, 1, 512)</i>	<i>0</i>
<i>flatten_1 (Flatten)</i>	<i>(None, 512)</i>	<i>0</i>
<i>dense_1 (Dense)</i>	<i>(None, 2)</i>	<i>1026</i>
=====		
<i>Total parameters: 20,025,410</i>		
<i>Trainable parameters: 1,026 , Non-trainable par parameters ams: 20,024,384</i>		

Using VGG19, we obtained 0.848 accuracy which is slightly better than VGG16, however, this network is deeper, and required more time to train the model.

Next, we decided to train an even deeper network called ResNet50. Very deep networks such as this network are essentially very hard to train and very prone to exploding and vanishing gradients (gradients becoming really large or very small, (Nielsen 2015)). To overcome this issue a residual learning framework is being used (Fung 2017). Figure 8-8 illustrates this network. Networks with residual learning are easier to optimize (He, Zhang et al. 2016) and each layer is not dependent on only one layer behind, but also receives input from few layers behind it (Fung 2017) as shown in Figure 8-8. The bypass abilities (Fung 2017) made this network very appealing. He and his group (He, Zhang et al. 2016) won 1st place in ILSVRC 2015 classification task using this Network. In this chapter, we also show the result of using this network for our prediction goal.

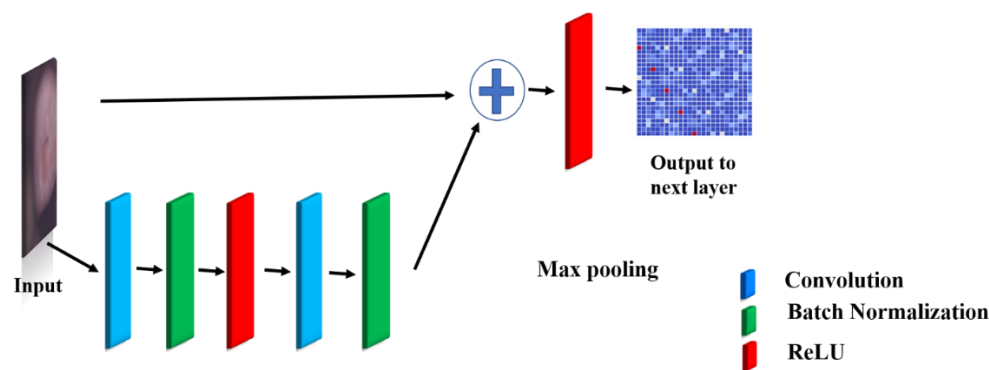


Figure 8-8 One block of ResNet-50. It has two learning paths, the bottom path contains layers of convolution, batch normalization, and ReLU, and the top path has no learning module.

We obtained 0.7871 in accuracy using the ResNet50 network (input images were from the first channel only). The accuracy dropped using this network and we suspected that it might be due to

exploding/vanishing gradients that we explained earlier. Since our input data is limited, we decided not to use a very deep network, and so we picked VGG16, which is relatively shallow compared with ResNet-50.

Table 8-5 shows the comparison between these four networks with a focus on accuracy and the training time.

Table 8-5 comparison between different networks

<i>Networks</i>	<i>Accuracy</i>	<i>Train time (hrs.)</i>
<i>Simple Model</i>	0.774	3
<i>VGG16</i>	0.847	14
<i>VGG19</i>	0.848	17
<i>ResNet-50</i>	0.7871	55

8.4.1 Per MDC Channel Performance Analysis

Based on the results we obtained from investigating the performance of different networks and the time required for training as well as the network complexity, we decided to select VGG16 for the rest of our analysis.

Using VGG16, for each channel, we recorded the loss¹⁰ and accuracy for each batch¹¹. Batch is a subset of samples¹² (images) that are being processed in parallel and independently (generally each batch approximates the input distribution)(Keras Training Documentation). We selected 32

¹⁰ Loss: Binary cross-entropy is used ((Raúl Gómez)

¹¹ Batch: A collection of sample data that are being processed independently and they represent the whole dataset (means they are assumed to have similar distribution) (Keras Training Documentation)

¹² Samples: usually refers to the images or in this study sliding windows (Keras Training Documentation)

samples per batch and ended up with the total number of 887 batches. The accuracy and loss plots over these 887 patches for 5 epochs¹³ (number of times to finish going through the training set, in our case 5 times) for channel 1 is shown in Figure 8-9. As illustrated in Figure 8-9 and Figure 8-10, we can observe that the accuracy and loss plateau at around 0.86 and 0.32 respectively. `acc_i` ($i=1:5$) and `loss_i` ($i=1:5$), are corresponded to accuracy per epoch and loss per epoch respectively.

For the compiled list of figures for all channels please refer to the Appendix.

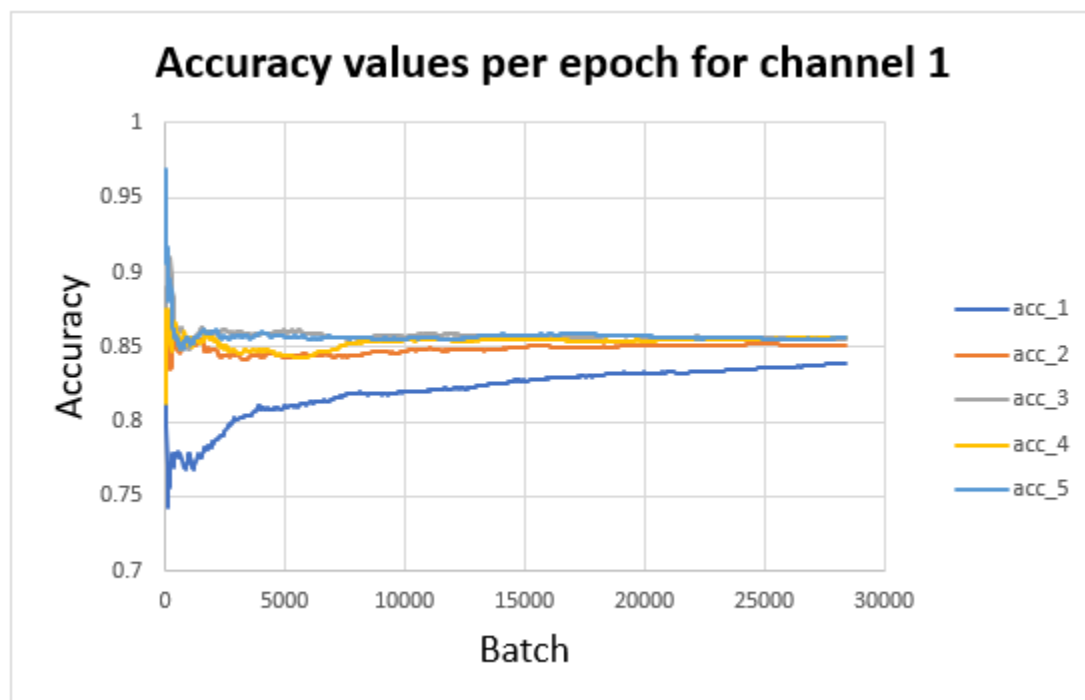


Figure 8-9 Accuracy values per epoch for channel 1. The accuracy level fluctuated at the beginning and eventually plateaus around 0.85. Each color represents one epoch.

¹³ Epoch: Usually refers to as one pass through the dataset. (Keras Training Documentation)

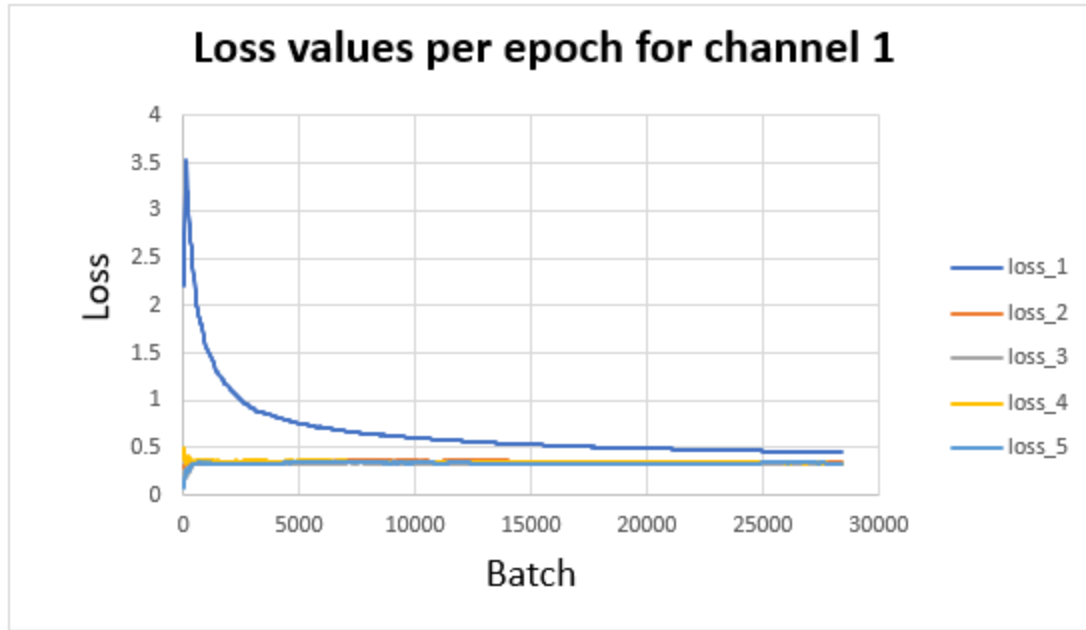


Figure 8-10 Loss values per epoch for channel 1. The loss level fluctuated at the beginning and eventually becomes plateau around zero. Each color represents one epoch.

To compare different channels, we calculated the average accuracy and loss per epoch. The result for accuracy and loss are shown in Figure 8-11 and Figure 8-12 respectively. Almost all channels show relatively similar behavior.

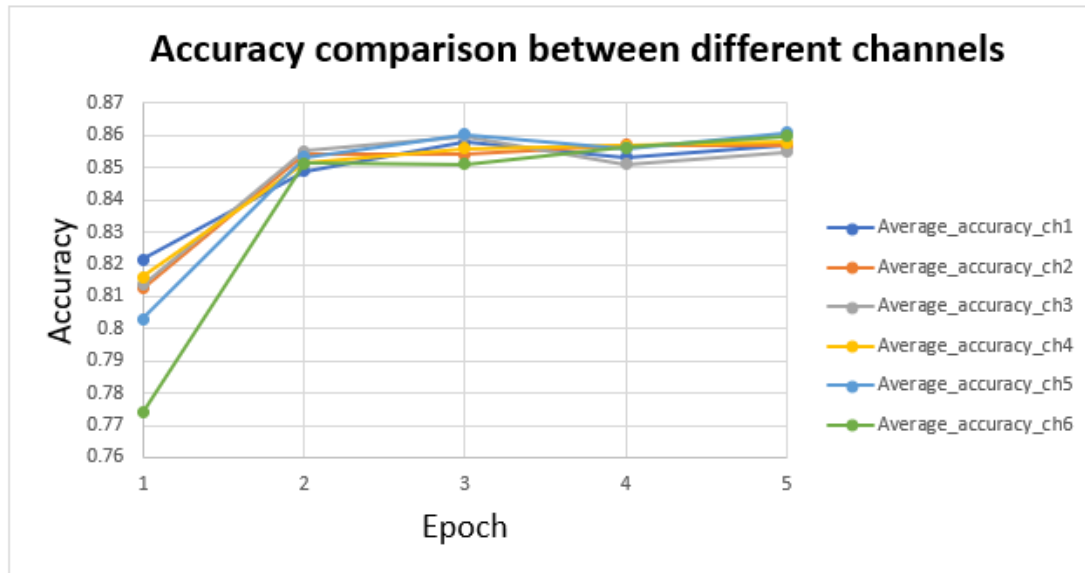


Figure 8-11 Accuracy values per channel for each epoch. The accuracy level increased at the beginning and became stable at around 0.85. Each color represents one channel. All channels show relatively similar performance.

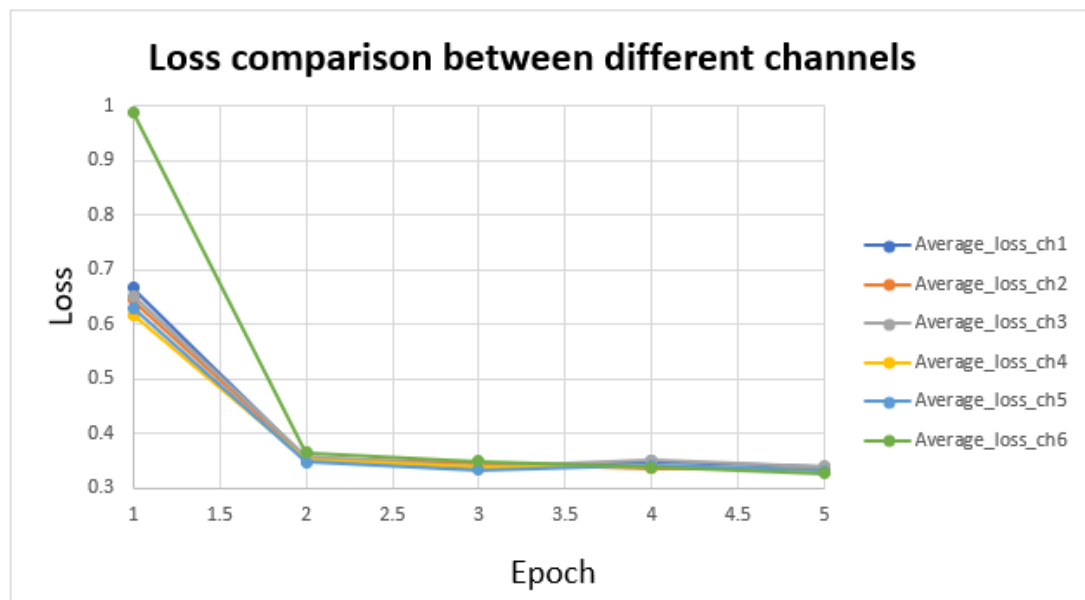


Figure 8-12 Loss values per channel for each epoch. The loss level decreased drastically at the beginning and became stable at around 0.25. Each color represents one channel. All channels show relatively similar performance.

To visualize the performance of the trained network (per channel) for one of the MDC images (contains 3600 samples, including the “don’t care” samples) refer to Figure 8-13. In this image you can observe that almost all channels show similar performance.

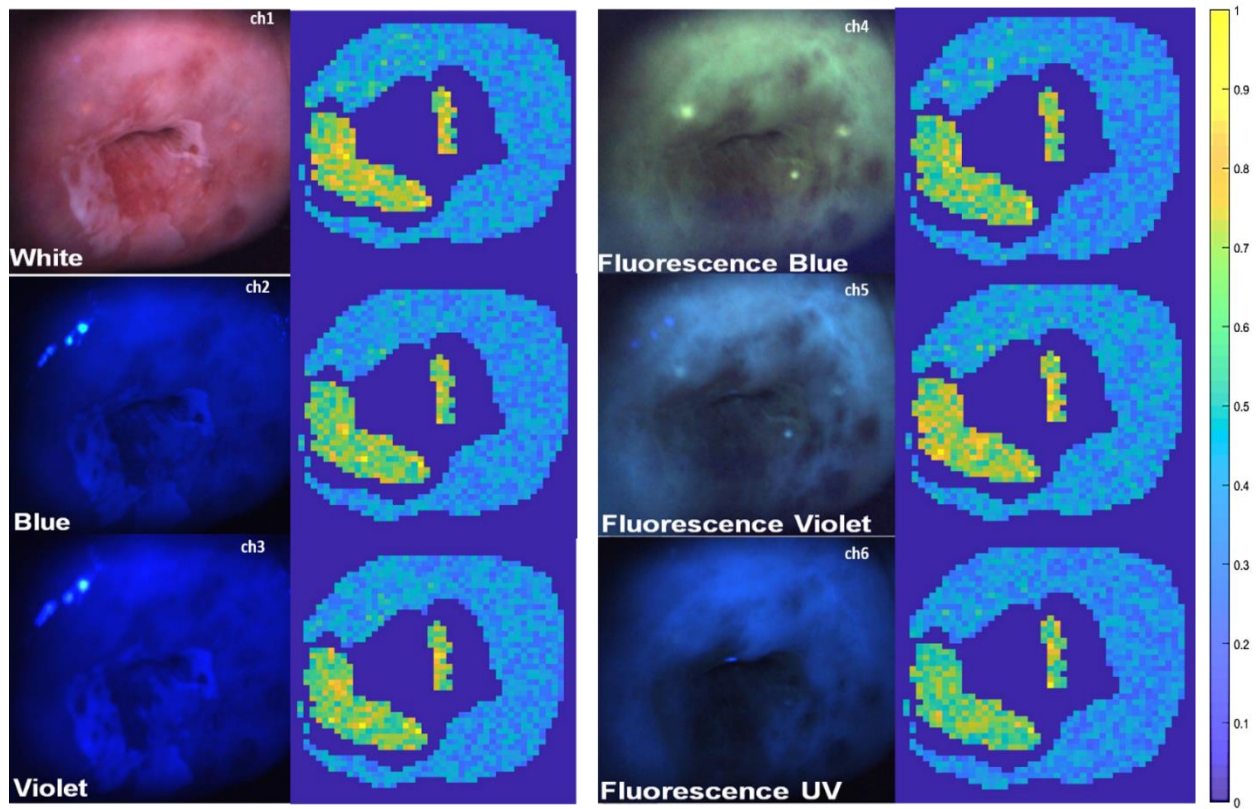


Figure 8-13 Prediction visualization for one whole image input. The prediction per sliding window is calculated and the whole image is reconstructed per channel (for 6 channels). All channels show relatively similar performance.

In order to increase the accuracy, we centered the data by normalizing the images (using the Z-score normalization method, (Feature scaling) and trained the network with the new dataset. The result is shown in Figure 8-14. This image compares normalized and not-normalized dataset. The grouped average accuracies on the top of Figure 8-14 are associated with normalized data and the grouped curved on the bottom is associated with non-normalized data.

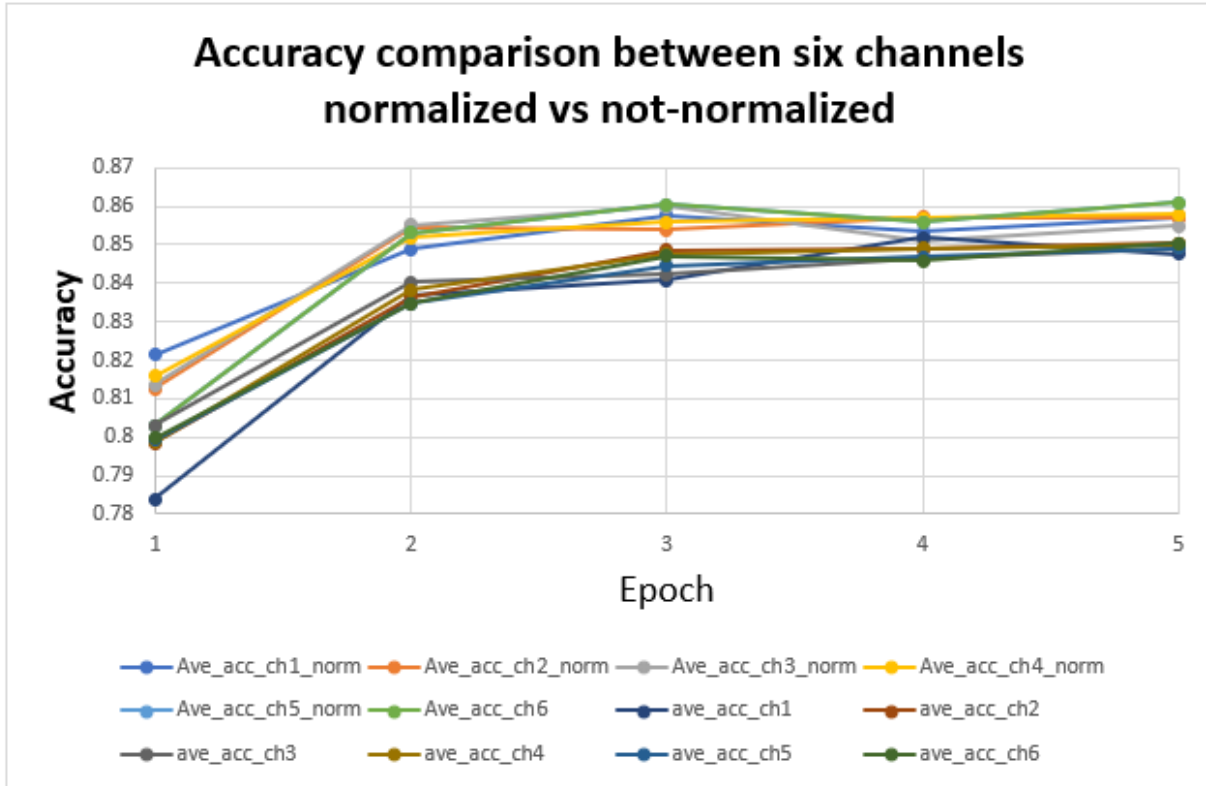


Figure 8-14 Comparison between normalized input and nonnormalized input. Accuracy values per channel for each epoch is shown for 10 inputs, half normalized and the other half non-normalized. The accuracy level increased at the beginning and became stable at around 0.85 for normalized channel and around 0.84 for the non-normalized channel. Each color represents one channel.

8.4.2 Improving the Performance

Leveraging our past studies on developing devices that use different light sources for cancer detection and diagnoses, we identified light channel ratios that might show better discriminative results (Zeng, MacAulay, et al. 1999). The ratios that we found are listed below where 1-3 showed the highest discriminative results.

1. Blue Excitation Green Emission (BG) / white Illumination Red Reflectance (wR)

2. UV Excitation Blue Emission (UB) / Violet Excitation Blue Emission (VB)
3. white Illumination Blue Reflectance (wB)/white Illumination Red Reflectance (wR)
4. Violet Excitation Green Emission (VG) / white Illumination Red Reflectance (wR)
5. Violet Excitation Green Emission (VG) / Blue Excitation Red Emission (BR)

The networks that we used so far accepts only 3 dimensional images, so we decided to construct an artificial RGB image using the top three ratios provided from the list above, we constructed a new RGB channel, shown in Figure 8-15. Channel R has BG/wR, channel G has UB/VB and channel B has wB/wR.

We used VGG16 as our CNN model and modified MDC images as the input (as explained above). We monitored the performance of our model on a larger epoch (25) as shown in Figure 8-16 (only one channel is shown here). We can observe that the accuracy is slightly improved at the beginning and plateaus around 0.848.

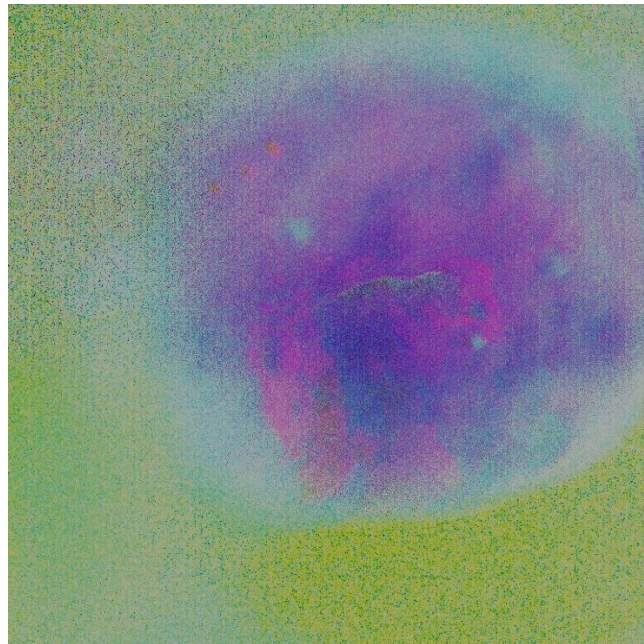


Figure 8-15 RGB images are constructed from a ratio image.

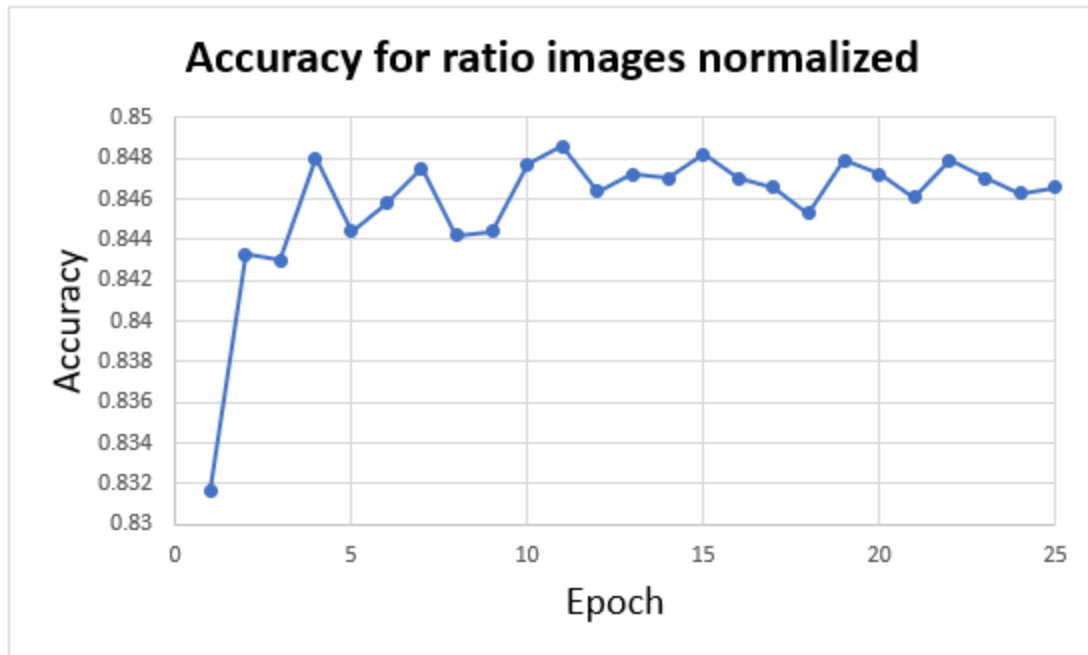


Figure 8-16 Monitoring the performance for a larger epoch for the normalized input. The accuracy value for channel 1 during 25 epochs are monitored. The accuracy level increased at the beginning and became stable at around 0.848.

8.5 Discussion

Convolutional Neural Networks (ConvNets) have played a very important role in large-scale image and video recognition. Machine Vision has entered a new era by advances in deep learning and high-speed GPUs. Researchers have attempted to improve the performance and accuracy of the original architecture presented by (Krizhevsky, Sutskever, et al. 2017) by using different window sizes, strides, filters, and deeper networks. Our goal was to improve the current infrastructure for cervical cancer screening by adapting and fine-tuning the existing implementation of deep learning algorithms. We are not aware of any work that uses VGG-like or ResNet-like architecture for abnormality detection in the cervix using MDC images. In summary,

we were able to show that it is possible to improve the performance by using another representation of the input (ratio images).

Sharma's team used classical ML for prediction of cervical cancer stage. His team obtained an accuracy of 67% using decision tree (Sunny Sharma 2016) .

Tomio's team recently published their study on application of deep learning on classification of colposcopy images. They obtained around 50% accuracy on the validation dataset. (Sato et al. 2018).

One of the advantage of our method is that it incorporates the feedback from 11 reviewers. Also the gold standard for our reviewers was the histopathological mapping.

Although by using CNN as our classifier we removed the dependency to hand-crafted features and improved the generalization capability of our classifier, we lost the feature interpretability of classical ML. This limitation exists by nature in our algorithm, it would be interesting to continue this research and implement interpretable CNN.

The other limitation of this study is treating each RGB image as an independent sample, the reason for this assumption is due to CNN architecture limitation that does not allow for higher dimensional input. We are aware of this limitation and we plan to implement custom made CNN to accept higher dimensional input.

Chapter 9 Conclusion and Future Work

In this chapter, we present a summary of our research and the contributions of this thesis. In addition, we discuss the potential future research directions and applications of our work.

9.1 Prostate Cancer: Gland Detection, Image Segmentation, and Prediction

Chapter 2 through 5 are focused on prostate cancer. In these chapters, we discussed prostate cancer and introduced methods for improved detection and prognoses.

Chapter 2:

In chapter 2 we introduced pixel-level techniques for glandular structures detection, these structures are important since they contain clinically relevant information regarding the disease status and Gleason score. In this chapter, we blended different techniques such as hyper-spectral imaging, PCA, clustering methods to develop a new automated technique for glandular detection. We obtained promising results on data obtained from 132 prostate TMA cores. In future studies, it would be beneficial to obtain larger datasets and investigate the effectiveness of this method on different datasets or different stains batches to determine the performance of our algorithm across the different H&E staining protocols that are in use clinically at different institutions.

Chapter 3:

In this chapter, we designed an automated system to recognize and localize cell nuclei in digitally scanned images. This is a very important step to enable quantitative digital pathology.

We constructed images of a Feulgen-Thionin-stained prostate cancer TMA from 33 prostate cancer patients. We developed a fully automated system to localize prostate cell nuclei and overcame implementation challenges such as tissue complexity, cell clusters, background noise, image and tissue inhomogeneities, etc. Using this method, we obtained a 96.6% true-positive (correct segmentation) rate and a 12% false-positive rate. Although this method was very successful it had some difficulties in the detection of faint nuclei.

Chapter 4:

To overcome some of the difficulties we encountered in chapter 3 and to improve the segmentation quality of our technique, we introduced an interactive segmentation algorithm. This method required minimum user interaction and produced a very good segmentation result. The images used here were selected from 33 prostate cancer patients as explained earlier.

Our algorithm requires less time to complete the segmentation process comparing with a human expert (45 min vs 6 min) with comparable accuracy. This shows that our algorithm is fast, efficient, and accurate which makes it suitable to be used in the clinical settings.

In a pilot study (on very small sample size) we studied other tissue types such as oral, cervix, lung and other stains such as H&E and obtained promising results using our proposed algorithm.

Chapter 5:

In this chapter, we discussed the development of a machine learning algorithm trained on the features extracted from cells/nuclei that were segmented using our technique as explained in the earlier chapters.

We constructed a large feature space and implemented a new classification method (LSVM algorithm) to use these features as the input. We implemented our algorithm and ran our proposed method on our dataset and achieved 87% accuracy for Gleason score prediction and 92% for aggressive vs non-aggressive prediction. Our method requires much less human interaction, a requirement for an automated Quantitative Digital Pathology (QDP) system.

We showed that the LSVM algorithm could serve as a predictive tool for prostate cancer progression prediction and reduce over-diagnosis/treatment. Combining our algorithm with a high-resolution imaging system could potentially improve the performance even more.

The method presented here can change clinical pathology by identifying aggressive prostate cancer lesions that require up-front treatments vs. non-aggressive lesions that need to be under active surveillance.

9.2 Renal Carcinoma: Vascular Structure Segmentation, Feature Extraction, and Classification

Chapter 6:

In this chapter, we presented the complete process of renal carcinoma classification from the image processing step to feature selection and development of machine learning techniques. In this chapter, we designed a new machine learning classifier to discriminate between different renal carcinoma subtypes. The data from this study were collected from 50 consecutive patients with renal cell carcinoma at Pasteur Hospital, Nice, France. These tissues were annotated by the study pathologist (Dr. Ambrosetti) into CCRCC, PRCC and mixed CCRCC/PRCC. We used advanced image processing techniques to extract the skeleton of renal blood vessels and then extracted

features from the processed vessels. These features were used as the input to our ML algorithm. We obtained a 95% training accuracy for classification of CCRCC and other tissue sites. The classification of CCP vs other tissue sites and PCC vs other tissue sites were promising as well.

Integrating other features (e.g. cellular- level features), applying other stains (e.g. H&E), using different de-convolution image processing methods and collecting more data are among the various interesting future works for this chapter that could potentially improve the proposed classification performance.

9.3 Cervical Cancer: Abnormalities Detection Using Classical and Convolutional Neural Network

Chapter 7:

In chapters 7 and 8 we designed techniques for cervical cancer abnormality detection at the macroscopic scale.

In chapter 7 we explained the whole workflow of designing an automated classification method starting from multispectral tissue imaging to the development of ML classifiers. The dataset used for this study were obtained from 49 patients. We captured reflectance and fluorescence images from these patients' cervix using our in-house Multispectral Digital Colposcope (MDC) device. We extracted features from histogram intensities of small sliding windows and applied various classification methods.

We obtained %93 and %91 accuracy when we applied a boosted-tree classifier to the training and test set, respectively. We also incorporated other features such as ratio features to improve classification accuracy.

When we applied our method on a newly acquired dataset the accuracy dropped which could be due to various reasons such as variation in the imaging device or overtraining of the method.

In the last chapter (chapter 8) we leveraged a convolutional neural network to improve the generalization limitation of our classical ML method.

Chapter 8:

Deep Neural Networks (DNN) has shown promising results, by automatically learning important features and eliminating the need to design hand-crafted features. In this chapter, we investigated the performance of different convolutional neural networks using MDC images to localize the at-risk cervical pre-cancer tissues. In this chapter, we implemented various convolutional neural networks and explained how it would fit as our solution.

We obtained promising results using the VGG16 network. The results were comparable with classical ML. As a future direction, it is possible to improve the performance by experimenting with different window sizes, strides, filters, and deeper networks and/or obtaining larger datasets.

9.4 Future Directions

Prostate Cancer (Segmentation): It is possible to improve the glandular segmentation process by taking advantage of neural networks. One of the famous networks that we can possibly use is UNet. UNet is a CNN that is designed for biomedical image segmentation. This Network uses data augmentation to compensate for the need for large annotated samples (Ronneberger, Fischer, et al. 2015). It would be interesting to see whether CNN could increase the segmentation accuracy. The other interesting approach would be to combine all images captured from different bandwidths and use this matrix as the input to the UNet.

A great extension of this study would be to investigate the performance of our algorithm on other tissue types (e.g breast tissue). We investigated and tested our proposed glandular detection algorithm on prostate cancer, however, our proposed method could be simply modified for gland detection of other tissues. It is necessary to fine-tune our method for the tissue under study (e.g. different stains and gland structure/size).

Furthermore, we tested our proposed cell segmentation methods on TMA images of prostate cancer cells (stained with Thionin). We also showed that it is possible to use this method to segment other cell types (e.g. lung, cervix, oral) and other stains (H&E). This generalization capability made our interactive method an essential core for developing an automated cell segmentation tool for different tissue types. More investigations on the generalization capability of our algorithm is essential.

Prostate Cancer (Classification):As we implemented a very comprehensive hand-crafted feature-set to accomplish the classification task, it seems natural to replace this process with an advanced CNN. However, by doing so, we would lose the feature interpretation which is still very important in the clinical settings. There are newer, more advanced CNN referred to as Interpretable CNNs (Zhang, et al, 2019) with comparable interpretability capability to the classical ML that one can possibly investigate.

Renal Carcinoma:We used the features obtained from blood vessels for Renal carcinoma classification. Blood vessels are considered an important biomarker and are used in differential diagnosis of different cancer types such as renal and skin. We conducted our study on renal carcinoma, however, a modified version of our algorithm could be used for the diagnostic purposes of other diseases where monitoring the blood structure is important. It is possible to improve the

performance by collecting more samples and constructing a pool of relatively even-balanced healthy and non-healthy tissues, extracting clinically relevant features such as (cell/nuclei features), collecting higher quality images, incorporate other imaging modalities and other stains. It would be interesting to see the performance of a deep learning algorithm on renal carcinoma classification (which eliminates the necessity to the design hand-crafted feature), however it requires access to a very large dataset.

Cervical cancer: The last cancer under study was cervical cancer. We presented an automated abnormality detection using both classical ML (chapter 7) and deep learning methods (chapter 8).

It would be interesting to see whether there would be any improvement in the performance if the classical and deep learning methods are combined together to form a hybrid model. To accomplish this task, there are two possible design approaches to explore:

- 1- First, apply classical ML techniques such as SVM, boosted tree, etc and obtain the probabilities of each pixel/window, next use this probability map as the input of a deep CNN.
- 2- Use an ensemble classifier, where both ML technique (classical and deep learning) would try to do the prediction in a parallel fashion and finally the ensemble classifier makes the final decision based on the results obtained from both classifiers (in this scenario it is possible to have multiple classical ML and multiple CNN).

The other avenue to explore is to observe the impact of combining all 18 channels together and use it as the input. In a pilot study, we explored this option. This required to redesign and modify the VGG16 Network since the current design only accepts 3 channels. We boosted the accuracy by about %5. This is a new and very interesting finding and more tuning and validation are needed.

General ways to improve the CNN performance would be to collect more MDC images, use the weights obtained from training medical images instead of ImageNet (as it would be more relevant to the subject under the study).

Similar approaches could be used to apply our methods to other cancer types by fine-tuning the existing implementation of deep learning algorithms to cancer/tissue under the study.

9.5 Significance and Application of This Study

The number of patients diagnosed with cancer increases daily. Being able to distinguish cancers requiring aggressive treatment from those which do not, would allow for the tailoring of more appropriate treatments to the individuals and would increase cancer survival rate and improve the quality of life of those living with cancer. Individualized treatments are the future of oncology and ML techniques are an effective solution to that. In this thesis, I presented new, cost-effective prognostic tools for cancer using advanced ML techniques. I introduced methods to apply ML techniques to microscopic and macroscopic images to improve our understanding of the neoplastic process, localization, and prediction of biological behavior of early cancers. Specifically, I presented techniques to apply ML techniques to digital pathology images of prostate and renal carcinoma, and macroscopic multispectral images of high grade cervical intraepithelial neoplasia lesions. These techniques have great potential to improve the healthcare environment by providing extra information/second opinion to the medical experts or to be used as the first line of screening/localization and detection.

References

- , CIELAB color space-05-23T21:22:37Z, 2019-last update.
Available: https://en.wikipedia.org/w/index.php?title=CIELAB_color_space&oldid=898482145 [May 31, 2019].
- , Color difference-09-24T09:09:32Z, 2019-last update.
Available: https://en.wikipedia.org/w/index.php?title=Color_difference&oldid=917541220 [Oct 19, 2019].
- Eccentricity (mathematics). 2019.
- Euler's number (mathematical constant). 2019.
- Linear discriminant analysis. 2019.
- , Prostate Cancer Statistics | CDC-01-18T10:16:26Z/, 2019-last update.
Available: <https://www.cdc.gov/cancer/prostate/statistics/index.htm> [May 2, 2019].
- , Rectifier (neural networks)-08-18T06:00:03Z, 2019-last update.
Available: [https://en.wikipedia.org/w/index.php?title=Rectifier_\(neural_networks\)&oldid=911329818](https://en.wikipedia.org/w/index.php?title=Rectifier_(neural_networks)&oldid=911329818) [Aug 25, 2019].
- , Softmax function-08-01T00:08:25Z, 2019-last update.
Available: https://en.wikipedia.org/w/index.php?title=Softmax_function&oldid=908788637 [Aug 25, 2019].
- Voronoi diagram. 2019.
- , Cervical cancer statistics-08-22T13:43:41+01:00, 2018-last update.
Available: <https://www.wcrf.org/dietandcancer/cancer-trends/cervical-cancer-statistics> [Dec 29, 2018].
- , Human papillomavirus infection-12-18T11:19:14Z, 2018-last update.
Available: https://en.wikipedia.org/w/index.php?title=Human_papillomavirus_infection&oldid=874301344 [Dec 29, 2018].
- , Kidney cancer statistics-08-22T14:21:31+01:00, 2018-last update.
Available: <https://www.wcrf.org/dietandcancer/cancer-trends/kidney-cancer-statistics> [May 11, 2019].
- , Binarize 2-D grayscale image or 3-D volume by thresholding.
Available: <https://www.mathworks.com/help/images/ref/imbinarize.html>.
- , Cancer of the Prostate - Cancer Stat Facts.
Available: <https://seer.cancer.gov/statfacts/html/prost.html> [May 2, 2019].
- , DRIVE: Digital Retinal Images for Vessel Extraction.
Available: <https://www.isi.uu.nl/Research/Databases/DRIVE/>.
- , Feature scaling. Available: https://en.wikipedia.org/wiki/Feature_scaling.

Available: https://en.wikipedia.org/wiki/Floyd%E2%80%93Warshall_algorithm.

, Generalized Linear Models — scikit-learn 0.20.2 documentation [1/6/, 2019].

, Haar 1-D wavelet transform - MATLAB haart.

Available: <https://www.mathworks.com/help/wavelet/ref/haart.html>.

, Introduction to Boosted Trees.

Available: <https://xgboost.readthedocs.io/en/latest/tutorials/model.html>.

, Keras Training Documentation. Available: <https://keras.io/getting-started/faq/#what-does-sample-batch-epoch-mean>.

, Key Statistics About Kidney Cancer. Available: <https://www.cancer.org/cancer/kidney-cancer/about/key-statistics.html> [May 11, 2019].

, Max-pooling / Pooling.

, Morphological operations on binary images.

Available: <https://www.mathworks.com/help/images/ref/bwmorph.html>.

, Morphology - Skeletonization/Medial Axis Transform.

Available: <https://homepages.inf.ed.ac.uk/rbf/HIPR2/skeleton.htm>.

, Principal component analysis of raw data.

Available: <https://www.mathworks.com/help/stats/pca.html#namevaluepairarguments>.

System and method for automatically detecting malignant cells and cells having malignancy-associated changes.

ADAMEK, T., O'CONNOR, N.E. and MURPHY, N., 2005. Region-based segmentation of images using syntactic visual features, 2005.

ADAMS, R. and BISCHOF, L., 1994. Seeded region growing. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 16(6), pp. 641-647.

ALEXE, B., DESELAERS, T. and FERRARI, V., June 2010. What is an object? June 2010, pp. 73-80.

ALOM, M.Z., YAKOPCIC, C., HASAN, M., TAHA, T.M. and ASARI, V.K., 2019. Recurrent residual U-Net for medical image segmentation. *Journal of Medical Imaging*, 6(1), pp. 1.

ALOMARI, R.S., ALLEN, R., SABATA, B. and CHAUDHARY, V., 2009. Localization of tissues in high-resolution digital anatomic pathology images, 2009, SPIE, pp. 726016-7260110.

BABAK EHTESHAMI BEJNORDI, 2017. Histopathological diagnosis of breast cancer using machine learning.

BEKTAS, C.T., KOCAK, B., YARDIMCI, A.H., TURKCANOGLU, M.H., YUCETAS, U., KOCA, S.B., ERDIM, C. and KILICKESMEZ, O., 2019. Clear Cell Renal Cell Carcinoma: Machine Learning-Based Quantitative Computed Tomography Texture Analysis for Prediction of Fuhrman Nuclear Grade. *European radiology*, 29(3), pp. 1153-1163.

- BENAVIDES, J., CHANG, S., PARK, S., RICHARDS-KORTUM, R., MACKINNON, N., MACAULAY, C., MILBOURNE, A., MALPICA, A. and FOLLEN, M., 2003. Multispectral digital colposcopy for in vivo detection of cervical cancer. *Optics express*, 11(10), pp. 1223.
- BENGTTSSON, E. and RODENACKER, K., 2003. A feature set for cytometry on digitized microscopic images. *Analytical Cellular Pathology*, 24(1), pp. 1.
- BENSCH, R. and RONNEBERGER, O., April 2015. Cell segmentation and tracking in phase contrast images using graph cut with asymmetric boundary costs, April 2015, pp. 1220-1223.
- BOTHA, M.H., 2015. Natural History of Human Papilloma Virus Infection of the Cervix. *Current Obstetrics and Gynecology Reports*, 4(4), pp. 189-194.
- BOUCHERON, L.E., 2008. Object- and spatial-level quantitative analysis of multispectral histopathology images for detection and characterization of cancer, ProQuest Dissertations Publishing.
- CAI, H., HUANG, Q., RONG, W., SONG, Y., LI, J., WANG, J., CHEN, J. and LI, L., 2019. Breast Microcalcification Diagnosis Using Deep Convolutional Neural Network from Digital Mammograms. *Computational and Mathematical Methods in Medicine*, 2019, pp. 1-10.
- CAMPBELL, S.C., RINI, B.I. and SPRINGERLINK EBOOKS - MEDICINE, 2012. Renal cell carcinoma: clinical management. 1. Aufl.; 2013 edn. New York: Humana Press.
- CANNY, J., 1986. A Computational Approach to Edge Detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, PAMI-8(6), pp. 679-698.
- CANZONIERI, V., MONFARDINI, S. and CARBONE, A., 1998. Defining prognostic factors in malignancies through image analysis. England: Elsevier Ltd.
- CHANG, S.K., FOLLEN, M., MALPICA, A., UTZINGER, U., STAERKEL, G., COX, D., NEELY ATKINSON, E., MACAULAY, C. and RICHARDS-KORTUM, R., 2002. Optimal excitation wavelengths for discrimination of cervical neoplasia. *IEEE Transactions on Biomedical Engineering*, 49(10), pp. 1102-1111.
- CHANG, S., ABDUL-KAREEM, S., MERICAN, A.F. and ZAIN, R.B., 2013. Oral cancer prognosis based on clinicopathologic and genomic markers using a hybrid of feature selection and machine learning methods. *BMC bioinformatics*, 14(1), pp. 170.
- CHERIET, M., SAID, J.N. and SUEN, C.Y., 1998. A recursive thresholding technique for image segmentation. *IEEE Transactions on Image Processing*, 7(6), pp. 918-921.
- COLEMAN, G.B. and ANDREWS, H.C., 1979. Image segmentation by clustering. *Proceedings of the IEEE*, 67(5), pp. 773-785.
- CRUZ, J.A. and WISHART, D.S., 2006. Applications of Machine Learning in Cancer Prediction and Prognosis. *Cancer Informatics*, 2, pp. 59-78.
- CUI, S., LUO, Y., TSENG, H., TEN HAKEN, R.K. and EL NAQA, I., 2019. Combining Handcrafted Features with Latent Variables in Machine Learning for Prediction of Radiation-Induced Lung Damage. *Medical physics*, .

DAVID FORMAN, CATHERINE DE MARTEL, CHARLES J LACEY, 2012. Global Burden of Human Papillomavirus and Related Diseases. *Vaccine*, 30, pp. F1-F23.

DE BREBISSON, A. and MONTANA, G., 2015. Deep Neural Networks for Anatomical Brain Segmentation.

DENG, J., DONG, W., SOCHER, R., LI, L., LI, K. and FEI-FEI, L., 2009. ImageNet: A large-scale hierarchical image database, 2009, IEEE, pp. 248-255.

DO, T. and ARTIÈRES, T., 2009. Large Margin Training for Hidden Markov Models with Partially Observed States, 2009, ACM, pp. 265–272.

DONG, Y.B., LI, M.J. and CHENG, G.L., 2014. Evaluation and Comparison of Thresholding Segmentation Techniques. *Applied Mechanics and Materials*, 519-520, pp. 689-692.

DOYLE, S., HWANG, M., SHAH, K., MADABHUSHI, A., FELDMAN, M. and TOMASZEWSKI, J., 2007. AUTOMATED GRADING OF PROSTATE CANCER USING ARCHITECTURAL AND TEXTURAL IMAGE FEATURES, 2007, pp. 1284-1287.

DOYLE, S., 2011. Computerized detection, segmentation and classification of digital pathology: Case study in prostate cancer, ProQuest Dissertations Publishing.

ENFIELD, K.S.S., MARTIN, S.D., MARSHALL, E.A., KUNG, S.H.Y., GALLAGHER, P., MILNE, K., CHEN, Z., NELSON, B.H., LAM, S., ENGLISH, J.C., MACAULAY, C.E., LAM, W.L. and GUILLAUD, M., 2019. Hyperspectral cell sociology reveals spatial tumor-immune cell interactions associated with lung cancer recurrence. *Journal for ImmunoTherapy of Cancer*, 7(1),.

FEHR, D., VEERARAGHAVAN, H., WIBMER, A., GONDO, T., MATSUMOTO, K., VARGAS, H.A., SALA, E., HRICAK, H. and DEASY, J.O., 2015. Automatic classification of prostate cancer Gleason scores from multiparametric magnetic resonance images. *Proceedings of the National Academy of Sciences of the United States of America*, 112(46), pp. E626-E6273.

FELZENSZWALB, P.F., GIRSHICK, R.B., MCALLESTER, D. and RAMANAN, D., 2010. Object Detection with Discriminatively Trained Part-Based Models. *IEEE Trans. Pattern Anal. Mach. Intell.*, 32(9), pp. 1627–1645.

FENNER, A., 2015. How to differenti-8 Gleason scores?: Prostate cancer. *Nature Reviews Urology*, 12(11), pp. 597.

FERRIS, D.G., LAWHEAD, R.A., DICKMAN, E.D., HOLTZAPPLE, N., MILLER, J.A., GROGAN, S., BAMBOT, S., AGRAWAL, A. and FAUPEL, M.L., 2001. Multimodal hyperspectral imaging for the noninvasive diagnosis of cervical neoplasia. *Journal of lower genital tract disease*, 5(2), pp. 65-72.

FU, H., QIU, G., SHU, J. and ILYAS, M., 2014. A Novel Polar Space Random Field Model for the Detection of Glandular Structures. *IEEE Transactions on Medical Imaging*, 33(3), pp. 764-776.

- FUNG, V., -07-17T09:59:13.909Z, 2017-last update, An Overview of ResNet and its Variants. Available: <https://towardsdatascience.com/an-overview-of-resnet-and-its-variants-5281e2f56035> [Aug 25, 2019].
- GAN, J. and TAO, Y., 2017. Dynamic Density Based Clustering, 2017, ACM, pp. 1493-1507.
- GONZALEZ, R.C. and WOODS, R.E., 2008. Digital image processing. 3rd edn. Upper Saddle River, NJ: Pearson/Prentice Hall.
- GONZALEZ, R.C., WOODS, R.E. and EDDINS, S.L., 2004. Digital Image processing using MATLAB. Upper Saddle River, N. J: Pearson Prentice Hall.
- GUNDUZ-DEMIR, C., KANDEMIR, M., TOSUN, A.B. and SOKMENSUER, C., 2010. Automatic segmentation of colon glands using object-graphs. Medical image analysis, 14(1), pp. 1-12.
- GUO, Y., LIU, Y., OERLEMANS, A., LAO, S., WU, S. and LEW, M.S., 2016. Deep learning for visual understanding: A review. Neurocomputing, 187, pp. 27-48.
- HALICEK, M., LITTLE, J.V., WANG, X., CHEN, A.Y. and FEI, B., 2019. Optical biopsy of head and neck cancer using hyperspectral imaging and convolutional neural networks. Journal of Biomedical Optics, 24(3), pp. 1.
- HAMILTON, P.W., ANDERSON, N.H., DIAMOND, J., BARTELS, P.H., GREGG, J.B., THOMPSON, D. and MILLAR, R.J., 1996. An interactive decision support system for breast fine needle aspiration cytology. Analytical and quantitative cytology and histology, 18(3), pp. 185.
- HANIN, B., 2018. Which Neural Net Architectures Give Rise To Exploding and Vanishing Gradients? .
- HASHIM AHMED, MANIT ARYA, FREEMAN, A. and EMBERTON, M., 2012. Do low-grade and low-volume prostate cancers bear the hallmarks of malignancy? Lancet Oncology, The, 13(11), pp. e50-e517.
- HE, K., ZHANG, X., REN, S. and SUN, J., 2016. Deep Residual Learning for Image Recognition, 2016, IEEE, pp. 770-778.
- HEINDL, A., NAWAZ, S. and YUAN, Y., 2015. Mapping spatial heterogeneity in the tumor microenvironment: a new era for digital pathology. Laboratory Investigation; a Journal of Technical Methods and Pathology, 95(4), pp. 377-384.
- HODGSON, S., 2007. Mechanisms of inherited cancer susceptibility. 浙江大学学报：B卷英文版, 9(1), pp. 1-4.
- IAKOVLEV, V.V., GABRIL, M., DUBINSKI, W., SCORILAS, A., YOUSSEF, Y.M., FARAGALLA, H., KOVACS, K., ROTONDO, F., METIAS, S., ARSANIOUS, A., PLOTKIN, A., GIRGIS, A.H.F., STREUTKER, C.J. and YOUSEF, G.M., 2012. Microvascular density as an independent predictor of clinical outcome in renal cell carcinoma: an automated image analysis study. Laboratory investigation; a journal of technical methods and pathology, 92(1), pp. 46-56.

IRSHAD, H., VEILLARD, A., ROUX, L. and RACOCEANU, D., 2014. Methods for Nuclei Detection, Segmentation, and Classification in Digital Histopathology: A Review-Current Status and Future Potential. *IEEE Reviews in Biomedical Engineering*, 7, pp. 97-114.

JALALIAN, AFSANEH|MASHOHOR, SYAMSIAH B.T.|MAHMUD, HAJJAH ROZI|SARIPAN, M. IQBAL B.|RAMLI,ABDUL RAHMAN B.|KARASFI, BABAK, 2013. Computer-aided detection/diagnosis of breast cancer in mammography and ultrasound: a review. *Clinical imaging*, 37(3), pp. 420-426.

KASS, M., WITKIN, A. and TERZOPOULOS, D., 1988. Snakes: Active contour models. *International Journal of Computer Vision*, 1(4), pp. 321-331.

KIM, W., KIM, K.S., LEE, J.E., NOH, D., KIM, S., JUNG, Y.S., PARK, M.Y. and PARK, R.W., 2012. Development of Novel Breast Cancer Recurrence Prediction Model Using Support Vector Machine. *Journal of Breast Cancer*, 15(2), pp. 230-238.

KINGMA, D.P. and BA, L.J., 2015. Adam: A Method for Stochastic Optimization, 2015, arXiv.org.

KOPP, R.P., PARSONS, J.K., SHIAU, J., WANG-RODRIGUEZ, J., PALAZZI-CHURAS, K., SILBERSTEIN, J.L., DERWEESH, I.H. and SAKAMOTO, K., 2011. Prostate atypia: clinical and pathological variables associated with cancer diagnosis on repeat biopsy. *Prostate Cancer and Prostatic Diseases*, 14(2), pp. 149-154.

KOUROU, K., EXARCHOS, T.P., EXARCHOS, K.P., KARAMOUZIS, M.V. and FOTIADIS, D.I., 2015. Machine learning applications in cancer prognosis and prediction. *Computational and Structural Biotechnology Journal*, 13(C), pp. 8-17.

KRIZHEVSKY, A., SUTSKEVER, I. and HINTON, G., 2017. ImageNet classification with deep convolutional neural networks. New York: ACM.

LALAOUI, L. and MOHAMADI, T., 2013. A comparative study of Image Region-Based Segmentation Algorithms. *International Journal of Advanced Computer Science and Applications*, 4.

LAM, S., MACAULAY, C., HUNG, J., LERICHE, J., PROFIO, A.E. and PALCIC, B., 1993. Detection of dysplasia and carcinoma in situ with a lung imaging fluorescence endoscope device. *The Journal of thoracic and cardiovascular surgery*, 105(6), pp. 1035.

LAM, S., MACAULAY, C., LERICHE, J.C. and PALCIC, B., 2000. Detection and localization of early lung cancer by fluorescence bronchoscopy. *Cancer*, 89(S11), pp. 2468-2473.

LANE, P.M., GILHULY, T., WHITEHEAD, P., ZENG, H., POH, C.F., NG, S., WILLIAMS, P.M., ZHANG, L., ROSIN, M.P. and MACAULAY, C.E., 2006. Simple device for the direct visualization of oral-cavity tissue fluorescence. *Journal of Biomedical Optics*, 11(2), pp. 24006.

LANE, P., POH, C.F., DURHAM, J.S., ZHANG, L., LAM, S.F., ROSIN, M. and MACAULAY, C., 2011. Fluorescence-guided surgical resection of oral cancer reduces recurrence, 2011, pp. 78832-6.

- LANGER, L., BINENBAUM, Y., GUGEL, L., AMIT, M., GIL, Z. and DEKEL, S., 2015. Computer-aided diagnostics in digital pathology: automated evaluation of early-phase pancreatic cancer in mice. *International Journal of Computer Assisted Radiology and Surgery*, 10(7), pp. 1043-1054.
- LATSON, L., SEBEK, B. and POWELL, K.A., 2003. Automated cell nuclear segmentation in color images of hematoxylin and eosin-stained breast biopsy. *Analytical and quantitative cytology and histology*, 25(6), pp. 321.
- LECUN, Y., BENGIO, Y. and HINTON, G., 2015. Deep learning. *Nature*, 521(7553), pp. 436-444.
- LEE, J.H., HA, E.J. and KIM, J.H., 2019. Application of deep learning to the diagnosis of cervical lymph node metastasis from thyroid cancer with CT. *European radiology*, , pp. 1-6.
- LG, A., AT, E., A, P., M, E. and AR, R., 2013. Using Three Machine Learning Techniques for Predicting Breast Cancer Recurrence. *Journal of Health & Medical Informatics*, 4(2), pp. 1-3.
- LIKAS, A., VLASSIS, N. and J. VERBEEK, J., 2003. The global K-means clustering algorithm. *Pattern Recognition*, 36(2), pp. 451-461.
- LISA TORREY, J.S., 2009. *Transfer Learning*.
- LONG, J., SHELHAMER, E. and DARRELL, T., June 2015. Fully convolutional networks for semantic segmentation, June 2015, pp. 3431-3440.
- MACAULAY, C., KEYES, M., HAYES, M., LO, A., WANG, G., GUILLAUD, M., GLEAVE, M., FAZLI, L., KORBELIK, J., COLLINS, C., KEYES, S. and PALCIC, B., 2017. Quantification of large scale DNA organization for predicting prostate cancer recurrence. *Cytometry Part A*, 91(12), pp. 1164-1174.
- MADABHUSHI, A. and LEE, G., 2016. Image analysis and machine learning in digital pathology: Challenges and opportunities. *Medical image analysis*, 33, pp. 170-175.
- MANIKANDAN, S., RAMAR, K., WILLJUICE IRUTHAYARAJAN, M. and SRINIVASAGAN, K.G., 2014. Multilevel thresholding for segmentation of medical brain images using real coded genetic algorithm. *Measurement*, 47, pp. 558-568.
- MAO, L. and XIE, M., 2013. Multiclass Image Segmentation Based on Pixel and Segment Level. *Research Journal of Applied Sciences, Engineering and Technology*, 11(9), pp. 2238-2244.
- MCCULLOUGH, D.P., GUDLA, P.R., HARRIS, B.S., COLLINS, J.A., MEABURN, K.J., NAKAYA, M.-, YAMAGUCHI, T.P., MISTELI, T. and LOCKETT, S.J., 2008. Segmentation of Whole Cells and Cell Nuclei From 3-D Optical Microscope Images Using Dynamic Programming. *IEEE Transactions on Medical Imaging*, 27(5), pp. 723-734.
- MEIBURGER, K.M., MOLINARI, F., WONG, J., AGUILAR, L., GALLO, D., STEINMAN, D.A. and MORBIDUCCI, U., 2016. Validation of the Carotid Intima–Media Thickness Variability: Can Manual Segmentations Be Trusted as Ground Truth? *Ultrasound in medicine & biology*, 42(7), pp. 1598-1611.

- MEIJERING, E., 2012. Cell Segmentation: 50 Years Down the Road [Life Sciences]. IEEE Signal Processing Magazine, 29(5), pp. 140-145.
- MILBOURNE, A., PARK, S.Y., BENEDET, J.L., MILLER, D., EHLEN, T., RHODES, H., MALPICA, A., MATISIC, J., VAN NIEKIRK, D., ATKINSON, E.N., HADAD, N., MACKINNON, N., MACAULAY, C., RICHARDS-KORTUM, R. and FOLLEN, M., 2005. Results of a pilot study of multispectral digital colposcopy for the in vivo detection of cervical intraepithelial neoplasia. Gynecologic oncology, 99(3), pp. S67-S75.
- MOURANT, J.R., BIGIO, I.J., BOYER, J., CONN, R.L., JOHNSON, T. and SHIMADA, T., 1995. Spectroscopic diagnosis of bladder cancer with elastic light scattering. Lasers in surgery and medicine, 17(4), pp. 350.
- NAFE, R., YAN, B., SCHLOTE, W. and SCHNEIDER, B., 2006. Application of different methods for nuclear shape analysis with special reference to the differentiation of brain tumors. Analytical and Quantitative Cytology and Histology, 28(2), pp. 69-77.
- NAIK, S., DOYLE, S., AGNER, S., MADABHUSHI, A., FELDMAN, M. and TOMASZEWSKI, J., 2008. Automated gland and nuclei segmentation for grading of prostate and breast cancer histopathology, 2008, IEEE, pp. 284-287.
- NARAYANAN RAMAKRISHNAN, H., 2010. Detection and estimation of image blur. Masters Theses, .
- NGUYEN, K., SABATA, B. and JAIN, A.K., 2012. Prostate cancer grading: Gland segmentation and structural features. Pattern Recognition Letters, 33(7), pp. 951-961.
- NIELSEN, M.A., 2015. Neural Networks and Deep Learning. Determination Press.
- NING, J., ZHANG, D., ZHANG, L. and WU, C., 2010. Interactive image segmentation by maximal similarity based region merging. Pattern Recognition, 43(2), pp. 445-456.
- NORDSTROM, R.J., BURKE, L., NILOFF, J.M. and MYRTLE, J.F., 2001. Identification of cervical intraepithelial neoplasia (CIN) using UV-excited fluorescence and diffuse-reflectance tissue spectroscopy. Lasers in surgery and medicine, 29(2), pp. 118-127.
- ORIGONI, M., PRENDIVILLE, W. and PARASKEVAIDIS, E., 2015. Cervical Cancer Prevention: New Frontiers of Diagnostic Strategies. BioMed research international, 2015, pp. 250917-2.
- OZGUL, E.B., SOKMENSUER, C. and GUNDUZ-DEMIR, C., 2011. Detection of colon glands using subgraph modeling, 2011, IEEE, pp. 726-729.
- PÄÄKKÖNEN, J., PÄIVINEN, N., NYKÄNEN, M. and PAAVONEN, T., 2015. An automated gland segmentation and classification method in prostate biopsies: an image source-independent approach. Machine Vision and Applications, 26(1), pp. 103-113.
- PAIK, E.S., LEE, J., PARK, J., KIM, J., KIM, M., KIM, T., CHOI, C.H., KIM, B., BAE, D. and SEO, S.W., 2019. Prediction of survival outcomes in patients with epithelial ovarian cancer using machine learning methods. Journal of Gynecologic Oncology, 30.

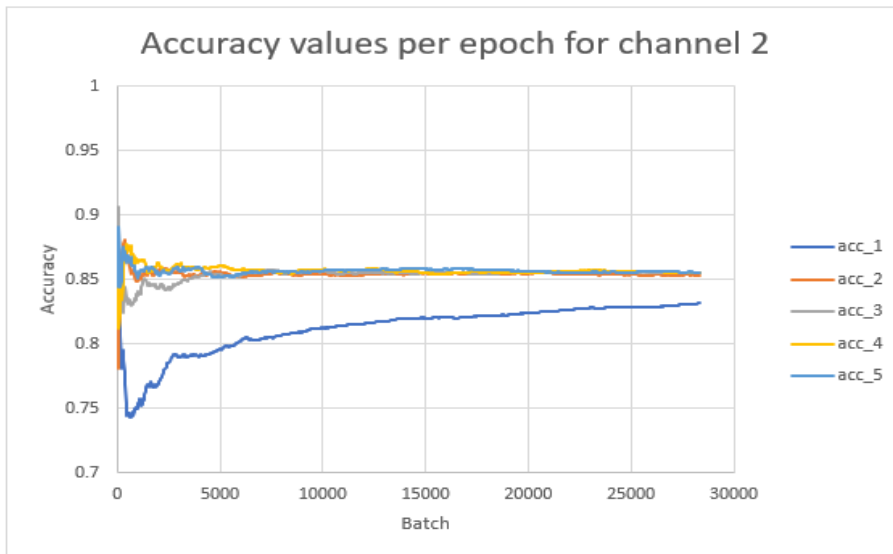
- PAPPAS, T.N., 1992. An adaptive clustering algorithm for image segmentation. *IEEE Transactions on Signal Processing*, 40(4), pp. 901-914.
- PARK, S.Y., COLLIER, T., AARON, J., MARKEY, M., RICHARDS-KORTUM, R., SOKOLOV, K., MACKINNON, N., MACAULAY, C., COGHLAN, L., MILBOURNE, A. and FOLLEN, M., 2005. Multispectral digital microscopy for in vivo monitoring of oral neoplasia in the hamster cheek pouch model of carcinogenesis. *Optics express*, 13(3), pp. 749.
- PELI, E., 1990. Contrast in complex images. *JOSA A*, 7(10), pp. 2032-2040.
- POH, C.F., NG, S.P., WILLIAMS, P.M., ZHANG, L., LARONDE, D.M., LANE, P., MACAULAY, C., ROSIN, M.P. and KRAUS, D.H., 2007. Direct fluorescence visualization of clinically occult high-risk oral premalignant disease using a simple hand-held device. *Head & neck*, 29(1), pp. 71-76.
- R, S., K, S. and HINDUSTHAN COLLEGE OF ENGINEERING AND TECHNOLOGY, 2014. A Survey of Image Segmentation Based On Multi Region Level Set Method. *International Journal of Computer Trends and Technology*, 18(6), pp. 252-256.
- RAMANUJAM, N., MITCHELL, M.F., MAHADEVAN, A., THOMSEN, S., MALPICA, A., WRIGHT, T., ATKINSON, N. and RICHARDS-KORTUM, R., 1996. Spectroscopic diagnosis of cervical intraepithelial neoplasia (CIN) in vivo using laser-induced fluorescence spectra at multiple excitation wavelengths. *Lasers in surgery and medicine*, 19(1), pp. 63.
- RAMANUJAM, N., MITCHELL, M.F., MAHADEVAN-JANSEN, A., THOMSEN, S.L., STAERKEL, G., MALPICA, A., WRIGHT, T., ATKINSON, N. and RICHARDS-KORTUM, R., 1996. Cervical precancer detection using a multivariate statistical algorithm based on laser-induced fluorescence spectra at multiple excitation wavelengths. *Photochemistry and photobiology*, 64(4), pp. 720-735.
- RAMANUJAM, N., MITCHELL, M.F., MAHADEVAN, A., WARREN, S., THOMSEN, S., SILVA, E. and RICHARDS-KORTUM, R., 1994. In vivo Diagnosis of Cervical Intraepithelial Neoplasia Using 337-nm- Excited Laser-Induced Fluorescence. *Proceedings of the National Academy of Sciences of the United States of America*, 91(21), pp. 10193-10197.
- RAÚL GÓMEZ, , Understanding Cross-Entropy Loss. Available: https://gombru.github.io/2018/05/23/cross_entropy_loss/.
- ROLI, F., KITTLER, J., SPRINGERLINK LECTURE NOTES IN COMPUTER SCIENCE, SPRINGERLINK EBOOKS - ENGLISH/INTERNATIONAL COLLECTION (ARCHIVE) and SPRINGERLINK (ONLINE SERVICE), 2001. Multiple classifier systems: third international workshop, MCS 2002, Cagliari, Italy, June 24-26, 2002 : proceedings. New York; Berlin: Springer.
- RONNEBERGER, O., FISCHER, P. and BROX, T., 2015. U-Net: Convolutional Networks for Biomedical Image Segmentation.
- SAIKIA, A.R., BORA, K., MAHANTA, L.B. and DAS, A.K., 2019. Comparative assessment of CNN architectures for classification of breast FNAC images. *Tissue and Cell*, 57, pp. 8-14.

- SEDGHI, A., PESTEIE, M., JAVADI, G., AZIZI, S., YAN, P., KWAK, J.T., XU, S., TURKBAY, B., CHOYKE, P., PINTO, P., WOOD, B., ROHLING, R., ABOLMAESUMI, P. and MOUSAVI, P., 2019. Deep neural maps for unsupervised visualization of high-grade cancer in prostate biopsies. *International journal of computer assisted radiology and surgery*, .
- SHAH, V., TURKBAY, B., MANI, H., PANG, Y., POHIDA, T., MERINO, M.J., PINTO, P.A., CHOYKE, P.L. and BERNARDO, M., 2012. Decision support system for localizing prostate cancer based on multiparametric magnetic resonance imaging. *Medical physics*, 39(7), pp. 4093-4103.
- SIEGEL, R.L., MILLER, K.D. and JEMAL, A., 2018. Cancer statistics, 2018. *CA: A Cancer Journal for Clinicians*, 68(1), pp. 7-30.
- SIMONYAN, K. and ZISSERMAN, A., 2014. Very Deep Convolutional Networks for Large-Scale Image Recognition.
- SINGH, M., KALAW, E.M., GIRON, D.M., CHONG, K., TAN, C.L. and LEE, H.K., 2017. Gland segmentation in prostate histopathological images. *Journal of Medical Imaging*, 4(2), pp. 027501.
- SLABAUGH, G., UNAL, G., WELS, M., FANG, T. and RAO, B., 2009. Statistical Region-Based Segmentation of Ultrasound Images. *Ultrasound in Medicine & Biology*, 35(5), pp. 781-795.
- SOILLE, P., SPRINGERLINK EBOOKS - ENGLISH/INTERNATIONAL COLLECTION (ARCHIVE) and SPRINGERLINK (ONLINE SERVICE), 2004. *Morphological Image Analysis Principles and Applications*. Second Correct second printing. edn. Berlin, Heidelberg: Springer Berlin Heidelberg.
- SOMMER, C., STRAEHLE, C., KOTHE, U. and HAMPRECHT, F.A., 2011. Ilastik: Interactive learning and segmentation toolkit, 2011, IEEE, pp. 230-233.
- SRINIVASA, G., FICKUS, M.C., GUO, Y., LINSTEDT, A.D. and KOVACEVIC, J., 2009. Active Mask Segmentation of Fluorescence Microscope Images. *IEEE Transactions on Image Processing*, 18(8), pp. 1817-1829.
- SRINIVASARAO, P., SURESH, K. and RAVI KIRAN, B., 2015. Image Segmentation using Clustering Algorithms. *International Journal of Computer Applications*, 120(14), pp. 36-38.
- TANNOCK, I.F. and HICKMAN, J.A., 2016. Limits to Personalized Cancer Medicine. *The New England journal of medicine*, 375(13), pp. 1289-1294.
- THILAGAMANI, 2011. Object Recognition Based on Image Segmentation and Clustering. *Journal of Computer Science*, 7(11), pp. 1741-1748.
- TSENG, C., LU, C., CHANG, C. and CHEN, G., 2014. Application of machine learning to predict the recurrence-proneness for cervical cancer. *Neural Computing and Applications*, 24(6), pp. 1311-1316.
- USLAN, V. and BUCAK, I., 2010. Microarray Image Segmentation Using Clustering Methods. *Mathematical and Computational Applications*, 15(2), pp. 240-247.

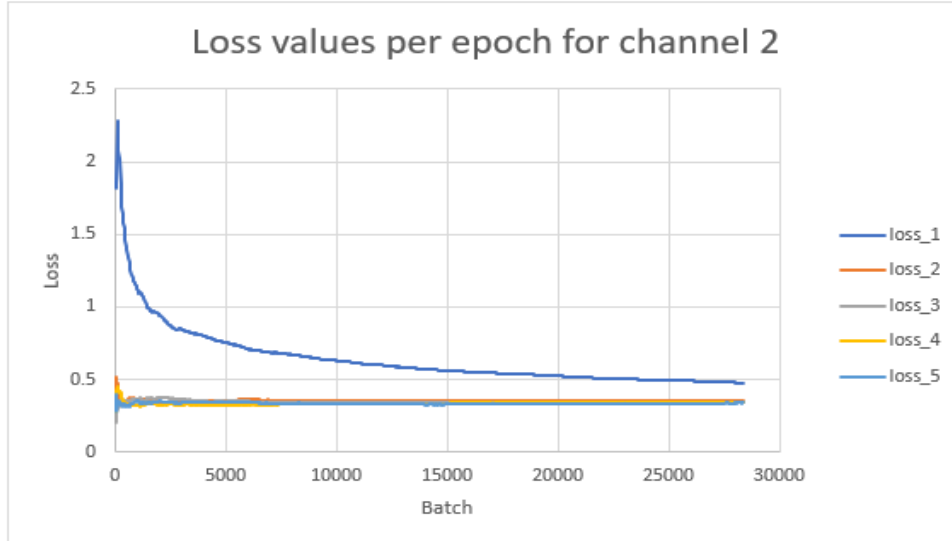
- VOS, P.C., HAMBROCK, T., HULSBERGEN- VAN DE KAA, C. A., FUTTERER, J.J., BARENTSZ, J.O. and HUISMAN, H.J., 2008. Computerized analysis of prostate lesions in the peripheral zone using dynamic contrast enhanced MRI. *Medical physics*, 35(3), pp. 888-899.
- WADDELL, M., PAGE, D. and SHAUGHNESSY, J., John, 2005. Predicting cancer susceptibility from single-nucleotide polymorphism data: a case study in multiple myeloma, 2005, ACM, pp. 21-28.
- WANG, S., BURTT, K., TURKBEY, B., CHOYKE, P. and SUMMERS, R.M., 2014. Computer aided-diagnosis of prostate cancer on multiparametric MRI: a technical review of current research. *BioMed research international*, 2014, pp. 789561-11.
- WANG, X., ZHANG, X., YANG, H. and BU, J., 2012. A pixel-based color image segmentation using support vector machine and fuzzy C-means. *Neural Networks*, 33, pp. 148-159.
- WEINGANDT, H., STEPP, H., BAUMGARTNER, R., DIEBOLD, J., XIANG, W. and HILLEMANN, P., 2002. Autofluorescence spectroscopy for the diagnosis of cervical intraepithelial neoplasia. *BJOG: An International Journal of Obstetrics and Gynaecology*, 109(8), pp. 947-951.
- WEYN, B., VAN DE WOUWER, G., VAN DAELE, A., SCHEUNDERS, P., VAN DYCK, D., VAN MARCK, E. and JACOB, W., 1998. Automated breast tumor diagnosis and grading based on wavelet chromatin texture description. *Cytometry*, 33(1), pp. 32-40.
- WINDEATT, T. and ARDESHIR, G., 2002. Boosted Tree Ensembles for Solving Multiclass Problems, 2002, Springer-Verlag, pp. 42–51.
- Freeberg, J. Adrian, J. L. Benedet, Calum MacAulay, Loyd A. West, and Michele Follen. 2007. "The Performance of Fluorescence and Reflectance Spectroscopy for the in Vivo Diagnosis of Cervical Neoplasia; Point Probe Versus Multispectral Approaches." *Gynecologic Oncology* 107 (1): S248-S255.
- WU, J. and SPRINGERLINK EBOOKS - COMPUTER SCIENCE, 2012. *Advances in K-means Clustering: A Data Mining Thinking*. 1. Aufl. edn. New York: Springer.
- XU, J., SPARKS, R., TOMASZEWSKI, J.E., FELDMAN, M.D. and MADABHUSHI, A., 2010. High-Throughput Prostate Cancer Gland Detection, Segmentation, and Classification from Digitized Needle Core Biopsies. Berlin, Heidelberg: Springer Berlin Heidelberg, pp. 77-88.
- XU, X., ZHANG, Y., ZOU, L., WANG, M. and LI, A., 2012. A gene signature for breast cancer prognosis using support vector machine, 2012, IEEE, pp. 928-931.
- YE, Z., HU, Z., LAI, X. and CHEN, H., 2012. Image Segmentation Using Thresholding and Swarm Intelligence. *Journal of Software*, 7(5),.
- YOUNG PARK, S., FOLLEN, M., MILBOURNE, A., RHODES, H., MALPICA, A., MACKINNON, N., MACAULAY, C., MARKEY, M.K. and RICHARDS-KORTUM, R., 2008. Automated image analysis of digital colposcopy for the detection of cervical neoplasia. *Journal of Biomedical Optics*, 13(1), pp. 14029.

- ZAHARA, E., FAN, S.S. and TSAI, D., 2005. Optimal multi-thresholding using a hybrid optimization approach. *Pattern Recognition Letters*, 26(8), pp. 1082-1095.
- ZAREI, N., BAKHTIARI, A., GALLAGHER, P., KEYS, M. and MACAULAY, C., 2017. Automated prostate glandular and nuclei detection using hyperspectral imaging, 2017, IEEE, pp. 1028-1031.
- ZAREI, N., BAKHTIARI, A., KORBELIK, J., CARRARO, A., KEYES, M., GUILLAUD, M. and MACAULAY, C., 2016. Automated Region-based Prostate Cancer Cell Nuclei Localization. Part of a Prognostic Modality Tool for Prostate Cancer Patients. *Analytical and quantitative cytopathology and histopathology*, 38(2), pp. 59.
- ZAREI, N. and DENNIS COX , PIERRE LANE , SCOTT CANTOR , NEELY ATKINSON , JOSE-MIGUEL YAMAL , LEONID FRADKIN , DANIEL SERACHITOPOL , SYLVIA LAM , DIRK NIEKERK , DIANNE MILLER , JESSICA MCALPINE, KAYLA CASTANEDA, FELIPE CASTANEDA, MICHELE FOLLEN, CALUM MACAULAY, 2016. Boosted Tree Classifier for in Vivo Identification of Early Cervical Cancer using Multispectral Digital Colposcopy .
- ZENG, H., MACAULAY, C., MCLEAN, D.I., PALCIC, B. and LUI, H., 1998. The dynamics of laser-induced changes in human skin autofluorescence--experimental measurements and theoretical modeling. *Photochemistry and photobiology*, 68(2), pp. 227-236.
- ZENG, H., MACAULAY, C.E., LAM, S. and PALCIC, B., 1999. Light-induced fluorescence endoscopy (LIFE) imaging system for early cancer detection, 1999, pp. 275-282.
- ZENG, H., WEISS, A., CLINE, R. and MACAULAY, C.E., 1998. Real-time endoscopic fluorescence imaging for early cancer detection in the gastrointestinal tract. *Bioimaging*, 6(4), pp. 151-165.
- ZHANG, Q., WU, Y.N. and ZHU, S., 2019. Interpretable CNNs.
- ZHANG, QUANSHIWU, YING NIANZHU, SONG-CHUN, 2019. Interpretable CNNs.
- ZUBIOLO, A., DEBREUVE, E., AMBROSETTI, D., POGNONEC, P. and DESCOMBES, X., 2016. Is the vascular network discriminant enough to classify renal cell carcinoma? 2016, IEEE, pp. 1-6.
- Sato, Masakazu, Koji Horie, Aki Hara, Yuichiro Miyamoto, Kazuko Kurihara, Kensuke Tomio, and Harushige Yokota. 2018. "Application of Deep Learning to the Classification of Images from Colposcopy." *Oncology Letters* 15 (3): 3518-3523.
- Sunny Sharma. 2016. "Cervical Cancer Stage Prediction using Decision Tree Approach of Machine Learning." *International Journal of Advanced Research in Computer and Communication Engineering* 5 (4). <https://www.ijarcce.com/upload/2016/april-16/IJARCCE%2088.pdf>.

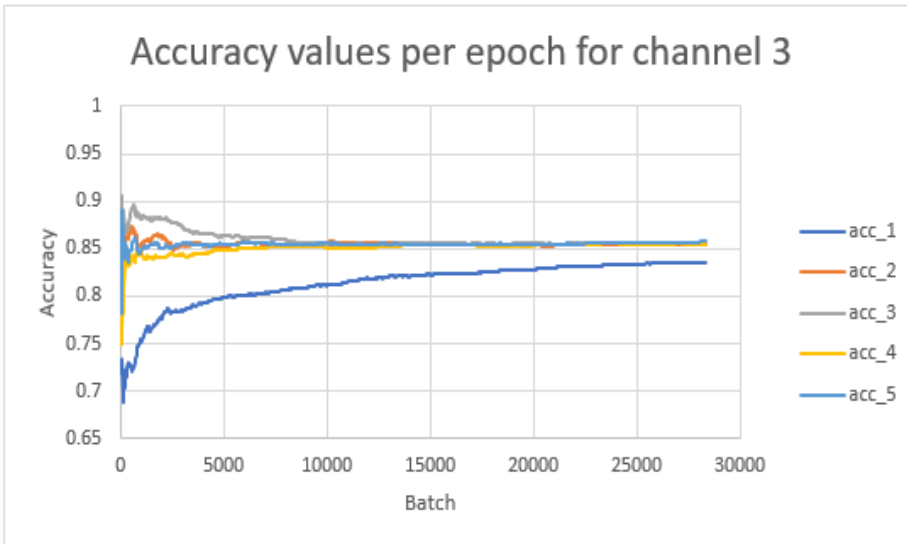
Appendix



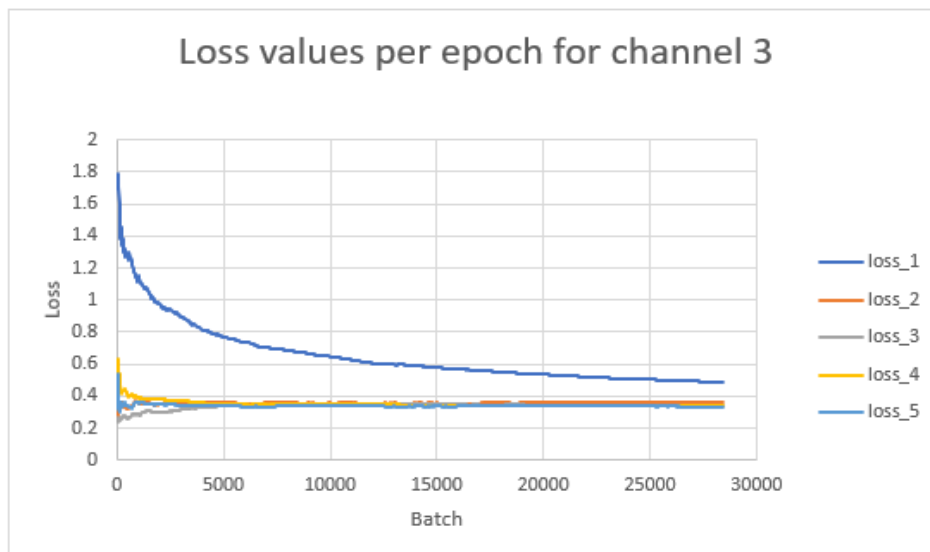
Accuracy values per epoch for channel 2. The accuracy level fluctuated at the beginning and eventually plateaus around 0.85. Each color represents one epoch.



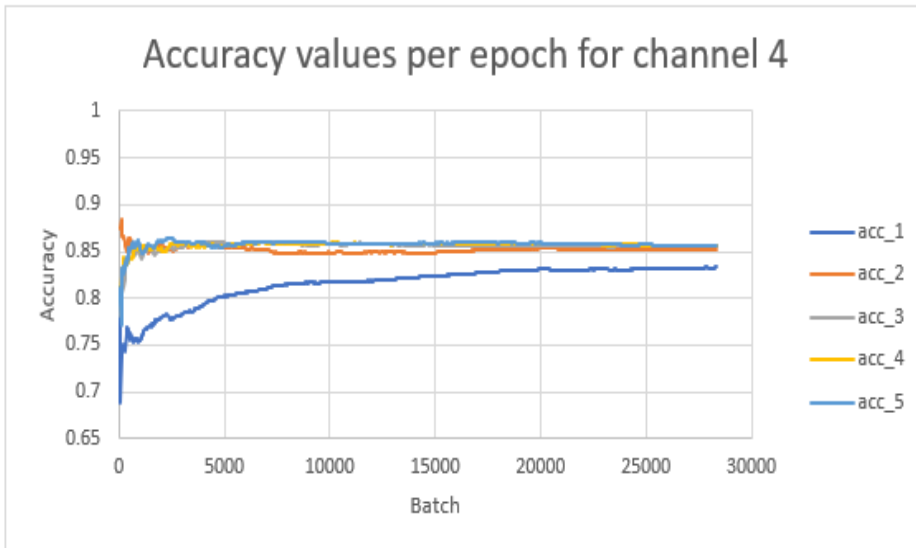
Loss values per epoch for channel 2. The loss level fluctuated at the beginning and eventually becomes plateau around zero. Each color represents one epoch.



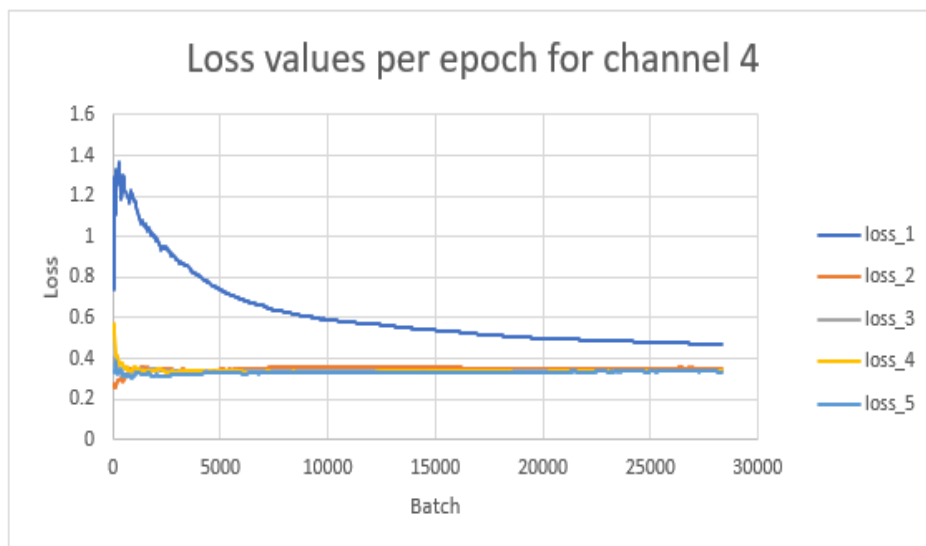
Accuracy values per epoch for channel 3. The accuracy level fluctuated at the beginning and eventually plateaus around 0.85. Each color represents one epoch.



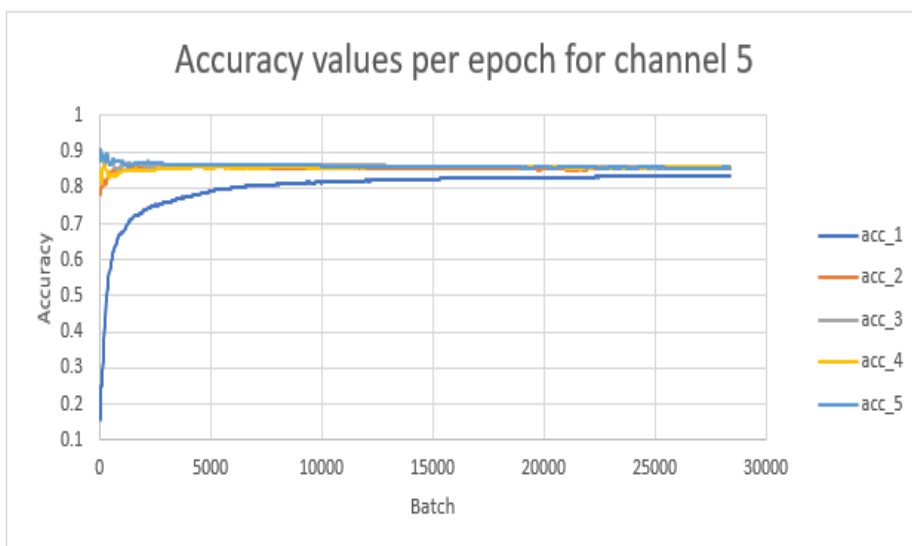
Loss values per epoch for channel 3. The loss level fluctuated at the beginning and eventually becomes plateau around zero. Each color represents one epoch.



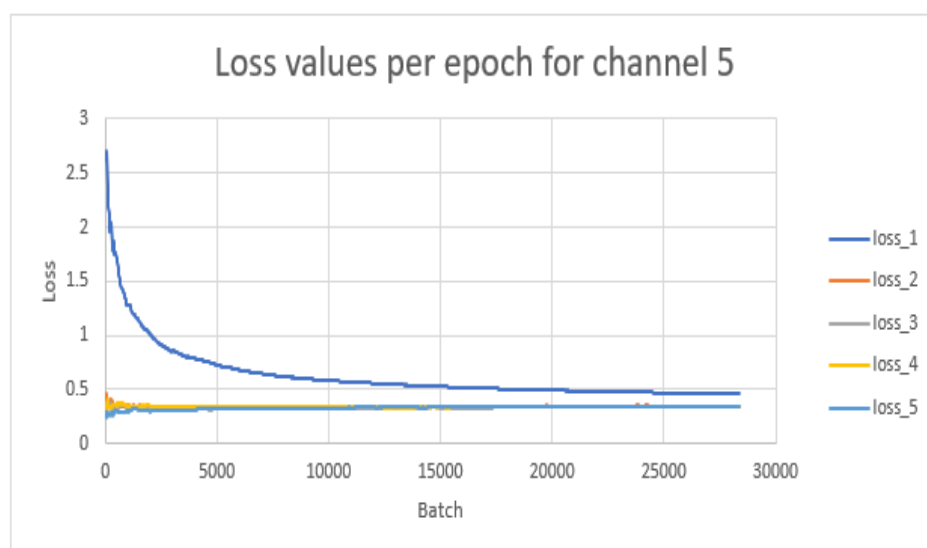
Accuracy values per epoch for channel 4. The accuracy level fluctuated at the beginning and eventually plateaus around 0.85. Each color represents one epoch.



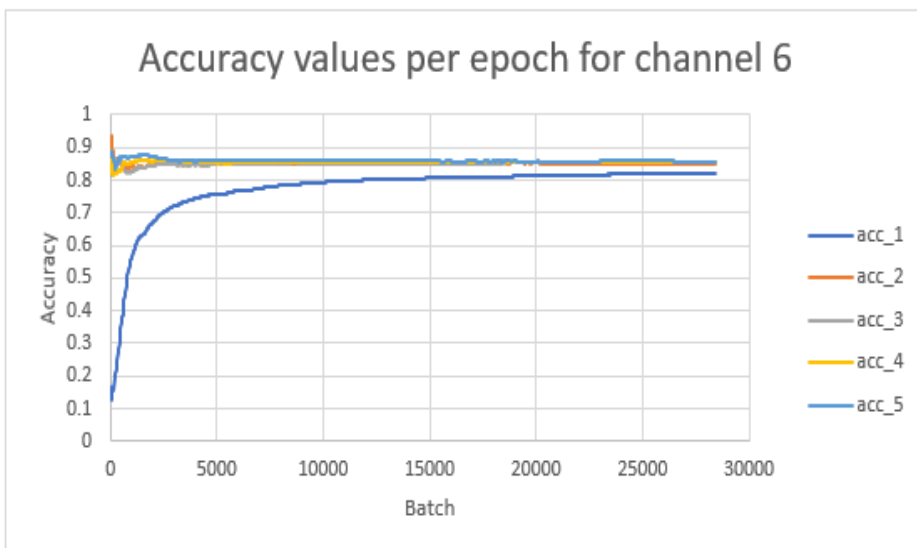
Loss values per epoch for channel 4. The loss level fluctuated at the beginning and eventually becomes plateau around zero. Each color represents one epoch.



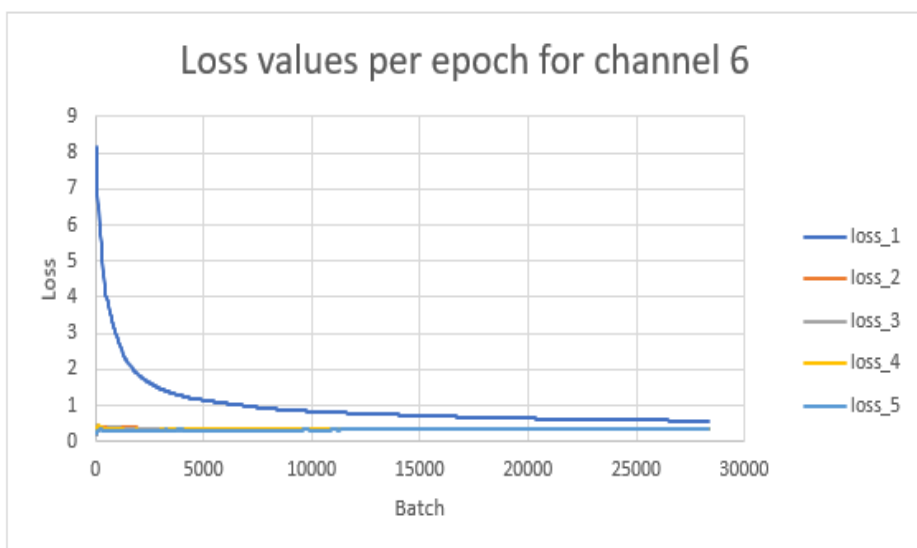
Accuracy values per epoch for channel 5. The accuracy level fluctuated at the beginning and eventually plateaus around 0.85. Each color represents one epoch.



Loss values per epoch for channel 5. The loss level fluctuated at the beginning and eventually becomes plateau around zero. Each color represents one epoch.



Accuracy values per epoch for channel 6. The accuracy level fluctuated at the beginning and eventually plateaus around 0.85. Each color represents one epoch.



Loss values per epoch for channel 6. The loss level fluctuated at the beginning and eventually becomes plateau around zero. Each color represents one epoch.