

Deep Learning Methods for Abnormality Detection and
Segmentation in Computed Tomography and Magnetic
Resonance Images

by

Fatemeh Zabihollahy

A PhD thesis submitted to the Faculty of Graduate and Postdoctoral
Affairs in partial fulfillment of the requirements for the degree of

PhD

in

Electrical and Computer Engineering

Carleton University

Ottawa, Ontario

© 2019, Fatemeh Zabihollahy

Short Abstract

In this dissertation, I describe development of novel deep learning (DL)-based methodologies for the detection and segmentation of clinical abnormalities including renal lesions, left ventricle (LV) scar, and prostate cancer (PCa) in 3D computed tomography (CT) scans and magnetic resonance (MR) images. In the first phase, I developed a decision fusion of patch-based convolutional neural network (CNN) for renal masses classification into cyst versus solid. The solid renal masses were then categorized into benign and malignant using an image-based CNN. These approaches were selected to capture local and global features of the renal masses including intensity, texture, shape, and size, which are key features used by radiologists to classify renal masses. Results demonstrated that automated assessment of renal mass with moderate-to-high degrees of accuracy is feasible. In the second phase, I designed a novel algorithm that comprehensively learns and integrates inter- and intra-slice features from 3D late gadolinium enhancement (LGE)-MR images and allows to accurately and efficiently delineate LV scar fully automatically. In the proposed method, three U-Nets were trained using LGE-MR images extracted from transversal, coronal, and sagittal directions to learn the description of LV scar from different views and the predicted results were combined through majority voting system for the final segmentation. This algorithm benefited from isotropic property of the voxels in 3D MRI that permits for multiplanar reformation. In the third phase, I described a U-Net-based methodology to segment prostate zones from T2-weighted (T2W) and apparent diffusion coefficient (ADC) map prostate MR images as a fundamental requirement for automated diagnosis of PCa. This work is the first attempt for prostate zonal segmentation using ADC map MR images. Furthermore, I presented an ensemble learning system for fully automated localization of peripheral zone PCa from the ADC map. The ensemble learning model allowed

accurate PCa detection, which was not possible using one network due to the complexity of the decision boundary. Our results confirmed that automated PCa detection and segmentation using ADC map MR image are feasible, highly sensitive and can be performed rapidly.

Long Abstract

Medical imaging is vital to non-invasive diagnosis and prognosis of medical abnormalities. Medical image acquisition has greatly advanced in terms of acquisition speed and the ability to resolve fine objects over the last decades. However, advances in technology have increased the size of the image and number of images to be interpreted by radiologists. Imaging studies of a single patient may now consist of hundreds of images, reformatted in multiple imaging planes with three-dimensional (3D) reconstruction. Detection of abnormalities, such as cancer or scar tissue, is an important part of disease diagnosis based on medical images. Abnormalities in tissue may manifest as differences in image intensity, contrast, and texture to the normal tissue in medical images. Currently, the medical images are interpreted manually by clinical experts, which is a tedious task and subject to large inter- and intra-operator variability due to observer limitations (e.g., constrained human visual perception, fatigue, or distraction) and the complexity of the clinical cases themselves (e.g., overlapping structures). Therefore, automated analysis of medical images is highly desirable. Parallel to the developments in imaging hardware, machine learning technologies, including deep learning (DL) methods, have evolved over the last decade and are providing exciting solutions in identification, classification, and quantification of abnormalities in medical images. In this dissertation, with the availability of the unique datasets of kidney computed tomography (CT) scans, prostate and cardiac magnetic resonance images (MRI) through a clinical collaboration with the Ottawa Hospital and the Libin Cardiovascular Institute of Alberta at the University of Calgary, I have focused on the computer-aided detection of kidney and prostate tumors and cardiac scar tissue as medical abnormalities. The goal of this dissertation is to describe the development of novel DL-based methodologies for the detection and segmentation of abnormalities in 3D CT scans and MR images for three high-impact clinical

applications. These applications are computer-aided detection of kidney tumors (renal masses) on CT scans, quantification of scar tissue in the heart in 3D cardiac MRI, and prostate tumor localization in multi-slice MRI. Our research has novelty in both methodology and clinical applications.

For the application of detecting kidney cancer, I developed a decision fusion of convolutional neural network (CNN)-based method for renal masses classification into cyst versus solid and then categorized solid renal masses into benign and malignant. For the application of detecting scar tissue in the heart, I designed a novel algorithm that comprehensively learns and integrates inter- and intra-slice features from 3D late gadolinium enhancement (LGE)-MR images and allows to accurately delineate LV scar fully automatically. For the application of detecting prostate cancer (PCa), I described a U-Net-based methodology to segment prostate zones from T2-weighted (T2W) and apparent diffusion coefficient (ADC) map prostate MR images as a fundamental requirement for automated diagnosis of PCa. Furthermore, I presented an ensemble learning system for fully automated localization of peripheral zone PCa from the ADC map that has not been described previously. In this dissertation the method developed for different applications progressed from a CNN to the cascaded multi-planar U-Net and ensemble learning system.

Keywords:

Renal mass, renal cell carcinoma; benign neoplasms; benign cyst, left ventricle myocardium, left ventricular scar, prostate whole gland, prostate central gland, prostate peripheral zone, prostate cancer, magnetic resonance image (MRI), T2 weighted MRI, apparent diffusion coefficient (ADC) MRI, computed tomography, late gadolinium enhancement, convolutional neural network, U-Net.

Acknowledgment

The completion of this dissertation was possible due to the generous support from numerous people whom I feel indebted to. Firstly, I would like to express my sincere gratitude to my supervisor, Dr. Eranga Ukwatta, who offered me the opportunity to study with him medical image processing at Carleton University in winter 2017. Eran is an extra-ordinary researcher and advisor, great mentor and patient leader. I highly doubt that a more productive study would have been possible for me anywhere else. I want to emphasize how grateful I am to him for his continuous support of my Ph.D. study and related research, for his patience, motivation, and immense knowledge and guidance.

I would also like to thank my clinical collaborator, Dr. Nicola Schieda, for his invaluable direction which helped me achieve key milestones. He gave me access to unique valuable annotated CT and MR images. He extended his support by providing invaluable reviews on numerous occasions despite tight deadlines and this is very much appreciated. His insightful comments and encouragement assisted me in achieving my goals. Without his precious support, it would not be possible to conduct this research.

I would also like to thank Dr. James A. White for his contribution and productive collaboration that led to interesting highlights in my research. I am very grateful for the valuable annotated images he provided us and also for his insightful comments on my papers.

I want to express my gratitude for all the council I have gotten from my advisory committee, Drs. Jagath Samarabandu, Emil M. Petriu, Sreeraman Rajan, Emily Heath, and James R. Green.

I acknowledge the funding support received from numerous sources including Natural Science and Engineering Research Council of Canada (NSERC), Ontario Graduate Scholarship (OGS), and Ontario Graduate Scholarships in Science and Technology (OGSST), which were all crucial for me to gain productivity and focus on research.

I am especially grateful to the reviewers, editors, and associate editors of Medical Physics, Magnetic Resonance Imaging, European Radiology, and IEEE Access journals. The wealth of the constructive comments they provided is very much appreciated.

I would like to thank my husband, Keikavoos, and my daughter, Yasmin, for all the support, during these years. Without them, I would not even have the chance to get this far. I am very grateful for all they have done for me. Also, I would appreciate the continuous encouragement of my parents during these years.

Table of Content

Short Abstract	ii
Long Abstract.....	iv
Acknowledgment	vii
List of Tables	xiv
List of Figures.....	xviii
List of Abbreviation.....	xxvii
Chapter 1: Introduction.....	1
1.1 Overview	3
1.2 Contributions.....	4
1.3 Thesis Organization	9
Chapter 2: Background.....	13
2.1 Medical Imaging Techniques	14
2.1.1 CT Scan.....	15
2.1.2 MRI.....	15
2.2 Medical Image Analysis.....	17
2.2.1 Object Detection and Image Segmentation.....	17
2.3 Regularization	34
2.3.1 Data Augmentation	34
2.3.2 Parameter Sharing.....	35
2.3.3 Dropout	35
2.3.4 Early Stopping	36
2.3.5 Batch Normalization	36
2.3.6 Ensemble Learning	36
2.4 Evaluation Metric for Object Detection and Segmentation	37

2.4.1	Object Detection Evaluation	38
2.4.2	Segmentation Evaluation	39
2.4.3	Statistical Tests	41
Chapter 3: Patch-Based Convolutional Neural Network for Differentiation of Cyst from Solid Renal		
Mass on Contrast-Enhanced Computed Tomography Images		
3.1	Introduction	49
3.2	Method	52
3.2.1	Patient	52
3.2.2	CT Technique	54
3.2.3	Manual Segmentation	55
3.2.4	Patch Extraction and Balancing Dataset	56
3.2.5	Network Architecture	57
3.2.6	Network Training and Testing	60
3.2.7	Evaluation Metrics	62
3.3	Results	63
3.4	Discussion	69
3.5	Conclusion	71
Chapter 4: Automated Classification of Solid Renal Masses on Contrast-Enhanced Computed		
Tomography Images using Convolutional Neural Network with Decision Fusion		
4.1	Introduction	73
4.2	Method	77
4.2.1	Semi-automated Method for Differentiation of RCC from Benign Solid Renal Masses..	78
4.2.2	Fully Automated Method for Differentiation of RCC from Benign Solid Renal Masses.	86
4.2.3	3D CNN-based Algorithm for Semi-automated Differentiation of RCC from Benign	
	Solid Renal Masses	86

4.2.4	Fully Automated Segmentation of Kidneys on CECT Images	87
4.2.5	Fully Automated Detection and Segmentation of Renal Masses on CECT Images	90
4.3	Evaluation Metrics	92
4.4	Results	92
4.5	Discussion	103
4.6	Conclusion.....	105
 Chapter 5: Fully Automated Segmentation of Left Ventricular Scar from 3D Late Gadolinium Enhancement Magnetic Resonance Imaging Using a Cascaded Multi-Planar U-Net (CMPU-Net)....		
5.1	Introduction	108
5.2	Method	112
5.2.1	Study Subjects and Data Acquisition.....	113
5.2.2	Manual Segmentation for Training and Testing	114
5.2.3	Architecture of CMPU-Net.....	115
5.2.4	Training the CMPU-Net	116
5.2.5	Comparison of Our Method to Alternative Methods	117
5.3	Results	119
5.4	Discussion	128
5.5	Conclusion.....	129
 Chapter 6: Automated Segmentation of Prostate Zonal Anatomy on T2W and ADC Map MR Images Using U-Nets		
6.1	Introduction	133
6.2	Method	136
6.2.1	Dataset	137
6.2.2	MRI.....	138
6.2.3	Manual Segmentation	139

6.2.4	U-Net Architecture	140
6.2.5	U-Nets Training	141
6.2.6	Comparison of Our Method to Alternative Method.....	144
6.3	Results	145
6.4	Discussion	155
6.5	Conclusion.....	157
Chapter 7: Fully Automated Localization of Prostate PZ Tumors on ADC Map MR Images Using an Ensemble Learning Method.....		161
7.1	Introduction	161
7.2	Materials and Methods.....	164
7.2.1	Study Subjects and Data Acquisition.....	164
7.2.2	Manual Segmentation of Prostate Boundaries and Tumors.....	167
7.2.3	Ensemble Model of U-Nets for PCa localization.....	169
7.2.4	Statistical Analysis.....	172
7.3	Results	173
7.4	Discussion	179
7.5	Conclusion.....	184
Chapter 8: Conclusion and Future Works		188
8.1	Summary	188
8.1.1	Differentiation of Benign Cyst from Solid Renal Mass using a Patch-Based CNN (Chapter 3)	189
8.1.2	Classification of Solid Renal Masses into Benign and RCC (Chapter 4)	190
8.1.3	Automated Identification of LV scar from 3D LGE MRI (Chapter 5)	191
8.1.4	Prostate Zonal Segmentation using T2W and ADC MR Images (Chapter 6)	192
8.1.5	Automated Localization of PCa in the PZ on ADC Map MRI (Chapter 7).....	193

8.2	Limitations and Future Works.....	194
8.2.1	Differentiation of Benign Cyst from Solid Renal Mass using a Patch-Based CNN (Chapter 3).....	195
8.2.2	Classification of Solid Renal Mass into Benign and RCC (Chapter 4)	196
8.2.3	Automated Identification of LV scar from 3D LGE MRI (Chapter 5).....	197
8.2.4	Prostate Zonal Segmentation using T2W and ADC MR Images (Chapter 6)	198
8.2.5	Automated Localization of PCa in the PZ on ADC Map MRI (Chapter 7).....	198
	Appendix A.....	201

List of Tables

Table 3-1. Configuration of the proposed convolutional neural network for classification of extracted patches into a cyst and solid renal masses. Reprinted with permission.	59
Table 3-2. The number and variety of CT systems used in our experiments for training and testing phases. Reprinted with permission.....	60
Table 3-3. Accuracy of classification of benign cysts from solid masses as evaluated using the independent segmentation of two observers. Reprinted with permission.....	67
Table 3-4. The performance of different techniques for benign cyst diagnosis on CECT test images. Reprinted with permission.....	69
Table 4-1. A summary statistic of the number of slices used in our experiments to train CNN on a per-patient basis. Reprinted with permission.	81
Table 4-2. Performance of the network for the different number of training epochs evaluated using a subset of the test set. Reprinted with permission.	85
Table 4-3. The number of CECT images from local patients with the different type of renal mass, used for the training and testing phases in our experiments. Reprinted with permission...	88
Table 4-4. Summary of evaluation results of the different networks for renal mass detection from CECT testing images. Mean values for sensitivity, specificity, and DSC are reported. Reprinted with permission.	91
Table 4-5. Results of subgroup analysis for RCC versus oncocytoma. Reprinted with permission.	102
Table 4-6. Summary of evaluation results of the different methods of classification of CECT images into RCC and benign. Reprinted with permission.....	102
Table 5-1. Image acquisition parameters. Reprinted with permission.....	114

Table 5-2. Summary of results of our proposed technique, CMPU-Net, for LV myocardium and LV scar segmentation as the mean value \pm SD, 25th and 75th percentiles on 3D LGE- MRI against expert manual segmentation. Reprinted with permission..... 122

Table 5-3. Performance of all 3D scar segmentation methods for LV scar segmentation on 3D LGE-MR images against expert manual segmentation. The asterisk denotes statistical significance using the Wilcoxon rank-sum test performed to compare the average value of DSC

reported from our method (CMPU-Net) against those of alternatives. Reprinted with permission.
..... 126

Table 6-1. Sequence parameters for multi-parametric MRI of the prostate protocol performed with pelvic surface coil^a at 3 Tesla^b. Reprinted with permission..... 139

Table 6-2. The number of train and test sets of images in each dataset used in this study from T2W and ADC map sequences. Reprinted with permission..... 142

Table 6-3. Causality experiment results for evaluating the impact of the cost function, pre- and post-processing for NetT2WWG. Reprinted with permission..... 144

Table 6-4. Summary of results of our proposed technique for prostate zonal segmentation as the median and mean values \pm standard deviation on T2W and ADC map MRI against expert manual segmentations. Reprinted with permission. 147

Table 6-5. The mean and range of prostate WG volume for the testing dataset calculated from manual segmentation of T2W MR and ADC map images. Reprinted with permission. 149

Table 6-6. Best and worst DSC achieved using our presented strategy for prostate zonal segmentation from T2W and ADC map MR test images against expert manual segmentations. Reprinted with permission. 150

Table 6-7. Performance of our proposed technique for prostate zonal segmentation as the median and mean values \pm standard deviation on T2W and ADC map MRI images against expert manual segmentations for the control group and patients with tumor. Reprinted with permission.
..... 152

Table 6-8. Summary of results of our proposed technique for prostate zonal segmentation at the base, apex, and middle craniocaudal slice levels as the median and mean values \pm standard

deviation on T2W and ADC map MRI compared to expert manual segmentations. Reprinted with permission.....	153
Table 6-9. Performance of RF-based method for prostate regional segmentation from T2W MR images as the median and mean values \pm standard deviation (SD) against expert manual segmentation. Reprinted with permission.....	155
Table 7-1. Summary of clinical features of 154 patients who underwent mp-MRI prior to radical prostatectomy in this study. Reprinted with permission.....	167
Table 7-2. The configuration of ADC-Nets used in our experiment for PCa localization. Reprinted with permission.....	170
Table 7-3. Summary of results obtained from the proposed model for fully automated PCa localization. Reprinted with permission.....	175
Table 7-4. Causality Experiment Results for the Proposed Model for Fully Automated PCa Localization. Reprinted with permission.....	177

List of Figures

- Figure 2-1. An example of a CT scan of the abdomen and pelvis, taken in the transversal plane, visualized in the standard abdominal window and level set of 400 and 50, respectively..... 15
- Figure 2-2. An example of a prostate T2-weighted MR image taken in the transversal plane. ... 16
- Figure 2-3. An example of object detection and image segmentation from left to right. In the left picture, a red box enclosed right ventricle (RV) and left ventricle (LV) that displays the location of RV and LV descriptions on MRI. In the right picture, RV, LV, and LV myocardium are segmented from the MR image and shown in blue, red, and green colors, respectively..... 17
- Figure 2-4. An overview of a simple neural network with one hidden layer. 25
- Figure 2-5. A typical CNN pipeline where Conv., detection, and pooling layers are applied to the input image to extract useful features from it. These features are then used as inputs of an ordinary ANN. 27
- Figure 2-6. a) A sample of the kernel that captures the horizontal line in an image is shown. b) An example of applying horizontal kernel to the image along with the output feature map are illustrated, where horizontal lines are highlighted in the feature map. c) The mathematical workflow of feature map generation in the Conv. layer is displayed, where the summation of the

element-wise dot product between the image and kernel at each and every position is calculated.
..... 28

Figure 2-7. An example of applying a maximum pooling layer of size (2×2) and stride of 2 that downsizes the feature map. Stride is the number of pixels shifts over the input matrix. 31

Figure 2-8. An example of CNN architecture, where two Conv. and one pooling layers are employed for feature extraction. 32

Figure 2-9. An overview of a typical U-Net with five stages. 34

Figure 3-1. An example of a) benign renal cyst b) solid renal mass on CECT images. The 2D CT images are visualized in the standard abdominal window and level set of 400 and 50, respectively. Reprinted with permission. 52

Figure 3-2. A diagram of our proposed method for patch extraction from renal mass tissue. Reprinted with permission. 57

Figure 3-3. A flow diagram of our proposed methodology, where the left and right block diagrams display training and testing procedures. Reprinted with permission. 61

Figure 3-4. Confusion matrix shows the performance of our proposed method for classification of cyst and solid renal masses on CECT images. Reprinted with permission. 64

Figure 3-5. ROC curve displays the performance of our designed patch-based CNN model for diagnosis of cyst versus solid renal masses, where the true positive rate (sensitivity) against false-positive rate (1 - specificity) has been plotted. Reprinted with permission. 64

Figure 3-6. Histogram of the ROC AUC values obtained from 10,000 bootstrapping of the AUC calculation. Reprinted with permission. 65

Figure 3-7. Example results of misclassified (a) benign cyst and (b) solid renal mass using our algorithm. The original CECT images are shown in the first column. The third column shows the

location of tumor where blue and red show benign cyst and solid renal mass, respectively. The CECT images are visualized in the standard abdominal window and level set of 400 and 50.

Reprinted with permission. 66

Figure 3-8. ROC curve displays the performance of conventional assessment for diagnosis of cyst versus solid renal masses, where true positive rate (sensitivity) against false-positive rate (1 - specificity) has been plotted. Reprinted with permission. 68

Figure 4-1. Flowchart illustrating strategy for patient selection and overall experimental design for differentiating RCC from benign solid renal masses at contrast-enhanced CT in the current study. Reprinted with permission. 78

Figure 4-2. Overview of semi-automated methodology for labeling a contrast-enhanced CT image as RCC or benign renal mass. It is notable that for the semi-automated approach manual segmentation of renal mass is not needed. Once the CECT images are cropped around the

kidneys, an expert needs to separate slices with tumors from the others to be passed to the trained CNN. Reprinted with permission.....	80
Figure 4-3. From left to right, an example of one training sample with benign renal mass and its deformed one are shown. Reprinted with permission.....	82
Figure 4-4. The architecture of the CNN used in our experiments for the classification of slices with solid renal masses into RCC or benign. In this figure, MP represents the maximum pooling layer. Reprinted with permission.	83
Figure 4-5. From left to right, the model accuracy and loss are plotted against the number of training epochs for train and validation sets. Reprinted with permission.....	85
Figure 4-6. An overview of our proposed methodology, where the top and bottom block diagrams display boundary segmentation and renal mass localization in both kidneys. Reprinted with permission.....	90
Figure 4-7. Example results of segmented kidneys from CECT images for five patients with different types of renal masses including from top to bottom clear cell RCC, chromophobe RCC, papillary RCC, fpAML, and oncocytomas. Columns 1 to 3 present three different slices from base to apex extracted from CT scans of each patient. On the source 2D images, the results from the U-Net-based method and manual segmentation of the kidney are shown in white and red, respectively. The 2D CT images are visualized in the standard abdominal window and level set of 400 and 50, respectively. The fourth and fifth columns indicate the 3D surfaces of the kidneys	

generated by our algorithm and expert manual segmentation, respectively. Reprinted with permission.	94
Figure 4-8. Boxplots of results generated by our proposed method for kidneys boundary segmentation on 125 and 60 CECT test images of institutional and KiTS19 datasets in comparison to manual segmentation. Reprinted with permission.	95
Figure 4-9. Example results from localized renal masses on CECT images in five patients with different types of renal masses including from top to bottom clear cell RCC, chromophobe RCC, papillary RCC, fpAML, and oncocytomas. In columns 1 to 4, 2D slices enclosing renal masses compiled over kidneys are shown. In all images, the renal masses contours computed from algorithm-generated and manual segmentations are shown in red and cyan, respectively. Reprinted with permission.	96
Figure 4-10. Boxplots of results generated by our proposed method for renal mass boundary segmentation on 125 and 60 CECT test images of institutional and KiTS19 datasets in comparison to manual segmentation. Sen., Spec., and DSC denote sensitivity, specificity, and Dice similarity coefficient, respectively. Reprinted with permission.	97
Figure 4-11. Boxplots of results generated by our proposed for subgroup analysis for localization of solid renal mass and benign cyst on 125 CECT images from the institutional dataset in comparison to expert manual segmentation. Sen., Spec., and DSC denote sensitivity, specificity, and Dice similarity coefficient, respectively. Reprinted with permission.	97
Figure 4-12. Example results from localized multiple renal masses on CECT images in five patients with different types of renal masses including from left to right clear cell RCC, chromophobe RCC, papillary RCC, fpAML, and oncocytomas. In all images, the renal masses	

contours computed from algorithm-generated and manual segmentations are shown in red and cyan, respectively. Reprinted with permission. 98

Figure 4-13. Boxplots of results generated by our proposed method for localization of renal mass on CECT images of the institutional dataset in comparison to expert manual segmentation for patients with multiple RM. Sen., Spec., and DSC denote sensitivity, specificity, and Dice similarity coefficient, respectively. Reprinted with permission. 99

Figure 4-14. ROC curve for the proposed method, where the true positive rate (sensitivity) against false-positive rate ($1 - \text{specificity}$) has been plotted. Reprinted with permission. 100

Figure 4-15. PR curve for the proposed method, where the precision against recall has been plotted. Reprinted with permission. 100

Figure 5-1. Overview of our proposed method, CMPU-Net, for fully automated myocardial scar segmentation. Reprinted with permission. 112

Figure 5-2. Block diagram of CMPU-Net framework. Reprinted with permission. 113

Figure 5-3. In the first four columns, exemplary results of segmented LV myocardium tissue in different slices extracted from 3D LGE-MR image of three subjects are displayed. The contours computed from the CMPU-Net method and manual segmentations are displayed in red and cyan. The fifth and sixth columns indicate the 3D view of the myocardial regions generated by our proposed technique and expert manual segmentation, respectively. Reprinted with permission. 120

Figure 5-4. In the first four columns, exemplary results of segmented scar tissue in different slices extracted from 3D LGE-MR image of five subjects are displayed. The contours computed from the CMPU-Net method and manual segmentations are displayed in red and cyan. The fifth and sixth columns indicate the 3D view of the myocardial scar regions generated by our

proposed technique and expert manual segmentation, respectively. Reprinted with permission.

..... 121

Figure 5-5. Bland-Altman plot of LV scar volume obtained from our proposed technique. The plot graphs the difference between the algorithm-generated and manual target volumes as a function of their mean. Reprinted with permission. 123

..... 123

Figure 5-6. Exemplary results of 3D scar segmentation methods applied for an example of 3D LGE-MR image, presented in orthogonal views (Column 1-3) and using a surface rendering. The last column indicates the 3D view of the myocardial scar regions. The results of each method are shown in the corresponding row. Reprinted with permission..... 125

..... 125

Figure 5-7. 3D view of LV scar in the LGE-MR test images. Reprinted with permission. 127

Figure 6-1. An example of axial T2-weighted (T2W) and apparent diffusion coefficient (ADC) map prostate MR images from left to right, where the boundary of prostate WG and CG are shown in yellow and red. PZ is the area enclosed between colored contours. Reprinted with

permission. 135

Figure 6-2. The block diagram of our proposed method for prostate regional segmentation of

T2W or ADC map MR images. Reprinted with permission..... 137

Figure 6-3. Example results from segmented prostate WG, CG, and PZ using T2W (two top rows) and ADC map MR images (two bottom rows) in four patients. On the source 2D images, the U-net-based method segmentation of the prostate CG and PZ for three slices presented from apex to base (left to right) extracted from MRI are shown in purple and red respectively and the contours computed from the expert manual are displayed in green and yellow. The fourth and

fifth picture elements depict a 3D surface rendering of algorithm-generated and expert manual segmentation of prostate zones, respectively. Reprinted with permission. 147

Figure 6-4. From top to bottom the best and worst segmentation results using our proposed methodology for prostate regional delineation from T2W and ADC map MR images are presented. The contours computed from the expert manual and algorithm-generated segmentations are displayed in cyan and red, respectively. The algorithm failed to detect prostate WG from ADC map MR in one of the test images. Reprinted with permission. 149

Figure 6-5. Scatter plot of DSC versus prostate WG volume reported by our proposed method on a) T2W and b) ADC map prostate MR images, along with the Pearson correlation coefficient (r-value). Reprinted with permission. 151

Figure 6-6. Exemplary results of prostate regional segmentation methods applied for an example of T2W MR image, where the contours computed from the expert manual and algorithm-generated segmentations are displayed in cyan and red, respectively. The results of each method are shown in the corresponding row. Reprinted with permission..... 154

Figure 7-1. Flowchart shows the strategy for patient selection and overall experimental design for localizing PCa on ADC map MR images. Reprinted with permission. 165

Figure 7-2. An example of axial (a) T2W and (b) ADC map prostate MR images and manual segmentation from left to right, where the surface of the prostate peripheral zone, prostate

transition zone, and PCa are shown in red, yellow, and blue, respectively. Reprinted with permission.....	168
Figure 7-3. The block diagram of our proposed methodology for PCa localization of ADC map MR images. Reprinted with permission.....	169
Figure 7-4. The architecture of ADC-Net1, 2 and 3 (acronyms are defined in the text). Reprinted with permission.	170
Figure 7-5. Example results from localized PCa on ADC map MR images in three patients from top to bottom. In each row, the 2D slices compiled over prostate WG from base to apex are shown from left to right. In all images, the PCa contours computed from manual and algorithm-generated segmentations are shown in cyan and yellow, respectively. Reprinted with permission.	175
Figure 7-6. ROC curve for ensemble U-Nets, where the true positive rate (sensitivity) against false-positive rate (1 – specificity) has been plotted. Reprinted with permission.	178
Figure 7-7. Histogram of the ROC AUC values obtained from 10,000 bootstrapping of the AUC calculation. Reprinted with permission.....	178
Figure 7-8. Example results of (a) undetected PCa of one patient with a significant tumor. The tumor may have been missed due to the system incorrectly attributing the tumor as representing a portion of the adjacent transition zone; (b) falsely labeled as PCa of one patient without significant tumor. The area marked as tumor represents the adjacent transition zone. The PCa contours computed from manual and algorithm-generated segmentations are shown in cyan and red, respectively. Reprinted with permission.....	179

List of Abbreviation

Abbreviation	Meaning
AVD	Absolute volume difference
ADC	Apparent diffusion coefficient
AI	Artificial intelligence
ANN	Artificial neural network
AUC	Area Under the Curve
CVD	Cardiovascular disease
CCU-Net	Cascaded conventional U-Net
CMPU-Net	Cascaded multi-planar U-Net-based method
CG	Central gland
CKD	Chronic kidney disease
CV	Coefficient-of-variation
CT	Computed tomography
CAD	Computer-aided diagnosis
CE	Contrast enhanced
CMF	Convex max flow
CNN	Convolutional neural network
CC	Cubic centimeter
DL	Deep learning
DSC	Dice similarity coefficient
DWI	Diffusion-weighted imaging
DICOM	Digital imaging and communication in medicine
DCU-Net	Direct conventional U-Net
DMPU-Net	Direct multiplanar U-Net
DCE	Dynamic contrast enhanced

FN	False negative
FP	False positive
fp-AML	Fat-poor angiomyolipoma
FWHM	Full width at half maximum
FCNN	Fully CNN
GPU	Graphics processing unit
HD	Hausdorff distance
HIPAA	Health insurance portability and accountability act
HMF	Hierarchical maximum flow
HU	Hounsfield units
IRB	Institutional review board
KNN	K nearest neighbour
LGE	Late gadolinium enhancement
LV	Left ventricle
LHIN	Local health integration network
ML	Machine learning
MRI	Magnetic resonance imaging
MDCT	Multidetector computer tomography
mp-MRI	Multiparametric-MRI
M2D	Multiple 2D
NAVD	Normalized absolute volume difference
PZ	Peripheral zone
PACS	Picture Archiving and Communication System
PPV	Positive predictive value
PET	Positron emission tomography
PR	Precision recall

PCa	Prostate cancer
PI-RADS	Prostate Imaging–Reporting and Data System
PSA	Prostate Specific Antigen
RF	Random forest
ROC	Receiver operating characteristic
ReLU	Rectified linear unit
ROI	Region of interest
RCC	Renal cell carcinoma
SI	Signal intensity
STRM	Signal threshold to reference mean
SPECT	Single photon emission computed tomography
SVM	Support vector machine
3D	Three-dimensional
T1W	T1-weighted
TN	True negative
TP	True positive
T2W	T2-weighted
2D	Two-dimensional
VOI	Volume of interest
WG	Whole gland

Chapter 1: Introduction

Medical imaging has evolved from film-based technology to big digital data representations over the last several decades and has been assisting clinicians with improved disease diagnosis and prognosis by detecting abnormalities in the structure of the image [1]. Due to the advances in image acquisition speeds and the ability to acquire high-resolution data, the mere size of images required to be interpreted has increased dramatically. Moreover, patients now undergo numerous medical imaging scans. These scans provide complementary information of a certain disease condition [2-4]. Despite the increase in size of the medical imaging data, medical images are still interpreted manually. Diagnostic radiologists are tasked with detecting and characterizing abnormalities in medical images. Manual interpretation of medical images is very time consuming, tedious and subject to high operator variability [5]. Automated interpretation of abnormalities is of extreme interest to the imaging community at present. This is not only due to the overwhelming amount of data in a particular examination but also due to a continued increase in the number of examinations being performed daily. Moreover, there are limitations to the ability of radiologists to detect diseases and to differentiate between disease states. Automated analysis of medical images such as abnormality detection, segmentation, and classification with artificial intelligence (AI) techniques has an enormous impact on disease prevention, cure, and care. Recently, a group of researchers demonstrated that an AI system outperformed all human readers in an independent study of six radiologists for breast cancer prediction [6]. This research revealed that the area under the receiver operating characteristic curve (AUC-ROC) for the AI system was greater than the AUC-ROC for the average radiologist by an absolute margin of 11.5%.

AI includes techniques that are built to emulate how decision making occurs in the human brain. Machine learning (ML), which is a sub-area of AI, has evolved substantially over the last several decades. ML contains algorithms and statistical models that computer systems use to learn from patterns instead of explicit instructions. ML algorithms use training data to build a statistical model to make decisions or predictions [7]. ML techniques have been widely used for classification, regression, clustering, feature selection and reduction, computer vision, image processing, object detection, and classification tasks [8]. Deep learning (DL), which is a branch of ML has been applied to many real-world problems in various domains such as manufacturing industry [9], finance [10], power and energy [11], agriculture [12], and healthcare [13]. DL is a class of ML algorithms that employs multiple layers of nonlinear processing units to solve a problem [14]. DL architectures such as deep neural networks, deep belief networks, recurrent neural networks and convolutional neural networks (CNN) have been applied to complex data with large feature sizes from different domains such as machine translation, bioinformatics, drug design, medical image analysis [15-16] and archived remarkable performance.

In this dissertation, I have chosen to investigate the ability of the DL methods to detect few abnormalities that are now feasible with the availability of large datasets and new imaging techniques and contrast agents that can differentiate abnormalities from normal tissue.

Abnormalities, such as tumors and scar tissue, describe a state, condition, or behavior that manifest differently from normal tissue because they may have more vasculature, extracellular space, or collagen [17]. Typically, most of the abnormalities cannot be detected using standard imaging techniques as there may not be any image contrast difference between normal and abnormal tissues [18]. To emulate image contrast to reflect the physical contrast of the tissues in the images, contrast agents need to be used.

Through data-sharing agreements with the Ottawa Hospital and the Libin Cardiovascular Institute of Alberta at the University of Calgary, unique datasets of several key abnormalities that are detected by radiologists in medical images were made available for this dissertation. These datasets include: 1) a 3D computed tomography (CT) images of 315 patients suffering from kidney cancer, 2) T2 weighted (T2W) and apparent diffusion coefficient (ADC) magnetic resonance images (MRI) of 225 patients suffering from prostate cancer (PCa), 3) 3D late gadolinium enhancement (LGE) MR images of the heart from 34 patients with ischemic cardiomyopathy. The availability of such large datasets and the moderate accuracy of previous non-ML-based methods for each application led to employ DL-based approach to detect the abnormalities in those increasingly complex data. Therefore the objective of this thesis is to describe the development of novel DL-based methodologies for abnormality detection and segmentation in medical images including CT scans, 3D MRIs in three high-impact clinical applications namely classification of renal mass on CT scans, quantification of scar tissue in the left ventricle (LV) myocardium using 3D MRI and localization of PCa in multi-slice MRI. In this chapter, I provide an overview of the thesis along with its objectives, contributions, and organization.

1.1 Overview

In this thesis, I investigated the automated detection of two different abnormalities in the body: tumors and scar tissues. I developed novel algorithms for automated evaluation of kidney tumors, localization of PCa, and characterization of scar in LV myocardium using CT and MR images. This dissertation consists of content from the research articles that have been published in, or submitted to, international peer-reviewed journals. Chapter 2 provides background details on the image processing technologies related to the thesis. The details of implementation and

validation of the developed detection and segmentation algorithms for the aforementioned applications are explained in Chapters 3 to 7. The dissertation ends with Chapter 8 containing a summary of the research work presented and future directions.

1.2 Contributions

The research work in this dissertation has novelty in both methodology and clinical applications.

The contributions of this thesis are as follows:

- Chapter 3: In this chapter, the development of a novel algorithm for differentiation of benign cyst from solid renal masses on contrast-enhanced (CE) CT images is presented. Automated differentiation of a benign cyst from the solid renal mass is clinically desirable and is an essential step for eventual classification into benign neoplasms and malignant solid renal mass (renal cell carcinoma (RCC)). However, to best of the author's knowledge differentiation of benign cysts from solid renal masses has not been studied to date. I described a decision fusion of patch-based CNN method to differentiate benign renal cysts from solid renal masses that benefits from the voting system for final prediction. This work is published in the journal of IEEE Access (*F. Zabihollahy, N. Schieda, and E. Ukwatta, "Patch-based convolutional neural network for differentiation of cyst from solid renal mass on contrast-enhanced computed tomography images," IEEE Access, 2020 [19]*).
- Chapter 4: This chapter describes the development of a semi- and fully automated methods to differentiate benign neoplasms from malignant (RCC) solid renal mass using CNN. Accurate imaging characterization of renal masses is important, because not all incidentally discovered renal masses are malignant and up to 20% of solid renal masses

are benign. I presented a semi- and fully automated methods based on decision fusion of a slice-based CNN model to differentiate malignant from benign tumors among solid renal masses on CECT. The proposed semi-automated method does not require manual segmentation of the kidneys and renal masses on CECT images. The proposed methodology benefits from the voting system for the final prediction of the renal mass type and has led to the most accurate results using the most unbiased dataset to date. The proposed method has been evaluated on a large test dataset of consecutive solid renal masses and has achieved a promising result compared to previous works, where a smaller number of renal masses in less well-balanced populations were studied. A key strength of our study is the use of a large institutional dataset of CT images from a well-designed patient cohort with different types of renal mass including all of the RCC subtypes and the two most common benign solid renal mass (namely fat-poor angiomyolipoma (fp-AML), and oncocytomas) acquired using CT systems developed by four different manufactures. Our results indicate that automated differentiation of RCC from benign neoplasms among solid renal masses encountered on CT is currently possible with moderate-to-high degrees of accuracy. This work is published in the Journal of European Radiology (*F. Zabihollahy, N. Schieda, S. Krishna, E. Ukwatta, "Automated classification of solid renal masses on contrast-enhanced computed tomography images using convolutional neural network with decision fusion," European Radiology, 2020 [20]*). Since the fully automated classification of renal mass was investigated in chapter 4, I also described a U-Net-based method to determine the boundaries of the renal mass on CECT images and demonstrated that this method is highly accurate using two different datasets. Our algorithm for renal mass detection yielded high sensitivity, such

that renal masses were detected in 100% and 86.67% of the institutional and kidney tumor segmentation (KiTS19) test images at the patient level. The algorithm also segmented the renal mass accurately, which would aid in future attempts to potentially classify renal mass into benign and malignant. Our designed model had the additional advantage of localizing multiple renal mass in patients who had more than one renal mass, which is commonly encountered in clinical practice. This work has been submitted to the Journal of Medical Physics (*F. Zabihollahy, N. Schieda, S. Krishna, "E. Ukwatta, Ensemble U-Net-Based Method for Fully Automated Detection and Segmentation of Renal Masses on Computed Tomography Images," J. Med. Phys., 2020 [21]*).

- Chapter 5: This chapter presents the development of a novel algorithm for fully automated detection and segmentation of LV scar from 3D LGE MRI. Accurate 3D segmentation of myocardial scar is emerging as a potentially valuable tool for risk stratification and procedural planning in patients with ischemic cardiomyopathy (i.e. heart attacks due to blockage of a coronary artery). I designed a novel DL-based method for fully automated segmentation of LV scar from 3D LGE MR images without any user interaction. To the best of the authors' knowledge, this is the first attempt to fully automatically segment LV scar from 3D LGE-MR images. This new strategy is able to comprehensively learn and integrate inter- and intra-slice features from 3D MR images, allowing for efficient context-based segmentation of target object. Inspired by human behavior, I tried to emulate human perception for 3D object recognition by integrating information from three orthogonal views. The designed methodology was evaluated by comparing the algorithm-generated segmentation with expert manual segmentation. Case-by-case investigation of the segmented scar on the test dataset revealed that our algorithm

is comparatively accurate in identifying scar description particularly at the apex, which is a challenging task due to its small size. The proposed algorithm is quite fast for LV myocardium and LV scar delineation from 3D LGE-MRI. It takes on an average 50 sec and 2 minutes for fully automated segmentation of LV myocardium and LV scar, versus 54 minutes and 42 minutes of expert manual segmentation, respectively on Intel Core i7, 2.8 GHz using a GPU-accelerated computing platform. This work is published in the *Journal of Medical Physics* (F. Zabihollahy, M. Rajchl, J.A. White, E. Ukwatta, “Fully Automated Segmentation of Left Ventricular Scar from 3D Late Gadolinium Enhancement Magnetic Resonance Imaging Using a Cascaded Multi-Planar U-Net (CMPU-Net)”, *J. Med. Phys.*, 2020 [22]).

- Chapter 6: This chapter contains details about the development of a fully automated algorithm for zonal segmentation of prostate anatomy on T2W and ADC map MR images. Accurate regional segmentation of the prostate boundaries on MR images is a fundamental requirement before an automated PCa diagnosis can be achieved. I developed and evaluated an automated method to segment the prostate whole gland (WG), central gland (CG) and peripheral zone (PZ) from T2W and ADC map MR images using a DL-based method, which is the first attempt for prostate zonal segmentation using ADC map MR images. Segmentation of prostate zonal anatomy using ADC map images is critical because it is considered one of the most important sequences for detection and grading of cancers in the PZ [23], [24] and, will be a fundamental step in potential computer-aided diagnosis (CAD) pipelines for detecting tumors in both PZ and transition zone (TZ). Our fully automated method yielded high accuracy in segmentation of the prostate WG and zonal anatomy for both T2W and ADC map images with substantially

less time compared to manual segmentation. Another contribution of our study is the combination of prostate MR images from a well-designed control group of patients without clinically significant tumors and in patients with significant tumors detectable on MRI. This distinction is important because if automated systems are to be used in clinical practice, the methods for segmentations should work equally well in patients with and without tumors. Our results showed similar segmentation accuracies regardless of the presence of tumor. This work is published in the Journal of Medical Physics (F.

Zabihollahy, N. Schieda, S. Krishna Jeyaraj, and E. Ukwatta, "Automated segmentation of prostate zonal anatomy on T2-weighted (T2W) and apparent diffusion coefficient (ADC) map MR images using U-Nets," Med. Phys., 2019 [25]).

- Chapter 7: Chapter 7 elucidates the development of a novel fully automated algorithm for detection and segmentation of PCa in PZ on ADC map MRI. Accurate detection and localization of PCa in men undergoing prostate MRI is a fundamental step for future targeted prostate biopsies and treatment planning. Fully automated localization of PCa in PZ using the ADC map might be clinically useful. I presented an ensemble learning-based model for fully automated PCa localization in the PZ from ADC map MR images. To this end, a novel combination of true positive and F1-score as a metric and loss function was used for the first time to training a deep network for PCa detection using ADC map MR images. It helped to enhance the true positive rate of the model while decreasing false-positive rates during network training that were crucial for abnormality detection, ameliorating both sensitivity and specificity of the suggested method. I suggested a tiling strategy for PCa detection, where the region of interest (ROI), PZ, is tiled in a window of fixed size. This strategy helped to excessively augment the data of

underrepresented class in the training phase. Additionally, in the testing phase, the tile-based summation strategy was applied that increased the probability of detecting suspicious tissue on ADC map MRI. The proposed framework does not require a full multiparametric (mp)-MRI examination and identifies the position of potential areas of PCa using only the ADC map MRI derived from diffusion-weighted images, which is the most important sequence for identification of tumors in PZ [23], [24]. Our presented methodology is highly sensitive, such that for all test cases with a clinically significant tumor, in at least one 2D slice that included tumor, PCa was detected. This translates into a 100% sensitivity at the level of dominant PZ tumor detection, which is ideal for a CAD system. The presented technique not only detects PCa but also discovers the boundaries of the tumor. To our knowledge, this is the first study evaluating prostate PZ tumor localization using only ADC map MR images employing the U-Net-based ensemble learning technique. This work is published in the Journal of Magnetic Resonance Imaging (*F. Zabihollahy, E. Ukwatta, S. Krishna, and N. Schieda, "Fully automated localization of prostate peripheral zone tumors on apparent diffusion coefficient map MR images using an ensemble learning method," J. Magn. Reson. Imaging, 2019 [26]*).

1.3 Thesis Organization

Chapter 1 In this chapter, we present the goals, overview, and contributions of this thesis.

Chapter 2 In this chapter, background information on the image processing technologies is first described. This chapter also provides an overview of abnormality detection and segmentation methods and algorithms and particularly provides general technical background on CNN and

CNN-based techniques.

Chapter 3 In this chapter, a semi-automated algorithm for the differentiation of cyst from solid renal masses on CECT images is presented. Moreover, the inter-observer variability of the algorithm is studied in this chapter and the results of the automated method are compared to those of thresholding based on CT number for benign cyst diagnosis.

Chapter 4 In this chapter, we describe semi- and fully automated methods for the classification of solid renal masses into malignant and benign and compare the obtained results with an alternative 3D-based algorithm.

Chapter 5 In this chapter, we introduce a novel U-Net-based technique for fully automated segmentation of LV myocardial scar from 3D LGE-MR images. The performance of the suggested methodology is investigated through comparison with several alternative methods including the most state-of-the-art DL-based methodologies.

Chapter 6 In this chapter, we present a method to segment prostate zones from T2W and ADC map MR images. I describe a coarse-to-fine localization approach for detecting and segmenting PCa on MR images, where first, the boundaries of the prostate regions including PZ and CG are accurately identified. Since T2W and ADC map MR images are the main sequences for detecting PCa in the CG and PZ respectively, the regional segmentation was performed for both types of MR images. The designed algorithm was then evaluated by comparing algorithm-generated segmentations to manual delineations performed by experts using a large local dataset.

Chapter 7 In this chapter, we describe an ensemble model to localize PCa in the PZ on ADC map MR images as one of the most important sequences for PCa detection. The accuracy of the designed model is investigated by comparing algorithm-generated results to the ground truth images.

Chapter 8 In this chapter, we conclude this thesis by summarizing the obtained results. It includes a list of limitations and recommendations for future research.

References

1. E. Bercovich and M. C. Javitt, "Medical Imaging: From Roentgen to the Digital Revolution, and Beyond," *Rambam Maimonides Med. J.*, 2018.
2. J. E. Thompson and P. D. Stricker, "Diagnostic accuracy of multi-parametric MRI and transrectal ultrasound-guided biopsy in prostate cancer," *The Lancet*. 2017.
3. B. M. Kline-Fath, D. I. Bulas, and R. Bahado-Singh, *Fundamental and advanced fetal imaging: Ultrasound and MRI*. 2014.
4. H. R. Torres, S. Queirós, P. Morais, B. Oliveira, J. C. Fonseca, and J. L. Vilaça, "Kidney segmentation in ultrasound, magnetic resonance and computed tomography images: A systematic review," *Computer Methods and Programs in Biomedicine*. 2018.
5. G. Lemaitre, R. Marti, and F. Meriaudeau, computer-aided diagnostic systems for prostate cancer detection challenges and methodologies, chapter 10, *Prostate Cancer Imaging: An Engineering and Clinical Perspective*, CRC Press, 2018.
6. S. M. McKinney, M. Sieniek, V. Godbole, J. Godwin, N. Antropova, H. Ashrafiyan, *et al.*, "International evaluation of an AI system for breast cancer screening," *Nature*, 2020.
7. C. M. Bishop, *Pattern Recognition and Machine Learning*, Springer, ISBN 978-0-387-31073-2. 2006.
8. H. Al-Sahaf, Y. Bi, Q. Chen, A. Lensen, Y. Mei, Y. Sun, B. Tran, B. Xue, and M. Zhang "A survey on evolutionary machine learning," *Journal of the Royal Society of New Zealand.*: 49:2, 205-228, 2019. DOI: 10.1080/03036758.2019.1609052.
9. G. Notte, M. Pedemonte, H. Cancela, and P. Chilibroste, "Resource allocation in pastoral dairy production systems: Evaluating exact and genetic algorithms approaches," *Agric. Syst.*, 2016.
10. S. Srinivasan and T. Kamalakannan, "Multi Criteria Decision Making in Financial Risk Management with a Multi-objective Genetic Algorithm," *Comput. Econ.*, 2018.
11. G. C. Liao and T. P. Tsao, "Application of a fuzzy neural network combined with a chaos genetic algorithm and simulated annealing to short-term load forecasting," *IEEE Trans. Evol. Comput.*, 2006.

12. B. B. Pal, S. Roy, and M. Kumar, "A Genetic Algorithm to Goal Programming Model for Crop Production with Interval Data Uncertainty," *Environmental and Agricultural Informatics*, 2016.
13. J. De Fauw, J. R. Ledsam, B. Romera-Paredes, S. Nikolov, N. Tomasev, S. Blackwell *et al.*, "Clinically applicable deep learning for diagnosis and referral in retinal disease," *Nat. Med.*, 2018.
14. Y. Lecun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*. 2015.
15. D. Cireşan, U. Meier, J. Masci, and J. Schmidhuber, "Multi-column deep neural network for traffic sign classification," *Neural Networks*, 2012.
16. A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," *Commun. ACM*, 2017.
17. NIH, National Cancer Institute, NCI Dictionary of Cancer Terms. Accessed on: Feb. 23, 2020. [Online]. Available: <https://www.cancer.gov/publications/dictionaries/cancer-terms/def/abnormal>.
18. Contrast Materials. July 23, 2018. Accessed on: Feb. 16, 2020. [Online]. Available: <https://www.radiologyinfo.org/en/pdf/safety-contrast.pdf>.
19. F. Zabihollahy, N. Schieda, and E. Ukwatta, "Patch-based convolutional neural network for differentiation of cyst from solid renal mass on contrast-enhanced computed tomography images," *IEEE Access*, 2020.
20. F. Zabihollahy, N. Schieda, S. Krishna, E. Ukwatta, "Automated classification of solid renal masses on contrast-enhanced computed tomography images using convolutional neural network with decision fusion," *European Radiology*, 2020.
21. F. Zabihollahy, N. Schieda, S. Krishna, E. Ukwatta, "Ensemble U-Net-Based Method for Fully Automated Detection and Segmentation of Renal Masses on Computed Tomography Images," *J. Med. Phys.*, 2020.
22. F. Zabihollahy, M. Rajchl, J.A. White, E. Ukwatta, "Fully Automated Segmentation of Left Ventricular Scar from 3D Late Gadolinium Enhancement Magnetic Resonance Imaging Using a Cascaded Multi-Planar U-Net (CMPU-Net)," *J. Med. Phys.*, 2020.
23. J. C. Weinreb, J. O. Barentsz, P. L. Choyke, F. Cornud, M. A. Haider, K. J. Macura *et al.*, "PI-RADS Prostate Imaging - Reporting and Data System: 2015, Version 2," *Eur. Urol.*, 2016.
24. J. O. Barentsz, J. C. Weinreb, S. Verma, H. C. Thoeny, C. M. Tempany, F. Shtern, A. R. Padhani, *et al.*, "Synopsis of the PI-RADS v2 Guidelines for Multiparametric Prostate Magnetic Resonance Imaging and Recommendations for Use," *European Urology*. 2016.
25. F. Zabihollahy, N. Schieda, S. Krishna Jeyaraj, and E. Ukwatta, "Automated segmentation of prostate zonal anatomy on T2-weighted (T2W) and apparent diffusion coefficient (ADC) map MR images using U-Nets," *Med. Phys.*, 2019.
26. F. Zabihollahy, E. Ukwatta, S. Krishna, and N. Schieda, "Fully automated localization of prostate peripheral zone tumors on apparent diffusion coefficient map MR images using an ensemble learning method," *J. Magn. Reson. Imaging*, 2019.

Chapter 2: Background

Medical abnormality manifest as a lesion or growth in the human body that may be benign (not cancer), precancerous or premalignant (likely to become cancer), or malignant (cancer) [1]. Since not all abnormalities are cancerous, it is essential to further examine abnormalities for treatment planning. Medical imaging that visualizes the internal body structures, is a non-invasive tool for diagnosis and prognosis of medical abnormalities [2]. Medical imaging establishes a collection of information for normal anatomy that can be used as a reference to identify abnormalities. Accordingly, abnormality in medical images is typically the suspicious structure in a medical image, which is different from normal tissue [3]. Standard imaging most of the time may not be able to differentiate them [4] and contrast agents needed to be used. Contrast agent is a substance used to increase the contrast difference between normal and abnormal tissues in medical imaging [5]. Different contrast agents are used for various imaging modalities. Gadolinium contrast media, a chemical substance, used in MRI scans. Gadolinium molecules shorten the spin-lattice relaxation time (T1) of voxels in which they are present. As a result, those voxels have a brighter signal on T1-weighted images [6]. This contrast agent is injected into a vein in the MRI scan and eventually gets eliminated from the body through the kidneys [7]. Contrast agent material such as iodine-based and barium-sulfate may be used in CT scans to improve the contrast of the abnormalities in CT images [4]. When these contrast agents are present in a specific area of the body, they limit the ability of x-rays to pass through. As a result, the appearance of the target tissue that temporarily contain contrast agent compounds changes on CT images.

Due to the advances in medical imaging technologies, medical imaging data now comprises of big data, multiple sequences with different contrasts and multiple dimensions. Although novel medical imaging techniques provide more accurate spatial representation and quantification of abnormal tissues, manual analysis and interpretation of big data is not feasible. Thus, developing CAD tools that could ease the diagnosis and prognosis process for at-risk populations, is of prominent importance. In this chapter, we first briefly describe CT and MRI imaging techniques. Different methods, used for automated detection, evaluation, and characterization of abnormalities in medical images are then reviewed.

2.1 Medical Imaging Techniques

Medical imaging technologies can be broadly categorized into anatomical and functional imaging. Widely used medical imaging techniques are x-ray radiography, nuclear medicine, ultrasound, MRI, and CT. Radiography employs X-rays radiation to visualize the internal form of the human body [8]. Nuclear medicine that has both diagnostic and treatment applications for various pathologies, uses certain properties of isotopes and the energetic particles emitted from radioactive materials. The two most common imaging modalities in nuclear medicine are single-photon emission computed tomography (SPECT) and positron emission tomography (PET) scans that use radioactive tracers (radiopharmaceuticals) to assess bodily functions and to diagnose and treat disease [9]. Ultrasound utilizes high-frequency acoustic waves (> 20 kHz) that are sent into the tissue and reflected to varying degrees depending on the composition of different tissues to produce images [10]. Although ultrasound provides less anatomical details compared to other imaging modalities like CT and MRI, it studies the function of moving structures in real-time, and emits no ionizing radiation. The focus of this dissertation is on CT and MRI, which will be described in more detail in the following sections.

2.1.1 CT Scan

CT is a type of imaging technique that produces 2D images of a thin section of the internal organs of the human body. In CT, beams of X-rays are emitted around an object and after penetrating the object from different angles, are collected by several detectors. The slices are then reconstructed using a reconstruction algorithm such as Radon transform [11]. The tomographic images of a CT scan are used for diagnostic and treatment planning in various medical applications [12]. CT has become an important tool in medical imaging to scan head and likewise to perform CT-guided stereotactic surgery and radiosurgery [13-14], to scan neck and thyroid [15-16], gain knowledge about cardiac or coronary anatomy and function [17], assess bronchial wall thickening in lung [18], and to diagnosis abdominal diseases. Figure 2-1 shows an example of a CT scan of the abdomen and pelvis.

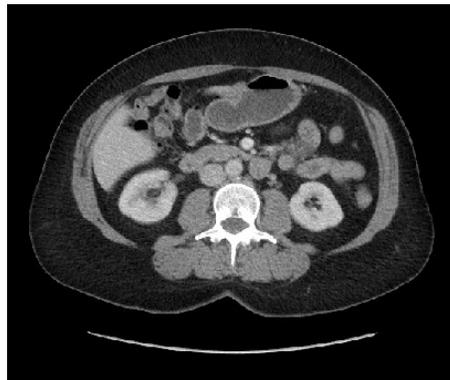


Figure 2-1. An example of a CT scan of the abdomen and pelvis, taken in the transversal plane, visualized in the standard abdominal window and level set of 400 and 50, respectively.

2.1.2 MRI

MRI introduces a powerful magnet to excite hydrogen nuclei (i.e., single protons) of water molecules in human tissue and align them with the direction of the static magnetic field [19].

MRI machine has radiofrequency coils that send pulse at the resonant frequency of the hydrogen atoms on water molecules to the area of the human body to be examined. The radio frequency signal is then absorbed by protons, align them in 180 degree with the direction of the primary magnetic field. When the radio frequency coils stop sending radiofrequency pulse, the protons relax and emit radio-waves, which are detected and reconstructed into an image. MRI uses a very strong static electromagnetic field to polarize the hydrogen nuclei (typically 1.5 to 3 Teslas). Most MRI techniques can be performed using either a 2D or 3D acquisition mode. To acquire 2D MR images, protons are shaken in a thin slice only (2D), and magnetic gradients are applied in two directions parallel to the xy-plane to create an image of a single slice (Figure 2-2). If we repeat the imaging process many times, we can create a stack of slices that together build a 3D volume, which is called multiple 2D (M2D). Although multiple 2D MRI requires a short time for imaging, it has a stair-step artifact due to the relatively thick, non-isotropic voxels. Recently 3D MRI has been introduced with high spatial resolution and high signal-to-noise ratio. 3D MR images have isotropic voxels that allow multiplanar reformation [20].

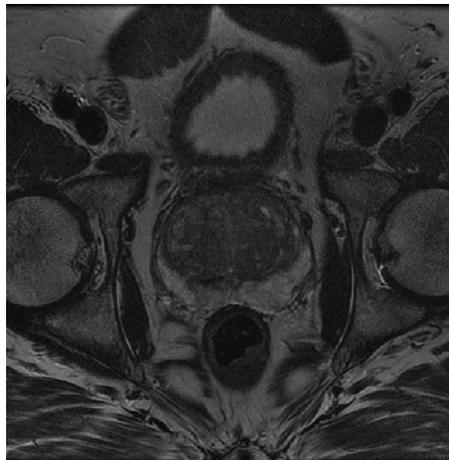


Figure 2-2. An example of a prostate T2-weighted MR image taken in the transversal plane.

2.2 Medical Image Analysis

The purpose of medical image analysis is to develop methods to interpret images for the diagnosis and prognosis of patients. Medical image analysis problems could be categorized as detection, segmentation, classification, registration (i.e. comparing different modalities/patients), reconstruction (i.e. making 3D-measurement), measuring, and visualizing results. In this research, our focus is on the abnormality detection using CT and MR images.

2.2.1 Object Detection and Image Segmentation

Object detection defines as detecting instances of semantic objects of a certain class in a digital image (Figure 2-3a) while image segmentation is partitioning an image into two or more meaningful regions (Figure 2-3b) [21-22].

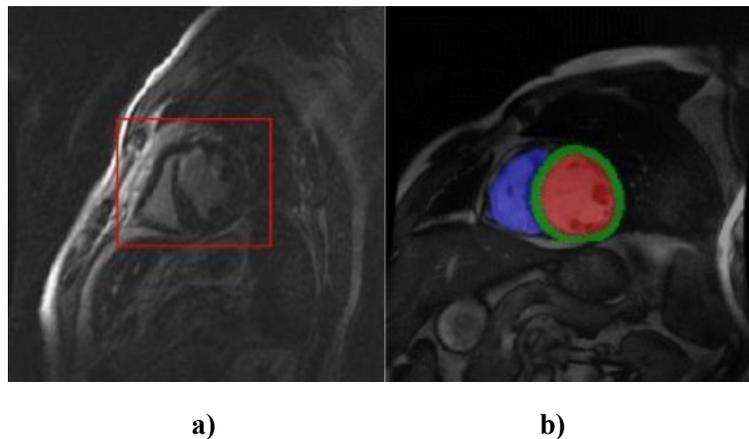


Figure 2-3. An example of object detection and image segmentation from left to right. In the left picture, a red box enclosed right ventricle (RV) and left ventricle (LV) that displays the location of RV and LV descriptions on MRI. In the right picture, RV, LV, and LV myocardium are segmented from the MR image and shown in blue, red, and green colors, respectively.

Object detection and image segmentation algorithms have been developed to examine images of scenes, videos, and 2D/3D/4D/nD medical images. Medical image segmentation algorithms are required in clinical applications to study anatomical structures; identify an ROI and locate tumors, lesions, and other abnormalities; measure tissue volume to measure growth/shrinking of tumor, and help in treatment planning prior to radiation therapy by calculating radiation dose [23].

Previous methods developed for object detection and image segmentation can be categorized into conventional, ML-based and DL-based algorithms. Conventional techniques are further divide into intensity-based thresholding and functional optimization methods (e.g. active contours, level sets, live wire, graph cuts, convex relaxation techniques, etc.) [24]. In this section, we first describe conventional and ML-based methods and then follow it up with DL-based methods. We then describe strategies that are useful to improve the performance of the DL methods.

2.2.1.1 Intensity-threshold-based Methods

Intensity-based thresholding methods are the simplest algorithms in which the image is directly divided into multiple or two (i.e. target and background) regions (called local and global thresholding, respectively), based on the intensity value of the different regions [25]. Intensity thresholding-based method such as Ostu [26], has been widely used for automated detection of abnormalities in medical images and has provided satisfactory results for images with a histogram of bimodal distribution [27]. Some modifications have been introduced in the literature to improve Ostu method for abnormality detection [28].

Another intensity thresholding-based algorithm is region growing (RG). The basic idea in RG is to cluster the pixels with similar properties into the same group to form regions on the image [29]. To do so, a seed pixel must be selected first and then the similar pixels around the seed pixel must be merged into the region where the seed pixel is located.

Some intensity thresholding-based methods such as full width at half maximum (FWHM) and signal threshold to reference mean (STRM) are cardiac-specific. These types of image segmentation methods are widely used in literature due to their efficiency and ease of use. At FWHM target is considered as a reference and all intensity values that are larger than half of its maximum intensity value, are defined as a target [30]. In STRM, background is considered as a reference and target is defined as its mean value plus two (STRM2), three (STRM3), four (STRM4), five (STRM5), or six (STRM6) standard deviation [31].

Although intensity-thresholding methods are very simple to implement, they fail when the image histogram is close to unimodal in binary segmentations. For abnormality detection applications, the intensity of the abnormal tissue can range from unimodal to bimodal. As these types of algorithms solely rely on intensity value as a discriminant feature, they yield poor results when a combination of shape, size, texture and intensity features must be considered to distinguish abnormality from medical images. Besides, these methods are highly influenced by image noise [32]. Furthermore, these methods are subject to high inter and intra-observer variability due to the need for manual interactions.

2.2.1.2 Energy Optimization Methods

Energy optimization methods first define a mathematical criterion for the “goodness” of a given segmentation that translates the formulation of the segmentation problem as an optimization problem under certain geometric constraints. Practically in functional optimization techniques, an initial contour is defined around or inside the target and evolved by minimizing an energy function. The energy function could be a combination of shape, length-based, and region-based energy terms [35] that reaches its minimum value when the contour lies on the boundary of the target [36-37]. In other words, functional optimization methods search for a unique contour that lies on the boundary of the target by minimizing energy function based on optimizing an objective function. The energy function can be defined as [35]:

$$E = \int_{inside(C)}^1 |I_s - c_1| dx dy + \int_{outside(C)}^1 |I_s - c_2| dx dy + Length(C) + Area(inside(C)), \quad (2-1)$$

which is a combination of shape, length-based, and region-based energy terms. C is the initial curve and constants c_1 and c_2 are the averages of intensities inside and outside C . I_s is the sub image and (x,y) are coordinates of image pixels.

While there are many energy minimization-based methods, here we describe three most popular methods, namely level set, graph cuts, and convex max-flow.

Level Sets

In the level sets method, the segmentation boundary is represented as the zero level of an implicit function and the segmentation boundary is iteratively evolved by optimizing an objective function to find the boundary of an object [35]. The level set has a volumetric representation in which the deformation of the surface is controlled by a speed function in the partial differential

equation. The level sets methods can change topology which is an advantage over conventional energy minimization methods such as, snakes. Level sets have been extensively used to segment medical structures from MR images [38], CT images and 3D Ultrasound (US) images [39-46] by modifying the objective function to add domain specific constraints.

Graph-Cuts

The Graph-cut method is an efficient energy minimization-based algorithm that has been used widely for image segmentation. Like other functional optimization methods, in this algorithm, the objective function contains data and regularization terms. Data term is a measure of conformity of the image data within the segmentation region along with the statistical model of the image data. Regularization term, on the other hand, controls the smoothness of the segmentation boundaries [47]. Minimizing objective function with a piecewise constant data term produces nearly global optima which is an advantage of this technique, makes it less sensitive to the initialization [34].

Convex Max-Flow

In convex relaxation-based approaches [48], a contour is propagated to its globally optimal position at each discrete time frame by solving a sequence of convex optimization problems, for which an efficient continuous max-flow algorithm is available [49]. In this method, the evolution can be summarized as expansion (C^+) and shrink (C^-) of contour (C) [50-52]. This method is

similar to the level-set with the difference that the max-flow algorithm is used for optimization instead of gradient descent. The object function is defined as:

$$C_{t+1} = \min \int_{C^+} C^+(x)dx + \int_{C^-} C^-(x)dx + \int_{\partial C} g(s)ds, \quad (2-2)$$

In this equation, the first two terms are regional, and $g(s)$ is a boundary term. Convex max-flow methods have been widely used for numerous clinical applications, such as vascular 3D MRI and US [53-55], prostate 3D MRI and US [56-60], and neonatal 3D US [61]. The performance of energy optimization-based methods plateaus despite the increased number of available images. However, the performance of ML- and DL-based methods ameliorate with the availability of more labeled data. Sun et al. demonstrated that the performance on vision tasks increases logarithmically based on the volume of training data size [62].

2.2.1.3 Machine Learning (ML)-based Methods

ML methods are a set of algorithms developed to learn meaningful patterns from example data aiming to minimize human interaction [63]. Selecting an appropriate algorithm and proper training allow the machine to learn patterns more efficiently. ML algorithms can be divided into two main categories: supervised and unsupervised. In supervised learning, we utilize labeled data to train a model whereas in unsupervised learning we allow the model to work on its own to discover information, and we utilize unlabelled data. For supervised learning-based methodologies, labeled data is divided into training and testing parts. The training samples are utilized to learn from and develop an algorithm to perform a task. Testing data are used to assess the performance of the trained model. Some popular ML algorithms are described in the following.

K-Means Clustering

K-means clustering is an example of unsupervised algorithms that has been widely used for image segmentation [64]. Like all clustering methods, the starting locations of the partitions are important for achieving optimal solution since they are susceptible to termination when achieving a local maximum. Incorporating anatomical and spatial information into segmentation is the main challenge for clustering algorithms in medical image segmentation applications.

Support Vector Machin (SVM)

Like other classifiers, SVM constructs a hyperplane or set of hyperplanes as a decision boundary to separate between different classes. A good separation is achieved when the decision boundary has the largest distance (margin) to the nearest training samples of any class. This larger margin leads to the lower generalization error of the classifier, a measure of how accurately an algorithm can predict outcome values for testing data (previously unseen data) [64]. SVM is efficient in high dimensional spaces and in cases where the number of dimensions is greater than the number of samples [65]. SVM is a type of supervised learning algorithm.

K Nearest Neighbor (KNN)

KNN is a non-parametric instance-based supervised learning method that does not construct a general model for classification. KNN stores training samples along with the labels in a feature space and classifies an object by a majority vote of its nearest neighbors [66]. Although this

algorithm does not require explicit training, KNN is quite sensitive to the local structure of the data.

Random Forest (RF)

The RF is an ensemble algorithm, which trains and combines multiple decision trees to produce a highly accurate classifier [67]. RF applies the general technique of bagging in the training phase where random subsets of features are selected to train decision trees on these samples. The prediction for a test sample is then performed by taking a vote from the predictions from all individual trees [68]. The bagging procedure leads to a better performance provided that decision trees are not highly correlated. It runs efficiently on large databases and is quite fast to be built and predict. RF is a type of supervised learning method.

Artificial Neural Network (ANN)

ANN is another type of supervised learning system that learns how to perform tasks by considering examples. ANN is made up of artificial neurons or nodes [69], where the connection between nodes is modeled as weights and learned from instances. A simple ANN is illustrated in Figure 2-4, where x_i and w_{ij} denote input features and weights, respectively.

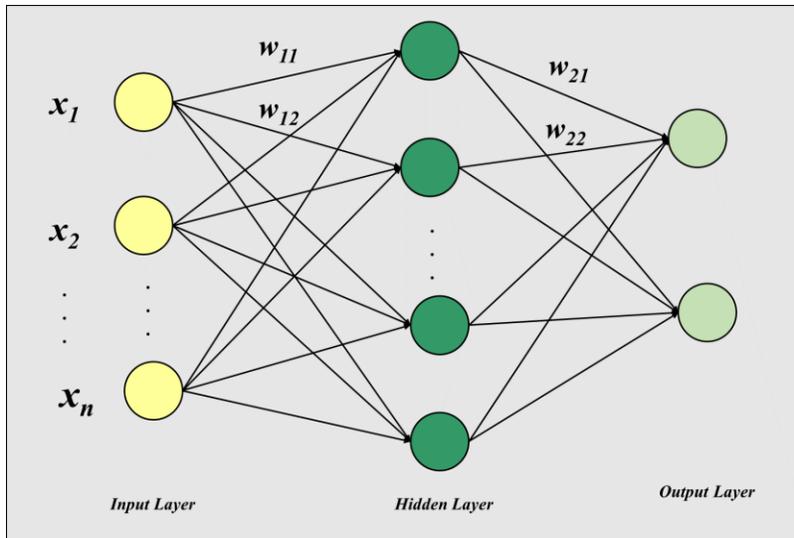


Figure 2-4. An overview of a simple neural network with one hidden layer.

In ANN, information (x_i) flows forward through the network to produce an output (forward propagation). The difference between produced output and the true label is then propagated backward through the network to compute the gradient and correct the weights (backpropagation algorithm [70]). This algorithm is used to efficiently train ANN following gradient descent approach by applying the chain rule. The backpropagation algorithm updates the weights via iteration technique to improve the network until it can perform the task for which it is being trained [71].

ANN can be used for image analysis to detect or localized an object, delineate the boundary of an object or classify images into different categories. As the input of ANN is in vector form, the image must be vectorized, where pixel intensities are directly converted into the feature space. Since the structural information among neighboring pixels or voxels is a source of information in a given image, this operation destroys spatial information [72], which is a major drawback of this method for image analysis.

2.2.1.4 Deep Learning Techniques

The availability of an enormous amount of data and improving of computational power made DL a default technique for medical image analysis. DL is a subset of ML algorithms, where ANN learns from large amounts of data. We refer to ‘deep learning’ because the ANN has several (deep) layers that enable learning. Among various DL techniques, CNN-based algorithms are the most popular ones for medical image analysis. Since 2D/3D convolutional layers are used in CNN-based methods to learn features, these types of networks are ideal for processing 2D/3D data (data with grid-like topology [71]) such as images. As we used CNN- and U-Net-based techniques in most of our research works, those methods are described in more detail in the following sections.

2.2.1.4.1 CNN

In the last few years, CNN-based algorithms have led to major advances in image analysis that has increased the interest in the applicability of these methods for problems in medical image analysis [73-75]. Recent works have shown breakthrough results on tasks as diverse as liver lesion segmentation [76], brain tumor diagnosis [77], left ventricle myocardium characterization [78], right arterial segmentation [79], muscularis propria delineation [80], diabetic retinopathy diagnosis [81-83], prostate tumor detection [84], and breast cancer classification [85].

The advantage of CNN compared to ANN is that it releases the constraint of immediate image vectorization by applying convolutional (Conv.) and pooling layers (having neurons arranged in 3 dimensions: width, height, and depth) to the input image [71], [86]. CNN, which is made up of Conv., detection, and pooling layers makes the explicit assumption that the inputs are images,

which allows us to encode certain properties into the architecture [87]. The special architecture of CNN helps to extract useful features from the image and use them as the input layer of the classifier (e.g. ANN) that is located at the end layer of the network. Figure 2-5 shows an overview of a typical CNN.

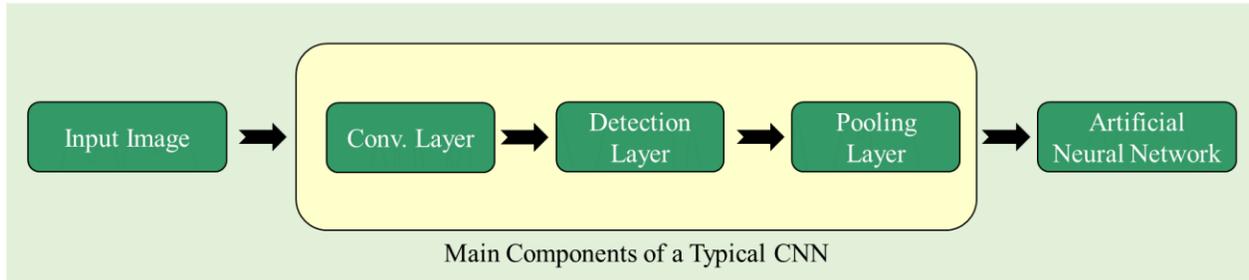


Figure 2-5. A typical CNN pipeline where Conv., detection, and pooling layers are applied to the input image to extract useful features from it. These features are then used as inputs of an ordinary ANN.

Convolutional Layer

In the convolutional layer, some kernels/filters ($K(i, j)$) convolve with the input image ($I(i, j)$) to produce the feature map ($F(i, j)$).

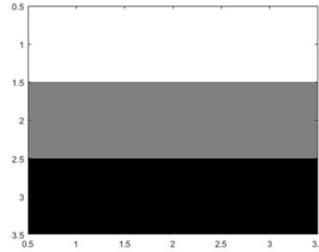
$$F(i, j) = (I * K)(i, j) = \sum_m \sum_n I(m, n)K(i - m, j - n) \quad (2-3)$$

i and j denote the position information of each pixel in a digital image of size $m \times n$. Conv. layer computes the dot product between the input image and kernel by moving the kernel across the image to detect local features at different positions. Mathematically, we get the large value in the feature maps where the template of the filter is found in the input image. By doing so, the image features which are similar to the kernel are captured. Figure 2-6 displays the convolution

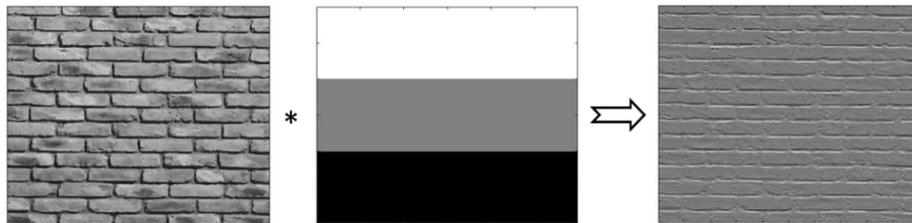
operation in the Conv. layer through an example. In CNN, filters are considered as weights and are set from the examples during network training via the backpropagation algorithm.

Image × Kernel = Feature Map

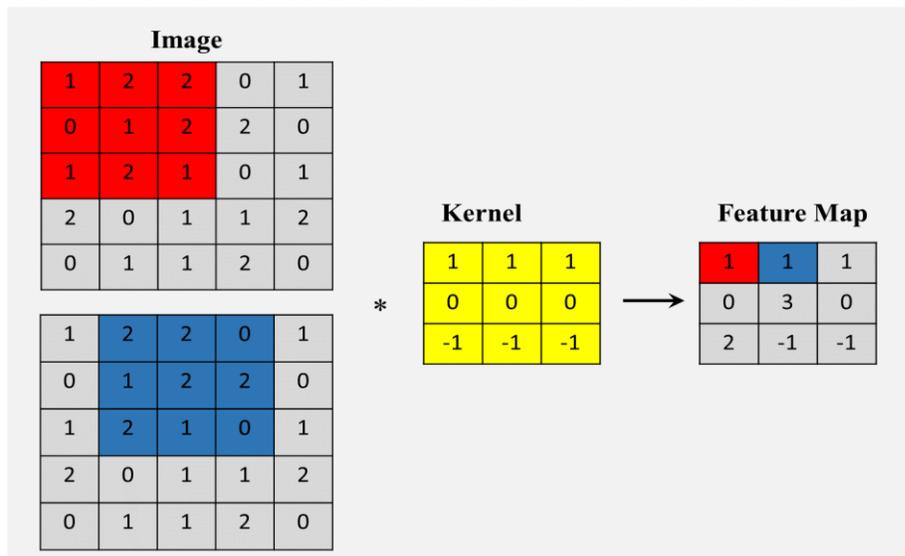
$$K = \begin{bmatrix} 1 & 1 & 1 \\ 0 & 0 & 0 \\ -1 & -1 & -1 \end{bmatrix}$$



a)



b)



c)

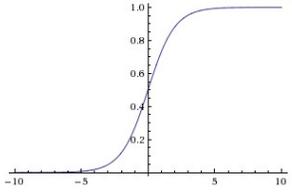
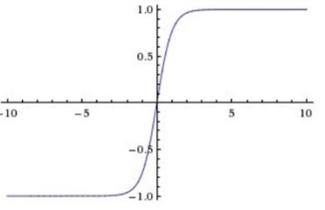
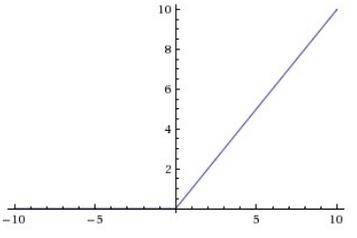
Figure 2-6. a) A sample of the kernel that captures the horizontal line in an image is shown. b) An example of applying horizontal kernel to the image along with the output feature map are illustrated, where horizontal

lines are highlighted in the feature map. c) The mathematical workflow of feature map generation in the Conv. layer is displayed, where the summation of the element-wise dot product between the image and kernel at each and every position is calculated.

Conv. layer improves the ML algorithm through sparse interactions, parameter sharing, and equivariant properties [71]. Despite traditional ANN, where every output unit interacts with every input unit (fully connected neural network), CNN has sparse interaction (sparse connectivity/weights) as the size of the kernel is smaller than that of the input image. The sparse interaction property not only reduces the memory requirements of the model but also improves its statistical efficiency (a measure of quality of an estimator/an experimental design where fewer observations for statistically efficient is needed compared to less efficient one to achieve a given performance [88]), which leads to fewer operations for output computation. In conventional ANN each element in the weight matrix is used once whereas, in CNN, each filter is utilized at every position of the input. In other words, in CNN parameter sharing is performed rather than learning a separate set of parameters for every location. Therefore, only one set of parameters is learned. This property further reduces the storage requirements of the model. Additionally, CNN is equivariance to translation, which means that if the input changes, the output changes in the same way. Put differently, moving an object in the input image will move its representation in the feature map the same amount.

Detection Layer

The detection layer (activation function) applies an elementwise activation function to the feature map, in which the size of the feature map does not change. Technically, this layer performs non-linear thresholding to the input image. This operation increases the nonlinear properties of the decision function as in most of the complex cases, data are not linearly separable. The most popular activation functions are tangent hyperbolic (tanh), sigmoid, and rectified linear unit (ReLU), which are defined as follow:

Sigmoid		$f(x) = \frac{1}{1 + \exp(-x)}$
tanh		$f(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}}$
ReLU		$f(x) = \begin{cases} 0 & \text{for } x < 0 \\ x & \text{for } x \geq 0 \end{cases}$

Pooling Layer

Pooling layers computes a summary statistic (maximum, mean, median, etc.) of some output nodes. Figure 2-7 illustrates an example of a maximum pooling layer. The advantage of this

layer is that it makes the model less sensitive to the local translation. Additionally, it yields a simpler model through down-sampling, which reduces the computational complexity of the next set of layers [89].

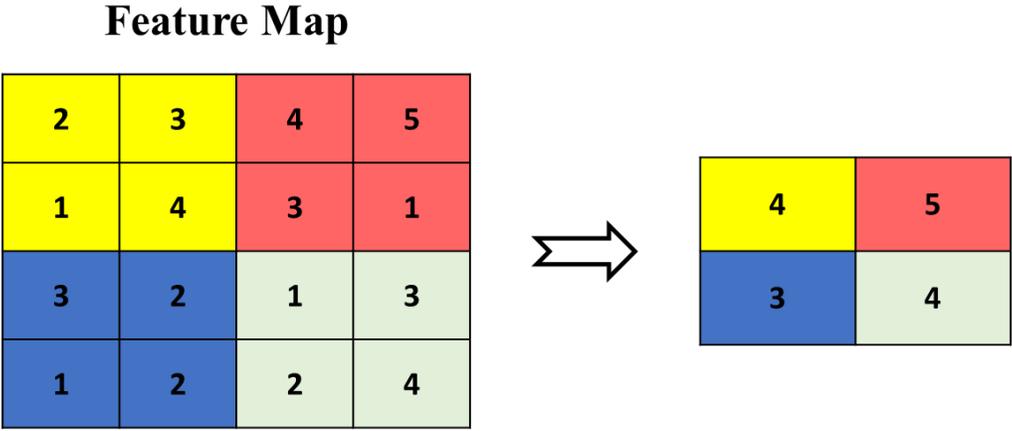


Figure 2-7. An example of applying a maximum pooling layer of size (2×2) and stride of 2 that downsizes the feature map. Stride is the number of pixels shifts over the input matrix.

Any combination of the Conv., detection, and pooling layers can be used to design an optimal CNN for a given problem. Figure 2-8 illustrates an example of CNN architecture in which two Conv. and detection layers followed by a pooling layer are applied to an image to extract useful features from that. Learned features are inputted to a fully connected ANN.

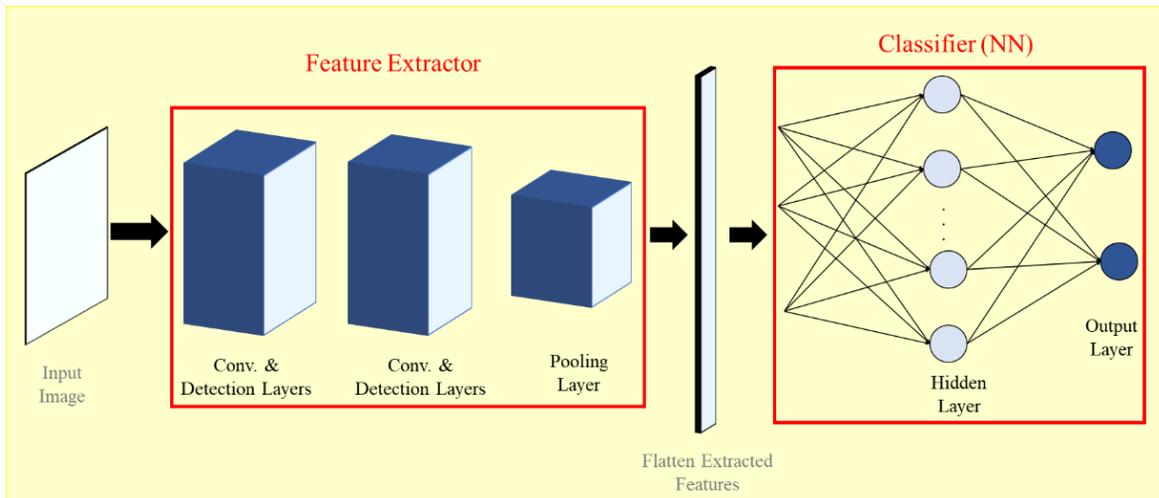


Figure 2-8. An example of CNN architecture, where two Conv. and one pooling layers are employed for feature extraction.

To employ CNN for segmentation tasks, a local region (patch) around each pixel is extracted in a sliding window fashion and passed through the network to be labeled as foreground or background. This method suffers from redundant computation due to the overlap exists between adjacent patches. Moreover, global features are not captured when small patches are extracted, and localization error is increased when large patches are used [90]. Finally, extracting image patches is time-consuming, requires prior constraint of the target tissue, and overall causes this process to remain slow. As the primary goal of applying AI in medical image analysis is to achieve both speed and user independence, this approach remains sub-optimal. U-Net has been introduced recently to address the limitations of CNN for image segmentation.

2.2.1.4.2 U-Net

U-Net architecture [88] was introduced explicitly with the segmentation of medical images in mind and used to produce state-of-the-art results on the international symposium of biomedical

imaging (ISBI) challenge for segmentation of neuronal structures in electron microscopic stacks as well as the ISBI cell tracking challenge 2015 [90]. The most important modification of this architecture is that convolutional layers are used in expanding path as well that allows learning features to propagate to the higher resolution layers. U-Net does not need patch extraction and generates a segmentation map with the corresponding size of the input image that accommodates images of arbitrary sizes.

Figure 2-9 shows the architecture of typical U-Net with five stages. The U-Net is made up of contracting and expanding paths where pooling and up-sampling layers are used in each way that yields a U-shaped architecture. The contraction path is identical to standard CNN in which convolutional layers along with pooling and activation layers are applied to the input data. In expanding path, pooling layers are replaced by up-sampling layers to expand the dimension of feature space. The output of up-sampling layers is merged with appearance feature representation learned from the corresponding layer in the shrinking path to localize high-resolution features. U-Net does not have a fully connected neural network at the end.

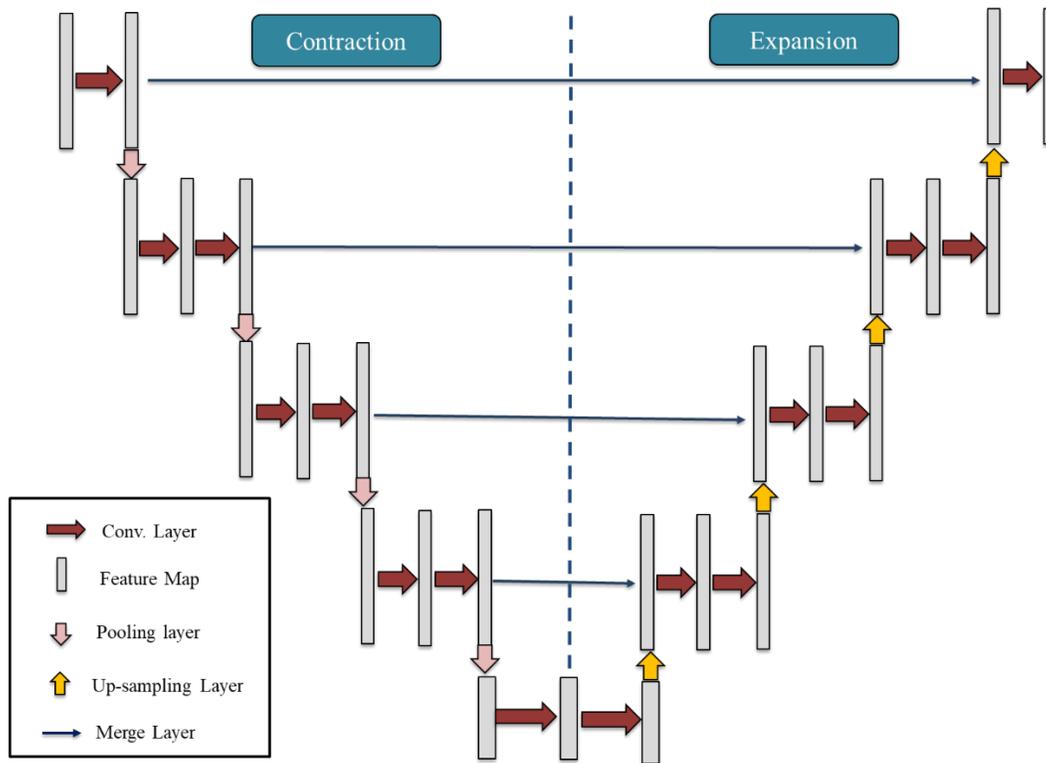


Figure 2-9. An overview of a typical U-Net with five stages.

2.3 Regularization

Regularization is the process of adding information in order to prevent overfitting in ML problems [91]. The foremost problem in ML is to develop an algorithm that not only performs well on training data but also on the new data. Different strategies help to reduce generalization error. These strategies are known collectively as regularization. Major forms of regularization that are available for DL practitioners are dataset augmentation, parameter sharing, dropout, early stopping, batch normalization, and ensemble methods.

2.3.1 Data Augmentation

The common approach for reducing generalization error in ML is to train the model on more data. The problem is that in practice the number of training instances is limited. To tackle this,

the artificial data can be created from the original data and augmented to the training set. For image data, artificial images can be generated through different transformations such as image shifting, rotating, and flipping. It has been shown that elastic deformation greatly improves generalization property [92]. Artificially augmenting training data is also beneficial for addressing class imbalance issue.

2.3.2 Parameter Sharing

By nature, an extensive parameter sharing occurs in CNN as each kernel in the convolutional layer is moved across the whole input data at every position. Thus, in CNN only one set of kernel is learned rather than learning a separate set of parameters for every location. Parameter sharing not only regularizes the parameters but also leads to a significant reduction in the required memory by storing only a subset of the parameters [71].

2.3.3 Dropout

The key idea in the dropout technique is to randomly drop some nodes along with their connections from the neural network in the training phase [93]. At every iteration, some nodes along with all their incoming and outgoing connections are randomly selected and removed from the network. So, each iteration has a different set of nodes (a different network) that results in a different set of outputs. During the testing phase, the average of all those networks' predictions is obtained.

2.3.4 Early Stopping

Investigating the train and validation errors during training a model shows that after a certain number of epochs (the number of times that the learning algorithm will work through the entire training dataset [94]), the validation error begins to increase while the training error steadily decreases. To obtain the best-trained model, we can store the model parameter with the lowest validation set error and use it for the test rather than the latest parameters, aiming to get a better result on the test set. This strategy is known as early stopping, which is one of the most common regularization techniques in DL due to its simplicity and efficiency.

2.3.5 Batch Normalization

Batch normalization is another regularization method that keeps the mean and standard deviation of the input data close to zero and one, respectively. This operation makes the learned function invariant to scaling of the weights [95]. Batch normalization makes the optimization significantly smoother, which in turn leads to more predictive and stable behavior of the gradient, allows for faster training [96].

2.3.6 Ensemble Learning

In ensemble learning, multiple classifiers of the same or different kinds are trained and combined for classification task to obtain a better predictive performance [97-99] or reduce the chance of selecting a poor model for a given classification task. The most popular algorithms in ensemble learning are Bagging, Boosting, Adaboost, and a mixture of experts [100] in which the outputs of several classifiers are combined through a linear rule. This rule could be an algebraic combiner

such as minimum, maximum, sum, average, etc. or a voting-based combiner, where majority voting or weighted majority voting is employed for final prediction. More comprehensive information is provided in chapter 7, section 7.4.

2.4 Evaluation Metric for Object Detection and Segmentation

The algorithms developed for a computer-aided detection task needs to be objectively evaluated for its accuracy and precision. When there are many alternative algorithms for a given task, we need to create performance criteria to choose the most appropriate algorithm. Additionally, the evaluation process reveals the strengths and limitations of suggested methods. The objectives of evaluation could be comparing two different methods or comparing one method with a well-known method (gold standard) or with ground truth. Since the actual boundaries of patient's organs and tissues cannot be identified, expert manual segmentation in medical images is often used as a surrogate for ground truth. This lack of ground truth is a clear difference from other image segmentation and object detection tasks in the computer vision domain. Ground truth in medical imaging refers to the process whereby an expert human operator (e.g., expert radiologist) segments/labels/registers an image by hand or detects an object in an image. The manual processing by an expert is subject to inter and intra-operator variability. To address this problem, the average statistics or a majority vote of several manual expert outcomes could be considered as a ground truth. An alternative would be to use a phantom with a known geometry to establish a ground truth and validate the accuracy of the developed algorithm; however, this is not a realistic approach as phantom does not simulate the characteristics of patients' real images with sufficient fidelity [101].

In this research work, various DL algorithms were employed to detect and segment abnormalities using different medical images. The results were compared to manual segmentations using the metrics and are described in the following sections to evaluate the performance of our developed algorithms. Additionally, the algorithm-generated results were compared to several alternative methods, where statistical tests were employed to assess the significance of the outcomes (section 2.4.3). The list of employed metrics for each application is provided in the relevant chapter under the section “Evaluation Metrics”.

2.4.1 Object Detection Evaluation

Precision, recall, and accuracy were estimated and reported to evaluate the performance of developed algorithms for object detection using the following equations:

$$Precision = \frac{TP}{TP+FP} , \quad (2-4)$$

$$Recall = \frac{TP}{TP+FN} , \quad (2-5)$$

$$Specificity = \frac{TN}{TN+FP} , \quad (2-6)$$

and,

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} , \quad (2-7)$$

Where TP, TN, FP, and FN denote true positive, true negative, false-positive, and false negative, respectively.

To evaluate the performance of the detection algorithms, the ROC and precision-recall (PR) curves might be generated. ROC graphs the true-positive rate (sensitivity) against the false-positive rate ($1 - \text{specificity}$) at various thresholds on the decision function, while in PR curve, precision is plotted against the recall. PR curve is much like the ROC curve; however, T. Saito et al. have shown that the precision-recall plot is more informative than the ROC plot when evaluating binary classifiers on imbalanced datasets [102]. ROC plays an important role in the assessment of the diagnostic ability of a method to find the optimal cut off values, and also to compare two alternative diagnostic methods applied on the same test population [94-96]. The area under the ROC curve (AUC) could be computed as a measure of the performance of a method for object detection for all possible classification thresholds. ROC and AUC are recognized as effective measures of accuracy with meaningful interpretation [103-105]. The 95% confidence interval and p-value for AUC can be measured as well. To calculate the 95% confidence interval for AUC value, bootstrapping and randomization testing is used. The AUC computations are bootstrapped (by sampling predicted and true labels for each test image across the prostate with replacement) 10,000 times. The lower and upper bounds at the 95% level of the histogram of the bootstrapped AUC values yield the confidence intervals. Given 0.5 as the null hypothesis value of AUC, the p-value is a matter of counting the proportion of time AUC is less than or equal to 0.5.

2.4.2 Segmentation Evaluation

We used volume/area-based and region-based metrics to evaluate the accuracy of the developed segmentation algorithms. Volume-based metrics such as (e.g. absolute volume difference (AVD)) capture the aspect of the segmentation that is of greatest clinical interest. Region-based

metrics evaluate the spatial fidelity of the segmentations to the ground truth. Most of the metrics used in the following chapters are described below.

As volume-based metrics, we used AVD, and normalized absolute volume difference (NAVD).

$$AVD = |V_A - V_M|, \quad (2-8)$$

$$NAVD = \frac{|V_A - V_M|}{V_M} \times 100\%, \quad (2-9)$$

V_A and V_M are the volumes of the object measured from the algorithm-generated and expert manually segmentations, respectively. We also used Bland Altman plots to analyze the agreement between V_A and V_M . It graphs the difference between V_A and V_M versus their mean and constructs limits of agreement [106].

Dice similarity coefficient (DSC) or Dice index (DI), which is a region-based metric, computes the spatial overlap between the surface of the algorithm generated (A) and manual (M) segmentations [107].

$$DSC = \frac{2(A \cap M)}{A + M} \quad (2-10)$$

This metric varies in the range [0,1], where the minimum and maximum values correspond to non-overlap and perfect match between A and M.

Hausdorff distance (HD) is another metric that is used to measure the segmentation accuracy. HD captures the deviations in segmentation by computing the distance between two contours generated from algorithm ($A = \{a_i: i = 1, 2, \dots, N_A\}$) and manual ($M = \{m_j: j = 1, 2, \dots, N_M\}$) segmentation. To this end, for each point in A (a_i) the Euclidean distance from all points in M are computed ($d(a_i, M)$) and the shortest distance is considered as a perpendicular distance (p_i) from A to M. HD measures the maximum of this set of the minimum distances.

$$p_i = \min\{d(a_i, M)\}, M = \{m_j: j = 1, 2, \dots, N_M\} , \quad (2-11)$$

$$HD = \max\{p_i\}, i = 1, 2, \dots, N_A \quad (2-12)$$

2.4.3 Statistical Tests

We used the t-test to determine if the difference between the mean value of two sets of observations obtained by employing two different image analysis methods on the same dataset is significant. In this statistical test, the null hypothesis assumes that the mean difference is zero. Conversely, the alternative hypothesis assumes that the mean difference between the paired samples is not equal, greater, or smaller than zero for a two-tailed, right-tailed, and left-tailed t-test, respectively.

As mostly the normal distribution for the sample population cannot be assumed, Wilcoxon rank-sum test was applied, where the significant difference between median values of two sets of observations at a 5% level of confidence was investigated.

References

1. NIH, National Cancer Institute, NCI Dictionary of Cancer Terms. Accessed on: Feb. 23, 2020. [Online]. Available: <https://www.cancer.gov/publications/dictionaries/cancer-terms/def/abnormal>
2. U.S. Food and Drug Administration, Medical Imaging. Aug. 28, 2018. Accessed Feb. 23, 2020. [Online]. Available: <https://www.fda.gov/radiation-emitting-products/radiation-emitting-products-and-procedures/medical-imaging>.
3. Abnormality. (n.d.) Miller-Keane Encyclopedia and Dictionary of Medicine, Nursing, and Allied Health, Seventh Edition. (2003). Accessed on Feb. 23, 2020. [Online]. Available: <https://medical-dictionary.thefreedictionary.com/abnormality>
4. Contrast Materials. July 23, 2018. Accessed on: Feb. 16, 2020. [Online]. Available: <https://www.radiologyinfo.org/en/pdf/safety-contrast.pdf>

5. Y. D. Xiao, R. Paudel, J. Liu, C. Ma, Z. S. Zhang, and S. K. Zhou, "MRI contrast agents: Classification and application (Review)," *International Journal of Molecular Medicine*. 2016.
6. A. Murphy and U. Bashir, Gadolinium contrast agents. Accessed on: Feb. 25, 2020. [Online]. Available: <https://radiopaedia.org/articles/gadolinium-contrast-agents>
7. Nick Ferris, Gadolinium Contrast Medium (MRI Contrast agents). June 26, 2017, Accessed on: Feb. 25, 2020. [Online]. Available: <https://www.insideradiology.com.au/gadolinium-contrast-medium/>
8. T. Adejoh, O. C. Ewuzie, J. K. Ogbonna, S. O. Nwefuru, and N. C. Onuegbu, "A Derived Exposure Chart for Computed Radiography in a Negroid Population," *Health (Irvine. Calif.)*, 2016.
9. J. L. Prince and J. Links, *Medical Imaging Signals and Systems*. 2014.
10. Dhawan, Atam P. *Medical Image Analysis*. Hoboken, NJ: Wiley-Interscience. ISBN 978-0-471-45131-0, 2003.
11. G. T. Herman, *Fundamentals of computerized tomography: Image reconstruction from projection*, 2nd edition, Springer, 2009.
12. Merriam-Webster Online Dictionary, Computed tomography. Accessed on Feb 23, 2020. [Online]. Available: <https://www.merriam-webster.com/dictionary/computed%20tomography>
13. R. L. Galloway, "Introduction and Historical Perspectives on Image-Guided Surgery," in *Image-Guided Neurosurgery*, 2015.
14. F. R. Khan and J. M. Henderson, "Deep brain stimulation surgical techniques," in *Handbook of Clinical Neurology*, 2013.
15. J. Haynes, K. R. Arnold, C. Aguirre-Oskins, and S. Chandra, "Evaluation of neck masses in adults," *Am. Fam. Physician*, 2015.
16. M. Bin Saeedan, I. M. Aljohani, A. O. Khushaim, S. Q. Bukhari, and S. T. Elnaas, "Thyroid computed tomography imaging: pictorial review of variable pathologies," *Insights into Imaging*. 2016.
17. J. Moss, M. R. Dweck, J. G. Dreisbach, M. C. Williams, S. M. Mak, T. Cartlidge *et al.*, "Complementary role of cardiac CT in the assessment of aortic valve replacement dysfunction," *Open Hear.*, 2016.
18. Y. Weerakkody. Bronchial wall thickening. Accessed on Feb. 25, 2020. [Online]. Available: <https://radiopaedia.org/articles/bronchial-wall-thickening>
19. R. W. Brown, Y. C. N. Cheng, E. M. Haacke, M. R. Thompson, and R. Venkatesan, *Magnetic Resonance Imaging: Physical Principles and Sequence Design: Second Edition*. 2014.
20. D. Saloner, "An introduction to MR angiography," *Radiographics*, 1995.
21. R. Szeliski, "Computer vision: algorithms and applications," Springer, 2010.

22. Y. Amit, "2D Object Detection and Recognition: Models, Algorithms and Networks," MIT Press, Cambridge, MA, 2002.
23. N. Sharma and L. M. Aggarwal, "Automated medical image segmentation techniques," *Journal of Medical Physics*. 2010.
24. E. Ukwatta, "Vascular Segmentation Algorithms for Generating 3D Atherosclerotic Measurements", PhD Dissertation, Western University, On, Canada, 2013.
25. L. S. Davis, A. Rosenfeld, and J. S. Weszka, "Region Extraction by Averaging and Thresholding," *IEEE Trans. Syst. Man Cybern.*, 1975.
26. N. Otsu, "A threshold selection method from gray-level histograms," *IEEE Trans. Syst. Man Cybern.*, 1996.
27. C. Malarvizhi, "Segmentation by Thresholding on Medical Imaging – A Survey". *IJRASET*. 5 (IX), 2017.
28. H. F. Ng, "Automatic thresholding for defect detection," *Pattern Recognit. Lett.*, 2006.
29. B. G. Batchelor, "Edge-Region-Based Segmentation of Range Images," *IEEE Trans. Pattern Anal. Mach. Intell.*, 1994.
30. M. Neizel, M. Katoh, E. Schade, T. Rassaf, G. A. Krombach, M. Kelm, H. P. Kühl, "Rapid and accurate determination of relative infarct size in humans using contrast-enhanced magnetic resonance imaging," *Clin. Res. Cardiol.*, 2009.
31. Kolipaka, G. P. Chatzimavroudis, R. D. White, T. P. O'Donnell, and R. M. Setser, "Segmentation of non-viable myocardium in delayed enhancement magnetic resonance images," *Int. J. Cardiovasc. Imaging*, 2005.
32. Y. Lu, Y. Yang, K. A. Connelly, G. A. Wright, and P. E. Radau, "Automated quantification of myocardial infarction using graph cuts on contrast delayed enhanced magnetic resonance images.," *Quant. Imaging Med. Surg.*, 2012.
33. T. F. Chan and L. A. Vese, "Active contours without edges," *IEEE Trans. Image Process.*, 2001.
34. Y. Boykov, O. Veksler, and R. Zabih, "Fast approximate energy minimization via graph cuts," *IEEE Trans. Pattern Anal. Mach. Intell.*, 2001.
35. T. F. Chan and L. A. Vese, "Active contours without edges," *IEEE Trans. Image Process.*, 2001.
36. M. Kass, A. Witkin, and D. Terzopoulos, "Snakes: active contour models," *Int'l J. Comp. Vis.*, 1987.
37. Amini, T. E. Weymouth, and R. C. Jain, "Using Dynamic Programming for Solving Variational Problems in Vision," *IEEE Trans. Pattern Anal. Mach. Intell.*, 1990.
38. J. Yuan, E. Ukwatta, W. Qiu, M. Rajchl, Y. Sun, X. C. Tai, and A. Fenster, "Jointly segmenting prostate zones in 3D MRIs by globally optimized coupled level-sets," in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 2013.

39. E. Ukwatta, J. Awad, A. D. Ward, D. Buchanan, G. Parraga, and A. Fenster, "Coupled level set approach to segment carotid arteries from 3D ultrasound images," in *Proceedings - International Symposium on Biomedical Imaging*, 2011.
40. E. Ukwatta, J. Yuan, D. Buchanan, B. Chiu, J. Awad, W. Qiu, G. Parraga, and A. Fenster., "Three-dimensional segmentation of three-dimensional ultrasound carotid atherosclerosis using sparse field level sets," *Med. Phys.*, 2013.
41. E. Ukwatta , J. Awad , A. D. Ward, D. Buchanan, J. Samarabandu, G. Parraga, and A. Fenster., "Three-dimensional ultrasound of carotid atherosclerosis: Semiautomated segmentation using a level set-based method," *Med. Phys.*, 2011.
42. W. Qiu, J. Yuan, E. Ukwatta, D. Tessier, and A. Fenster, "Three-dimensional prostate segmentation using level set with shape constraint based on rotational slices for 3D end-firing TRUS guided biopsy," *Med. Phys.*, 2013.
43. E. Ukwatta, J. Awad, D. Buchanan, G. Parraga, and A. Fenster, "Three-dimensional semi-automated segmentation of carotid atherosclerosis from three-dimensional ultrasound images," in *Medical Imaging 2012: Computer-Aided Diagnosis*, 2012.
44. W. Qiu, J. Yuan, J. Kishimoto, Y. Chen, M. Rajchl, E. Ukwatta, et al., "Automatic 3D US brain ventricle segmentation in pre-term neonates using multi-phase geodesic level-sets with shape prior," in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 2015.
45. E. Ukwatta, J. Awad, A. D. Ward, J. Samarabandu, A. Krasinski, G. Parraga, and A. Fenster., "Segmentation of the lumen and media-adventitia boundaries of the common carotid artery from 3D ultrasound images," in *Medical Imaging 2011: Computer-Aided Diagnosis*, 2011.
46. W. Qiu, J. Yuan, E. Ukwatta, D. Tessier, and A. Fenster, "Three-dimensional prostate segmentation using level set with shape constraint based on rotational slices for 3D end-firing TRUS guided biopsy," *Med. Phys.*, 2013.
47. Y. Y. Boykov and M. P. Jolly, "Interactive graph cuts for optimal boundary & region segmentation of objects in N-D images," in *Proceedings of the IEEE International Conference on Computer Vision*, 2001.
48. J. Yuan, E. Ukwatta, X.-C. Tai, A. Fenster, and C. Schnoerr, "A fast global optimization-based approach to evolving contours with generic shape prior," *UCLA, Technical report CAM-12-38*, 2012.
49. J. Yuan, E. Bae, and X. Tai, "A study on continuous max-flow and min-cut approaches," *CVPR*, 2010.
50. J. Yuan, E. Ukwatta, X.C. Tai, A. Fenster, C. Schnorr, "A Fast Global Optimization-Based Approach to Evolving Contours with Generic Shape Prior," in *IEEE TPAMI*, 2012.
51. E. Ukwatta, J. Yuan, W. Qiu, M. Rajchl, and A. Fenster, "Efficient convex optimization-based curvature dependent contour evolution approach for medical image segmentation," in *Medical Imaging 2013: Image Processing*, 2013.

52. E. Ukwatta, J. Yuan, W. Qiu, M. Rajchl, B. Chiu, and A. Fenster, "Joint segmentation of lumen and outer wall from femoral artery MR images: Towards 3D imaging measurements of peripheral arterial disease," *Med. Image Anal.*, 2015.
53. E. Ukwatta, J. Yuan, M. Rajchl, W. Qiu, D. Tessier, and A. Fenster, "3-D carotid multi-region MRI segmentation by globally optimal evolution of coupled surfaces," *IEEE Trans. Med. Imaging*, 2013.
54. R. Zhou, F. Guo, M. R. Azarpazhooh, J. D. Spence, J. D. Spence, E. Ukwatta, M. Ding, and A. Fenster, "A Voxel-based Fully Convolution Network and Continuous Max-flow for Carotid Vessel-wall-volume Segmentation from 3D Ultrasound Images," *IEEE Trans. Med. Imaging*, 2020.
55. E. Ukwatta, J. Yuan, M. Rajchl, and A. Fenster, "Efficient global optimization based 3D carotid AB-LIB MRI segmentation by simultaneously evolving coupled surfaces," in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 2012.
56. W. Qiu, J. Yuan, E. Ukwatta, and A. Fenster, "Rotationally resliced 3D prostate TRUS segmentation using convex optimization with shape priors," *Med. Phys.*, 2015.
57. W. Qiu, M. Rajchl, F. Guo, Y. Sun, E. Ukwatta, A. Fenster, and J. Yuan, "3D prostate TRUS segmentation using globally optimized volume-preserving prior," in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 2014.
58. W. Qiu, J. Yuan, E. Ukwatta, Y. Sun, M. Rajchl, and A. Fenster, "Prostate segmentation: An efficient convex optimization approach with axial symmetry using 3-D tRUS and MR images," *IEEE Trans. Med. Imaging*, 2014.
59. W. Qiu, J. Yuan, E. Ukwatta, Y. Sun, M. Rajchl, and A. Fenster, "Fast globally optimal segmentation of 3D prostate MRI with axial symmetry prior," in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 2013.
60. W. Qiu, J. Yuan, E. Ukwatta, Y. Sun, M. Rajchl, and A. Fenster, "Efficient 3D multi-region prostate MRI segmentation using dual optimization," in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 2013.
61. W. Qiu, Y. Chen, J. Kishimoto, S. de Ribaupierre, B. Chiu, A. Fenster, B. K. Menon, J. E. Yuan, "Longitudinal Analysis of Pre-Term Neonatal Cerebral Ventricles from 3D Ultrasound Images Using Spatial-Temporal Deformable Registration," *IEEE Trans. Med. Imaging*, 2017.
62. C. Sun, A. Shrivastava, S. Singh, and A. Gupta, "Revisiting Unreasonable Effectiveness of Data in Deep Learning Era," in *Proceedings of the IEEE International Conference on Computer Vision*, 2017.

63. G. S. Fu, Y. Levin-Schwartz, Q. H. Lin, and D. Zhang, "Machine Learning for Medical Imaging," *Journal of Healthcare Engineering*. 2019.
64. R. M. Gray and D. L. Neuhoff, "Quantization," *IEEE Trans. Inf. Theory*, 1998.
65. Support Vector Machines. Accessed on Feb. 7, 2020. [Online]. Available: <https://scikit-learn.org/stable/modules/svm.html>
66. N. S. Altman, "An introduction to kernel and nearest-neighbor nonparametric regression," *Am. Stat.*, 1992.
67. T. K. Ho, "Random Decision Forests Tin Kam Ho Perceptron training," *Proc. 3rd Int. Conf. Doc. Anal. Recognit.*, 1995.
68. T. K. Ho, "A data complexity analysis of comparative advantages of decision forest constructors," *Pattern Anal. Appl.*, 2002.
69. J. J. Hopfield, "Neural networks and physical systems with emergent collective computational abilities.," *Proc. Natl. Acad. Sci. U. S. A.*, 1982.
70. H. J. KELLEY, "Gradient Theory of Optimal Flight Paths," *ARS J.*, 1960.
71. Goodfellow I, Bengio Y, Courville A. "Convolutional Networks" in *Deep Learning*, MIT Press, 2016.
72. D. Shen, H. Greenspan, and S. K. Zhou, "deep learning for medical image analysis," Academic Press, 2017.
73. G. Hinton, "Deep learning-a technology with the potential to transform health care," *JAMA - Journal of the American Medical Association*. 2018.
74. L. Carin and M. J. Pencina, "On deep learning for medical image analysis," *JAMA - Journal of the American Medical Association*. 2018.
75. G. Litjens, T. Kooi, B. Ehteshami Bejnordi, A. A. A. Setio, F. Ciompi, M. Ghafoorian *et al.*, "A survey on deep learning in medical image analysis," *Medical Image Analysis*. 2017.
76. P. F. Christ, M. E. A. Elshaer, F. Ettliger, S. Tatavarty, M. Bickel, P. Bilic *et al.*, "Automatic liver and lesion segmentation in CT using cascaded fully convolutional neural networks and 3D conditional random fields," in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 2016.
77. L. Zhao and K. Jia, "Multiscale CNNs for Brain Tumor Segmentation and Diagnosis," *Comput. Math. Methods Med.*, 2016.
78. F. Zabihollahy, J. A. White, and E. Ukwatta, "Fully automated segmentation of left ventricular myocardium from 3D late gadolinium enhancement magnetic resonance images using a U-net convolutional neural network-based model," *SPIE Medical Imaging* 2019.
79. C. Wang, E. Ukwatta, M. Rajchl, and A. Chan, "An ensemble of U-Net architecture variants for left atrial segmentation," *SPIE Medical Imaging, San Diego, California, United States*, 2019.

80. F. Zabihollahy, C. McKeen, J. Kurian, A. D. C. Chan, D. E. Demellawy, and E. Ukwatta, "Machine learning-based approach for fully automated segmentation of muscularis propria from histopathology images of intestinal specimens," *SPIE Medical Imaging, San Diego, California, United States*, 2019.
81. H. Pratt, F. Coenen, D. M. Broadbent, S. P. Harding, and Y. Zheng, "Convolutional Neural Networks for Diabetic Retinopathy," in *Procedia Computer Science*, 2016.
82. F. Zabihollahy, A. Lochbihler, and E. Ukwatta. "Deep Learning Based Approach for Fully Automated Segmentation of Hard Exudate from Retinal Images," *SPIE Medical Imaging*, 2019.
83. F. Zabihollahy, and E. Ukwatta. "Fully-Automated Segmentation of Optic Disk from Retinal Images Using Deep Learning Techniques," *SPIE Medical Imaging*, 2019.
84. F. Zabihollahy, E. Ukwatta, S. Krishna, and N. Schieda, "Fully automated localization of prostate peripheral zone tumors on apparent diffusion coefficient map MR images using an ensemble learning method," *J. Magn. Reson. Imaging*, 2019.
85. Z. Han, B. Wei, Y. Zheng, Y. Yin, K. Li, and S. Li, "Breast Cancer Multi-classification from Histopathological Images with Structured Deep Learning Model," *Sci. Rep.*, 2017.
86. Zhou SK, Greenspan H, Shen D, Deep Learning for Medical Image Analysis. Academic Press, 2017.
87. "CS231n: Convolutional Neural Networks for Visual Recognition," Accessed on: May 20, 2019. [Online]. Available: <http://cs231n.github.io/convolutional-networks/>
88. Nikulin, M.S., "Efficiency of a statistical procedure," in Hazewinkel, Michiel, *Encyclopedia of Mathematics*, Springer, ISBN 978-1-55608-010-4, 2001.
89. Y. Lecun, Y. Bengio Y, and G. Hinton; "Deep learning," *Nature*. 2015.
90. O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 2015.
91. P. Bühlmann and S. van de Geer, "Statistics for high-dimensional data: Methods, Theory and Applications," *Journal of the Japan Statistical Society*, 2014.
92. P. Y. Simard, D. Steinkraus, and J. C. Platt, "Best practices for convolutional neural networks applied to visual document analysis," in *Proceedings of the International Conference on Document Analysis and Recognition, ICDAR*, 2003.
93. N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, "Dropout: A simple way to prevent neural networks from overfitting," *J. Mach. Learn. Res.*, 2014.
94. Jason Brownlee, "Difference Between a Batch and an Epoch in a Neural Network," Oct. 26, 2019. [Online]. Available: <https://machinelearningmastery.com/difference-between-a-batch-and-an-epoch/> [Accessed on: March 5,2020.].
95. T.V. Laarhoven, "L2 Regularization versus Batch and Weight Normalization." ArXiv e-prints, Accessed on: 2017 Available: ui.adsabs.harvard.edu.

96. S. Santurkar, D. Tsipras, A. Ilyas, and A. Madry, "How Does Batch Normalization Help Optimization? (No, It Is Not About Internal Covariate Shift)," *Adv. Neural Inf. Process. Syst.* 31, 2018.
97. D. Opitz and R. Maclin, "Popular Ensemble Methods: An Empirical Study," *J. Artif. Intell. Res.*, 1999.
98. R. Polikar, "Ensemble based systems in decision making," *IEEE Circuits and Systems Magazine*. 2006.
99. L. Rokach, "Ensemble-based classifiers," *Artif. Intell. Rev.*, 2010.
100. R. A. Jacobs, M. I. Jordan, S. J. Nowlan, and G. E. Hinton, "Adaptive Mixtures of Local Experts," *Neural Comput.*, 2008.
101. E. Ukwatta, J. Awad, A. D. Ward, D. Buchanan, J. Samarabandu, G. Parraga, and A. Fenster, "Three-dimensional ultrasound of carotid atherosclerosis: Semiautomated segmentation using a level set-based method," *Med. Phys.*, 2011.
102. T. Saito and M. Rehmsmeier, "The precision-recall plot is more informative than the ROC plot when evaluating binary classifiers on imbalanced datasets," *PLoS One*, 2015.
103. J. A. Hanley and B. J. McNeil, "The meaning and use of the area under a receiver operating characteristic (ROC) curve," *Radiology*, 1982.
104. J. A. Hanley and B. J. McNeil, "A method of comparing the areas under receiver operating characteristic curves derived from the same cases," *Radiology*, 1983.
105. R. Kumar and A. Indrayan, "Receiver operating characteristic (ROC) curve for medical researchers," *Indian Pediatrics*. 2011.
106. D. Giavarina, "Understanding Bland Altman analysis," *Biochem. Medica*, 2015.
107. P. Zijdenbos, B. M. Dawant, R. A. Margolin, and A. C. Palmer, "Morphometric Analysis of White Matter Lesions in MR Images: Method and Validation," *IEEE Trans. Med. Imaging*, 1994.

Chapter 3: Patch-Based Convolutional Neural Network for Differentiation of Cyst from Solid Renal Mass on Contrast-Enhanced Computed Tomography Images

In this chapter, we describe the development and evaluation of a patch-based CNN algorithm for automated classification of renal masses into benign cyst and solid using a local dataset of CECT images of 315 patients suffering from kidney cancer [1].

3.1 Introduction

The kidneys are a pair of bean-shaped organs on either side of the spine that play an important role in many aspects of human physiology including maintenance of the body's homeostasis through filtering the blood, removing wastes, controlling the body's fluid balance, and maintaining electrolyte balances [2]. Among other key functions, the kidneys maintain blood pressure, play an important role in erythropoiesis (i.e. red blood cell creation) and help regulate bone mineral density [3]. Medical renal diseases are diseases of the kidney which attack the filtering units of the kidneys (nephrons) and damage their ability to eliminate wastes and excess fluids. Chronic kidney disease (CKD) which describes as the presence of kidney damage or a decreased level of kidney function for a period of three months or more, can range from mild to severe and in some cases, lead to kidney failure [4]. Research shows that CKD has a massive cost to healthcare and human life and imposes a profound economic burden on the US population [5].

Aside from medical renal disease, kidney cancer, is another common, costly and potentially lethal medical condition of the kidneys. According to the American cancer society, kidney cancer (kidney tumor/renal mass), which is an abnormal growth in the kidney is among the 10

most common cancers in both men and women and overall, the lifetime risk for developing kidney cancer in men and women is about 1 in 48 and 1 in 83, respectively [6]. Most kidney tumors are actually benign. For example, a kidney tumor might be a simple renal cyst that does not require treatment or follow up [7]. Kidney cysts are ubiquitous in the human population but may pose problems in diagnostic medical imaging when they are detected incidentally because they must be characterized as benign and not assumed to be malignant. Even in solid renal neoplasms, approximately 20% of solid tumors are benign (most commonly oncocytomas, and fat poor renal angiomyolipoma (fpAML)). Among the malignant kidney tumors, RCC is the most common and this tumor is typically divided into three subtypes: clear cell, papillary, chromophobe, with clear cell being the most common and most aggressive [6]. Kidney cancer diagnosis is an exceptionally difficult task with only modest accuracy of current techniques and variable inter-observer agreement using subjective assessment. Automated evaluation of renal masses is highly desirable as it might improve the diagnosis accuracy. In this chapter, we focus on the differentiation of cyst from solid renal masses. Classification of renal masses into benign and RCC will be discussed in the next chapter.

A benign cyst can be diagnosed at CT when it is homogeneous, well-circumscribed and measures water attenuation (<20 Hounsfield Units [HU]). When all of these three criteria are met, a cyst can be diagnosed with a high degree of accuracy [8]. In many instances, a cyst cannot be diagnosed at time of initial CT which may occur due to: internal complexity from protein or hemorrhage or from pseudo-enhancement (which is the artificial increase in attenuation of a cyst due to beam hardening artifact from adjacent iodine in the kidney on single-phase enhanced CT scans) [9]. Indeterminate renal lesions that may be cysts require further imaging with dedicated multi-phase CT or MRI or biopsy for definitive diagnosis. The further workup of potentially

benign incidental imaging findings is costly and increases patient morbidity. Automated differentiation of a benign cyst from solid mass is clinically desirable, as discussed above, and, is an essential step for automated evaluation of solid renal masses for eventual classification into benign neoplasms and RCC that to the best knowledge of the author, has not been evaluated so far. In this chapter, we describe the procedure of the design and evaluation of a patch-based CNN model for the differentiation of benign renal cyst from solid renal mass (Figure 3-1). In an image-based approach the network is trained and tested using the whole image whereas, in a patch-based approach, small images are extracted around each pixel in the image and used for CNN training and testing. We evaluated the developed algorithm on the largest cohort of renal masses at CECT studied to date for automated diagnosis that was acquired from four CT scanners that assure us of technique robustness and generalizability of results. Moreover, we computed the inter-observer variability of the algorithm. Additionally, we compared the results of the CNN to those of thresholding based on CT number for benign cyst diagnosis.

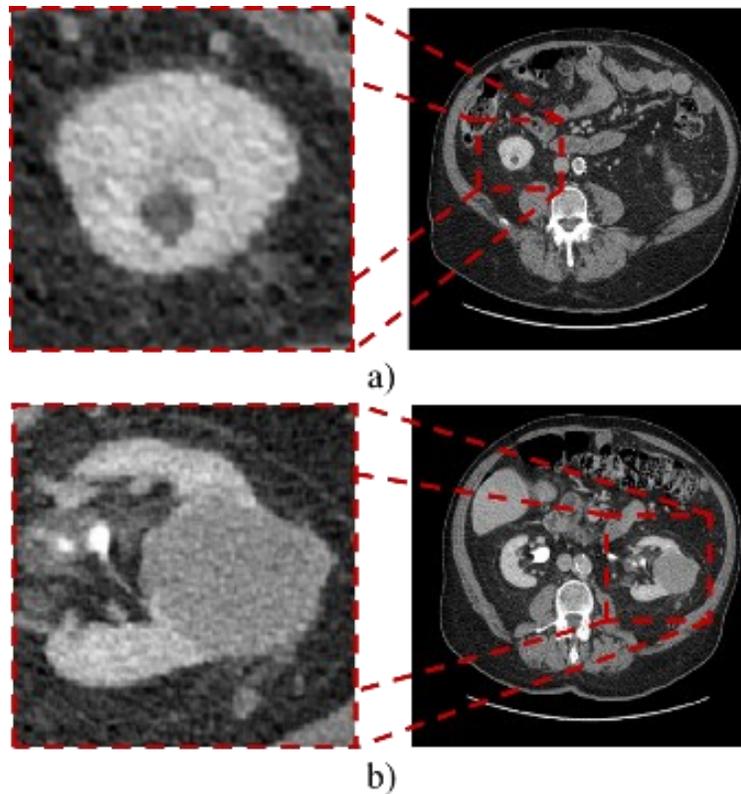


Figure 3-1. An example of a) benign renal cyst b) solid renal mass on CECT images. The 2D CT images are visualized in the standard abdominal window and level set of 400 and 50, respectively. Reprinted with permission.

3.2 Method

3.2.1 Patient

This retrospective study was approved by the Ottawa hospital research ethics board. The informed consent was waived by the institutional review board (IRB) and anonymized data was shared through a data-sharing agreement between institutions. This study is a health insurance portability and accountability act (HIPAA) compliant. We have included consecutive patients who underwent partial or total nephrectomy for a renal mass with pre-operative CT and without intervening chemo- or radiotherapy between January 2015 and December 2017. All RCC,

oncocytomas, and fpAMLs were confirmed histologically by fellowship-trained genitourinary pathologists using the World Health Organization standards for analysis of cystic and solid [10]. Our database consisted of 315 patients with solid renal mass (123 clear cell RCC, 69 papillary RCC, 46 chromophobe RCC, 57 oncocytomas and 20 fpAMLs). The mean renal mass volume was $10.39 \pm 24.19 \text{ cm}^3$.

Patients with solid renal mass also commonly demonstrated incidental synchronous benign simple or minimally complex renal cysts, which as discussed, are a ubiquitous imaging finding at CT [11]. We included simple and minimally complex (Bosniak 1 or 2) cysts in our study when they measured ≥ 10 mm in size and fulfilled diagnostic imaging criteria enabling accurate diagnosis of a Bosniak 1 or 2 cysts. Renal hypodensities measuring < 10 mm are considered too small to accurately characterize with CT and almost always benign cysts. Therefore, they are not typically of clinical concern [12] and were not included in the present study. To be included in the study, the diagnosis of a Bosniak Type 1 or 2 was established using the following stringent imaging criteria where an incidental cyst must fulfill one of the following imaging criteria: 1) Ultrasound; a homogeneously anechoic round or oval lesion with increased through transmission and a thin imperceptible wall +/- a few thin hairlike septae, 2) CT; a homogeneously hypoattenuating lesion with smooth imperceptible wall measuring < 20 Hounsfield Units [HU] or a homogeneously hyperattenuating lesion measuring > 70 HU at non-enhanced CT or a homogeneously hypoattenuating lesion with smooth imperceptible wall measuring < 20 HU and showing few hairline thin septae with or without calcification, 3) MRI; a homogeneously hyperintense T2-weighted (T2W) lesion with signal intensity (SI) similar to cerebrospinal fluid +/- a few thin hairline septae or a homogeneously markedly increased T1-weighted (T1W) lesion with SI 2.5x greater than the adjacent ipsilateral renal cortical parenchyma and showing no

enhancement at subtraction imaging [13]. In our cohort of 315 patients with solid RM, 77 simple or minimally complex benign Bosniak Type 1 or 2 cysts were identified.

3.2.2 CT Technique

CT examinations were performed using a variety of multi-detector systems at Ottawa Hospital and also at referral institutions within our Local Health Integration Network (LHIN) with all images available for review in our Picture Archiving and Communication System (PACS, (Horizon Medical Imaging v10.3, McKesson Corporation). The institutional renal mass protocol consists of an unenhanced CT acquisition (slice thickness 2.5-5 mm) followed by a corticomedullary phase acquisition (slice thickness 2.5-5 mm) and a delayed nephrographic phase acquisition (slice thickness 2.5-5 mm). A fixed 120-kVp technique is used for all three phases with automated tube current modulation and a variable 100-500-mAs (matched for all three phases of the study). The renal mass CT protocol utilizes a fixed 120-kVp technique, which provides the ability to compare mean attenuation numbers across machines [13]. Although automated tube current modulation is performed, the variable tube current affects the standard deviation (noise) of the attenuation measurement but not the mean attenuation within a measured region [14]. Acquisition is helical with a rotation time of 0.8 sec and a Noise Index of 41.4. The timing of the corticomedullary phase is established by bolus tracking (timing of approximately 30 seconds after injection). The nephrographic phase is acquired 80 seconds after the corticomedullary acquisition (approximately 110-120 seconds). Patients are given 105 mL of non-ionic contrast material (Iohexol [Omnipaque], GE Healthcare, prior to 2008; or Iopamidol [Isovue], Bracco Healthcare, after 2008) at a fixed intravenous rate of 3.5 mL/second using a power injector followed by a saline flush. This study evaluated only the nephrographic phase

enhanced images, where the kidneys enhance maximally, and renal masses are most commonly detected [9]. We also included CT examinations with similar technical acquisition parameters to the described CT renal mass protocol but where phase of enhancement differed slightly from the dedicated renal mass protocol (for example, using the late portal-venous phase (early nephrographic phase) defined as ~70 seconds post-injection and also the late nephrographic phase (early urographic phase) defined as ~3-5 minutes post-injection). In all included CT examinations, a fixed 120-kVp technique was used. The images used in this study were taken using four CT machines including GE Medical Systems (Lightspeed 64 and Discovery 750 HD), Toshiba (Aquilion 64 and 320), Siemens (Sensation 64), and Philips (Brilliance 64).

3.2.3 Manual Segmentation

For each patient, anonymized images in digital imaging and communication in medicine (DICOM) format were exported for axial CECT from our PACS. Images were evaluated by a fellowship-trained genitourinary radiologist with 13 years of experience in renal CT working in conjunction with a research assistant experienced in analysis and segmentation of CT and MRI images using ITK-SNAP v3.2 (Philadelphia, PA). For each patient, the radiologist and research assistant identified each solid renal mass and correlated the location on CT images to the surgical and pathology reports. The radiologist and research assistant then defined the boundaries of the kidneys (excluding the retroperitoneal fat and hilar structures such as the vessels and the collecting systems) and then manually segmented the kidney, including the renal cortex and medulla. For solid renal masses which extended beyond the outer margin of the renal cortex (into the retroperitoneal fat) or inner margin of the renal cortex (into the renal hilum) the entire mass was included in the segmentation. For solid renal masses showing extension of tumor into the

renal veins or collecting system, only the mass proper was segmented. Before completing the segmentation, the radiologist reviewed all the images for other renal cystic or solid masses in the kidneys present on CT images. When additional solid renal masses were identified without histological confirmation, the patient was excluded. When incidental cystic renal masses were encountered that measured ≥ 10 mm with adequate reference standard to confirm benign Bosniak 1 or 2 simple cyst, described above, the cysts were also segmented. When a cystic renal mass was encountered measuring ≤ 10 mm in size it was ignored and for cystic renal mass ≥ 10 mm without adequate reference standard for diagnosis, the patients were excluded.

3.2.4 Patch Extraction and Balancing Dataset

For both training and testing phases, renal masses were manually delineated from CECT scans as explained in the previous section. A local image (patch) of size 40×40 pixels around voxels in the renal mass tissue was then derived in the transversal direction with the stride of 5 pixels and the density values were normalized into the range $[0,1]$ across each patch (Figure 3-2). To label patches as a cyst or solid, the corresponding patch extracted from manual segmentation was used as a reference. As the number of cysts was less than the number of solid renal masses in our dataset, we had fewer patches labeled as cyst than those assigned to the solid renal mass class that caused a class imbalance in the training set. To address this problem, the elastic transformation was applied to the instances from underrepresented class to expand the training dataset. The method described by Simard et al. [15] was followed to perform elastic deformation.

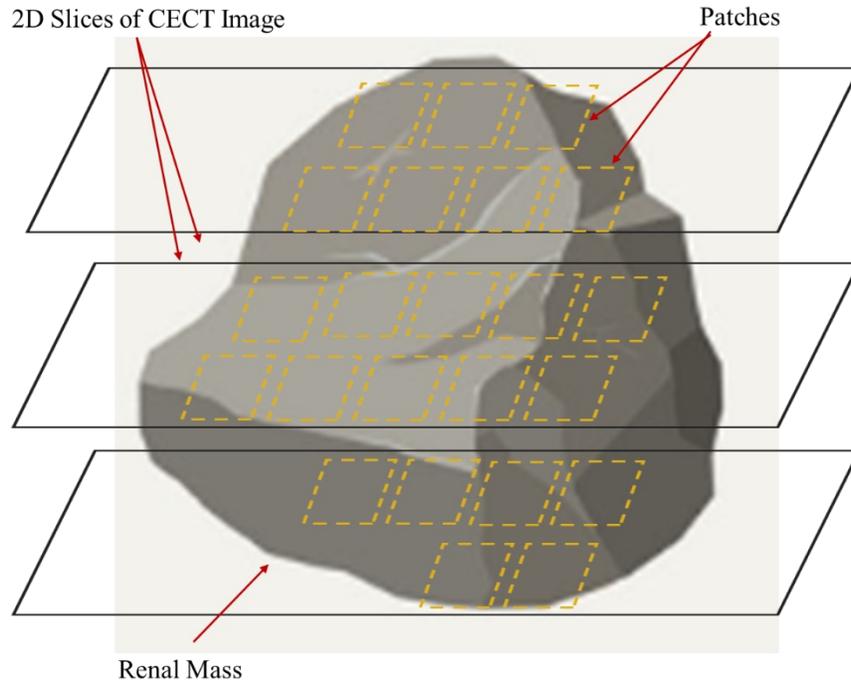


Figure 3-2. A diagram of our proposed method for patch extraction from renal mass tissue. Reprinted with permission.

3.2.5 Network Architecture

The goal of this study was to develop an automated method for the classification of renal masses on CECT images. In practice, radiologists differentiate renal mass based on the difference of texture, and intensity that exists between benign cyst and solid renal masses. A simple cyst on CT is completely homogeneous and measure < 20 HU in density [8]. We chose CNN architecture for this application as it helps to learn different features including shape, texture, intensity. enabling the model to distinguish between benign cyst and solid renal masses. Small patches were extracted and classified using a CNN. A voting system was then applied for the final prediction that allows for a more confident diagnosis of renal mass type. Despite similar works on kidney tumor classification where the middle slice with the largest surface area of the

tumor is considered for classification, in the proposed algorithm, patches are extracted around each voxel in the renal mass that benefits from the volumetric data.

In designing the CNN model, the number of layers (including convolutional, maximum pooling, and Dropout), number of filters, number of layers and nodes in the fully connected layer, number of training images, patch size, optimizer, and loss functions were varied. For each set of parameters, a separate network was trained, which was then tested on a subset of unseen images to compare the results and the CNN model that yields the highest classification accuracy was chosen.

Table 3-1 displays the configuration of the CNN used in our experiments to distinguish a cyst from solid renal masses. All parameters of the network were adjusted experimentally aimed at raising the network performance. The developed CNN is made up of 6 convolutional (Conv.) layers of size 3×3 with 16, 16, 32, 32, 64, and 64 filters in each layer. Rectified Linear Unit (ReLU) and batch normalization were applied after each Conv. layer. After every two Conv. layers, one maximum pooling (Max. Pool) layer of size 2×2 was employed that calculates the maximum of 4 output nodes. A conventional neural network with two hidden layers of size 64 neurons was included to the end of our proposed network in which three dropout layers were added before hidden and output layers. Flatten layer vectorizes learned features from the previous layers to be presented to the conventional neural network. We used Softmax as an activation function at the last layer of our suggested CNN to map the non-normalized output of a network to a probability distribution over predicted output classes. Categorical cross-entropy and Adadelta were used as loss function and optimizer to train the CNN, respectively.

Table 3-1. Configuration of the proposed convolutional neural network for classification of extracted patches into a cyst and solid renal masses. Reprinted with permission.

Layer	Type	Number of Filters	Fully Connected Unit	Layer	Type	Number of Filters	Fully Connected Unit
1	Convolutional 2D	16	-	17	Batch Normalization	-	-
2	Activation (Relu)	-	-	18	Convolutional 2D	64	-
3	Batch Normalization	-	-	19	Activation (Relu)	-	-
4	Convolutional 2D	16	-	20	Batch Normalization	-	-
5	Activation (Relu)	-	-	21	Maximum Pooling	-	-
6	Batch Normalization	-	-	22	Flatten	-	-
7	Maximum Pooling	-	-	23	Dropout	-	-
8	Convolutional 2D	32	-	24	Fully Connected	-	64
9	Activation (Relu)	-	-	25	Activation (Relu)	-	-
10	Batch Normalization	-	-	26	Dropout	-	-
11	Convolutional 2D	32	-	27	Fully Connected	-	64
12	Activation (Relu)	-	-	28	Activation (Relu)	-	-
13	Batch Normalization	-	-	29	Dropout	-	-
14	Maximum Pooling	-	-	30	Fully Connected	-	2
15	Convolutional 2D	64	-	31	Activation (Softmax)	-	-
16	Activation (Relu)	-	-				

3.2.6 Network Training and Testing

Our dataset comprised of 315 CECT images of size $512 \times 512 \times N$ ($N: [31,555]$). The images were randomly divided into the training ($N=40$), and testing ($N=275$) sets. Table 3-2 shows the number of CECT images used in our experiments broken by the CT systems for training and testing phases, separately. As seen, a combination of data from multiple CT scanners was used for training and testing. This strategy may be limited because it does not account for small differences between CT scanner types; however, using the same technical parameters across systems not only allowed the network to learn the signal characteristics of all scanners during training but also assured us that the results could be more generalizable and reproducible between systems and centers [16].

Table 3-2. The number and variety of CT systems used in our experiments for training and testing phases.

Reprinted with permission.

CT Scan System	GE Medical Systems	Toshiba	Siemens	Philips	Total
Train	18	17	1	4	40
Test	159	93	3	20	275
Total	177	110	4	24	315

As we used the patch-based approach for this study, the number of patches derived from 40 images was enough for training the suggested network. Figure 3-3 shows an overview of our proposed method for renal masses assessment. The network was trained using patches that extracted from training CECT scans and artificially augmented that results in a total of 65287 patches (32837 and 32450 patches labeled as cyst and solid renal mass respectively) for training. We implemented the algorithm in Python using Keras library built on top of Tensorflow. The

network was trained for 250 epochs on Intel Core i7, 2.8 GHz on a graphics processing unit (GPU)-accelerated computing platform, which took approximately 110 minutes. In the testing phase, the extracted patches around each voxel from manually segmented renal mass tissue were passed to the trained CNN to be labeled as benign cyst or solid renal mass. For each renal mass, if the number of predicted patches with the label of benign cyst was more than that of solid, the renal mass was categorized as cyst and vice versa. For cases with an equal number of patches labeled with benign cyst and solid renal masses, the patch-extraction was repeated with the different stride value (3 or 4 pixels) to assure of existing unequal number of cyst and solid labels as a prerequisite for applying majority voting rule.

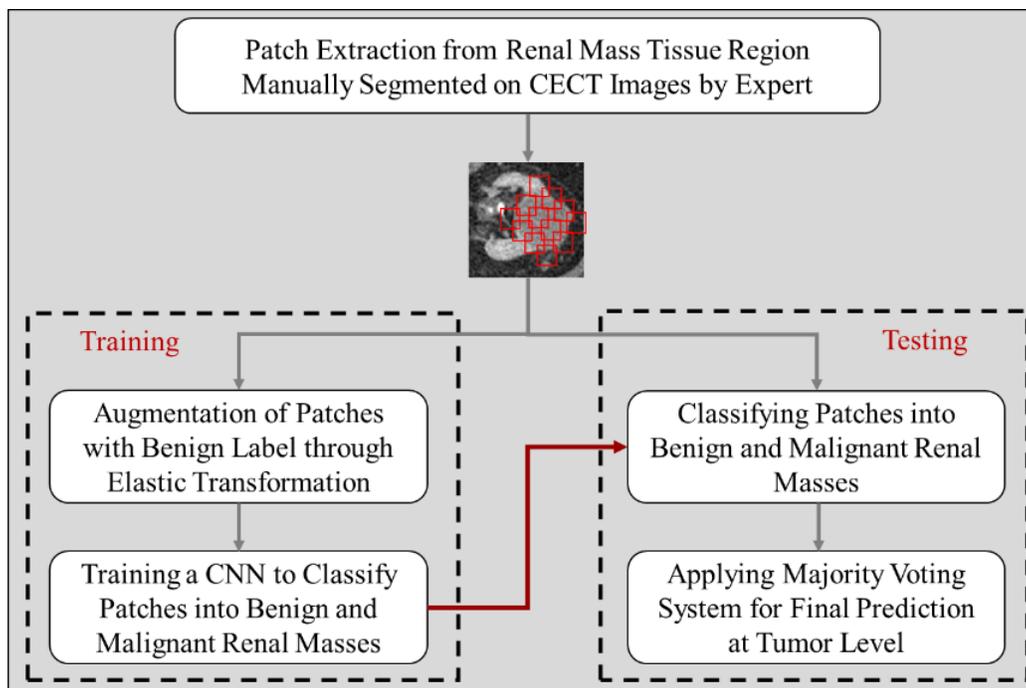


Figure 3-3. A flow diagram of our proposed methodology, where the left and right block diagrams display training and testing procedures. Reprinted with permission.

3.2.7 Evaluation Metrics

We used accuracy, precision, and recall to assess the performance of our algorithm for renal mass classification (i.e., benign cyst vs. solid renal mass). We also computed the accuracy of renal mass classification on a per-patient basis. To this end for patients with multiple tumors, the mean accuracy was computed for each patient and the average accuracy across the whole test dataset was reported. Moreover, the ROC curve and AUC were reported. We measured 95% confidence interval and p-value for AUC.

We used the coefficient-of-variation (CV) [17] in accuracy of the algorithm in classifying renal masses into benign cyst and solid renal masses on a per-patient basis to evaluate the variability among users.

To diagnose a simple cyst on CT, the lesion should be completely homogeneous and measure < 20 HU in density [8]. A cyst measuring ≥ 20 HU is technically indeterminate on CT and would require further characterization typically with a follow-up CT or MRI performed with and without intravenous contrast (IV) contrast to assess for the presence of enhancement. Follow-up imaging results in unnecessary health care expenditure and increases patient morbidity. Since many cysts will spuriously measure above 20 HU at single-phase CECT (due to pseudo-enhancement) higher attenuation thresholds have been proposed to improve accuracy including 30 HU [18] and more recently 40 HU [19]. A limitation of using higher attenuation thresholds to diagnose simple cysts is the potential false-positive results and misclassification of solid masses which show low levels of enhancement and measure < 40 HU. We calculated the number of cysts and solid masses with average HU ≥ 20 HU, ≥ 30 HU, and ≥ 40 HU to determine what proportion of cysts and masses would be accurately classified using only attenuation measurements. Since

attenuation measurements are performed in clinical practice by manual placement of a circular ROI in the center of a lesion encompassing approximately 2/3 of its surface area and avoiding the edges to prevent inclusion of adjacent structures, we trimmed the boundary of the cysts and tumors for three iterations using the erosion operation in python to sample the center of tumor for density computation. The HU was then estimated at the tumor level. The ROC for HU was also plotted, where the diagnosis of a benign cyst was true positive when $HU < 20$, and false negative when $HU \geq 20$. The diagnosis of solid mass was considered as false-positive for $HU < 20$, and true negative otherwise.

3.3 Results

Our proposed method for differentiating benign cyst from solid renal masses yielded mean accuracy, precision, and recall of 88.96% (291/327), 95.64%, and 91.64% on 275 CECT test images including 327 renal masses (39 benign cysts and 288 solid renal masses) on a per-renal mass basis. The confidence interval for accuracy, precision, and recall at a 95% level was (0.859 - 0.917), (0.935 - 0.975), and (0.888 - 0.942) respectively. Our methodology reported accuracy of $91.21\% \pm 25.88\%$ as mean \pm standard deviation at the patient level. The average running time required to classify a manually segmented renal masses from a typical CECT test image on Intel Core i7, 2.8 GHz using a graphics processing unit (GPU)-accelerated computing platform was 14.46 ± 12.30 sec. Figure 3-4 shows the confusion matrix for the developed method where the number of misclassified cases are shown for each type of renal masses. The average of HU for misclassified cystic renal masses using our developed method was 51.

N = 327	Predicted Cyst	Predicted Solid
Cyst	27	12
Solid	24	264

Figure 3-4. Confusion matrix shows the performance of our proposed method for classification of cyst and solid renal masses on CECT images. Reprinted with permission.

The ROC curve was graphed for our proposed methodology for diagnosis of benign cyst versus solid renal masses as shown in Figure 3-5. The AUC was estimated as 0.804.

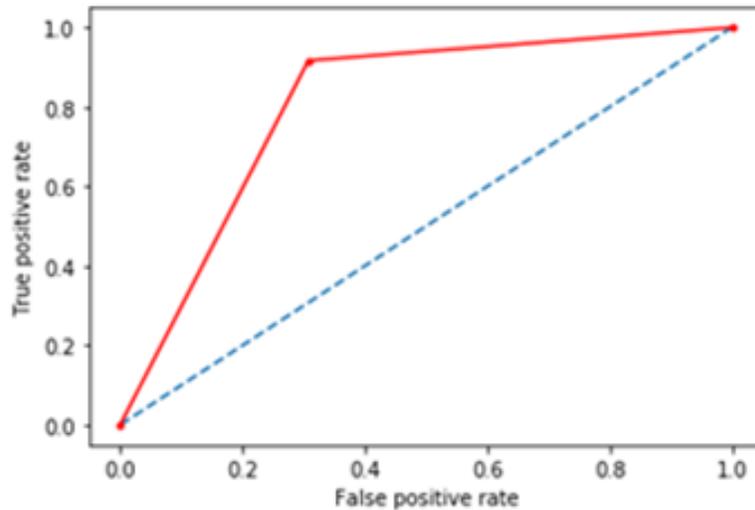


Figure 3-5. ROC curve displays the performance of our designed patch-based CNN model for diagnosis of cyst versus solid renal masses, where the true positive rate (sensitivity) against false-positive rate (1 - specificity) has been plotted. Reprinted with permission.

Figure 3-6 displays the histogram of the ROC AUC values obtained from 10,000 bootstrapping of the AUC calculation. The confidence intervals at the 95% level was (0.739 - 0.866) with p-value < 0.0001.

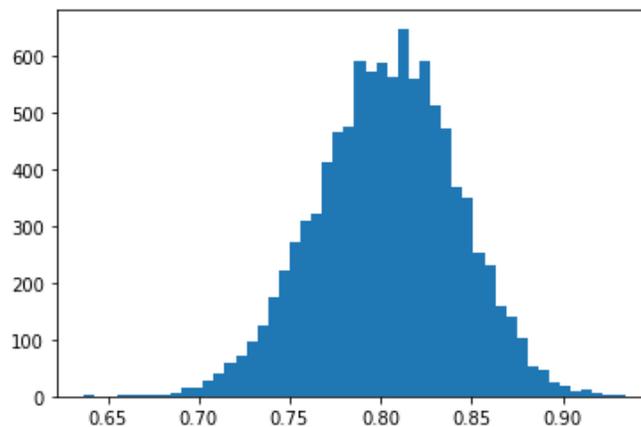


Figure 3-6. Histogram of the ROC AUC values obtained from 10,000 bootstrapping of the AUC calculation. Reprinted with permission.

Figure 3-7 shows examples of cases in which our developed model failed to classify tumors correctly. In order to investigate whether there is a correlation between tumor size and misclassification, the volume of the renal masses was measured. Although the majority of small renal masses were correctly classified, the size of approximately 80% of the misclassified cases was less than one-fifth of the average tumor volume, which suggests that the proposed algorithm is relatively less accurate in identifying smaller renal lesions.

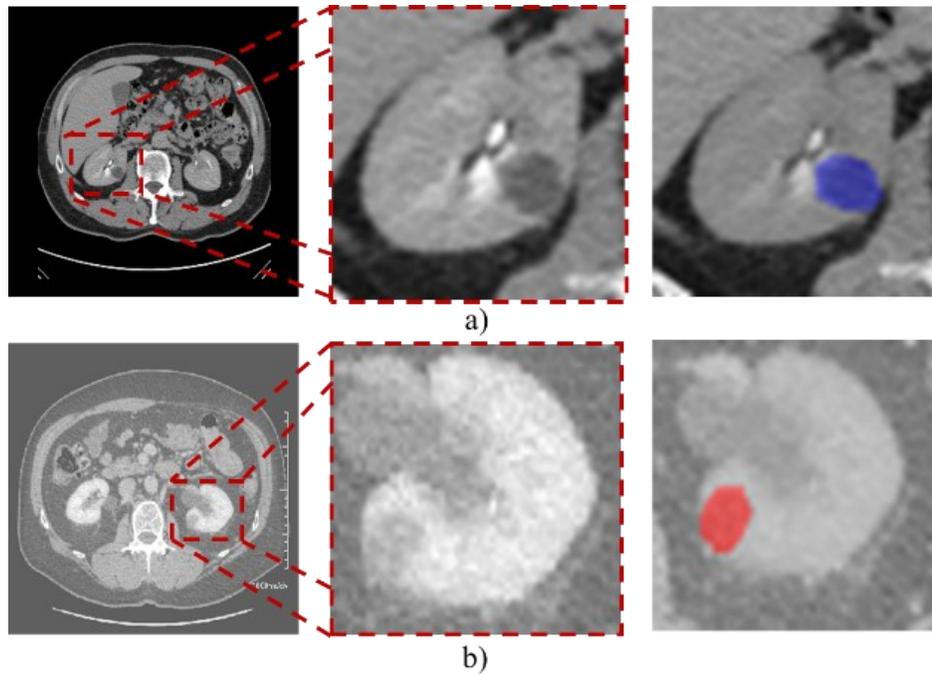


Figure 3-7. Example results of misclassified (a) benign cyst and (b) solid renal mass using our algorithm. The original CECT images are shown in the first column. The third column shows the location of tumor where blue and red show benign cyst and solid renal mass, respectively. The CECT images are visualized in the standard abdominal window and level set of 400 and 50. Reprinted with permission.

A subset of our test dataset was used to evaluate the inter-observer variability of our algorithm. We randomly selected 30 images from patients with different renal mass subtypes including oncocytomas, fpAml, clear cell RCC, papillary RCC, and chromophobe RCC (six from each subtype). Two observers participated in this study. In addition to the initial segmentations performed (described earlier) by observer 1, a second observer (observer 2) who was a non-radiologist shown only where the location of renal masses was in each patient, and independently segmented the tumors on all slices in which they appeared. The inter-observer CV was computed as 8.8%. As seen, the algorithm yielded a very small inter-observer CV. Besides, the results for

the classification of benign cysts from solid masses performed equally well using segmentations from both observers (Table 3-3). This can be expected because in our method, small patches are extracted from the ROI and the majority voting system is applied for final prediction. Therefore, minor discrepancies in manual segmentation at the boundary of the renal masses do not substantially affect the accuracy of classification at the patient level.

Table 3-3. Accuracy of classification of benign cysts from solid masses as evaluated using the independent segmentation of two observers. Reprinted with permission.

Observer (O)	Accuracy (%)	Precision (%)	Recall (%)
O1	96.97	100.0	96.55
O2	93.94	96.55	96.55

Our analysis of the average attenuation of benign cysts in the test dataset revealed that 71.79% of cysts measured ≥ 20 HU, 48.72% ≥ 30 HU and 38.46% ≥ 40 HU. Figure 3-8 shows the ROC curve for the diagnosis of a simple cyst from solid renal masses using attenuation with the different threshold values of 20 HU and 30 HU. AUC was 0.639 and 0.743 for HU ≥ 20 and 30, respectively.

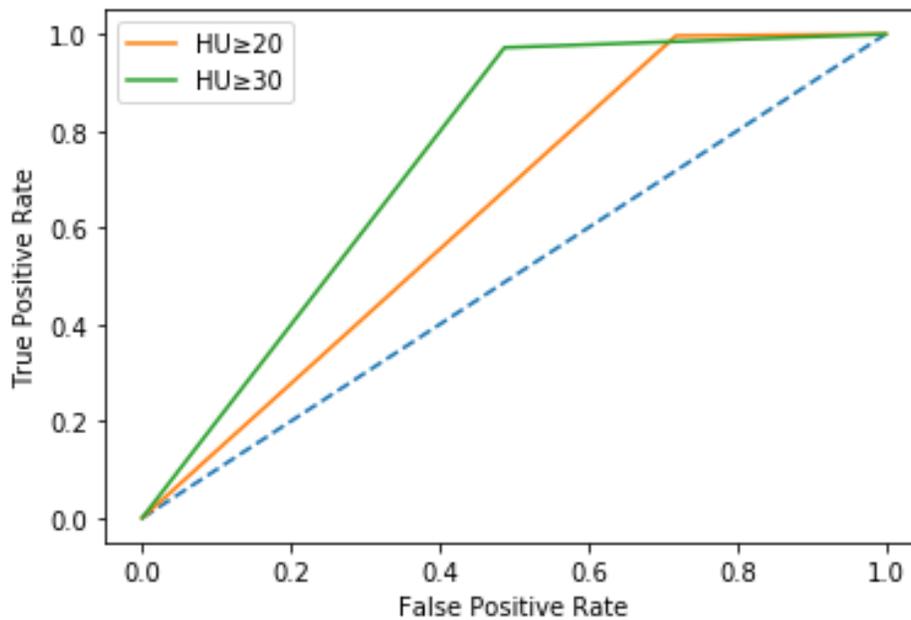


Figure 3-8. ROC curve displays the performance of conventional assessment for diagnosis of cyst versus solid renal masses, where true positive rate (sensitivity) against false-positive rate (1 - specificity) has been plotted. Reprinted with permission.

The results of our study revealed that the automated diagnosis of benign cyst is more challenging than that of solid tumors. We examined the performance of CNN for cyst cases and compared the results with the traditional method, where benign cysts are differentiated from solid tumors based on the HU threshold. Table 3-4 shows the results of different methods for cyst detection, where the CNN approach is more accurate for cyst identification than thresholding based on CT numbers. Remarkably, we measured the CT numbers in HU in 3D while in practice the slice-wise measurement is performed. Therefore, it is likely to have a greater number of incorrectly diagnosed benign cyst using HU thresholding-based technique.

Table 3-4. The performance of different techniques for benign cyst diagnosis on CECT test images.

Reprinted with permission.

Method	Correctly Classified Benign Cysts	Incorrectly Classified Benign Cysts
CNN	27	12
$HU \leq 20$	14	25
$HU \leq 30$	22	17

3.4 Discussion

In this chapter, we described a CNN-based method to differentiate benign renal cysts from solid renal masses. Our proposed method was evaluated on a relatively large test dataset compared to previous studies have done on automated evaluation of renal masses and achieved highly accurate results. Using the patch-based CNN model for classification, our proposed algorithm benefits from the majority voting system for the final prediction that improves the performance of the designed model for renal mass evaluation. The results may be useful for groups actively investigating automated diagnosis of renal masses since an initial step after kidney segmentation is to differentiate benign cysts (which are ubiquitous in the population) from any solid renal mass. The results may also be clinically useful, since a proportion of benign cysts may appear complex and indeterminate in CT, requiring further costly work-up in many instances and our results indicate that the proposed model can be highly accurate to diagnose renal cysts from solid masses at single-phase enhanced CT exams.

In training our CNN model, we had substantially fewer images with cysts. We employed elastic transformation to artificially augment data from training samples of patches with cyst label that

was the key concept in designing our method. Elastic transformation not only enhanced the available data for training instances but also reduced the generalization error as this operation allows the network to learn invariance to such deformation which is a common variation in CT images due to the dynamic motion of abdominal tissues. The low inter-observer CV of our algorithm suggests high reproducibility. This is particularly useful in multi-center clinical trials and when using large registries of tumors, where segmentation could be performed by multiple observers.

We evaluated the average density of benign cysts in our test dataset which demonstrated that the majority of cysts measured greater or equal than 20 HU which, when applied strictly, would require that these lesions be further characterized with additional dedicated imaging with renal mass protocol CT or MRI. The increased attenuation of benign cysts in our study is likely due to the fact that only single-phase enhanced CT images were evaluated where pseudo-enhancement spuriously increases attenuation values. Comparing the ROC curves for HU and our proposed methodology shows the potential clinical value of our method which may help to characterize a subset of benign cysts which may otherwise be considered indeterminate. In the recently published Bosniak version 2019 guideline, an upper threshold of 30 HU has been proposed to diagnose benign cysts from solid masses and to partially mitigate the effects of pseudo-enhancement [18]; however, this proposed threshold requires further validation. When applied in our cohort, 48.72% of cysts measured >30 HU, whereas 2.78% of solid masses measured <30 HU.

The main limitation of our proposed method is that it requires expert manual identification of renal masses on CECT images. Future studies are required to develop a fully automated

approach including renal mass detection and segmentation. State-of-the-art methods might release the constraint of the manual segmentation. Although the sample size in this study was quite large and images were acquired from several CT scanners, images were taken from a single institution retrospective cohort that may create the potential for population bias.

3.5 Conclusion

We described a DL-based model that differentiates renal lesions on CECT as benign cysts compared to solid masses. The results indicate that our algorithm is highly accurate in characterizing benign cysts from solid masses and maybe clinically valuable to prevent unnecessary imaging follow-up for characterization in a proportion of patients.

References

1. F. Zabihollahy, N. Schieda, and E. Ukwatta, "Patch-based convolutional neural network for differentiation of cyst from solid renal mass on contrast-enhanced computed tomography images," *IEEE Access*, 2020.
2. Matthew Hoffman, "Picture of the Kidneys,". Accessed on: Sep. 06, 2019. [Online]. Available: <https://www.webmd.com/kidney-stones/picture-of-the-kidneys#1>.
3. The official foundation of the American urological association, "What is a Renal Mass and What is a Localized Renal Tumor?". Accessed on: Sep. 06, 2019. [Online]. Available: <https://www.urologyhealth.org/urologic-conditions/renal-mass-and-localized-renal-tumors>.
4. The kidney foundation of Canada, "What is kidney disease?", Accessed on: Sep. 06, 2019. [Online]. Available: <https://www.kidney.ca/kidney-disease>.
5. M. N. Ozieh, K. G. Bishu, C. E. Dismuke, and L. E. Egede, "Trends in healthcare expenditure in United States adults with chronic kidney disease: 2002-2011," *BMC Health Serv. Res.*, 2017.
6. American society cancer, "Key Statistics About Kidney Cancer," Jan. 2018. Accessed on: Sep. 06, 2019. [Online]. Available: <https://www.cancer.org/cancer/kidney-cancer/about/key-statistics.html>.

7. C. Ellimoottil, K. A. Greco, S. Hart, T. Patel, M. M. Sheikh, T. M. Turk, R. C. Flanigan, "New Modalities for evaluation and surveillance of complex renal cysts," *Journal of Urology*. 2014.
8. S. D. O'Connor, S. G. Silverman, I. K. Ip, C. K. Maehara, and R. Khorasani, "Simple cyst-appearing renal masses at unenhanced CT: Can they be presumed to be benign?," *Radiology*, 2013.
9. S. Krishna, C. A. Murray, M. D. McInnes, R. Chatelain, M. Siddaiah, O. Al-Dandan, *et al.*, "CT imaging of solid renal masses: pitfalls and solutions," *Clinical Radiology*. 2017.
10. J. Eble, G. Sauter, J. Epstein, and I. Sesterhenn, "Pathology and Genetics of Tumours of the Urinary System and Male Genital Organs.," *World Heal. Organ. Classification Tumours* 2004;
11. H. J. Meyer, A. Pfeil, D. Schramm, A. G. Bach, and A. Surov, "Renal incidental findings on computed tomography: Frequency and distribution in a large non selected cohort," *Med. (United States)*, 2017.
12. B. R. Herts Stuart, G. Silverman, N. M. Hindman, R. G. Uzzo, R. P. Hartman, G. M. Israel *et al.*, "Management of the Incidental Renal Mass on CT: A White Paper of the ACR Incidental Findings Committee," *J. Am. Coll. Radiol.*, 2018.
13. M. Siddaiah, S. Krishna, M. D. F. McInnes, J. S. Quon, W. M. Shabana, D. Papadatos, and N. Schieda, "Is ultrasound useful for further evaluation of homogeneously hyperattenuating renal lesions detected on CT?," *Am. J. Roentgenol.*, 2017.
14. B. D. Pooler, P. J. Pickhardt, S. D. O'Connor, R. J. Bruce, S. R. Patel, and S. Y. Nakada, "Renal cell carcinoma: Attenuation values on unenhanced CT," *Am. J. Roentgenol.*, 2012.
15. P. Y. Simard, D. Steinkraus, and J. C. Platt, "Best practices for convolutional neural networks applied to visual document analysis," in *Proceedings of the International Conference on Document Analysis and Recognition, ICDAR*, 2003.
16. T. Hodgdon, M. D. F. McInnes, N. Schieda, T. A. Flood, L. Lamb, and R. E. Thornhill, "Can quantitative CT texture analysis be used to differentiate fat-poor renal angiomyolipoma from renal cell carcinoma on unenhanced CT images?," *Radiology*, 2015.
17. M. Bland, *How Should I Calculate a Within-Subject Coefficient of Variation?* Oct. 16, 2006. Accessed on: Dec. 12, 2019. [Online]. Available: <https://www-users.york.ac.uk/~mb55/meas/cv.htm>
18. S. G. Silverman, I. Pedrosa, J. H. Ellis, N. M. Hindman, N. Schieda, A. D. Smith, *et al.*, "Bosniak classification of cystic renal masses, version 2019: An update proposal and needs assessment," *Radiology*, 2019.
19. E. M. Hu, J. H. Ellis, S. G. Silverman, R. H. Cohan, E. M. Caoili, and M. S. Davenport, "Expanding the Definition of a Benign Renal Cyst on Contrast-enhanced CT: Can Incidental Homogeneous Renal Masses Measuring 21–39 HU be Safely Ignored?," *Acad. Radiol.*, 2018.

Chapter 4: Automated Classification of Solid Renal Masses on Contrast-Enhanced Computed Tomography Images using Convolutional Neural Network with Decision Fusion

In this chapter, we describe the development and evaluation of an image-based-CNN algorithm for automated classification of solid renal masses into RCC and benign on CECT images using the same dataset used in Chapter 3 [1-2]. The proposed method is similar to what is described in the previous chapter, with the difference that image-based CNN was used instead of patch-based CNN. We chose this approach because the shape of the renal mass is an important feature for the identification of RCC from benign. By using patch-based strategy, intensity and texture features which are local features, are preserved whereas global features such as shape are lost. Image-based CNN helps to capture both local and global features including intensity, texture, shape, and size. Quantitative results demonstrated that image-based-CNN is more accurate for the differentiation of RCC from benign renal mass compared to patch-based-CNN (83.75% versus 72.63%). Therefore, semi-, fully automated, and 3D image-based-CNN approaches are selected and explained in this chapter and their performance for the classification of renal masses into RCC and benign is compared together.

4.1 Introduction

Renal masses are commonly detected on CT and it is estimated that between to 15-40% of patients undergoing CT will have an incidentally discovered renal mass [3-4]. Accurate imaging characterization of renal masses is important because not all incidentally discovered renal masses are RCC. Up to 20% of solid renal masses measuring <4 cm in size are benign and the number of

benign diagnoses increases considering <2 cm mass [5]. Oncocytomas and fat poor renal angiomyolipomas (fpAML) are most common benign renal masses [6]. Though CT has been studied extensively for the purpose of subtyping renal masses, achievements pertain mainly to the ability to differentiate between clear cell and papillary RCC [7-8] and to suggest a diagnosis of fat-poor angiomyolipoma (fp-AML) [7]; however, differentiating between renal oncocytoma and RCC remains limited [10-12]. Although percutaneous biopsy can establish a diagnosis in the majority of cases, it is relatively invasive and can be inconclusive in a minority of cases [6]. Therefore, non-invasive imaging diagnosis remains desirable.

Quantitative CT analysis of radiomic features, mainly first- and second-order texture features in renal masses is well studied [13-16]. The use of ML of CT radiomic features for diagnosis in renal masses is also actively under investigation. A summary of related works is given here.

L. Zhou et al. investigated the effect of transfer learning (InceptionV3 model, pretrained on the ImageNet dataset) for benign versus malignant classification on CT images using a dataset comprised of 192 cases (malignant/benign = 134 (117 clear-cell RCCs, 8 papillary RCCs, and 9 other RCC subtypes) /58 (50 renal cysts and 8 renal AMLs)) of renal tumors [17]. They established five image-level models for each of the slice, ROI, and rectangular box region (RBR) datasets. Then, two patient-level models were built based on the optimal image-level models. Their suggested method reported the accuracy of 0.69 on 39 (benign/malignant = 25/14) test images. The accuracy of the ROI and RBR dataset was 0.97 and 0.93, respectively. This study has not included any renal oncocytoma, which is by far the most difficult benign tumor to differentiate from RCC with overlapping features with clear cell and chromophobe RCC in particular.

S.P. Raman et al. [18] applied RF to the parameters derived using CT texture analysis software to distinguish between several common renal masses. First, the regions of interest were drawn around each mass on multiple slices in the arterial, venous, and delayed phases on renal mass protocol CT scans on a dataset comprised of 20 clear cell RCCs, 20 papillary RCCs, 20 oncocytomas, and 20 renal cysts. The RF-based model was tested on 19 new images and correctly classified oncocytomas, clear cell RCCs, cysts, and papillary RCCs in 89%, 91%, 100%, and 100% of cases, respectively.

In another research work, an automatic deep feature classification method for distinction of benign angiomyolipoma without visible fat (AMLwvf) from malignant clear cell RCC from 80 (AMLwvf/clear cell RCC = 39/41) CECT images was developed. They used a combination of 71-dimensional hand-crafted texture and shape features extracted from small renal mass (SRM) contours, and 1000–4000-dimensional deep features extracted from the ImageNet pre-trained DL model with the SRM image patches [19]. An RF classifier was trained using concatenated features to classify the types of SRMs. Their proposed method yielded an accuracy of $76.6 \pm 1.4\%$ on the dataset using a leave-one-out cross-validation technique.

H. Lee et al. [20] developed a computer-aided classification system to differentiate benign fp-AML from malignant clear cell RCC using quantitative feature classification on histogram and texture patterns on a dataset of 50 CT multidetector computer tomography (MDCT) images (clear cell RCC/fp-AML = 25/25). The tumors were manually delineated from images by an expert radiologist. 64 features were extracted from ROIs and several feature selection methods were applied to select a group of useful features. Then various machine-learning algorithms including logistic regression, KNN, SVM, and RF were trained on the selected features to

differentiate benign fp-AML from malignant clear cell RCC on CE MDCT images. A different combination of feature selection and ML methods were tested using fivefold cross-validation technique. KNN and SVM classifiers with ReliefF feature selection demonstrated the best performance among other choices of feature selection and classification methods (ReliefF+KNN and ReliefF+SVM achieved the accuracy of $72.3 \pm 4.6\%$ and $72.1 \pm 4.2\%$, respectively).

H. Coy et al. [21] used the Inception model to distinguish clear cell RCC from oncocytoma utilizing radiomic features extracted from routine four-phase MDCT (unenhanced, corticomedullary, nephrographic, and excretory phases) in 179 patients (clear cell RCC/oncocytoma = 128/51) with pathologically confirmed renal masses. In this study, the entire renal mass was contoured in the axial plane in each of the four phases, which resulted in a 3D volume of interest (VOI) representative of the entire renal mass. Thirteen different approaches were then investigated to slice the VOI into 2D images that were used for network training. 10% of the data designated to validate the method. Their proposed methodology achieved the specificity and positive predictive value (PPV) of 52.9% and 80.3%, respectively in the EX phase when the single mid-slice of the tumor in the axial, coronal and sagittal plane was analyzed. The classifier performance improved when the number of mid-slices of the tumor was raised to three (accuracy, sensitivity, and PPV of 75.4%, 88.3%, and 79.6%, respectively). When the entire tumor volume was shown, the classification performance was best in the EX phase with an accuracy, sensitivity, and PPV of 74.4%, 85.8%, and 80.1%, respectively. Their developed method yielded the highest PPV of 82.5% when the entire tumor volume, plus mid-slices from all phases and all planes presented as tiled images.

The main limitation of all previous studies is the use of relatively small test datasets that may do not represent the real test population. Another limitation of previous works in the domain of ML diagnosis of renal masses is that study samples were not reflective of the distribution of disease in clinical practice but were selected to study specific classification questions (e.g. RCC versus fpAML or RCC versus oncocytoma). These strategies, though important, may potentially bias results by overestimating accuracy of the classification and ultimately do not address an important clinical question, which is “whether CT can differentiate between RCC and benign tumors (including both of the most common subtypes, namely fpAML and oncocytoma)?”.

The purpose of the present study was to design and evaluate a decision fusion of CNN-based model to assess a large population of consecutively identified solid renal masses, reflective of natural disease distribution in clinical practice, on CECT images. More specifically, we sought to determine the accuracy of a semi-automated CNN-based algorithm to diagnose RCC from benign solid renal masses (e.g. oncocytoma and fpAML). A secondary objective was to evaluate and compare accuracy to a fully automated approach and a 3D CNN-based algorithm as an alternative for this application.

4.2 Method

For this study we used the same dataset that was explored for differentiation of cyst from solid renal masses. The comprehensive explanation of the patient, CT technique, and manual segmentation can be found in the Section 3.2.1.

4.2.1 Semi-automated Method for Differentiation of RCC from Benign Solid Renal Masses

Our dataset comprised of 315 CECT images of size $512 \times 512 \times N$ ($N: [31,555]$). The images were randomly divided into the training ($N=155$), and testing ($N=160$) sets. Figure 4-1 indicates our strategy for patient selection and overall experimental design for the model training, and testing.

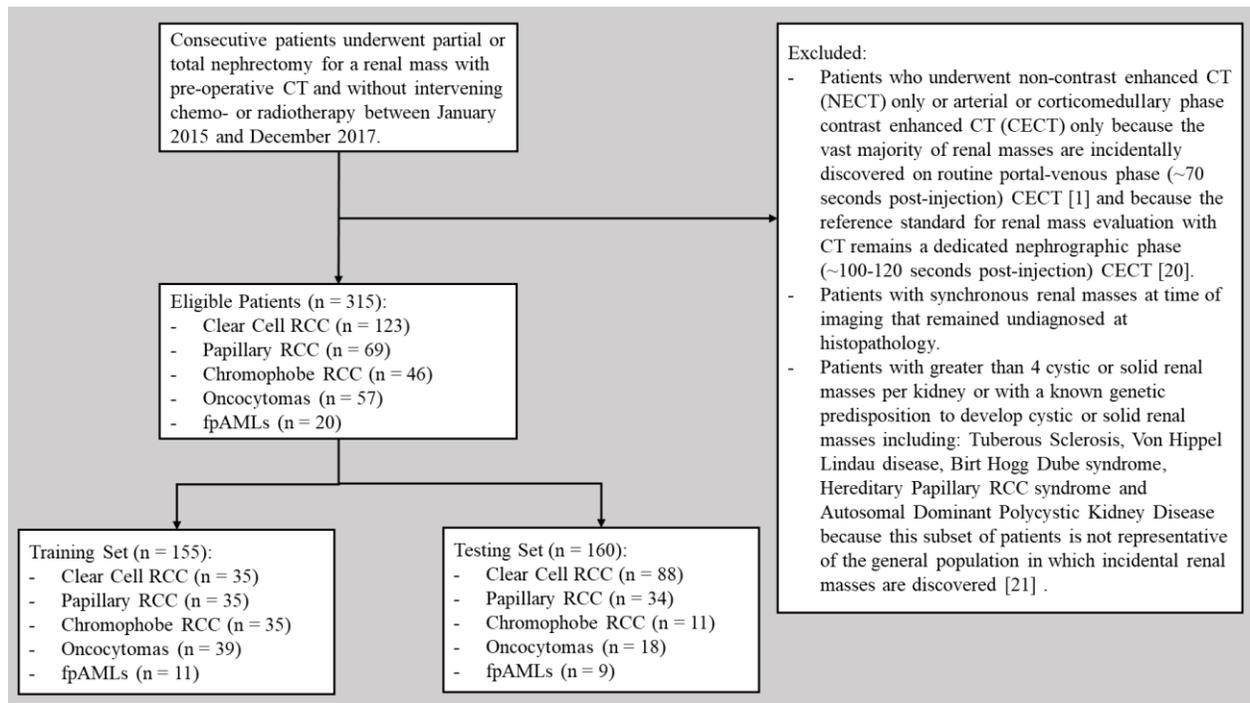


Figure 4-1. Flowchart illustrating strategy for patient selection and overall experimental design for differentiating RCC from benign solid renal masses at contrast-enhanced CT in the current study. Reprinted with permission.

Dividing images into train and test is the main step in the experiment design that greatly affects the network performance for renal mass classification. To split the data the following factors were considered:

1) The number of training samples must be proportional to the trainable parameters of the network. Accordingly, by increasing the capacity of CNN, the number of training samples was increased. For CNN design, we applied a growing approach, where a simple CNN including a convolutional layer along with a fully connected network with one hidden layer and few nodes was trained. To gradually improve the accuracy of renal mass classification, some layers were incrementally added, and the parameters were tuned manually to achieve higher classification accuracy. During the training phase, we witnessed that using the architecture given in Figure 4-4 gives the best classification results. Given this network, the number of training samples was varied, and it was observed that using more than 155 training images does not improve the network performance for renal mass classification, substantially. Thus, as can be seen from Figure 4-1, 155 and 160 images were used for train and test, respectively. It is noteworthy that it was desirable to keep a large part of the dataset for the test as evaluating the method with a greater number of unseen images could have assured us of the algorithm robustness for automated renal mass evaluation.

2) Another important factor in dividing images for train and test in the dataset with imbalanced class is to keep approximately the same ratio for each class in the test set as the test set must be a real representative of the dataset. Otherwise, the results might not be valid. The ratio of benign to RCC renal masses in our dataset was 77/238 ($\approx 1:3$) and in the test set, we had 27/133 ($\approx 1:4$).

Figure 4-2 shows an overview of our methodology for renal masses assessment. We first fully-automatically discovered the boundaries of the kidneys on CECT images using the method described in section 4.2.4. Once the boundaries of the kidneys were discovered, CT images were cropped in a window of size 128×128 pixels around the central point of the left and right kidneys. This procedure was needed because comparing the size of CT images, the size of the renal masses was relatively small and cropping the image allows us to focus on the tumor as ROI. In other words, by employing this strategy, the images are zoomed in around the ROI and the remainder of the CT images are discarded.

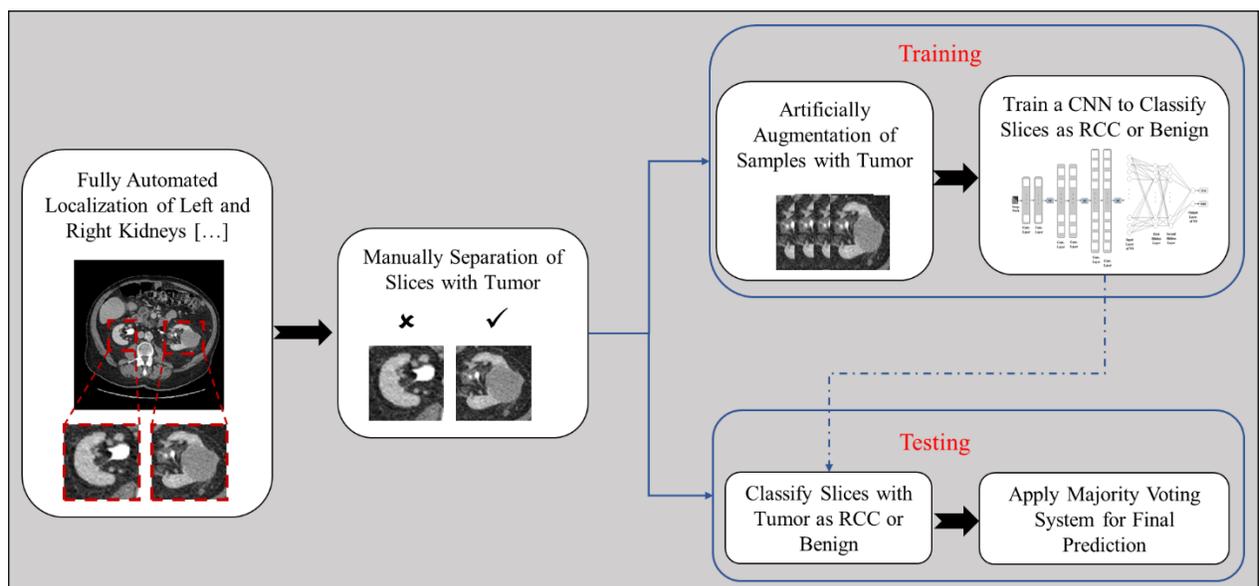


Figure 4-2. Overview of semi-automated methodology for labeling a contrast-enhanced CT image as RCC or benign renal mass. It is notable that for the semi-automated approach manual segmentation of renal mass is not needed. Once the CECT images are cropped around the kidneys, an expert needs to separate slices with tumors from the others to be passed to the trained CNN. Reprinted with permission.

The cropped images were compiled manually to separate slices with the renal mass for training the network that resulted in 2085 slices including 475 and 1610 with benign and RCC solid

masses. Then histogram equalization was performed to enhance the contrast. To this end, the normalized sum of the image histogram was calculated, and a remapping procedure was used to obtain the density values of the equalized image. The CT density values (D) were then normalized to the range of $[0, 1]$ using the following equation:

$$D_N = (D - D_{min}) / (D_{max} - D_{min}) \quad (4-1)$$

Table 4-1 shows a summary statistic of the number of slices used in our experiments to train CNN on a per-patient basis. Since the number of slices labeled with benign renal mass was less than that labeled with RCC, the images of the underrepresented class were augmented artificially to address the class imbalance problem in the training set. We used elastic deformation and generated 3 images from each image labeled as benign and included them to our training samples (total of 3511 images). The elastic transformation was applied again to the whole images to increase the number of training instances (total of 7023 images). To perform elastic deformation the method described by Patrice Y. Simard et al. [24] was followed. Elastic deformation as described in [24] is controlled by two parameters, the elasticity coefficient σ , which was set to 19.2, and a scaling factor, α , which controls the intensity of the deformation and was set to 256. Those parameters were adjusted experimentally such that the generated artificial and original images remain similar in order to share similar features. Examples are provided in Figure 4-3. As seen position, size and intensity of different tissues are changed. The same transformation was applied to the corresponding manual segmentation, simultaneously.

Table 4-1. A summary statistic of the number of slices used in our experiments to train CNN on a per-patient basis. Reprinted with permission.

	Total no. of 2D slices extracted from training 3D CT images	Mean	Standard Deviation	Median	Minimum	Maximum
Train	2085	14	12	10	2	76

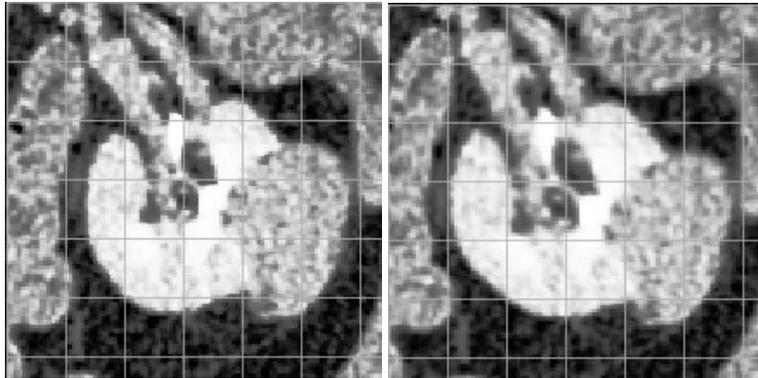


Figure 4-3. From left to right, an example of one training sample with benign renal mass and its deformed one are shown. Reprinted with permission.

Figure 4-4 shows the architecture of the CNN used in our experiments to distinguish RCC from benign renal masses. All parameters of the network were adjusted experimentally aiming to raise the network performance.

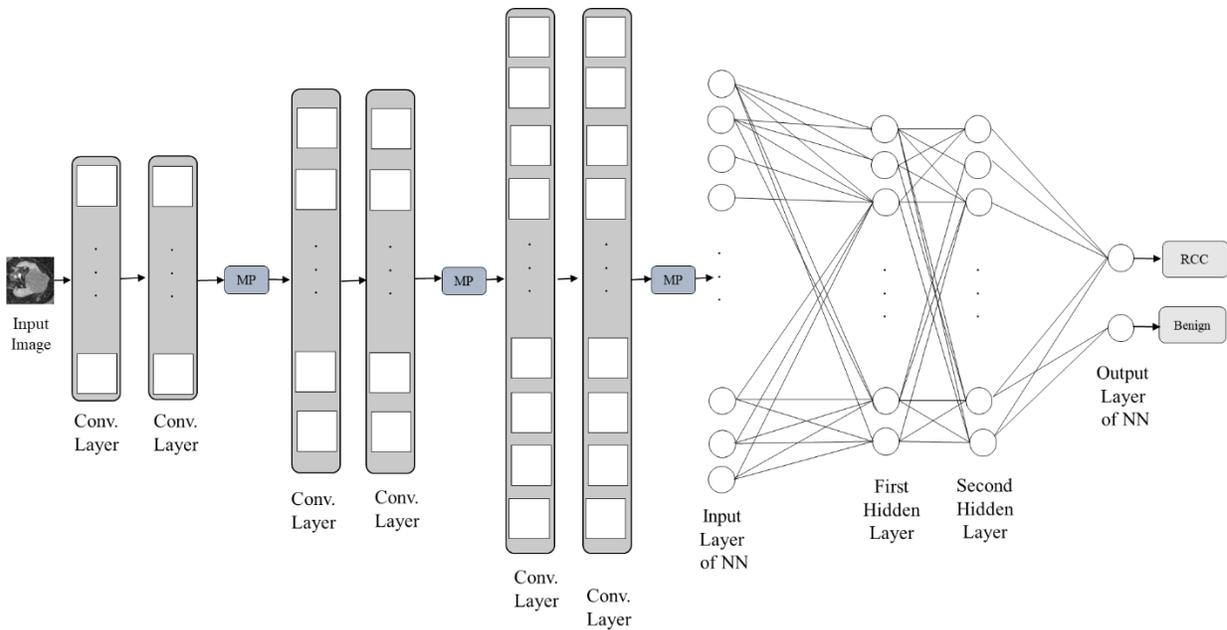


Figure 4-4. The architecture of the CNN used in our experiments for the classification of slices with solid renal masses into RCC or benign. In this figure, MP represents the maximum pooling layer. Reprinted with permission.

The developed CNN is made up of 6 convolutional (Conv.) layers of size 3×3 with 32, 32, 64, 64, 128, and 128 filters in each layer. Rectified Linear Unit (ReLU) and batch normalization were applied after each Conv. layer. After every two Conv. layers, one maximum pooling layer of size 2×2 was employed that calculates the maximum of 4 output nodes. A neural network with two hidden layers of size 128 neurons was included to the end of our proposed network in which three dropout layers were added before hidden and output layers. We used Softmax as an activation function at the last layer of our suggested CNN to map the non-normalized output of a network to a probability distribution over predicted output classes. Categorical cross-entropy and Adadelta were used as loss function and optimizer to train the CNN, respectively.

We implemented the algorithm in Python using Keras library built on top of Tensorflow. The network was trained for 150 epochs on Intel Core i7, 2.8 GHz on a graphics processing unit (GPU)-accelerated computing platform, which took a total of approximately 90 minutes. The network was trained for 50, 100, 150, and 200 epochs and the performance of the network in terms of accuracy was calculated for a subset of the test set (50 CECT test images). Table 4-2 shows the results of this experiment. As seen, the accuracy does not improve after 150 epochs. To make sure that increasing the number of training epochs more than 150 does not improve the performance of the method for classification of RCC and benign renal masses, the network trained for 200 epochs was evaluated using the whole test images. We got accuracy, precision, and recall of 80.63%, 90.48%, and 85.71%, which were less than what we got with a network trained for 150 epochs. Figure 4-5 plots the model accuracy and loss value against the number of training epochs for train and validation sets. The validation set was 20% of the training images. As it is evident from the plots, the model accuracy and loss does not improve substantially for the training set after 100 epochs whereas for the testing set (Table 4-2), increasing the number of training epochs, improved the network performance till 150 and after that, increasing the number of training epochs deteriorated the classification accuracy. Thus, the network was trained for 150 epochs.

Table 4-2. Performance of the network for the different number of training epochs evaluated using a subset of the test set. Reprinted with permission.

No. of Training Epochs	Accuracy (%)
50	69.39
100	68.00
150	71.43
200	71.43

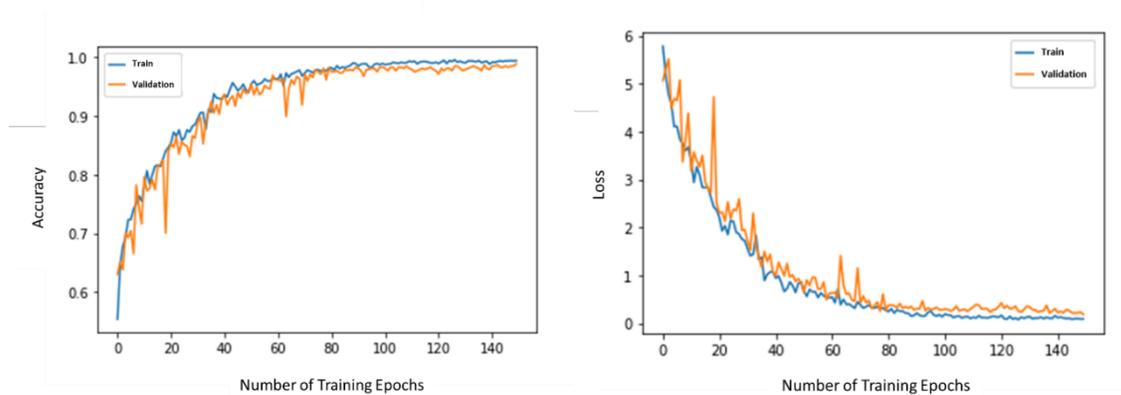


Figure 4-5. From left to right, the model accuracy and loss are plotted against the number of training epochs for train and validation sets. Reprinted with permission.

In the testing stage, the majority voting system was employed for classifying CT images into RCC versus benign in which the predicted labels for slices extracted from CT scans were combined and the CT image was assigned to the class with the majority label predicted. For the cases with an equal number of predicted labels in both classes, the image was labeled as RCC.

4.2.2 Fully Automated Method for Differentiation of RCC from Benign Solid Renal Masses

The fully automated approach was similar to the semi-automated except that after delineating the kidneys, renal masses were localized on CECT employing the method described in Section 4.2.5, where an ensemble of three U-Net models was utilized to detect and segment renal masses. Next, the slices with the predicted renal mass were automatically detached from the cropped images and passed to the trained CNN to be classified as RCC or benign. The majority voting system was used for final prediction and labeling the CECT image as RCC or benign.

4.2.3 3D CNN-based Algorithm for Semi-automated Differentiation of RCC from Benign Solid Renal Masses

We used the same number of images to train and test a 3D CNN. 3D CNN must be fed with a volume of data of the same size in the X, Y, and Z directions. Renal masses have different sizes and volumes. In the semi-automated method used in our study, the images which included the renal mass were cropped around each kidney with a fixed size of 128×128 pixels in the X and Y directions. However, the number of slices including tumors for each patient was highly variable and ranged from 2 to 88 with a mean and median of 15 and 11 in our dataset. To apply a 3D CNN, smaller volumes (sub-volumes) of size $128 \times 128 \times 4$ voxels were extracted from CECT images. We picked out $n = 4$ because; 1) it is big enough to capture global features in the Z direction; 2) the amount of memory and time are required to train a 3D network is rational; and 3) up-sampling is required for few cases (14 images), where the number of slices across renal mass is less than four. We used the same network (Figure 4-4) with the difference that all 2D

operations were replaced by 3D counterparts and the last pooling layer was removed from the network. Assuming N_{rm} denotes the number of images including renal mass for each CECT in the dataset. For cases with $N_{rm} < 4$, up-sampling by duplicating the slice with the larger tumor was performed to create a volume of size $128 \times 128 \times 4$ voxels. For cases with $N_{rm} > 4$, if $N_{rm} = 4k$, sub-volumes without overlap was created from the cropped images. Otherwise, the slice with the larger tumor was duplicated until $N_{rm} = 4k$ and then sub-volumes of size $128 \times 128 \times 4$ voxels were generated. Then the sub-volumes were augmented using elastic deformation to address the class-imbalance issue in the training dataset and the network was trained using 2568 sub-volumes. In the testing phase, if applicable, the majority vote strategy was taken for final label prediction.

4.2.4 Fully Automated Segmentation of Kidneys on CECT Images

We trained a U-Net-based classifier to delineate the boundaries of the kidney on CECT images. Our U-Net model has 5 stages, where 32 kernels of size 3×3 with the stride of 1×1 were moved across the image in the first stage and the number of kernels was doubled in the subsequent layers that resulted in a total of 7,852,517 trainable parameters. The size and stride of pooling/up-sampling layers were set to 2×2 . Zero padding was performed such that the output feature map had the same size as the original input layer. At the final stage, a convolutional layer of size 1×1 followed by sigmoid activation layer was applied to map all feature maps and generate a probability map as the same size of the input image. The probability map was then thresholded such that each pixel with a probability of greater than 0.5 was considered as within the kidney boundary to create the desired segmentation map for each slice of the CT image. We

have chosen the parameters based on previous studies [19-20] and then optimized by experimental observation.

First, we down sampled all the slices, extracted in the transversal direction, by the factor of two to reduce the computational complexity. Next, the histogram equalization was performed to enhance the contrast. To this end, the normalized sum of the image histogram was calculated, and a remapping procedure was used to obtain the density values of the equalized image. The CT density values were then normalized to the range of [0, 1] to adjust for differences in density across each slice. The images were then randomly divided into the training (N=190) and testing (N=125) sets and the network was trained using 14480 slices compiled from the training CECT images. 20% of the training images were considered as a validation set to evaluate the performance of the network during training. Table 4-3 shows the number of images in each subgroup used in our experiments for the model training and testing.

Table 4-3. The number of CECT images from local patients with the different type of renal mass, used for the training and testing phases in our experiments. Reprinted with permission.

	Clear Cell RCC	Chromophobe RCC	Papillary RCC	fpAML	Oncocytoma	Total
Training set	75	30	40	10	35	190
Testing set	48	16	29	10	22	125
Total per group	123	46	69	20	57	315

DSC and its negative value were used as the metric to assess the accuracy of segmentation and loss function during training. AdaDelta was employed as an optimizer, which is robust against noisy gradient information, differences in the network architecture, various data modalities, and selection of hyper-parameters [25]. The main advantage of this optimizer compared to those which require learning rate initializations such as stochastic gradient descent is that it does not require manual tuning of a learning rate as choosing a small value for learning rate may slow down the optimization while selecting a large value causes oscillation around minima. The negative value of the DSC was defined as a loss function to be minimized during network training.

In the testing phase, images were pre-processed similar to training data. The trained U-Net was then used to classify the voxels within the kidney on the CECT images. The boundary of the segmented kidney was smoothed through mathematical erosion and dilation operations in Python. The first operation erodes the boundaries of the kidney while dilation increases the segmented area as erosion shrinks the target object in the segmentation map. Employing a combination of these two morphological image processing actions removed the patchy and sharp edges while avoiding target to become smaller. To evaluate the performance of the designed network for kidney delineation from CT images, we compared algorithm segmentation to the expert manual segmentation.

We implemented the algorithm in Python using Keras library, on top of Tensorflow. We trained each network for 70 epochs on an Intel Core i7, 2.8 GHz using a GPU-accelerated computing platform, which took a total of approximately 110 minutes.

4.2.5 Fully Automated Detection and Segmentation of Renal Masses on CECT Images

Figure 4-6 displays an overview of our developed algorithm for renal mass localization. Once the kidneys boundaries were segmented on the CECT images, the images were cropped around the kidney in a window of size 128×128 voxels. Extracted patches were used as an ROI for renal mass detection and segmentation. The histogram equalization was applied to the extracted patches and the density values were normalized to the range of $[0, 1]$.

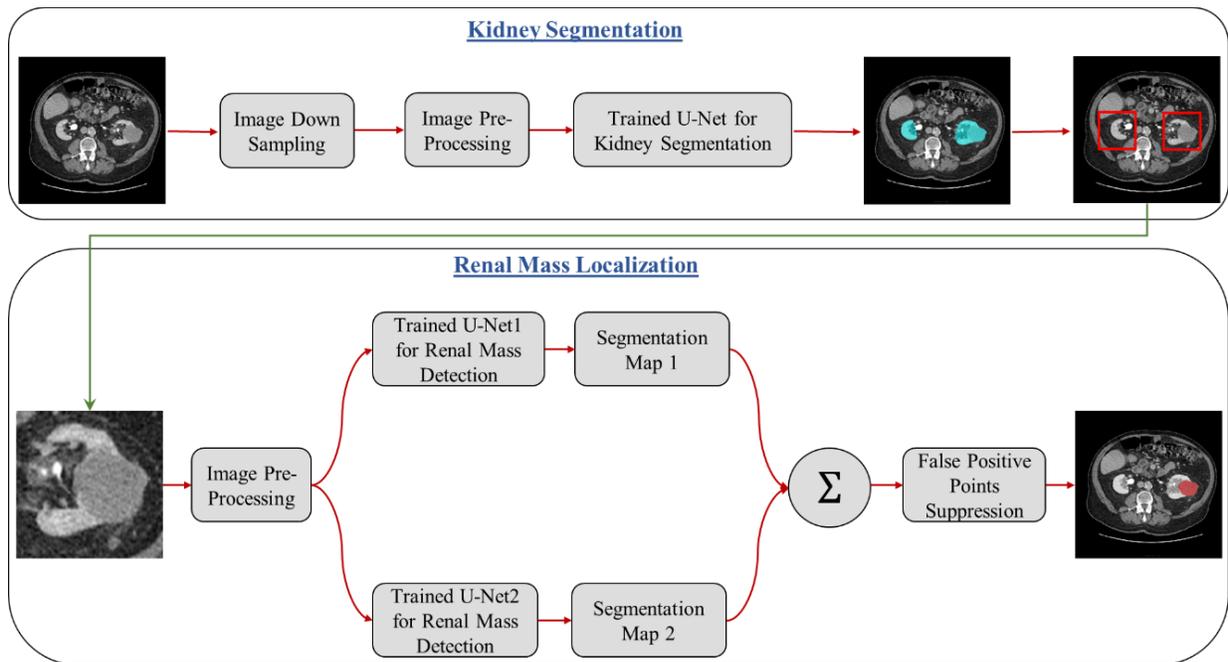


Figure 4-6. An overview of our proposed methodology, where the top and bottom block diagrams display boundary segmentation and renal mass localization in both kidneys. Reprinted with permission.

To choose an appropriate architecture for U-Net, we trained and tested several U-Nets with a varying number of filters in the first stage using extracted patches for renal mass localization and evaluated the performance of the network for each configuration. Table 4-4 shows a summary of results for each network on the testing dataset. The architecture of the network used for renal

mass detection was similar to that of the one used for kidney segmentation, except the number of kernels used in the first stage of the networks was set to 8, 10,12, 16, and 32. The same optimizer and loss function were used to train the networks. The number of kernels was doubled in the next stages for each network. For training, the U-Nets for renal mass localization patches derived from the same training images (N=190) used in the previous phase for kidney segmentation network training. As the number of patches with renal mass was less than those without RM, extracted patches with renal mass were flipped left-right and up-down and augmented to the training set, which resulted in a total of 20950 and 8984 training samples for institutional and KiTS19 datasets. We investigated the combination of different networks (2 and 3 networks) utilizing different combination rules (adding and majority vote system) and observed that adding up the prediction outcomes of U-Net3 and U-Net4 yields the highest sensitivity, specificity, and DSC for renal mass detection. We elected to trade off increased sensitivity for lower specificity, to preferentially reduce the risk of undetected renal mass.

Table 4-4. Summary of evaluation results of the different networks for renal mass detection from CECT testing images. Mean values for sensitivity, specificity, and DSC are reported. Reprinted with permission.

	No. of the Filter in the First Stage of U-Net	Sensitivity (%)	Specificity (%)	DSC (%)
U-Net1	8	66.80	89.48	90.72
U-Net2	10	68.51	92.43	92.05
U-Net3	12	71.42	89.33	91.34
U-Net4	16	75.95	82.53	89.38
U-Net5	32	68.80	90.53	91.74
U-Net2 + U-Net3	10,12	82.55	88.01	90.64
U-Net3 + U-Net4 (proposed)	12,16	87.41	81.82	88.65

In the testing phase, small patches of size 128×128 pixels were extracted around kidneys, segmented automatically using our proposed method, and after pre-processing (down sampling, histogram equalization, and density normalization) were presented to the trained networks. The aggregate of the regions proposed by networks as the target was considered as renal mass and the boundary of the segmented kidney was used as an ROI to remove false positives on non-kidney tissues. We then compared the refined segmentation map to manual segmentation by using the following evaluation metrics.

4.3 Evaluation Metrics

To evaluate the performance of our proposed methodology for renal mass assessment, the confusion matrix was created for the whole test dataset and then accuracy, precision, and recall along with the 95% confidence interval were measured. We also generated the receiver operating ROC and PR curves. The area under both ROC and PR curves are reported as well. To assess the performance of the proposed methodology for kidney and renal mass segmentation, DSC, AVD, precision, and recall were reported.

4.4 Results

We used the proposed method to segment the boundaries of the kidneys from 125 and 60 patients in the institutional and KiTS19 test sets, respectively. Figure 4-7 displays the segmentation results for four patients from both test datasets, where the algorithm-generated segmentation closely matches the manual segmentation. Boxplots of results on the test population are shown in Figure 4-8. Results demonstrate that our algorithm is highly accurate and efficient for kidney

segmentation as the boundaries of the kidneys using a typical CECT image of size 512×512 pixels are segmented with a DSC of $95.79\% \pm 5.16\%$ (mean \pm SD) and this is accomplished rapidly in an average of 8.43 ± 3.75 sec (mean \pm SD) using our above-stated hardware specifications. Our suggested methodology discovered the boundaries of the kidneys on 60 test images from KiTS19 with mean DSC of $96.25\% \pm 3.37\%$ in 9.62 ± 10.92 sec. The AVDs were 1.84 ± 4.02 cubic centimeter (cc) and 1.79 ± 2.17 cc for institutional and KiTS19 test cases, respectively.

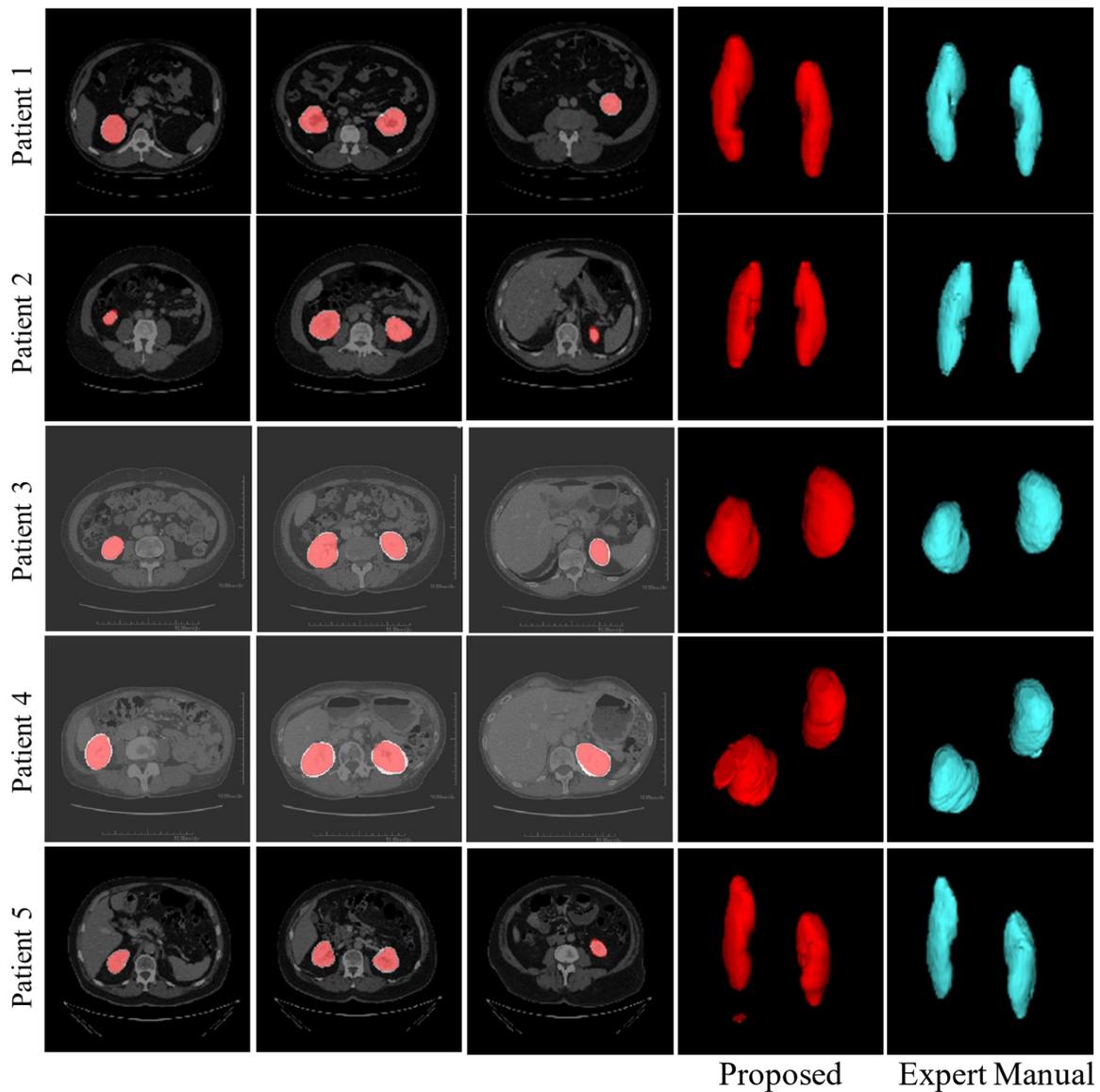


Figure 4-7. Example results of segmented kidneys from CECT images for five patients with different types of renal masses including from top to bottom clear cell RCC, chromophobe RCC, papillary RCC, fpAML, and oncocytomas. Columns 1 to 3 present three different slices from base to apex extracted from CT scans of each patient. On the source 2D images, the results from the U-Net-based method and manual segmentation of the kidney are shown in white and red, respectively. The 2D CT images are visualized in the standard abdominal window and level set of 400 and 50, respectively. The fourth and fifth columns indicate the 3D surfaces of the kidneys generated by our algorithm and expert manual segmentation, respectively. Reprinted with permission.

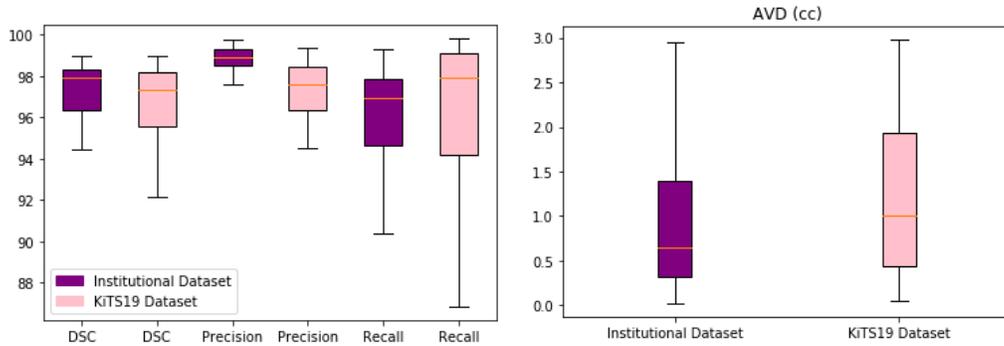


Figure 4-8. Boxplots of results generated by our proposed method for kidneys boundary segmentation on 125 and 60 CECT test images of institutional and KiTS19 datasets in comparison to manual segmentation.

Reprinted with permission.

We used a summation ensemble model to detect renal mass on the CECT test images using 125 cases and 60 cases from institutional and KiTS19 datasets, respectively. Figure 4-9 shows some example results, where localized renal masses were superimposed on test images of four patients randomly selected from both test dataset. As shown in Figure 4-9, our method not only detected the renal mass with a high degree of accuracy but also accurately segmented its boundary compared to manual segmentation. Using our method, renal masses were detected in 125 and 52 test images from institutional and KiTS19 datasets, which translates into 100% and 86.67% sensitivity at patient level (missing rate of zero and 13.33%), respectively. AUC for detecting lesion from institutional and KiTS19 datasets using our proposed method was 0.865 and 0.741, respectively. Figure 4-10 shows the boxplot of obtained results for renal mass detection and segmentation using our suggested methodology on the test populations of the Institutional and KiTS19 datasets. Overall, the proposed method yielded higher segmentation accuracy for the institutional dataset.

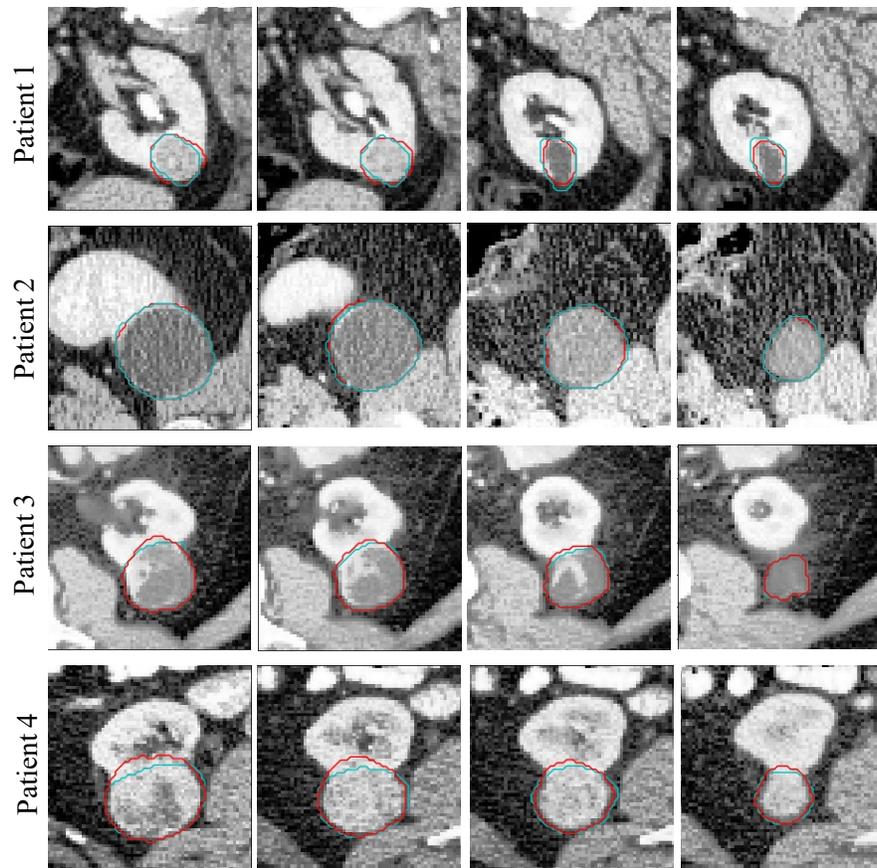


Figure 4-9. Example results from localized renal masses on CECT images in five patients with different types of renal masses including from top to bottom clear cell RCC, chromophobe RCC, papillary RCC, fpAML, and oncocytomas. In columns 1 to 4, 2D slices enclosing renal masses compiled over kidneys are shown. In all images, the renal masses contours computed from algorithm-generated and manual segmentations are shown in red and cyan, respectively. Reprinted with permission.

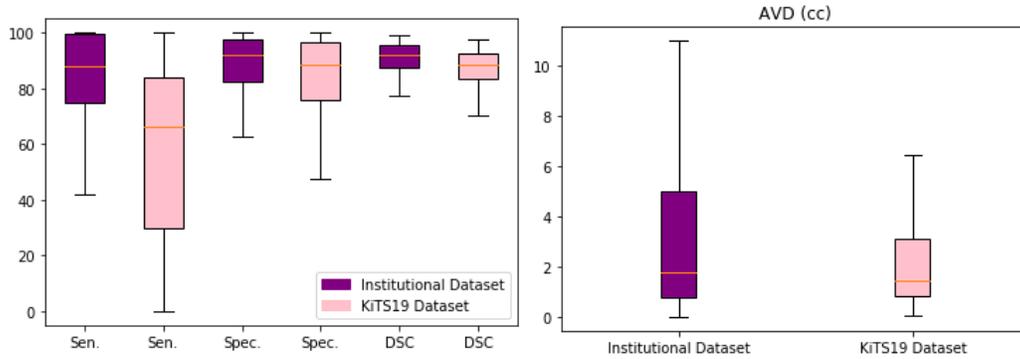


Figure 4-10. Boxplots of results generated by our proposed method for renal mass boundary segmentation on 125 and 60 CECT test images of institutional and KiTS19 datasets in comparison to manual segmentation. Sen., Spec., and DSC denote sensitivity, specificity, and Dice similarity coefficient, respectively. Reprinted with permission.

Furthermore, the accuracy of the proposed methodology for localization of cyst and solid renal mass was investigated on the institutional test cohort including 185 renal mass (60 benign cysts and 125 solid RM), separately. Figure 4-11 indicates the result of the subgroup analysis for the institutional dataset. Overall, our method yielded higher localization accuracy for solid RM.

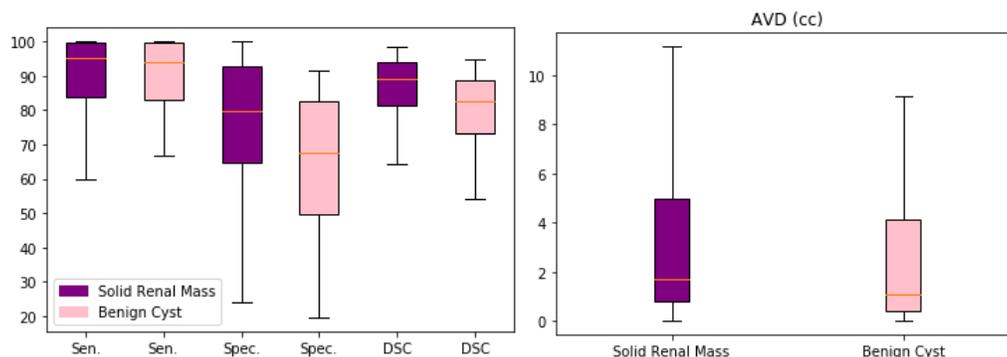


Figure 4-11. Boxplots of results generated by our proposed for subgroup analysis for localization of solid renal mass and benign cyst on 125 CECT images from the institutional dataset in comparison to expert

manual segmentation. Sen., Spec., and DSC denote sensitivity, specificity, and Dice similarity coefficient, respectively. Reprinted with permission.

Our method can also localize multiple renal mass in a single slice. We investigated the performance of our method for renal mass segmentation on the images with multiple tumors (34 out of 125 cases) on the institutional dataset. Examples of a slice with more than one tumor for five patients with different types of renal mass are displayed in Figure 4-12. As in Figure 4-13, the segmentation accuracy of renal mass for patients with multiple renal mass does not differ from that of the entire testing sample. The mean value of sensitivity for renal mass detection is higher than that for the whole test cases. Our method yielded a patient-wise sensitivity of 100% with multiple tumor (zero missing rate).

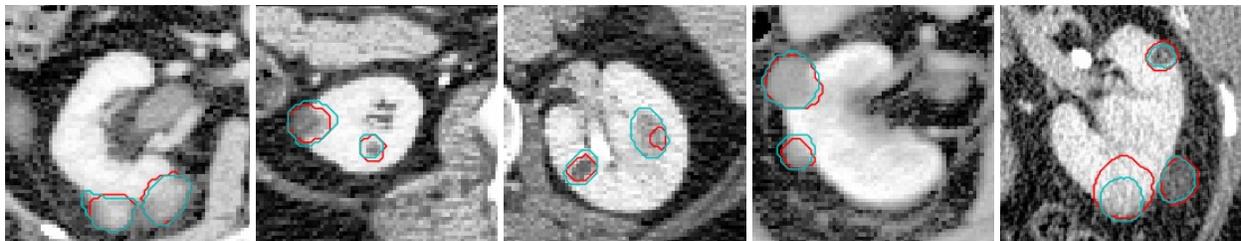


Figure 4-12. Example results from localized multiple renal masses on CECT images in five patients with different types of renal masses including from left to right clear cell RCC, chromophobe RCC, papillary RCC, fpAML, and oncocytomas. In all images, the renal masses contours computed from algorithm-generated and manual segmentations are shown in red and cyan, respectively. Reprinted with permission.

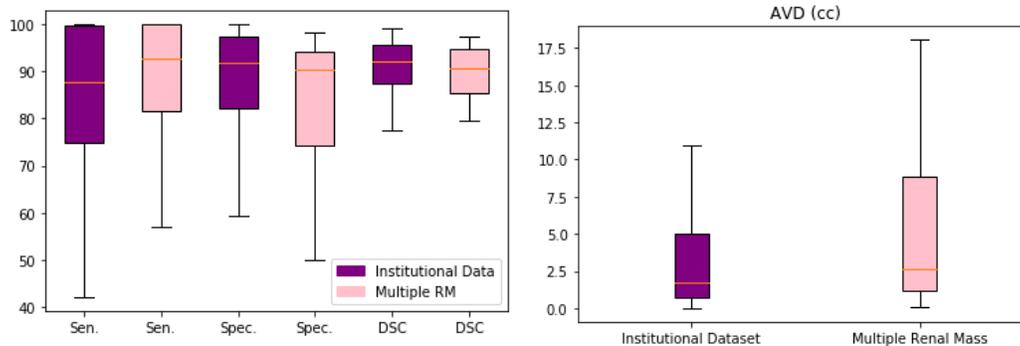


Figure 4-13. Boxplots of results generated by our proposed method for localization of renal mass on CECT images of the institutional dataset in comparison to expert manual segmentation for patients with multiple RM. Sen., Spec., and DSC denote sensitivity, specificity, and Dice similarity coefficient, respectively.

Reprinted with permission.

The semi-automated majority voting-based CNN algorithm for diagnosis of RCC versus benign among solid renal masses yielded accuracy, precision, and recall of 83.75%, 89.05%, and 91.73% on 160 CECT test images (benign/ RCC= 27/133), respectively. Using this methodology, 89.05% of the RCC renal masses have been correctly classified (PPV).

Figure 4-14 shows the ROC curve for the diagnosis of RCC using our method. The AUC is reported as 0.670. The PR curve plot in Figure 4-15 displays the precision-recall for each threshold for the proposed model compared to a no skill one. In the PR curve, no-skill classifier changes based on the distribution of the positive to negative classes. It is a horizontal line with the value of the ratio of RCC cases in the dataset. The area under the PR curve is 0.938.

Generally, the closer a ROC and PR curves are to the upper left and right corners, the better the classifier is. Accordingly, the proposed method differentiates RCC from benign renal masses well but not perfectly, which is due to the poor performance of the ML approach for the diagnosis of fpAML. It might be related to the fact that we only used single-phase assessment at

portal-venous or nephrographic phase enhanced CT. Previously Hodgdon et al. showed that fpAML could be differentiated from RCC at unenhanced CT (using texture analysis) and it is known that fpAML avidly enhances and then washes out on multi-phase CECT studies [13], [26].

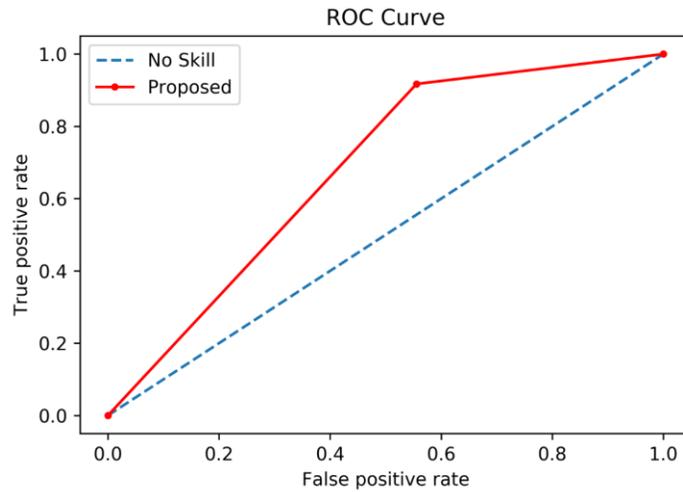


Figure 4-14. ROC curve for the proposed method, where the true positive rate (sensitivity) against false-positive rate (1 – specificity) has been plotted. Reprinted with permission.

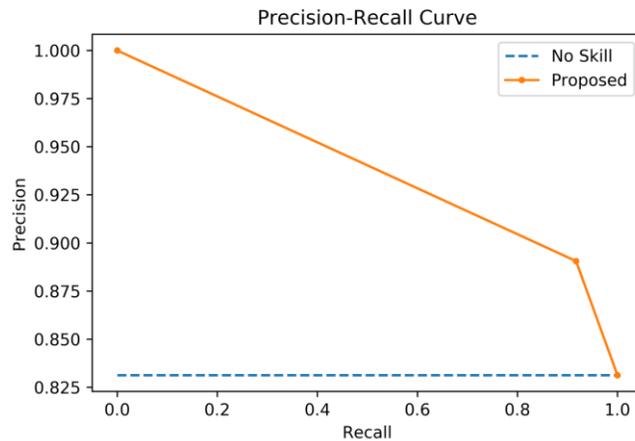


Figure 4-15. PR curve for the proposed method, where the precision against recall has been plotted. Reprinted with permission.

We labeled cases with equal probability of predicted labels in each class as RCC. This strategy was selected to maximize the sensitivity of the diagnosis of malignancy. Case by the case study of the results revealed that equal probability occurred for only two test cases within the papillary RCC cohort. Thus, the converse strategy of labeling ties as benign would have only slightly reduced the accuracy (83.75% for labeling in favor of RCC vs 82.50% labeling in favor of benign renal mass).

Moreover, we studied the intratumor heterogeneity to see how the slices compiled across a tumor have been labeled using the suggested methodology. The results revealed that the predicted labels for slices extracted from 63.13% of the test cases are homogeneous from that 80.20% of the cases have been correctly classified. From the rest part (36.87%) with heterogeneous labels, 89.83% have been correctly classified that further demonstrated the usefulness of applying a decision fusion system for final classification of the renal masses into RCC and benign particularly for uncertain cases.

We also trained a separate network using 120 training samples including oncocytoma, clear cell, chromophobe, and papillary (30 from each) to evaluate the performance of the proposed method for differentiation of RCC from oncocytoma. The results for 175 unseen test images are given in Table 4-5.

Table 4-5. Results of subgroup analysis for RCC versus oncocytoma. Reprinted with permission.

	Accuracy (%) [95% CI]	Precision (%) [95% CI]	Recall (%) [95% CI]
Semi-automated 2D CNN	79.31 [72.99 - 85.10]	86.49 [83.44 - 93.80]	88.89 [80.82 - 91.70]

Comparatively, the fully automated method achieved accuracy, precision, and recall of 76.88%, 85.29%, and 87.22% on the test images for categorizing solid renal masses into RCC and benign. The 3D CNN-based method yielded, accuracy, precision, and recall of 79.24%, 90.32%, and 84.21% for semi-automated classification of RCC and benign renal masses on CECT images. Table 4-6 shows a summary of the results for each method.

Table 4-6. Summary of evaluation results of the different methods of classification of CECT images into RCC and benign. Reprinted with permission.

Method	Accuracy (%) [95% CI]	Precision (PPV) (%) [95% CI]	Recall (%) [95% CI]
Semi-automated 2D CNN	83.75 [78.75 - 88.10]	89.05 [84.44 - 93.30]	91.73 [87.56 - 95.50]
Fully automated 2D CNN	77.36 [71.70 - 83.00]	85.92 [80.85 - 90.80]	87.22 [82.31 - 91.90]
Semi-automated 3D CNN	79.24 [73.59 - 84.30]	90.32 [85.71 - 94.50]	84.21 [78.91 - 89.30]

4.5 Discussion

In this study, we describe semi- and fully automated methods based on decision fusion of a slice-based CNN model to differentiate RCC from benign tumors among solid renal masses on CECT. Since we combined the predictions of slice-based CNN for classification, our proposed algorithm benefits from the majority voting system for final prediction. Also, we evaluated the performance of a 3D CNN for semi-automated classification of renal masses. Our proposed method has been evaluated on a large test dataset of consecutive solid renal masses and has achieved a promising result compared to previous works, where a smaller number of renal masses in less well-balanced populations were studied. We acknowledge that the absolute number of tumors in each category (for example fpAML); however, may be less than in previously reported studies. This is a limitation of the consecutive sample study design and reflects the low prevalence of fpAML (only 5% of all AML and representing <10% of small solid renal masses overall) seen in clinical practice. A strength of our work is an unbiased dataset reflective of clinical practice (i.e. tumors were identified consecutively); however, this strategy results in a greater proportion of ccRCC > pRCC > chRCC and oncocytoma > fpAML which is a reflection of disease distribution in clinical practice.

In our study, the semi-automated method outperformed the fully automated approach and 3D CNN algorithm for classifying CECT images into RCC or benign. The lower accuracy of the fully automated algorithm is likely due to false-positive renal masses predicted on CECT images using the method proposed in section 4.2.5. Using their suggested methodology, the sensitivity and specificity for automated renal mass localization reported as approximately 90% and 80% that demonstrates renal mass has been falsely detected on 20% of the slices, which in turn negatively influences the renal mass classification process. As in the proposed method, the slices

with tumor need to be separated and passed to the network to be labeled as RCC or benign, introducing slices without tumor to the CNN interferes with the network performance and majority voting system for final prediction. There is a cumulative penalty in accuracy by combining the automated renal mass detection and diagnosis, which is definitely not unexpected.

In recent years, automated assessment of renal masses has been evaluated using ultrasound, MRI, and CT images. Zhou et al. investigated the effect of transfer learning for RCC versus benign classification on CT images using a dataset comprised of 192 cases [17]. Their method reported the accuracy and Matthews correlation coefficient of 0.69 and 0.45 on 39 test images; however, the study is limited because it has not included any renal oncocytoma, which is by far the most difficult benign tumor to differentiate from RCC. The results from our study are more reflective of clinical practice since we included a consecutive sample of patients with both oncocytoma and fat poor AML in the benign cohort of solid masses. Our appropriate inclusion of renal oncocytoma almost certainly explains the reason for lower accuracy in our study compared to those reported by other authors; since it has been shown that fp-AML are less similar to RCC than oncocytoma [5], [10], [11], and [21] and therefore, including only fp-AML in the benign group of an experiment comparing RCC to benign solid renal masses biases the outcome. A strength of our study was the sample size and consecutive inclusion of solid renal masses encompassing RCC and the two most common benign histologies encountered since our sample was quite larger than comparable studies [13], [18] and [19-21].

Our study has some limitations. To minimize inter-observer variability, only one experienced radiologist performed manual segmentation in this study. Further studies are required to examine the inter-reader variability of humans. We only studied renal masses on single-phase enhanced

CT images, future studies might determine what benefit there is, if any, for summing information from a renal mass protocol CT which includes both an unenhanced, corticomedullary and nephrographic phase enhanced CT images.

4.6 Conclusion

A semi-automated method for classification of solid renal masses into benign and malignant subtypes was presented. Obtained results indicate that automated evaluation of renal mass on CT is currently possible with moderate-to-high degrees of accuracy. Also, we described a DL-based method for fully automated segmentation of kidneys and renal mass on CECT images without any user interaction. Evaluating the developed method on two large test datasets demonstrated that this algorithm is capable of accurate localization of renal mass on CECT images. Our results are clinically important since fully automated detection of renal mass is a fundamental step for automated assessment and classification of RM.

References

1. F. Zabihollahy, N. Schieda, S. Krishna, E. Ukwatta, “Automated classification of solid renal masses on contrast-enhanced computed tomography images using convolutional neural network with decision fusion,” *European Radiology*, 2020.
2. F. Zabihollahy, N. Schieda, S. Krishna, “E. Ukwatta, Ensemble U-Net-Based Method for Fully Automated Detection and Segmentation of Renal Masses on Computed Tomography Images,” *J. Med. Phys.*, 2020.
3. H. J. Meyer, A. Pfeil, D. Schramm, A. G. Bach, and A. Surov, “Renal incidental findings on computed tomography: Frequency and distribution in a large non selected cohort,” *Med. (United States)*, 2017.
4. S. D. O’Connor, P. J. Pickhardt, D. H. Kim, M. R. Oliva, and S. G. Silverman, “Incidental finding of renal masses at unenhanced CT: Prevalence and analysis of features for guiding management,” *Am. J. Roentgenol.*, 2011.

5. C. S. Lim, N. Schieda, and S. G. Silverman, "Update on indications for percutaneous renal mass biopsy in the era of advanced CT and MRI," *American Journal of Roentgenology*. 2019.
6. N. Schieda, R. S. Lim, M. D. F. McInnes, I. Thomassin, R. Renard-Penna, S. Tavolaro, F. H. Cornelis *et al.*, "Characterization of small (<4 cm) solid renal masses by computed tomography and magnetic resonance imaging: Current evidence and further development," *Diagnostic and Interventional Imaging*. 2018.
7. A. Udare, D Walker, S. Krishna, R. Chatelain, M. D. McInnes, T. A. Flood, N. Schieda, "Characterization of clear cell renal cell carcinoma and other renal tumors: evaluation of dual-energy CT using material-specific iodine and fat imaging," *Eur. Radiol.*, 2019.
8. S. A. Lee-Felker. E. R. Felker, N. Tan, D. J. Margolis, J. R. Young, J. Say J, S. S. Raman, "Qualitative and quantitative MDCT features for differentiating clear cell renal cell carcinoma from other solid renal cortical masses," *Am. J. Roentgenol.*, 2014.
9. R. S. Lim, T. A. Flood, M. D. F. McInnes, L. T. Lavalley, and N. Schieda, "Renal angiomyolipoma without visible fat: Can we make the diagnosis using CT and MRI?," *Eur. Radiol.*, 2018.
10. S. K. Kang, W. C. Huang, P. V. Pandharipande, and H. Chandarana, "Solid renal masses: What the numbers tell us," *American Journal of Roentgenology*. 2014.
11. N. Schieda, M. Al-Subhi, T. A. Flood, M. El-Khodary, and M. D. F. McInnes, "Diagnostic accuracy of segmental enhancement inversion for the diagnosis of renal oncocytoma using biphasic computed tomography (CT) and multiphase contrast-enhanced magnetic resonance imaging (MRI)," *Eur. Radiol.*, 2014.
12. Schieda N, Al-Subhi M, Flood TA, et al (2014) Diagnostic accuracy of segmental enhancement inversion for the diagnosis of renal oncocytoma using biphasic computed tomography (CT) and multiphase contrast-enhanced magnetic resonance imaging (MRI). *Eur Radiol*. <https://doi.org/10.1007/s00330-014-3310-y>.
13. T. Hodgdon, M. D. F. McInnes, N. Schieda, T. A. Flood, L. Lamb, and R. E. Thornhill, "Can quantitative CT texture analysis be used to differentiate fat-poor renal angiomyolipoma from renal cell carcinoma on unenhanced CT images?," *Radiology*, 2015.
14. N. Schieda, R. E. Thornhill, M. Al-Subhi, M. D. McInnes, W. M. Shabana, C. B. van der Pol, T. A. Flood, "Diagnosis of sarcomatoid renal cell carcinoma with CT: evaluation by qualitative imaging features and texture analysis," *Am. J. Roentgenol.*, 2015.
15. N. Schieda, R. S. Lim, S. Krishna, M. D. F. McInnes, T. A. Flood, and R. E. Thornhill, "Diagnostic accuracy of unenhanced CT analysis to differentiate low-grade from high-grade chromophobe renal cell carcinoma," *Am. J. Roentgenol.*, 2018.
16. K. Sasaguri, N. Takahashi, D. Gomez-Cardona, S. Leng, G. D. Schmit, R. E. Carter, *et al.*, "Small (< 4 cm) renal mass: Differentiation of oncocytoma from renal cell carcinoma on biphasic contrast-enhanced CT," *Am. J. Roentgenol.*, 2015.
17. L. Zhou, Z. Zhang, Y. C. Chen, Z. Y. Zhao, X. D. Yin, and H. B. Jiang, "A Deep Learning-Based Radiomics Model for Differentiating Benign and Malignant Renal Tumors," *Transl. Oncol.*, 2019.
18. S. P. Raman, Y. Chen, J. L. Schroeder, P. Huang, and E. K. Fishman, "CT texture analysis of renal masses: Pilot study using random forest classification for prediction of pathology," *Acad. Radiol.*, 2014.

19. H. Lee, H. Hong, J. Kim, and D. C. Jung, "Deep feature classification of angiomyolipoma without visible fat and renal cell carcinoma in abdominal contrast-enhanced CT images with texture image patches and hand-crafted feature concatenation," *Med. Phys.*, 2018.
20. H. S. Lee, H. Hong, D. C. Jung, S. Park, and J. Kim, "Differentiation of fat-poor angiomyolipoma from clear cell renal cell carcinoma in contrast-enhanced MDCT images using quantitative feature classification," *Med. Phys.*, 2017.
21. H. Coy, K. Hsieh, W. Wu, M. B. Nagarajan, J.R. Young, M. L. Douek, *et al.*, "Deep learning and radiomics: the utility of Google TensorFlow™ Inception in classifying clear cell renal cell carcinoma and oncocytoma on multiphasic CT," *Abdom. Radiol.*, 2019.
22. S. Krishna, C. A. Murray, M. D. McInnes, R. Chatelain, M. Siddaiah, O. Al-Dandan, *et al.*, "CT imaging of solid renal masses: pitfalls and solutions," *Clinical Radiology*. 2017.
23. M. Czarniecki, R. Gautam, P. L. Choyke, and B. Turkbey, "Imaging findings of hereditary renal tumors, a review of what the radiologist should know," *European Journal of Radiology*. 2018.
24. P. Y. Simard, D. Steinkraus, and J. C. Platt, "Best practices for convolutional neural networks applied to visual document analysis," in *Proceedings of the International Conference on Document Analysis and Recognition, ICDAR*, 2003.
25. M. D. Zeiler, "ADADELTA: an adaptive learning rate method.," *arXiv:1212.5701v1* 2012.
26. T. Saito and M. Rehmsmeier, "The precision-recall plot is more informative than the ROC plot when evaluating binary classifiers on imbalanced datasets," *PLoS One*, 2015.

Chapter 5: Fully Automated Segmentation of Left Ventricular Scar from 3D Late Gadolinium Enhancement Magnetic Resonance Imaging Using a Cascaded Multi-Planar U-Net (CMPU-Net)

In Chapter 3, we introduced a CNN-based method for semi-automated evaluation of renal masses on CT scans. Previously Zabihollahy et al. used a similar patch-based algorithm using 3D CNN to segment LV scar from pre-segmented LV myocardium [1]. Although their proposed method [1] is highly accurate for the segmentation of LV scar on 3D LGE-MR images, it requires the manual segmentation of the LV myocardium, which is tedious, time-consuming (≈ 54 minutes [2]) and subject to observer variability. To address this problem, a novel U-Net based algorithm is introduced and evaluated using a dataset of the patients with chronic ischemic cardiomyopathy to delineate the boundary of LV myocardium and LV scar tissues from 3D LGE MR images [3]. 3D MR images have isotropic voxels that allow multiplanar reformation [4]. We exploit this property of 3D MR images and introduce a novel DL-based method to delineate LV scar from 3D LGE-MR images.

5.1 Introduction

Cardiovascular disease (CVD) is the leading cause of death worldwide [5], ischemic heart disease being a dominant contributor [6]. These patients suffer irreversible myocardial necrosis from ischemia leading to the replacement of contractile muscle with a kinetic scar, inherently reducing cardiac function and leading to heart failure. Approximately one-third of these patients will also die suddenly from malignant ventricular arrhythmias, most commonly provoked by electrical re-entry circuits surrounding regions of myocardial scar. Recognition of these

relationships has expanded the emphasis on myocardial scar imaging to guide the management of patients with chronic ischemic cardiomyopathy. In particular, characterizing myocardial scar volume, as well as its heterogeneity and geographic distribution, may aid clinicians in determining the appropriateness and procedural approach to percutaneous ablation (aimed at eliminating electrical channels) and cardiac resynchronization therapy [7-8]. Recently, computational modeling of patient with ischemic cardiomyopathy has emerged as a promising non-invasive tool that allows clinicians to conduct a patient-specific diagnosis and treatment of associated rhythm disorders [9-10]. However, accurate remodeling of myocardial structural requires incorporation of intact geometry of the scar region.

2D LGE MR imaging is the established reference standard for myocardial scar imaging, identifying the latter through enhancement from the retention of gadolinium-based contrast agents [11]. However, the use of 3D acquisition techniques is rapidly growing due to advancements in both hardware technology and experience, as well as strong interest from the electrophysiology community to use such images to guide interventions [12]. This has been catalyzed by the capacity of 3D LGE-MR images to provide improved spatial resolution while enabling volumetric reconstruction of scar geometry [13-16]. Infarct masses determined by improved spatial resolution have also been shown to be better predictor of clinical outcomes [17]. Compared to 2D techniques, 3D LGE-MR images have been shown to provide higher signal intensity and contrast for myocardial scar with reduced overall acquisition times [18-20]. However, segmentation of scar from 3D LGE-MR images is highly tedious, time-consuming (≈ 42 minutes [20]) and subject to high observer variability. Therefore, a fully automated method is quite desirable for this task. However, this is particularly challenging due to complex cardiac

structure that is variable across basal, mid and apical regions of the LV, poor image contrast at the myocardial-blood boundary, as well as a higher prevalence of image artifacts [21-22].

Previous studies evaluating LV scar segmentation from 3D LGE-MR images can be divided into three main categories: i) intensity thresholding-based (i.e., full width at half maximum (FWHM) or signal threshold to reference mean (STRM)), ii) energy minimization-based (i.e. hierarchical maximum flow (HMF) and convex max flow (CMF)), and iii) DL-based (i.e., CNN) methods. Except for the HMF, all the other methods require manual segmentation of the myocardial borders to constrain an ROI for scar segmentation. Once myocardial boundaries are defined, the FWHM approach identifies the maximum intensity value of LV scar tissue (I_r) and considers this as a reference with scar classified as all myocardial tissue with signal intensity greater than $I_r / 2$ [23]. In the STRM-based approach, the scar is considered as any regional signal above a mean intensity value of remote myocardium plus two (STRM2), three (STRM3), four (STRM4), five (STRM5), or six (STRM6) standard deviations [24]. Although the intensity thresholding-based methods are very simple to implement, they maintain reliance on pre-segmentation of the myocardial borders. In contrast, the HMF technique, a multi-label convex optimization technique introduced by Rajchl et al. [18], [25] identifies scar without such requirements by using an intensity log-likelihood criterion as a data term. This said it does still require the manual application of reference ROIs to inform what the expected signal histogram is of normal myocardium and all visible scarred regions, taking on average seven minutes of expert time and being subject to high-operator variability [13]. Ukwatta et al. introduced CMF in which the infarct segmentation is expressed as a continuous min-cut optimization problem, which is then solved using its dual formulation [26-27]. The CMF is also semi-automated and requires manual segmentation of LV myocardium.

Zabihollahy et al. previously described a method based on 3D CNN to segment LV scar from 3D LGE-MR images [28-29]. Their method uses manually segmented LV myocardium as a ROI and employs a local region (patch) around each voxel that is extracted in a sliding window fashion and passed through the network to be labeled. The assigned label to the patch is considered as a voxel label to generate the output segmentation map for a given image. This patch-based segmentation addresses concerns of the limited number of training labeled datasets, which is a common concern for DL techniques in medical imaging. However, this method suffers from redundant computation due to the overlap between adjacent patches. Moreover, global features are not captured when small patches are extracted, and localization error is increased when large patches are used [30]. Finally, extracting image patches is time-consuming, requires prior constraint of the target tissue (i.e. delineation of the myocardial boundaries), and overall causes this process to remain slow. As the primary goal of medical image analysis is to achieve both speed and user independence, this approach remains sub-optimal.

In this chapter, we describe a cascaded multi-planar U-Net-based method (CMPU-Net) for rapid, fully automated segmentation of myocardial scar from 3D LGE-MR images with no user interaction. U-Net was introduced by O. Ronneberger et al. [30] in which global and local features of an image are learned concurrently to retain high spatial consistency. This model is built using fully CNN (FCNN) and has been shown to be more accurate for numerous organ segmentation tasks across different modalities [31-33]. Moreover, cascaded 2D and 3D FCNNs have been used for regional segmentation of several anatomical structures (e.g. prostate, artery, portal vein, lungs, liver, spleen, stomach, gallbladder, and pancreas) [34].

Our main contribution is the development of a new joint-feature extraction strategy to comprehensively learn and integrate inter- and intra-slice features from 3D LGE-MR images, allowing for efficient context-based segmentation of myocardial tissue and scar.

5.2 Method

Figure 5-1 shows the workflow of CMPU-Net, which consists of two cascaded subnets (Myo-Net and Scar-Net) as the scar is enclosed within the LV myocardium. By segmenting the LV myocardium first, we may potentially be able to exclude other hyper-enhanced regions of the image, such as pericardial fat and blood pool.

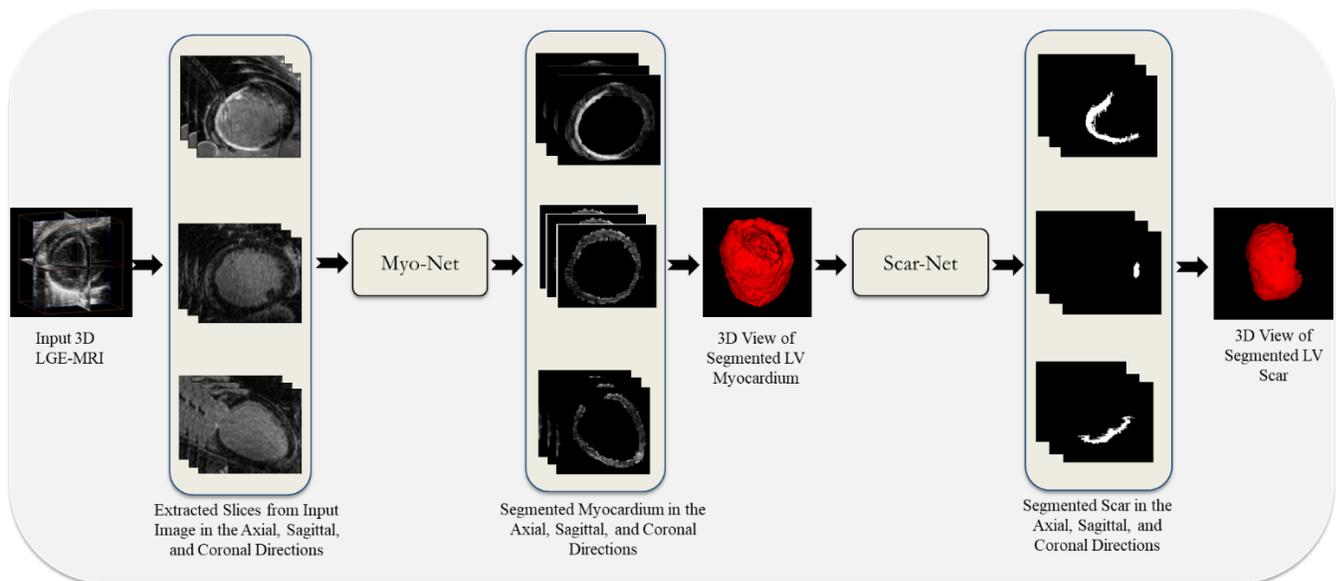


Figure 5-1. Overview of our proposed method, CMPU-Net, for fully automated myocardial scar segmentation. Reprinted with permission.

A schematic diagram of the CMPU-Net framework is given in Figure 5-2. In the training phase, three U-Nets were trained using 2D slices extracted from 18 LGE-MR images in three orthogonal views. In the testing phase, 2D slices extracted from test images were fed to the

trained network separately to generate the corresponding segmentation maps, which were then combined through a majority voting system for prediction of output label.

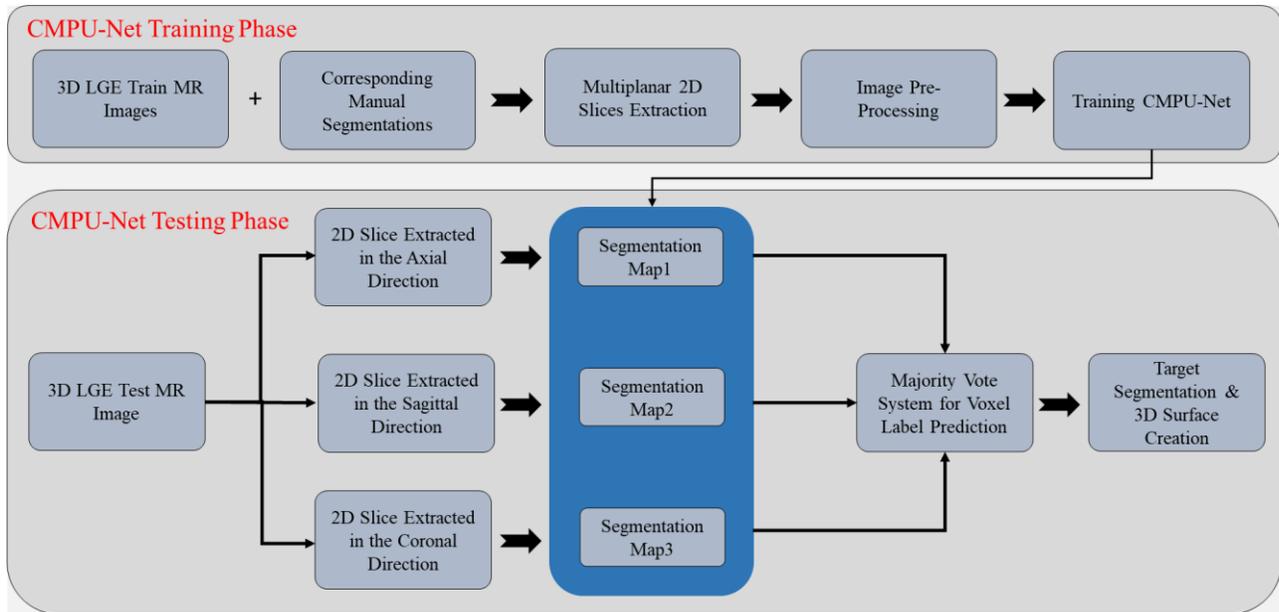


Figure 5-2. Block diagram of CMPU-Net framework. Reprinted with permission.

5.2.1 Study Subjects and Data Acquisition

Our study consisted of 34 subjects with known chronic ischemic cardiomyopathy. By definition, these subjects were a minimum of 3-months from any clinically recognized acute coronary syndrome. Therefore, all myocardial scar visualized by LGE imaging was reasonably assumed to be “mature” (i.e. healed). All subjects provided written informed consent and the study protocol was approved by the Research Ethics Board at Western University. All subjects underwent a standardized imaging protocol inclusive of cine imaging, Clinical Evaluation coronary MRI and post-contrast 3D LGE at the Robarts Research Institute (London, ON, Canada). The latter two image datasets were acquired using a whole-heart, respiratory navigated, 3D inversion-recovery gradient echo pulse sequence performed during and 30 minutes following infusion of 0.2

mmol/kg Gadovist (Bayer, Toronto, ON, Canada), as previously described [35]. The imaging hardware and pulse sequence parameters are described in Table 5-1.

Table 5-1. Image acquisition parameters. Reprinted with permission.

Manufacturer	Siemens Medical
Model	MAGNETOM Trio
Field Strength	3 Tesla
Echo Time	1.3 ms
Flip Angle	20°
Pulse Sequence	Inversion-recovery gradient echo
Voxel size	0.625 x 0.625 x 1.3 mm

The average age and body mass of the population, who completed the imaging protocol was 51.5 ± 12.6 years and 28.2 ± 4.7 kg/m². The mean of heart rate and the ventricle ejection fraction (LVEF) at the time of imaging were 67.1 ± 11.2 beats per minute (BPM) and $32.1 \pm 12.7\%$, respectively. The mean volume of LV myocardium and myocardial scar in the test population was 257.43 ± 55.91 cm³ and 45.97 ± 26.41 cm³, respectively.

5.2.2 Manual Segmentation for Training and Testing

The boundaries of myocardial and scar in all images in the dataset were manually segmented by an expert using a 3D brush in a multi-planar view in ITK-SNAP Software [36] that took 54 and 42 minutes respectively [20]. An interactive algorithm [37] was employed to segment the border of epicardial and endocardial as the manual delineation of myocardial tissue on the isotropic dataset was impractical. The segmentation results were then refined by an experienced

radiologist. Papillary muscles, and/or mural thrombus, and valvular tissue, if present, were excluded from the segmentation.

5.2.3 Architecture of CMPU-Net

Each subnet was trained separately. The number of trainable parameters for each network in Myo-Net and Scar-Net was $\approx 1.9M$ and $\approx 7.8M$. Although the number of parameters for a network within each subnet was the same, the weights were different as each network was trained using slices from one of the orthogonal views. The U-Net consists of contracting and expanding paths, where pooling and up-sampling layers are used in each way that yields a U-shaped architecture. The contraction path is identical to that of standard CNN in which convolutional layers along with pooling and activation layers are applied to the input data. In expanding path, pooling layers are replaced by up-sampling layers to expand the dimension of feature space. The output of up-sampling layers is merged with appearance feature representation learned from the corresponding layer in the shrinking path to localize high-resolution features. The most important modification of this architecture is that convolutional layers are used in expanding path as well that allows learning features to propagate to the higher resolution layers [30]. U-Net does not have a fully connected layer at the end but generates a probability output to the size of the input image [30].

The contraction path of our U-Net consists of five stages in each two convolutional layers of size 3×3 followed by a rectified linear unit (ReLU) activation layer and one max pooling layer of size 2×2 were applied. 32 filters were used in the first level of Myo-Net and the number of filters was doubled in each subsequent stage. In the Scar-Net, 16 filters were employed in the first stage of the network and similarly, the number of filters was doubled in the next stages. To maintain the

activation function invariant to scaling of the weights [38], we conducted batch normalization with a batch size of 10 after each convolutional layer. Batch normalization also leads to faster training by enabling smoother convergence in optimization [39]. We used a dropout layer at the end of the contraction path to minimize overfitting, where neurons are randomly dropped out of the network during training [40]. At the final stage, a convolutional layer of size 1×1 followed by a sigmoid activation layer was applied to all feature maps to generate a probability map to the size of the input image. The probability map was then thresholded such that each voxel with a probability of greater than 0.5 was labeled as the desired object (i.e., myocardium for Myo-Net and scar for Scar-Net).

5.2.4 Training the CMPU-Net

As the first step, a median filter of size 3×3 pixels was applied to each slice to minimize noise. Pixel intensities of the images were then normalized to a range of [0-1]. Image normalization was performed for all 2D slices extracted from 3D training images in axial, sagittal, and coronal directions. The 3D LGE-MR images were then divided into 18 and 16 subjects including a total of 8430 and 7250 2D slices for training and testing the networks, respectively.

We used the DSC, which quantifies the spatial overlap between algorithm-generated and manual segmentations. The DSC is used both as a loss function and evaluation metric for segmentation. The DSC as a loss function has led to fast and accurate segmentation [41]. We used AdaDelta as the optimizer, which employs an adaptive learning rate method for gradient descent as choosing a small value for learning rate slows down the algorithm and large values cause parameters oscillation around local minima [42].

Morphological cleaning was applied to the output segmentation map to remove false positives that are small in size. All sparse small segmented objects fewer than 100 pixels (39.06 mm^2) were removed from the segmented results. To reduce the sharp edges, segmented scar 3D scar volume was smoothed via mathematical erosion operation followed by dilation. Employing a combination of these removes the patchy and sharp edges while avoiding target to become smaller.

We implemented the algorithm in Python using Keras library, on top of Tensorflow. We trained each network for 50 epochs on Intel Core i7, 2.8 GHz on a GPU-accelerated computing platform. We monitored DSC on a validation set that was 20% of training data samples during training and saved the latest best model according to the monitored quantity. In our experiments, the best-saved model performed better compared to the models saved after 50 epochs.

5.2.5 Comparison of Our Method to Alternative Methods

We compared the results of our method to those of several alternative techniques, including cascaded conventional U-Net (CCU-Net), direct multiplanar U-Net (DMPU-Net), direct conventional U-Net (DCU-Net), 3D U-Net, hierarchical max-flow (HMF), FWHM, and STRM. The LV scar was segmented using those methods from the same test dataset. Out of all the alternatives, only the HMF, FWHM, and STRM methods have been previously reported for scar segmentation.

We chose the CCU-Net as an alternative method, which has been successfully applied to numerous segmentation tasks [33], [43], to compare uniplanar vs multi-planar networks for scar segmentation. The CCU-Nets is a fully automated pipeline in which two U-Nets were used in

series to segment the myocardium and scar in a cascaded manner [44]. In CCU-Net, the network was trained and tested using 2D slices in the axial direction.

In the manual delineation of scar from LGE-MR images, surrounding information outside the myocardial boundary influences the expert's decisions. To investigate whether such information influences the automated algorithm to the same extent, we included direct segmentation methods in our comparison. DMPU-Net and DCU-Net are similar in network architecture to CMPU-Net and CCU-Net but have been applied to directly segment scar from LGE-MR images without using pre-segmented LV myocardium. Furthermore, in the DMPU-Net approach, a voxel-by-voxel summation was employed to combine the results of each orthogonal view to establish a final label prediction. In this rule, the voxel was labeled as scar when at least one of the constituent networks predicted the label of the voxel as a scar.

We also used 3D U-Net to directly segment the scar. The 3D U-Net is a widely used network for segmentation, which is an extension of the U-Net architecture from Ronneberger et al. in which all 2D operations are replaced with their 3D counterparts [30]. We used a 3D U-Net with five stages in which 16 filters of size $3 \times 3 \times 2$ were used in the first stage of the network and the number of applied filters was doubled in the four subsequent layers. As this method is computationally expensive, we ran all our experiments on images down sampled by a factor of two. The network was trained for 30 epochs within 150 minutes.

We used the same dataset and the results were post-processed similar to what was applied for CMPU-Net, where the smallest allowable size for the segmented myocardium/myocardial scar was set to 100 pixels and the boundary of segmented regions was smoothed. Identical to the

evaluation of our method, the algorithm-generated results were compared to manual segmentations.

5.3 Results

Figures 5-3 and 5-4 show the LV myocardium and scar segmentation results of the CMPU-Net for three and five subjects, randomly selected from the test dataset. The algorithm-generated segmentations closely match the manually segmented boundaries for both LV myocardium and scar. Table 5-2 shows a summary of the results of our method. The test results suggested that CMPU-Net is capable of accurately segmenting LV myocardium with a mean DSC, AVD, and HD of $85.14\% \pm 3.36\%$, $43.72 \pm 27.18 \text{ cm}^3$, and $19.21 \pm 4.74 \text{ mm}$, respectively. In testing, our method also reported DSC, AVD, and HD of $88.61\% \pm 2.54\%$, $9.33 \pm 7.24 \text{ cm}^3$, and $17.04 \pm 9.93 \text{ mm}$ for LV scar segmentation for the test images, respectively. Once trained, the average running times required to segment LV myocardium and scar in a typical 3D LGE-MR test image were $49.96 \pm 9.76 \text{ sec}$ and $120.45 \pm 23.34 \text{ sec}$, which are substantially faster than manual segmentation (54 and 42 minutes, respectively [20]).

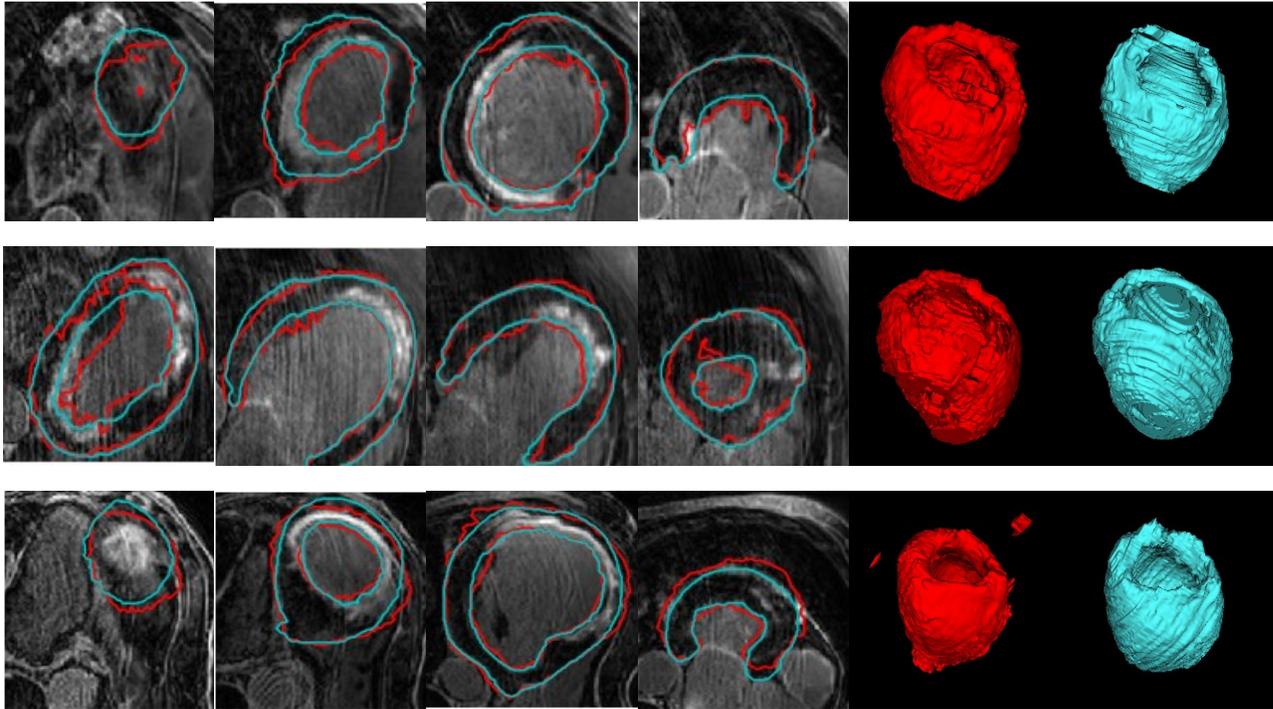


Figure 5-3. In the first four columns, exemplary results of segmented LV myocardium tissue in different slices extracted from 3D LGE-MR image of three subjects are displayed. The contours computed from the CMPIU-Net method and manual segmentations are displayed in red and cyan. The fifth and sixth columns indicate the 3D view of the myocardial regions generated by our proposed technique and expert manual segmentation, respectively. Reprinted with permission.

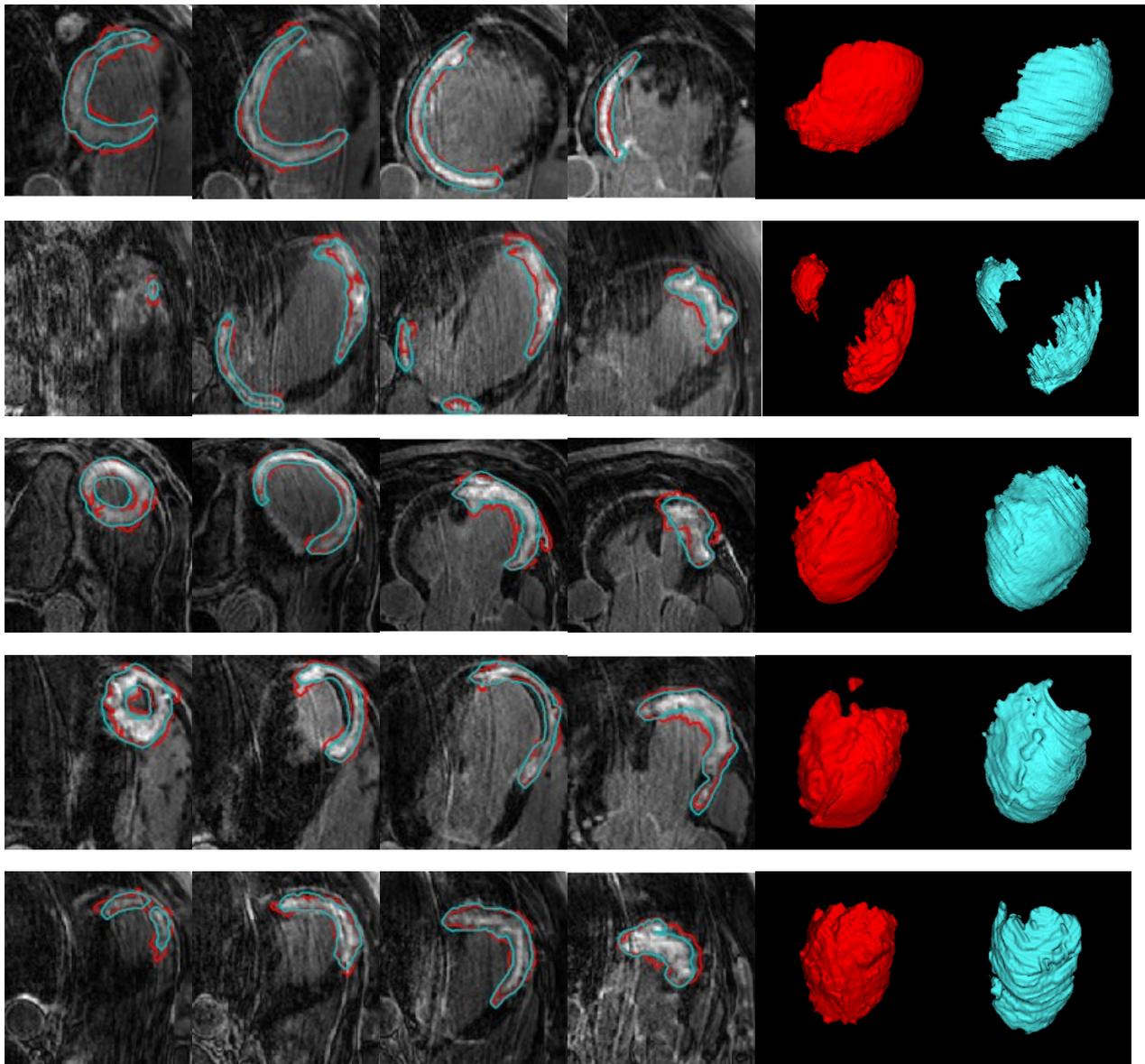


Figure 5-4. In the first four columns, exemplary results of segmented scar tissue in different slices extracted from 3D LGE-MR image of five subjects are displayed. The contours computed from the CPMU-Net method and manual segmentations are displayed in red and cyan. The fifth and sixth columns indicate the 3D view of the myocardial scar regions generated by our proposed technique and expert manual segmentation, respectively. Reprinted with permission.

Table 5-2. Summary of results of our proposed technique, CMPU-Net, for LV myocardium and LV scar segmentation as the mean value \pm SD, 25th and 75th percentiles on 3D LGE- MRI against expert manual segmentation. Reprinted with permission.

	LV Myocardium			LV Scar		
	Mean \pm SD	25 th Percentile	75 th Percentile	Mean \pm SD	25 th Percentile	75 th Percentile
DSC (%)	85.14 \pm 3.36	84.18	86.37	88.61 \pm 2.54	87.24	90.95
Acc. (%)	92.30 \pm 2.71	91.28	93.79	98.30 \pm 0.71	97.70	98.69
Prec. (%)	84.43 \pm 2.56	83.27	86.76	87.79 \pm 3.35	85.81	90.00
Rec. (%)	87.36 \pm 4.51	85.70	90.66	90.68 \pm 2.46	88.90	92.54
HD (mm)	19.21 \pm 4.74	16.80	21.21	17.04 \pm 9.93	11.39	19.40
AVD (cm³)	43.72 \pm 27.18	25.90	54.71	9.33 \pm 7.24	5.73	10.35
Required Time (s)	49.96 \pm 9.76	44.93	57.94	120.45 \pm 23.34	43.55	57.34

The volume of LV scar computed from the CMPU-Net algorithm and expert manual segmentations were compared and the p-value was estimated employing Wilcoxon rank-sum as 0.4856. Accordingly, we failed to reject the null hypothesis at the 5% significance level.

Therefore, proposed CMPU-Net could be employed as a replacement of manual segmentation for LV scar volume measurement. This conclusion is valid considering high segmentation accuracy reported using DSC, which penalizes for the discrepancy in shape and spatial location.

Figure 5-5 shows the Bland Altman plots for CMPU-Net versus manual LV scar volume. In this graph, the solid blue line, upper and lower dashed lines represent the mean value of the difference, the upper and lower limits of agreement, respectively. In comparison to manually

computed volume, the average bias of our method in estimating LV scar volume was 7.58 cm^3 . Thus, our proposed technique slightly overestimated the LV scar volume.

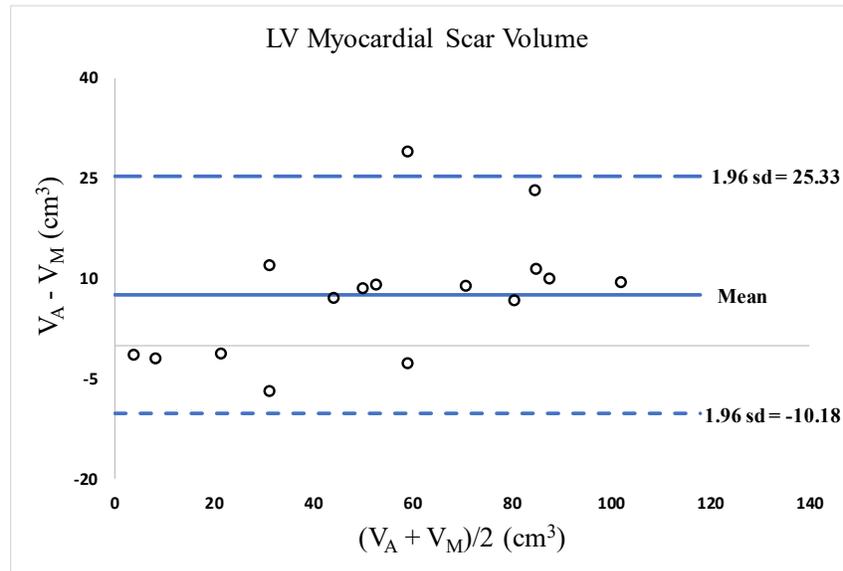


Figure 5-5. Bland-Altman plot of LV scar volume obtained from our proposed technique. The plot graphs the difference between the algorithm-generated and manual target volumes as a function of their mean. Reprinted with permission.

We computed the Pearson correlation coefficient (r) between DSC of LV myocardium and LV scar segmentation on the test population to quantitatively assess the impact of myocardium segmentation on the final scar segmentation. To this end, the covariance of DSC of myocardial and scar was divided into the product of their standard deviations. r -value was reported as 0.503, which indicates a moderate positive correlation between those variables. Accordingly, the segmentation accuracy of myocardial boundaries somewhat influences the segmentation accuracy of the scar. Moreover, we studied the effects of scar presence on myocardium segmentation. Case-by-case study of algorithm-generated segmentation showed that decrease in

myocardium segmentation accuracy is not caused by misclassification in regions where scar is observed. Therefore, using our proposed algorithm, despite potential poor segmentation outside the myocardium, the scar tissue can be correctly identified.

Our alternative methods include CCU-Net, DMPU-Net, DCU-Net, 3DU-Net, HMF, FWHM, and STRM>2SD to 6SD. Figure 5-6 shows the segmented LV scar for all the tested methods for a 3D LGE-MR image chosen randomly. 2D slices were selected from basal, apical and middle level to provide a level-based visual comparison of the different algorithms as well. We included only the results for STRM>5SD as it yielded the best DSC among the STRMs. In comparison to other methods, our method segmented scar more accurately, particularly at the basal and apical slices. Table 5-3 summarizes the quantitative results. Our proposed method yielded significantly higher DSC than that of alternatives for fully automated segmentation of myocardial scar. The next best DSC was achieved using the CCU-Net technique ($85.69\% \pm 4.20\%$). Image threshold-based methods reported the most inferior DSC.

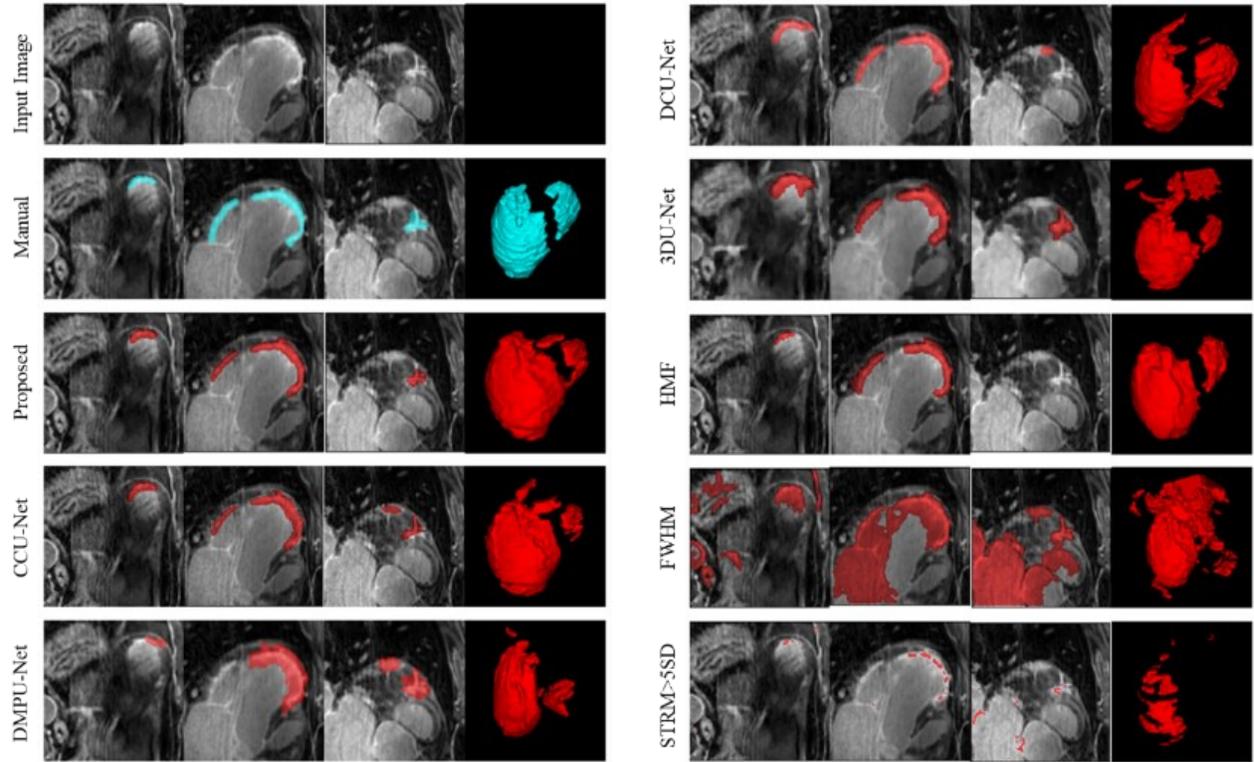


Figure 5-6. Exemplary results of 3D scar segmentation methods applied for an example of 3D LGE-MR image, presented in orthogonal views (Column 1-3) and using a surface rendering. The last column indicates the 3D view of the myocardial scar regions. The results of each method are shown in the corresponding row. Reprinted with permission.

Table 5-3. Performance of all 3D scar segmentation methods for LV scar segmentation on 3D LGE-MR images against expert manual segmentation. The asterisk denotes statistical significance using the Wilcoxon rank-sum test performed to compare the average value of DSC reported from our method (CMPU-Net) against those of alternatives. Reprinted with permission.

		CMPU-Net (Proposed)	CCU-Nets	DMPU-Net	DCU-Net	3DU-Net	HMF	FWHM	STRM>5SD
DSC (%)	Mean	88.61	*85.69	*85.19	*83.48	*72.39	*77.41	*64.16	*64.07
	SD	2.54	4.20	4.59	4.81	7.02	3.05	7.65	7.14
	25th Percentile	87.24	83.60	82.96	80.35	67.13	75.88	57.17	61.88
	75th Percentile	90.95	88.37	88.32	87.33	77.23	80.24	71.39	69.02
AVD (cm³)	Mean	9.33	7.96	6.96	7.37	31.26	12.05	91.96	74.16
	SD	7.24	5.41	5.92	5.96	21.27	5.49	77.46	102.22
	25th Percentile	5.73	4.22	2.97	3.20	11.63	9.54	22.91	28.47
	75th Percentile	10.35	11.28	10.34	8.84	46.97	17.11	128.44	62.11
HD (mm)	Mean	17.04	17.49	13.12	12.62	19.61	10.75	42.76	33.85
	SD	9.93	11.38	4.80	11.78	10.74	5.45	20.52	26.68
	25th Percentile	11.39	9.36	9.75	7.38	12.14	7.93	27.60	10.47
	75th Percentile	19.40	22.31	16.44	10.98	25.19	9.79	47.40	53.52

Although we had a limited number of test images, the LV scar was sufficiently variable in terms of myocardial scar volume, shape, location, and signal characteristics (Figure 5-7).

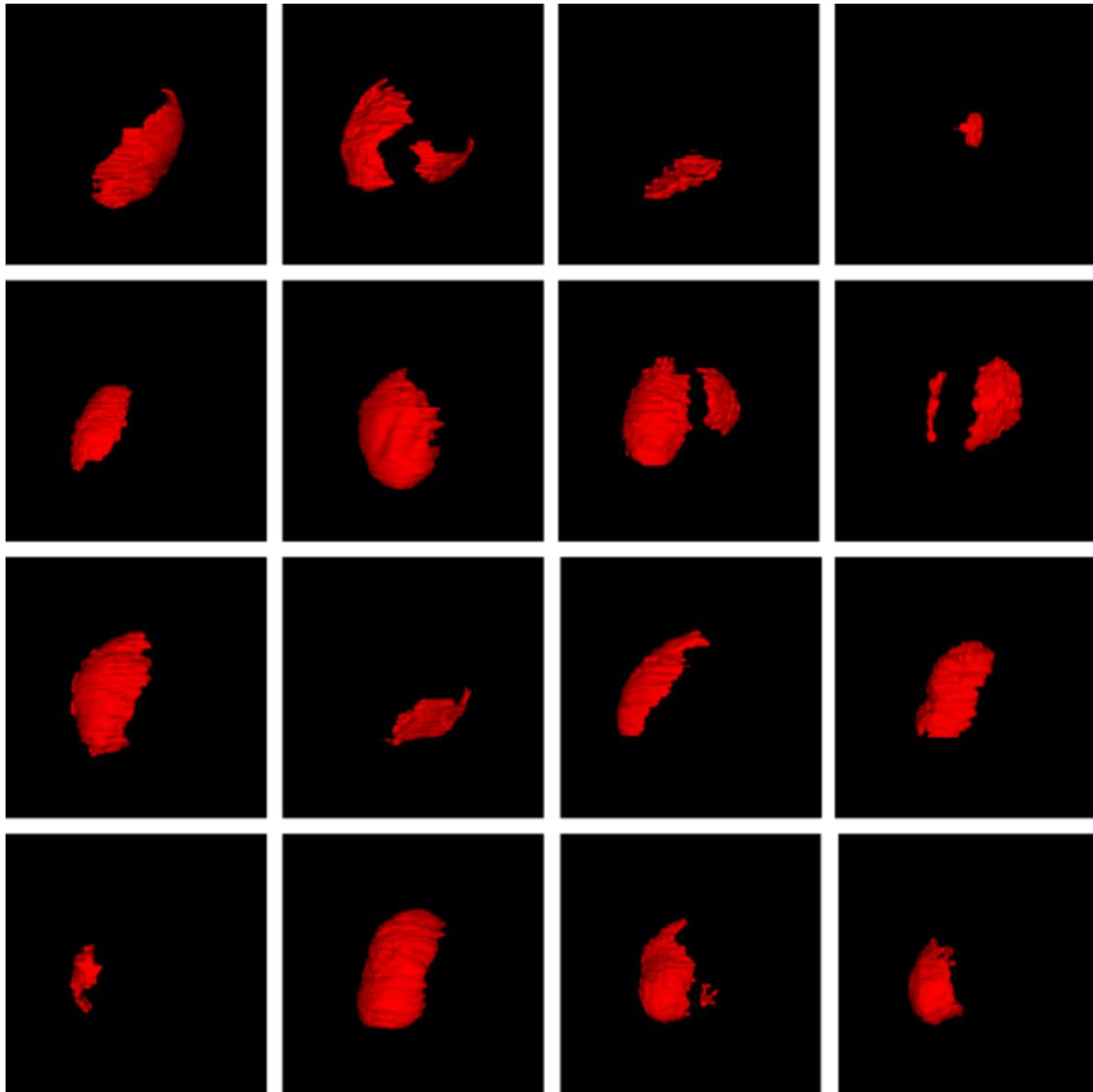


Figure 5-7. 3D view of LV scar in the LGE-MR test images. Reprinted with permission.

5.4 Discussion

In this chapter, we described CMPU-Net, a method to fully automatically segment LV scar by integrating information along three orthogonal planes of a 3D LGE-MR image. As compared to numerous alternative methods, our method reported significantly higher DSC. As 2D U-Net takes a single slice as input, it inherently fails to leverage context from adjacent slices. Our suggested methodology addresses this issue by applying 2D convolutional kernels to the image slices from three orthogonal views to make segmentation predictions for a volumetric scan with a computational cost as a result of the increased number of parameters used by the model.

In this work, we also developed a fully automated method for LV myocardium segmentation. Although several methods have been reported for fully automated myocardial segmentation in 3D LGE MRI, fully automated segmentation of scar in 3D LGE MRI has not been reported. Tao et al. developed an algorithm based on the global registration of the Cine-MRI and LGE-MRI volumes using a cohort of 50 subjects (25 ischemic) to segment LV myocardium from 3D LGE-MRI [45]. Their method yielded the DSC of $81\% \pm 7\%$ and $83\% \pm 9\%$ for the two observers who manually segmented the LV in LGE-MR images. Kurzendorfer et al. utilized a two-step registration-based method to delineate the LV myocardium and evaluated their proposed method on 30 clinical 3D LGE-MRI datasets from individual subjects obtained at two different clinical sites, which reported DSC of 83% and 80% for the endocardium and epicardium, respectively [46].

We used a majority voting system for final label prediction from three orthogonal views that resulted in more confident predictions. This is particularly useful in ambiguous cases, where suppression of false-positive errors can be achieved. Our results demonstrated that multi-planar

slice extraction not only helps to maintain spatial consistency across adjacent slices but also increases the number of extracted 2D slices for training. Combining this architecture with the cascaded U-Net model further improves the localization accuracy for segmentation tasks.

One limitation of our proposed method is that it is slower than other conventional algorithms as extracting slices from three different directions of a 3D image takes time. Furthermore, as we used cascaded U-Nets to segment scar within segmented LV myocardium, incorrect segmentation of LV myocardium will adversely influence the segmentation accuracy of scar.

5.5 Conclusion

We described a novel DL-based method for fully automated segmentation of LV scar from 3D LGE-MR images without any user interaction. Through simultaneous delineation of global myocardial architecture and scar, this paradigm is relevant to the clinical translation of image-guided therapeutics that are reliant upon the reproducible representation of scar distribution for their success.

References

1. F. Zabihollahy, J. A. White, and E. Ukwatta, "Convolutional neural network-based approach for segmentation of left ventricle myocardial scar from 3D late gadolinium enhancement MR images," *Med. Phys.*, 2019.
2. M. Rajchl, J. Yuan, J. A. White, E. Ukwatta, J. Stirrat, C. M. Nambakhsh, *et al.*, "Interactive hierarchical-flow segmentation of scar tissue from late-enhancement cardiac mr images," *IEEE Trans. Med. Imaging*, 2014.
3. F. Zabihollahy, M. Rajchl, J.A. White, E. Ukwatta, "Fully Automated Segmentation of Left Ventricular Scar from 3D Late Gadolinium Enhancement Magnetic Resonance Imaging Using a Cascaded Multi-Planar U-Net (CMPU-Net)", *J. Med. Phys.*, 2020.
4. D. Saloner, "An introduction to MR angiography.", *Radiographics* 1995.

5. D. Mozaffarian, "Global Scourge of Cardiovascular Disease: Time for Health Care Systems Reform and Precision Population Health," *Journal of the American College of Cardiology*. 2017.
6. A. Schmidt, C. F. Azevedo, A. Cheng, S. N. Gupta, D. A. Bluemke, T. K. Foo, *et al.*, "Infarct tissue heterogeneity by magnetic resonance imaging identifies enhanced cardiac arrhythmia susceptibility in patients with left ventricular dysfunction," *Circulation*, 2007.
7. D. Bello, D. S. Fieno, R. J. Kim, F. S. Pereles, R. Passman, G. Song, *et al.*, "Infarct morphology identifies patients with substrate for sustained ventricular tachycardia," *J. Am. Coll. Cardiol.*, 2005.
8. R. J. Kim, D. S. Fieno, T. B. Parrish, K. Harris, E. L. Chen, O. Simonetti, *et al.*, "Relationship of MRI delayed contrast enhancement to irreversible injury, infarct age, and contractile function," *Circulation*, 1999.
9. E. Ukwatta, P. Nikolov, F. Zabihollahy, N. A. Trayanova, and G. A. Wright, "Virtual electrophysiological study as a tool for evaluating efficacy of MRI techniques in predicting adverse arrhythmic events in ischemic patients," *Phys. Med. Biol.*, 2018.
10. E. Ukwatta, P. Nikolov, N. Trayanova, and G. Wright, "Computational heart modeling for evaluating efficacy of MRI techniques in predicting appropriate ICD therapy," in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 2018.
11. E. Ukwatta, H. Arevalo, M. Rajchl, J. A. White, F. Pashakhanloo, A. Prakosa, *et al.*, "Image-based reconstruction of three-dimensional myocardial infarct geometry for patient-specific modeling of cardiac electrophysiology," *Med. Phys.*, 2015.
12. K. Kawaji, A. Tanaka, M. B. Patel, H. Wang, F. Maffessanti, T. Ota, *et al.*, "3D late gadolinium enhanced cardiovascular MR with CENTRA-PLUS profile/view ordering: Feasibility of right ventricular myocardial damage assessment using a swine animal model," *Magn. Reson. Imaging*, 2017.
13. M. Rajchl, J. Stirrat, M. Goubran, J. Yu, D. Scholl, T. M. Peters, *et al.*, "Comparison of semi-automated scar quantification techniques using high-resolution, 3-dimensional late-gadolinium-enhancement magnetic resonance imaging," *Int. J. Cardiovasc. Imaging*, 2015.
14. E. Ukwatta, H. Arevalo, K. Li, J. Yuan, W. Qiu, P. Malamas, *et al.*, "Myocardial Infarct Segmentation from Magnetic Resonance Images for Personalized Modeling of Cardiac Electrophysiology," *IEEE Trans. Med. Imaging*, 2016.
15. F. Usta, W. Gueaieb, J. A. White, C. McKeen, and E. Ukwatta, "Comparison of myocardial scar geometries generated from 2D and 3D LGE MRI," *Proc. SPIE 10578, Medical Imaging 2018: Biomedical Applications in Molecular, Structural, and Functional Imaging*, 2018.
16. M. Viallon, A. Jacquier, C. Rotaru, B. M. Delattre, N. Mewton, F. Vincent, *et al.*, "Head-to-head comparison of eight late gadolinium-enhanced cardiac MR (LGE CMR) sequences at 1.5 tesla: From bench to bedside," *J. Magn. Reson. Imaging*, 2011.
17. N. A. Farrag, V. Ramanan, G. A. Wright, and E. Ukwatta, "Effect of T1-mapping technique and diminished image resolution on quantification of infarct mass and its ability in predicting appropriate ICD therapy," *Med. Phys.*, 2018.
18. K. Morita, D. Utsunomiya, S. Oda, M. Komi, T. Namimoto, T. Hirai, *et al.*, "Comparison of 3D Phase-Sensitive Inversion-Recovery and 2D Inversion-Recovery MRI at 3.0 T for

- the Assessment of Late Gadolinium Enhancement in Patients with Hypertrophic Cardiomyopathy,” *Acad. Radiol.*, 2013.
19. R. Goetti, S. Kozerke, O. F. Donati, D. Sürder, P. Stolzmann, P. A. Kaufmann, *et al.*, “Acute, subacute, and chronic myocardial infarction: Quantitative comparison of 2D and 3D late gadolinium enhancement MR imaging,” *Radiology*, 2011.
 20. M. Rajch, J. Yuan, J. A. White, E. Ukwatta, J. Stirrat, C. M. Nambakhsh, *et al.*, “Interactive hierarchical-flow segmentation of scar tissue from late-enhancement cardiac mr images,” *IEEE Trans. Med. Imaging*, 2014.
 21. X. Zhuang, “Challenges and methodologies of fully automatic whole heart segmentation: A review,” *Journal of Healthcare Engineering*. 2013.
 22. C. Petitjean and J. N. Dacher, “A review of segmentation methods in short axis cardiac MR images,” *Medical Image Analysis*. 2011.
 23. M. Neizel, M. Katoh, E. Schade, T. Rassaf, G. A. Krombach, M. Kelm, *et al.*, “Rapid and accurate determination of relative infarct size in humans using contrast-enhanced magnetic resonance imaging,” *Clin. Res. Cardiol.*, 2009.
 24. A. Kolipaka, G. P. Chatzimavroudis, R. D. White, T. P. O’Donnell, and R. M. Setser, “Segmentation of non-viable myocardium in delayed enhancement magnetic resonance images,” *Int. J. Cardiovasc. Imaging*, 2005.
 25. M. Rajchl, J. Yuan, J. A. White, C. M. S. Nambakhsh, E. Ukwatta, *et al.*, “A fast convex optimization approach to segmenting 3D scar tissue from delayed-enhancement cardiac MR images,” in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 2012.
 26. E. Ukwatta, J. Yuan, W. Qiu, K. C. Wu, N. Trayanova, and F. Vadakkumpadan, “Myocardial infarct segmentation and reconstruction from 2D late-gadolinium enhanced magnetic resonance images,” *Med. Image Comput. Comput. Assist. Interv.*, 2014.
 27. F. Usta, W. Gueaieb, J. A. White, and E. Ukwatta, “3D scar segmentation from LGE-MRI using a continuous max-flow method,” *Proc. SPIE 10578, Medical Imaging 2018: Biomedical Applications in Molecular, Structural, and Functional Imaging*, 2018.
 28. F. Zabihollahy, J. A. White, E. Ukwatta, “Myocardial scar segmentation from magnetic resonance images using convolutional neural network.”, *Proc. SPIE 10575, Medical Imaging 2018: Computer-Aided Diagnosis*, 2018.
 29. F. Zabihollahy, J. A. White, and E. Ukwatta, “Convolutional neural network-based approach for segmentation of left ventricle myocardial scar from 3D late gadolinium enhancement MR images,” *Med. Phys.*, 2019.
 30. O. Ronneberger, P. Fischer, and T. Brox, “U-net: Convolutional networks for biomedical image segmentation,” in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 2015.
 31. Ö. Çiçek, A. Abdulkadir, S. S. Lienkamp, T. Brox, and O. Ronneberger, “3D U-net: Learning dense volumetric segmentation from sparse annotation,” in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 2016.
 32. J. Li, K. V. Sarma, K. Chung Ho, A. Gertych, B. S. Knudsen, and C. W. Arnold, “A Multi-scale U-Net for Semantic Segmentation of Histological Images from Radical Prostatectomies,” *AMIA Annu. Symp. proceedings. AMIA Symp.*, 2017.

33. Y. Zhu, R. Wei, G. Gao, L. Ding, X. Zhang, X. Wang *et al.*, “Fully automatic segmentation on prostate MR images based on cascaded fully convolution network,” *J. Magn. Reson. Imaging*, 2019.
34. H. R. Roth, H. Oda, X. Zhou, N. Shimizu, Y. Yang, Y. Hayashi, *et al.*, “An application of cascaded 3D fully convolutional networks for medical image segmentation,” *Comput. Med. Imaging Graph.*, 2018.
35. J. A. White, N. Fine, L. J. Gula, R. Yee, M. Al-Admawi, Q. Zhang, *et al.*, “Fused whole-heart coronary and myocardial scar imaging using 3-T CMR. Implications for planning of cardiac resynchronization therapy and coronary revascularization.,” *JACC. Cardiovasc. Imaging*, 2010.
36. P. A. Yushkevich, J. Piven, H. C. Hazlett, R. G. Smith, S. Ho, J. C. Gee, G. Gerig, “User-guided 3D active contour segmentation of anatomical structures: Significantly improved efficiency and reliability,” *Neuroimage*, 2006.
37. M. Rajchl, J. Yuan, E. Ukwatta, and P. M. Peters, “Fast interactive multi-region cardiac segmentation with linearly ordered labels,” in *Proceedings - International Symposium on Biomedical Imaging*, 2012.
38. T. V. Laarhoven, “L2 Regularization versus Batch and Weight Normalization,” arXiv:1706.05350, 2017.
39. S. Santurkar, D. Tsipras, A. Ilyas, and A. Madry, “How Does Batch Normalization Help Optimization? (No, It Is Not About Internal Covariate Shift),” *Adv. Neural Inf. Process. Syst.* 31, 2018.
40. N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, “Dropout: A simple way to prevent neural networks from overfitting,” *J. Mach. Learn. Res.*, 2014.
41. F. Milletari, N. Navab, and S. A. Ahmadi, “V-Net: Fully convolutional neural networks for volumetric medical image segmentation,” in *Proceedings - 2016 4th International Conference on 3D Vision, 3DV 2016*, 2016.
42. M. D. Zeiler, ADADELTA: an adaptive learning rate method. arXiv:1212.5701v1; 2012.
43. P. F. Christ *et al.*, “Automatic liver and lesion segmentation in CT using cascaded fully convolutional neural networks and 3D conditional random fields,” in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 2016.
44. F. Zabihollahy, J. A. White, and E. Ukwatta, “Fully automated segmentation of left ventricular myocardium from 3D late gadolinium enhancement magnetic resonance images using a U-net convolutional neural network-based model,” *SPIE Medical Imaging, San Diego, California, United States*, 2019.
45. Q. Tao, S. R. D. Piers, H. J. Lamb, and R. J. Van Der Geest, “Automated left ventricle segmentation in late gadolinium-enhanced MRI for objective myocardial scar assessment,” *J. Magn. Reson. Imaging*, 2015.
46. T. Kurzendorfer, C. Forman, M. Schmidt, C. Tillmanns, A. Maier, and A. Brost, “Fully automatic segmentation of left ventricular anatomy in 3-D LGE-MRI,” *Comput. Med. Imaging Graph.*, 2017.

Chapter 6: Automated Segmentation of Prostate Zonal Anatomy on T2W and ADC Map MR Images Using U-Nets

In the previous chapter, a novel method for automated identification of scar tissue from 3D MR images was introduced that benefits from isotropic voxel properties in 3D images [1]. The fact is that the number of 3D images in clinical trials is very limited and for most applications, multiple 2D images are acquired with lower spatial resolution. The thick, non-isotropic voxel in multiple 2D (M2D) images prevents multi-planar reformation. Therefore, other methods must be investigated for abnormality recognition in M2D images. We used a dataset of M2D prostate MR images to develop a novel algorithm for automated diagnosis of PCa. To achieve fully automated non-invasive PCa diagnosis using MR images, two steps must be accomplished: 1) the boundary of prostate WG and its regions including CG and PZ must first be identified and accurately segmented from MR images (Figure 6-1), and 2) renal PCa must be accurately detected and segmented. In this chapter, a DL-based method for the zonal segmentation of the prostate is described. The next chapter pertains to the methodology developed for PCa detection in PZ.

6.1 Introduction

Prostate regional segmentation is a challenging task due to the large variability of prostate anatomy across patients, image noise and artifacts, and poor image contrast at boundary layers. Prostate WG segmentation has been widely investigated for MR and ultrasound images [2-9]. Prior methods for semi/fully automated segmentation of prostate WG from T2W MR images can be categorized into four different groups: atlas-based registration, shape and appearance models, deformable model-based, and DL-based [10-14]. Although atlas-based registration methods have

the ability to segment the image with no well-defined relationship between regions and pixels' intensities, this method, requires time to construct atlas or a complex nonrigid registration wherever iterative procedure is incorporated within it [15]. Moreover, the prostate segmentation accuracy produced using this method in terms of DSC does not exceed $78.70 \pm 5.51\%$ [16]. The main limitation of the shape-based method is that assuming a circular or oval shape for the prostate is not strictly definitive in most of the cases, resulting in the inaccurate segmentation of prostate boundaries [17]. Qiu et al. introduced a method for prostate zonal segmentation using a dual optimization technique in 18 and 25 prostate T2W images obtained with a body-coil and endo-rectal coil, respectively. Their approach yielded DSC of $89.3 \pm 3.2\%$ for the WG, $82.2 \pm 3.0\%$ for the CG, and $69.1 \pm 6.9\%$ for the PZ using 3D body-coil MR images and; $89.2 \pm 3.3\%$ for the WG, $83.0 \pm 2.4\%$ for the CG, and $70.0 \pm 6.5\%$ for the PZ using 3D endo-rectal coil MR images [18]. However, the dataset used by Qiu et al. was small. Several DL-based methods were previously published delineating prostate WG and zonal anatomy on T2W [19-20] and in one study using trace diffusion-weighted imaging (DWI) images [21].

To our knowledge segmentation on ADC map images has not been evaluated. ADC is a pixel by pixel depiction of the slope of the exponential decay in plotted signal intensity from low to high b values, a factor that reflects the strength and timing of the gradients in DWI generation [22]. ADC is the most important determining sequence for interpretation of the PZ using Prostate Imaging–Reporting and Data System (PI-RADS) version 2 and also a key feature in TZ tumor detection [23-24]). Previously, Clark et al. performed automated segmentation on trace b_0 mm^2/sec DWI single-shot echo-planar images, which are effectively a poor spatial resolution fat-suppressed T2W image [25]. Clark et al. reported a median DSC of 93% and 88% for the

prostate gland and TZ, respectively. Since it poorly depicted in ADC images, compared to the zonal anatomy in DWI, leads to a more challenging segmentation task.

The objective of this study is, therefore, to evaluate the feasibility of fully automated segmentation of prostate WG, CG, and PZ, on T2W and ADC map images using the U-Net CNN-based method.

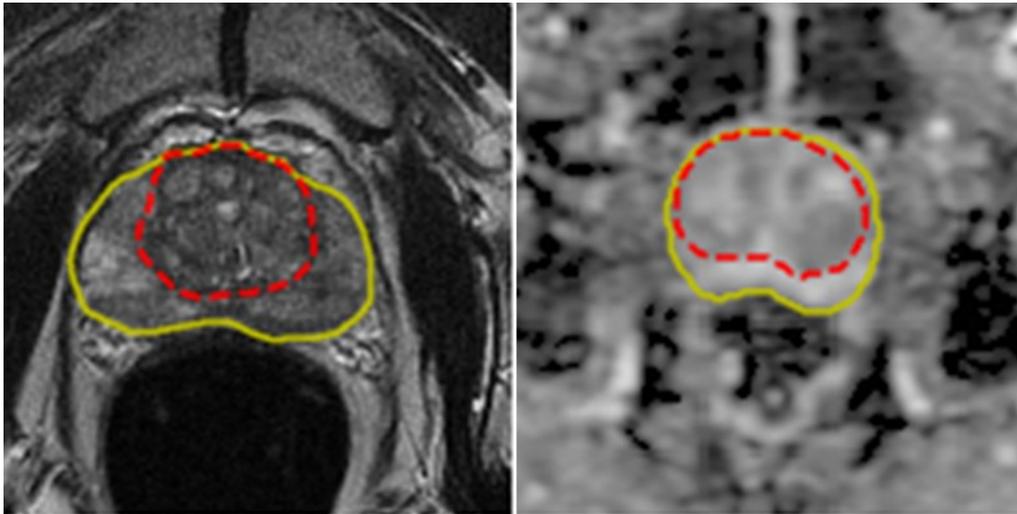


Figure 6-1. An example of axial T2-weighted (T2W) and apparent diffusion coefficient (ADC) map prostate MR images from left to right, where the boundary of prostate WG and CG are shown in yellow and red. PZ is the area enclosed between colored contours. Reprinted with permission.

The advantages of using U-Net for segmentation tasks and its architecture are described in the previous chapter. The main contribution of this work is developing a new algorithm for delineating prostate zonal anatomy using both T2W and ADC map images, which is: 1) fully-automated without requiring any user input, thus completely reproducible, 2) highly accurate for delineating prostate zones across all prostate levels (i.e. at the basal, apical, and mid-levels) for different sizes of the prostate for patients with and without clinically significant tumors

detectable on MRI, and 3) quite efficient compared to manual segmentation (i.e. less than 5 seconds versus approximately 30 minutes). A thorough search of the relevant literature did not yield any related article on quantification of prostate zones from ADC map MR images using deep-learning-based techniques. An advantage of our study is that the dataset used in this research was relatively large which assured us of our technique robustness.

6.2 Method

Figure 6-2 illustrates an overview of our proposed pipeline. As the first step of the pipeline, we normalized the image intensities to adjust for global contrast change across each slice. We then created two separate U-Nets to segment prostate WG and CG. Segmentation for the PZ was obtained by subtracting the segmentation map of CG from that of WG. As we had multiple 2D slices for each patient, 3D surfaces were created from the slice-wise segmentation maps via interpolation technique to visualize the surface of prostate geometry.

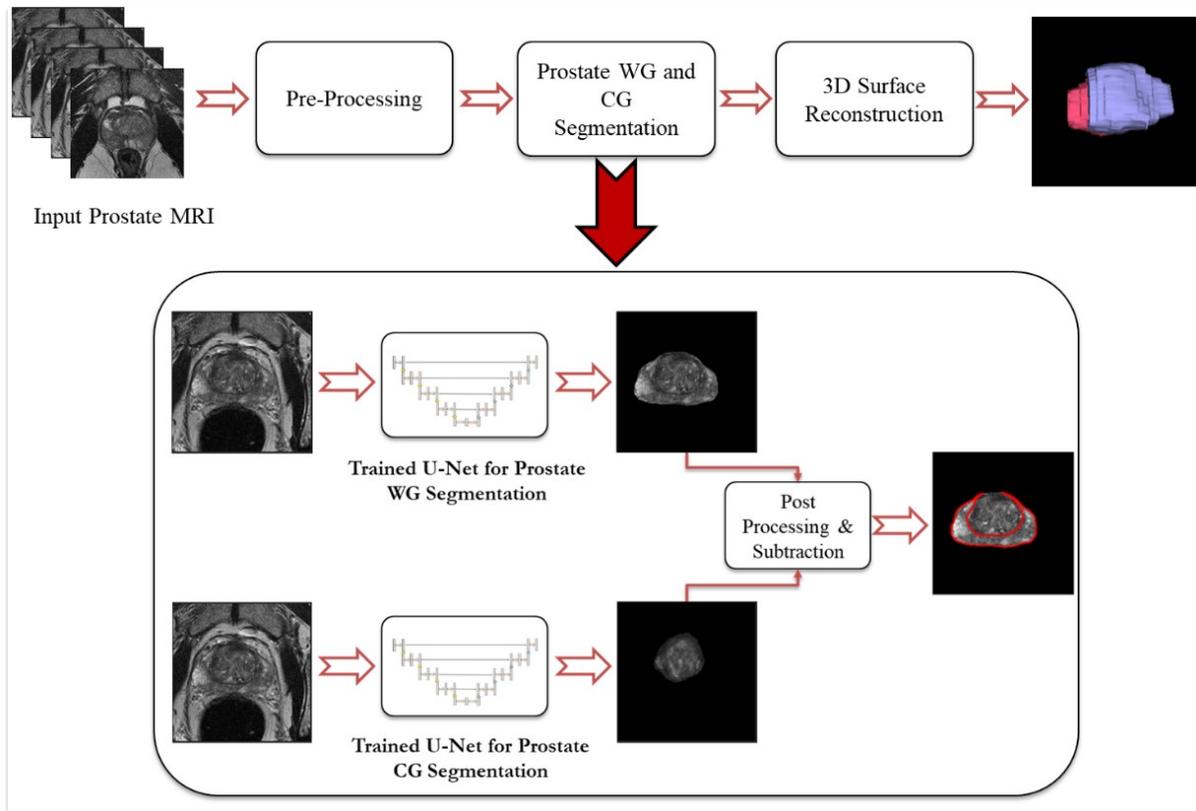


Figure 6-2. The block diagram of our proposed method for prostate regional segmentation of T2W or ADC map MR images. Reprinted with permission.

6.2.1 Dataset

This study was approved by joint institutional review boards of Carleton and Ottawa Universities, who waived the need for informed consent and also approved the exchange of data between institutions through a data-sharing agreement. Between the dates of July 2012 and December 2017, we retrieved 225 patients who underwent mp-MRI of the prostate at 3-Tesla and radical prostatectomy (RP) at a single institution referral center for PCa. Every patient in this cohort, which formed the case group, had a pre-existing mp-MRI-RP map created where the dominant significant tumor foci (those measuring ≥ 0.5 cc in size) were identified at RP and

mapped onto T2W and ADC map images by a dedicated Genitourinary Pathologist and Radiologist both with 13 years of experience in PCa. For the control group (patients with mp-MRI without significant tumors at time of mp-MRI) we identified 72 patients on AS who underwent mp-MRI after at least one round of random biopsies showing only low volume Gleason 6 cancers without any abnormality detected on mp-MRI and with a repeat negative template biopsy performed after mp-MRI as well as with stable clinical parameters (digital rectal exam (DRE) and prostate specific antigen (PSA) metrics) for at least one year following mp-MRI with no clinical suspicion of an occult significant tumor that may have been missed on mp-MRI or biopsy.

6.2.2 MRI

Every patient underwent mp-MRI using the same clinical 3-Tesla system, Discovery 750W (GE Healthcare, Milwaukee WI) using multi-array surface coils. Endorectal coils were not utilized in any patients since these are generally no longer required at 3T to perform prostate MRI. A summary of the mp-MRI protocol utilized at our institution is provided in Table 6-1. The voxel size of T2W MR and DW images was $0.3906 \times 0.3906 \times 3.4$ mm and $1.1719 \times 1.1719 \times 5$ mm, respectively.

Table 6-1. Sequence parameters for multi-parametric MRI of the prostate protocol performed with pelvic surface coil^a at 3 Tesla^b. Reprinted with permission.

	Imaging Plane	Field of View (mm)	Matrix Size	Slice Thickness/Gap (mm)	TR/TE (msec)	Echo Train Length	Flip Angle	Acceleration Factor	Receiver Bandwidth (Hz/Voxel)	Acquisition time (min)	Number of signals averaged
T1 3D Dual Echo GRE ^d	Axial	240 x 240	292 x 224	4.0/1.0	4.8/1.1;TE1 2.2;TE2	N/A	12	2	558	Breath Hold	1
T2 TSE ^c	Coronal	220 x 220	320 x 256	4.0/0	3890/105	27-35	111	N/A	122	4 min	1-2
	Sagittal			3.0/0						4 min	
	Axial			3.0/0						4 min	
DWI ^e	Axial	280 x 280	128 x 80	4-5.0/0	4200/90	1	90	2	1950	5 min	4-10
T1 GRE ^f Dynamic Contrast	Axial	220 x 220	128 x 128	4.0/0	4.3/1.3	N/A	12	2	488	2-6 min	1

- a) Integrated pelvic surface coils (16 channels) with activated spine coils (8 channels).
- b) Clinical 3 Tesla systems: Discovery 750W (General Electric, Milwaukee WI).
- c) Turbo/Fast Spin Echo
- d) Gradient Recalled Echo
- e) DWI = Diffusion-weighted imaging performed with spectral fat suppression echo-planar imaging with tridirectional motion probing gradients and B values of 0,500,1000,1500 with automatic ADC map generation.
- f) Dynamic fast spoiled 2D GRE performed with a temporal resolution of 10 seconds after injection of 0.1 mmol/kg of gadobutrol (Gadovist, Bayer Inc. Toronto, ON) at a rate of 3 mL/sec.

6.2.3 Manual Segmentation

For each patient, anonymized images in Digital Imaging and Communication in Medicine (DICOM) format were exported for axial T2W and ADC map images from our PACS (Horizon Medical Imaging version 13.1, McKesson Corporation, San Francisco CA). Images were evaluated by a team of four fellowship-trained abdominal and pelvic Radiologists with between 5-13 years of experience in prostate mp-MRI using ITK-SNAP version 3.2 (Philadelphia, PA). The images were divided into four groups and each part of the images was segmented by one

radiologist. Therefore, each image was segmented only once. Prior to segmentation, the four Radiologists met in person for a training and practice session and together segmented two example patients to achieve a similar methodology for the remainder of the dataset. For each patient, Radiologists manually segmented the whole prostate gland, the PZ and the CG (TZ + central zone) first on axial T2W and then on axial ADC map images. The T2W images were segmented first to provide a better delineation of the zonal anatomy for later segmentation on ADC map images where the zonal anatomy is subjectively less well depicted than on T2W images [26]. In addition to segmentation of the PZ, Radiologists also divided and labeled the segmentations into the PZ base, middle and apical regions, as described previously [23-24], in order to perform a sub-group analysis of segmentation accuracy by cranio-caudal level of the prostate. Lastly, for patients with significant tumors identified at RP, Radiologists segmented tumors using existing mp-MRI-RP maps.

6.2.4 U-Net Architecture

We created four U-Nets and trained to segment prostate WG and CG from T2W and ADC map images, separately, hereinafter called NetT2WWG, NetT2WCG, NetADCWG, and NetADCCG, respectively. The network contains shrinking (left) and expanding (right) paths. A shrinking path includes 5 stages in each two convolutional layers of size 3x3 followed by the ReLU activation layer and one max pooling layer of size 2x2 were applied. We used 64 filters in the convolutional layers of the first stage of NetT2WWG and NetADCWG and 32 filters in the convolutional layers of the first stage of NetT2WCG and NetADCCG. The number of filters is duplicated in the subsequent stages. In expanding path, two convolutional layers of size 3x3

followed by the Relu activation layer and one up-sampling layer of size 2×2 at each stage were used, which are the most widely used filter and pooling/up-sampling size in CNN-based architectures. We set the batch size equal to 10 and performed batch normalization after each convolutional layer to keep the mean and standard deviation of the activation layer close to zero and one, respectively. This operation made the learned function invariant to scaling of the weights [27]. A Drop-out layer was employed at the end of the contraction path to randomly set half of the input units to zero at each update during training time. A convolutional layer of size 1×1 followed by a sigmoid activation layer was used at the final layer of expanding path to map all feature maps and generate a probability map with the same size of the input image. The probability map was then thresholded such that each pixel with a probability higher than 0.5 was considered as a proper region (WG or CG) to create the desired segmentation map for each slice of input MR image. 20% of the training images were used to validate the performance of the trained model on unobserved samples and fine-tune the network. All parameters were then adjusted experimentally based on the segmentation results obtained on the validation set.

6.2.5 U-Nets Training

Our dataset consists of 225 ADC map images and 225 T2W MR images from 225 patients. T2W and ADC map images were then cropped into size 256×256 and 128×128 pixels around the center pixel of each slice. The T2W images were down sampled by the factor of 2 in the transversal direction to reduce computational costs. All images were normalized into the range $[0,1]$ by subtracting the minimum intensity value from each pixel and then dividing the result by the difference between the highest and lowest intensity values. We performed image normalization to make the intensity range consistent. The dataset was then divided into the train (N=100) and test (N=125) sets. Since the ratio of patients with and without significant tumors in

our study was 2 to 1, we kept approximately the same ratio in train and test sets. Table 6-2 indicates the number of patients and compiled 2D slices utilized in our experiments in more detail.

Table 6-2. The number of train and test sets of images in each dataset used in this study from T2W and ADC map sequences. Reprinted with permission.

MR Image Type	Total number of patients in the dataset	Train			Test		
		Number of patients	The ratio of patients with/without significant tumor	Total No. of compiled 2D slices	Number of patients	The ratio of patients with/without significant tumor	Total No. of compiled 2D slices
T2W	225	100	70/30	1154	125	83/42	1587
ADC map	225	100	70/30	812	125	83/42	917

We used DSC as the evaluation metric for network training and the negative value of DSC was used as a loss function. It has been shown that using DSC as a loss function leads to fast and accurate results on prostate volume segmentation from MRI [28]. We employed Adadelta as an optimization method, which is a robust extension of Adagrad that does not need learning rate initialization, which is the main advantage of this technique [29] as choosing a small value for learning rate may slow down the optimization while selecting a large value causes oscillation around minima.

The domain knowledge about the prostate was used to refine the segmentation results. Since prostate WG is relatively large but a single anatomic region on MR images, all sparsely segmented regions smaller than 100 and 20 pixels in T2W and ADC map images were considered as false-positive and removed from the algorithm-generated segmentation map.

Similarly, the smallest allowable size for segmented CG was set up to 100 and 5 for T2W and ADC map images respectively. All holes inside the remaining mask were filled as the WG and CG are integrated regions in MR images. The boundary of segmented objects was smoothed using mathematical erosion operation.

To understand the individual impact of a customized cost function, image pre- and post-processing, the NetT2WWG was trained with different cost functions and without pre- and post-processing and the performance of the network was re-evaluated. Table 6-3 displays the DSC as a mean \pm standard deviation on the test dataset for each experiment. We conducted the Wilcoxon signed-rank test ($\alpha=0.05$) to investigate whether post-processing significantly improves the DSCs. There is a significant difference ($p<0.0001$) between the median values of DSCs before and after the post-processing step.

We implemented the algorithm in Python using Keras library, on top of TensorFlow. We trained U-Nets for 150 epochs on Intel Core i7, 2.8 GHz on a GPU-accelerated computing platform, that required 48, 18, 34, and 12 minutes for NetT2WWG, NetT2WCG, NetADCWG, and NetADCCG, respectively. The Dice coefficient on a validation set that was 20% of training data samples was monitored and the latest best model according to the monitored quantity was saved as well.

Table 6-3. Causality experiment results for evaluating the impact of the cost function, pre- and post-processing for NetT2WWG. Reprinted with permission.

	Image Pre-Processing	Image Post-Processing	Loss Function	Metric	DSC (%)
NetT2WWG-1	-	Morphological cleaning, filling the holes inside prostate WG, smoothing the boundary of the segmented area	Negative DSC	DSC	92.53 ± 8.97
NetT2WWG-2	Normalization	-	Negative DSC	DSC	92.59 ± 7.66
NetT2WWG-3	Normalization	Morphological cleaning, filling the holes inside prostate WG, smoothing the boundary of the segmented area	Binary Cross-Entropy	Accuracy	92.30 ± 8.38
NetT2WWG (Proposed)	Normalization	Morphological cleaning, filling the holes inside prostate WG, smoothing the boundary of the segmented area	Negative DSC	DSC	92.96 ± 7.77

6.2.6 Comparison of Our Method to Alternative Method

Employed algorithms for prostate regional delineation can be categorized into energy minimization (i.e. active appearance models, registration-based, probabilistic active shape model, convex optimization approach) and ML (i.e. RF) based. To provide a benchmark, we compared our obtained results on T2W MR images with another state-of-the-art technique that was reported on the prostate challenge PROMISE2012 [13]. As our approach was also ML-based, we chose RF to perform prostate regional segmentation on our dataset. We trained two different classifiers to delineate prostate WG and CG. The segmentation maps were then compared to

determine the PZ region on T2W MR images. Similarly, images of 100 and 125 patients were down-sampled by a factor of two and utilized to train and test RF-based classifiers.

To this end, small patches were extracted around each pixel of the prostate T2W MR images and labeled as target (prostate WG/CG) and non-target. To label each patch, a patch of the same size was derived from the corresponding manual segmentation. If the center pixel was labeled as the target or more than half of the pixels were labeled as the target, the patch was labeled as target and vice-versa. As the target was a relatively small area in the T2W MR image, the extracted patches from non-target regions were more than those derived from the target, which caused a huge class imbalance in the training samples. To address this problem, some patches from the non-target class were randomly selected and removed from the dataset to not only create a balance between training instances but also introduce minimal computational complexity to the system as a large number of patches were extracted from our training images. In our experiments, we witnessed that extracting patches of size 50×50 pixels and using 70 decision tree classifiers yield the best segmentation results. The intensity of patches was normalized into the range of $[0,1]$. As deriving patch took a long time due to a large number of images in our dataset and the size of T2W MR images, patches were extracted with the stride of 10. The output segmentation maps were compared to the expert manual segmentation to examine the performance of the RF-based method for prostate zonal segmentation.

6.3 Results

The proposed method was used to delineate the boundary of prostate WG and its sub-regions from T2W MR and ADC map test images in 125 patients for evaluation purposes. Figure 6-3

illustrates segmentation results for different image types in four patients (randomly selected from our test dataset). As depicted in Figure 6-3 the boundary of algorithm-generated segmented regions tightly matches the boundary of manual segmentations. Experimental results demonstrated that the proposed method is capable of efficiently and accurately extracting prostate WG, CG, and PZ from the input prostate MR images, with a median DSC of $95.33 \pm 7.77\%$ for the WG, 93.75 ± 8.91 for the CG, and $86.78 \pm 3.72\%$ for the PZ using T2W MR images; $92.09 \pm 8.89\%$ for the WG, $89.89 \pm 10.69\%$ for the CG, and $86.1 \pm 9.56\%$ for the PZ in ADC map images. Table 6-4 shows a summary of the results of our method. The high value of other metrics including precision, recall, and accuracy furthermore confirmed the accuracy of our proposed automated method for prostate zonal segmentation.

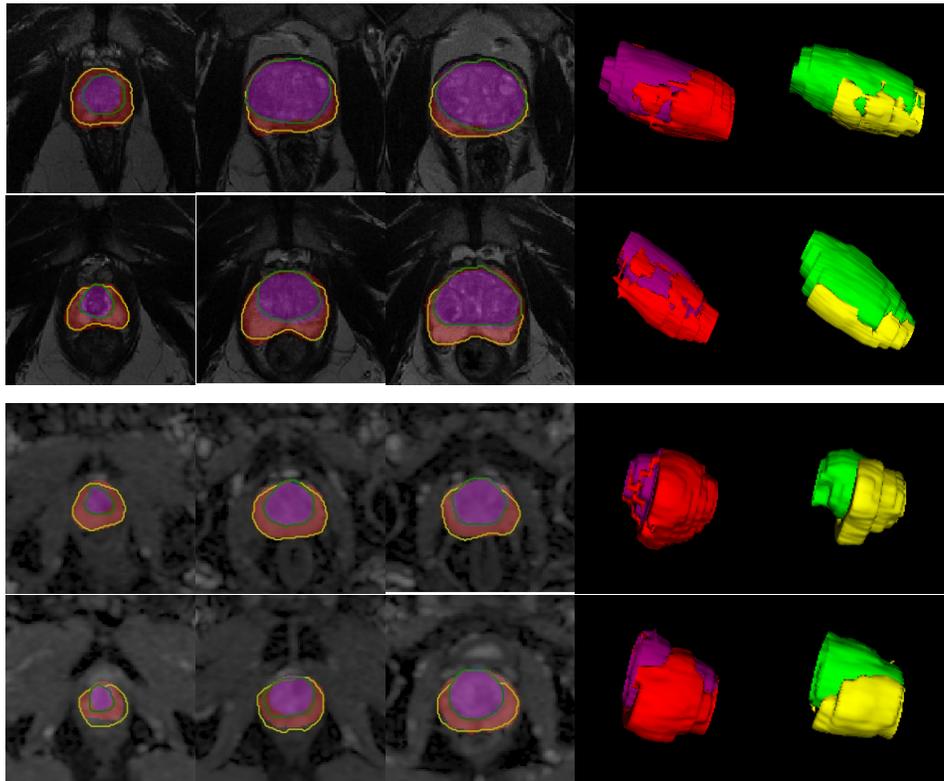


Figure 6-3. Example results from segmented prostate WG, CG, and PZ using T2W (two top rows) and ADC map MR images (two bottom rows) in four patients. On the source 2D images, the U-net-based method segmentation of the prostate CG and PZ for three slices presented from apex to base (left to right) extracted from MRI are shown in purple and red respectively and the contours computed from the expert manual are displayed in green and yellow. The fourth and fifth picture elements depict a 3D surface rendering of algorithm-generated and expert manual segmentation of prostate zones, respectively. Reprinted with permission.

Table 6-4. Summary of results of our proposed technique for prostate zonal segmentation as the median and mean values \pm standard deviation on T2W and ADC map MRI against expert manual segmentations.

Reprinted with permission.

MR Image	Segmented Region	Average	DSC (%)	AVD (cc)	Precision (%)	Recall (%)	Accuracy (%)
T2W	WG	Median	95.33 \pm 7.77	5.97 \pm 5.51	96.51 \pm 7.87	95.34 \pm 8.08	98.05 \pm 2.36
		Mean	92.96 \pm 7.77	6.61 \pm 5.51	94.00 \pm 7.87	93.01 \pm 8.08	97.39 \pm 2.36
	CG	Median	93.75 \pm 8.91	2.95 \pm 7.23	95.95 \pm 8.97	93.56 \pm 9.6	98.37 \pm 2.29
		Mean	91.07 \pm 8.91	4.51 \pm 7.23	93.07 \pm 8.97	90.64 \pm 9.6	97.80 \pm 2.29
	PZ	Median	86.78 \pm 3.72	10.78 \pm 4.46	85.57 \pm 4.38	90.87 \pm 4.25	96.16 \pm 1.3
		Mean	86.22 \pm 3.72	10.97 \pm 4.46	84.57 \pm 4.38	89.52 \pm 4.25	95.89 \pm 1.3
ADC map	WG	Median	92.09 \pm 8.89	8.74 \pm 12.90	93.93 \pm 9.92	94.96 \pm 9.90	98.11 \pm 1.87
		Mean	89.71 \pm 8.89	12.01 \pm 12.90	90.17 \pm 9.92	91.04 \pm 9.90	97.62 \pm 1.87
	CG	Median	89.89 \pm 10.69	5.48 \pm 12.15	93.68 \pm 11.52	90.15 \pm 12.25	98.66 \pm 1.6
		Mean	86.33 \pm 10.69	8.40 \pm 12.15	89.10 \pm 11.52	86.78 \pm 12.25	98.26 \pm 1.6
	PZ	Median	86.10 \pm 9.56	18.80 \pm 8.41	86.65 \pm 10.57	86.05 \pm 10.47	96.89 \pm 1.98
		Mean	83.30 \pm 9.56	18.13 \pm 8.41	85.10 \pm 10.57	83.57 \pm 10.47	96.45 \pm 1.98

We also assessed the correlation in segmentation results between T2W and ADC images. The Pearson correlation coefficients were computed as 0.1558, 0.4467, and 0.4535 as a measure of a linear relationship between DSC for T2W and ADC map MR images for prostate WG, CG, and PZ. We conducted t-tests ($\alpha=0.05$) to evaluate the significant correlation in DSCs for T2W and ADC images. For the prostate WG segmentations, the test failed to show a significant correlation ($p = 0.0828$) of DSCs between the T2W and ADC images. For the prostate CG and PZ segmentations, there is a significant linear relationship between DSCs for T2W and ADC map MR images ($p<0.0001$).

Overall, the results indicate that our algorithm is more accurate in segmenting T2W images than segmenting ADC map images, most likely due to poor contrast at the prostate boundaries in ADC map images.

The average running time required to delineate the boundary of prostate WG, CG, and PZ using the developed U-Net models from a typical T2W and ADC map test image on Intel Core i7, 2.8 GHz using a GPU-accelerated computing platform was 2.39 ± 0.34 s and 2.17 ± 0.48 s, respectively. This time is significantly faster than manual segmentation, which was not precisely measured but required on average approximately 15 minutes for a set of MR images. The automated segmentation process included pre-processing, segmentation, post-processing, and 3D surface creation.

The mean of prostate WG volume for the testing dataset calculated from manual segmentation using T2W MR and ADC map images along with the volume range are shown in Table 6-5. The lowest error of volume-based metric was obtained for segmentation results of T2W MR images.

Table 6-5. The mean and range of prostate WG volume for the testing dataset calculated from manual segmentation of T2W MR and ADC map images. Reprinted with permission.

	T2W	ADC map
	Prostate WG	Prostate WG
Mean \pm sd (cc)	48.20 \pm 27.85	63.06 \pm 33.17
Minimum (cc)	13.47	14.75
Maximum (cc)	169.44	171.76

The segmentation results correspond to the highest and lowest DSCs (see Table 6-6) are shown in Figure 6-4.

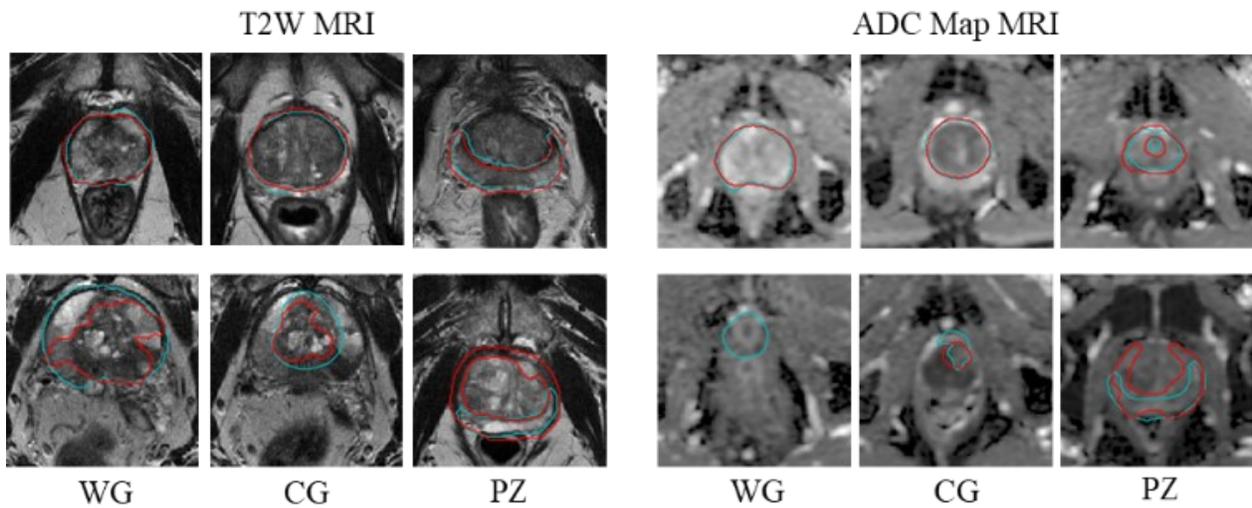
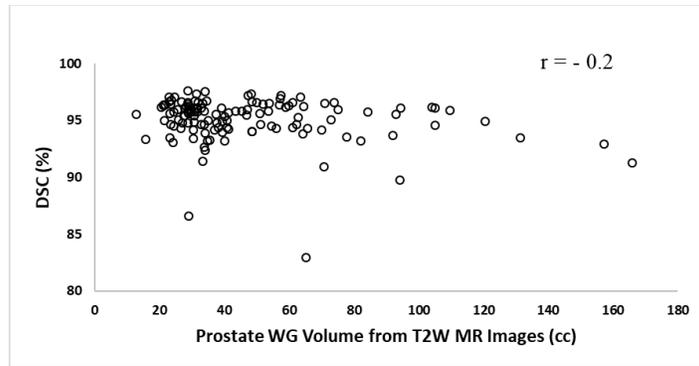


Figure 6-4. From top to bottom the best and worst segmentation results using our proposed methodology for prostate regional delineation from T2W and ADC map MR images are presented. The contours computed from the expert manual and algorithm-generated segmentations are displayed in cyan and red, respectively. The algorithm failed to detect prostate WG from ADC map MR in one of the test images. Reprinted with permission.

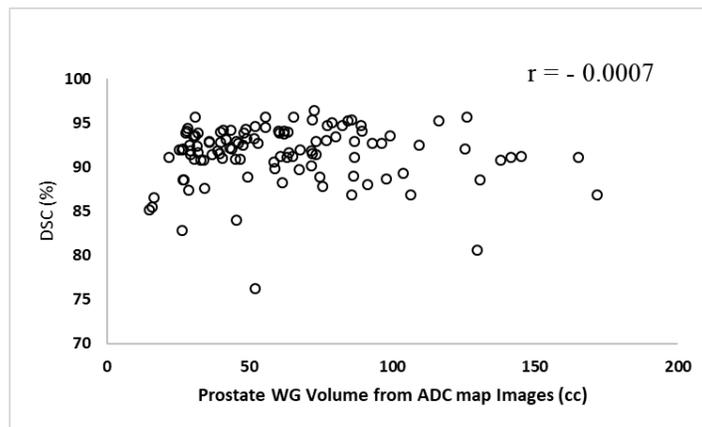
Table 6-6. Best and worst DSC achieved using our presented strategy for prostate zonal segmentation from T2W and ADC map MR test images against expert manual segmentations. Reprinted with permission.

Segmented Zone	T2W MRI			ADC Map MRI		
	Prostate WG	Prostate CG	Prostate PZ	Prostate WG	Prostate CG	Prostate PZ
Maximum DSC (%)	97.09	95.54	92.22	96.19	95.37	92.22
Minimum DSC (%)	77.08	70.79	60.81	48.71	63.73	49.59

To explore the relationship between prostate WG size and segmentation accuracy, the Pearson correlation coefficient was measured by dividing the covariance of prostate WG volume and DSC to the product of their standard deviations, which yielded - 0.2 and - 0.0007 for T2W MR and ADC map images. Figure 6-5 illustrates the scatter diagrams for each set, where DSC is plotted against prostate WG volume. There is not any linear relationship between those variables. Accordingly, our presented strategy segmented the prostate WG with high accuracy from both T2W and ADC map images (DSC of $95.33 \pm 7.77\%$ and $92.09 \pm 8.89\%$) regardless of prostate WG size.



a)



b)

Figure 6-5. Scatter plot of DSC versus prostate WG volume reported by our proposed method on a) T2W and b) ADC map prostate MR images, along with the Pearson correlation coefficient (r-value). Reprinted with permission.

We further conducted a subgroup analysis of the prostate WG, CG and PZ segmentation accuracy for patients with and without verified clinically significant tumor detectable on MRI separately to see if the presence of tumor affects the network performance. As shown in Table 6-7, the accuracy of segmentation in terms of DSC, precision, recall, accuracy, and AVD across WG, CG, and PZ from both types of MR images do not differ significantly for the control group and in patients with tumor.

Table 6-7. Performance of our proposed technique for prostate zonal segmentation as the median and mean values \pm standard deviation on T2W and ADC map MRI images against expert manual segmentations for the control group and patients with tumor. Reprinted with permission.

MR Image	Subjects	Segmented Region	Average	DSC (%)	AVD (cc)	Precision	Recall	Accuracy
T2W	Control	WG	Median	95.53 \pm 8.96	6.72 \pm 8.01	96.58 \pm 9.00	95.64 \pm 9.23	97.86 \pm 3.02
			Mean	92.54 \pm 8.96	8.15 \pm 8.01	93.77 \pm 9.00	92.37 \pm 9.23	96.97 \pm 3.02
		CG	Median	94.54 \pm 9.79	3.57 \pm 12.00	96.74 \pm 9.56	94.14 \pm 10.25	98.09 \pm 3.08
			Mean	91.41 \pm 9.79	6.54 \pm 12.00	93.70 \pm 9.56	90.66 \pm 10.25	97.29 \pm 3.08
		PZ	Median	86.17 \pm 3.75	12.55 \pm 5.68	83.67 \pm 4.96	90.65 \pm 3.95	95.69 \pm 1.17
			Mean	85.07 \pm 3.75	12.87 \pm 5.68	82.71 \pm 4.96	89.45 \pm 3.95	95.54 \pm 1.17
	Patients with Tumor	WG	Median	95.25 \pm 6.92	5.53 \pm 3.64	96.47 \pm 7.09	95.32 \pm 7.24	98.14 \pm 1.8
			Mean	93.23 \pm 6.92	5.91 \pm 3.64	94.14 \pm 7.09	93.40 \pm 7.24	97.65 \pm 1.8
		CG	Median	93.32 \pm 8.29	2.90 \pm 2.71	95.55 \pm 8.56	93.22 \pm 9.17	98.54 \pm 1.53
			Mean	90.86 \pm 8.29	3.58 \pm 2.71	92.67 \pm 8.56	90.63 \pm 9.17	98.12 \pm 1.53
		PZ	Median	87.66 \pm 3.58	10.42 \pm 3.42	86.03 \pm 3.79	90.99 \pm 4.38	96.34 \pm 1.33
			Mean	86.75 \pm 3.58	10.10 \pm 3.42	85.42 \pm 3.79	89.56 \pm 4.38	96.06 \pm 1.33
ADC-map	Normal	WG	Median	92.83 \pm 7.09	8.72 \pm 14.30	95.13 \pm 8.11	95.68 \pm 8.12	98.18 \pm 2.04
			Mean	91.08 \pm 7.09	12.58 \pm 14.30	91.75 \pm 8.11	92.51 \pm 8.12	97.61 \pm 2.04
		CG	Median	91.79 \pm 9.53	6.77 \pm 16.81	94.88 \pm 9.65	92.39 \pm 11.01	98.54 \pm 2.01
			Mean	88.61 \pm 9.53	11.48 \pm 16.81	91.27 \pm 9.65	89.14 \pm 11.01	98.05 \pm 2.01
		PZ	Median	86.04 \pm 9.2	20.15 \pm 8.89	86.00 \pm 10.0	86.29 \pm 10.05	96.84 \pm 2.06
			Mean	83.25 \pm 9.2	20.17 \pm 8.89	84.76 \pm 10.0	83.79 \pm 10.05	96.30 \pm 2.06
	Patients with Tumor	WG	Median	91.38 \pm 9.98	8.74 \pm 11.92	91.78 \pm 10.97	94.6 \pm 11.09	98.05 \pm 1.78
			Mean	88.54 \pm 9.98	11.49 \pm 11.92	88.77 \pm 10.97	90.57 \pm 11.09	97.57 \pm 1.78
		CG	Median	87.80 \pm 11.34	4.85 \pm 7.37	91.69 \pm 12.64	87.59 \pm 13.05	98.68 \pm 1.23
			Mean	84.47 \pm 11.34	6.49 \pm 7.37	87.39 \pm 12.64	84.94 \pm 13.05	98.36 \pm 1.23
		PZ	Median	86.11 \pm 9.75	17.28 \pm 7.84	86.84 \pm 10.88	85.83 \pm 10.69	96.95 \pm 1.93
			Mean	83.33 \pm 9.75	16.86 \pm 7.84	85.29 \pm 10.88	83.45 \pm 10.69	96.53 \pm 1.93

Moreover, we investigated the performance of designed models for prostate WG, CG, and PZ delineation at basal, apical, and mid-slice levels separately. The summary of results separated by the anatomic craniocaudal level of the prostate is provided in Table 6-8. For level-based segmentation, the highest DSC for both T2W MR and ADC map images was obtained for mid-slices, except for CG segmentation from the ADC map, where the best result was seen for basal slices.

Table 6-8. Summary of results of our proposed technique for prostate zonal segmentation at the base, apex, and middle craniocaudal slice levels as the median and mean values \pm standard deviation on T2W and ADC map MRI compared to expert manual segmentations. Reprinted with permission.

MR Image	Segmented Region	Slice Level	Average	DSC (%)	Precision	Recall	Accuracy
T2W	WG	Base	Median	95.68 \pm 4.16	96.78 \pm 2.57	95.18 \pm 5.2	97.61 \pm 2.25
			Mean	94.54 \pm 4.16	96.03 \pm 2.57	93.81 \pm 5.2	96.99 \pm 2.25
		Middle	Median	96.72 \pm 2.19	97.4 \pm 1.82	96.66 \pm 3.18	98.36 \pm 1.34
			Mean	96.19 \pm 2.19	96.91 \pm 1.82	95.74 \pm 3.18	98.01 \pm 1.34
		Apex	Median	94.71 \pm 6.89	95.99 \pm 7.07	95.06 \pm 7.73	98.55 \pm 1.48
			Mean	93.03 \pm 6.89	94.10 \pm 7.07	92.83 \pm 7.73	98.17 \pm 1.48
ADC map	WG	Base	Median	92.28 \pm 8.51	94.63 \pm 8.67	95.07 \pm 9.74	97.42 \pm 2.45
			Mean	90.09 \pm 8.51	91.24 \pm 8.67	91.16 \pm 9.74	96.76 \pm 2.45
		Middle	Median	94.14 \pm 6.92	95.84 \pm 8.17	95.74 \pm 7.60	98.10 \pm 1.49
			Mean	92.53 \pm 6.92	92.75 \pm 8.17	93.58 \pm 7.60	97.77 \pm 1.49
		Apex	Median	90.61 \pm 8.76	92.10 \pm 9.92	94.49 \pm 10.47	98.39 \pm 0.99
			Mean	88.62 \pm 8.76	89.06 \pm 9.92	90.44 \pm 10.47	98.22 \pm 0.99
T2W	CG	Base	Median	94.74 \pm 6.93	96.92 \pm 6.97	94.25 \pm 7.46	97.99 \pm 1.94
			Mean	92.75 \pm 6.93	94.73 \pm 6.97	92.04 \pm 7.46	97.51 \pm 1.94
		Middle	Median	94.84 \pm 5.58	96.75 \pm 5.18	94.15 \pm 6.39	98.52 \pm 1.56
			Mean	93.33 \pm 5.58	95.13 \pm 5.18	92.51 \pm 6.39	98.12 \pm 1.56
		Apex	Median	93.11 \pm 8.99	94.04 \pm 9.56	93.56 \pm 10.23	99.23 \pm 1.23
			Mean	90.26 \pm 8.99	91.47 \pm 9.56	90.49 \pm 10.23	98.82 \pm 1.23
ADC map	CG	Base	Median	91.56 \pm 9.07	94.30 \pm 9.45	92.80 \pm 10.57	98.24 \pm 1.84
			Mean	88.73 \pm 9.07	90.65 \pm 9.45	89.59 \pm 10.57	97.77 \pm 1.84
		Middle	Median	90.93 \pm 9.11	95.16 \pm 9.88	90.48 \pm 10.90	98.52 \pm 1.35
			Mean	88.15 \pm 9.11	91.36 \pm 9.88	88.05 \pm 10.90	98.21 \pm 1.35
		Apex	Median	84.21 \pm 11.97	88.70 \pm 14.06	85.67 \pm 14.01	99.13 \pm 0.88
			Mean	81.50 \pm 11.97	84.49 \pm 14.06	82.44 \pm 14.01	98.86 \pm 0.88
T2W	PZ	Base	Median	85.35 \pm 6.36	83.00 \pm 5.87	88.84 \pm 7.92	95.64 \pm 1.91
			Mean	83.79 \pm 6.36	82.65 \pm 5.87	86.55 \pm 7.92	95.15 \pm 1.91
		Middle	Median	88.19 \pm 3.87	86.62 \pm 5.03	91.23 \pm 4.22	96.01 \pm 1.39
			Mean	87.30 \pm 3.87	85.63 \pm 5.03	90.21 \pm 4.22	95.73 \pm 1.39
		Apex	Median	86.92 \pm 5.67	85.18 \pm 6.74	90.80 \pm 5.97	97.00 \pm 1.15
			Mean	85.76 \pm 5.67	84.20 \pm 6.74	89.53 \pm 5.97	96.77 \pm 1.15
ADC map	PZ	Base	Median	83.08 \pm 9.59	83.20 \pm 9.96	83.58 \pm 11.11	95.78 \pm 1.91
			Mean	79.65 \pm 9.59	81.73 \pm 9.96	80.63 \pm 11.11	95.45 \pm 1.91
		Middle	Median	86.72 \pm 6.45	87.24 \pm 8.17	87.95 \pm 7.48	96.69 \pm 1.39
			Mean	85.81 \pm 6.45	86.66 \pm 8.17	86.39 \pm 7.48	96.48 \pm 1.39
		Apex	Median	86.69 \pm 8.21	87.75 \pm 9.23	86.73 \pm 9.11	97.69 \pm 1.19
			Mean	84.34 \pm 8.21	85.88 \pm 9.23	84.59 \pm 9.11	97.42 \pm 1.19

Lastly, we segmented the prostate WG and its region from T2W MR images using the RF-based model. Figure 6-6 displays segmented prostate WG, CG, and PZ for one patient randomly selected from the test dataset using different methods, where our method determined the prostate

WG, CG, and PZ boundaries more accurately. Table 6-9 summarizes the quantitative results for the RF method. Compared to the results of Table 6-4, the highest DSC was reported for our approach.

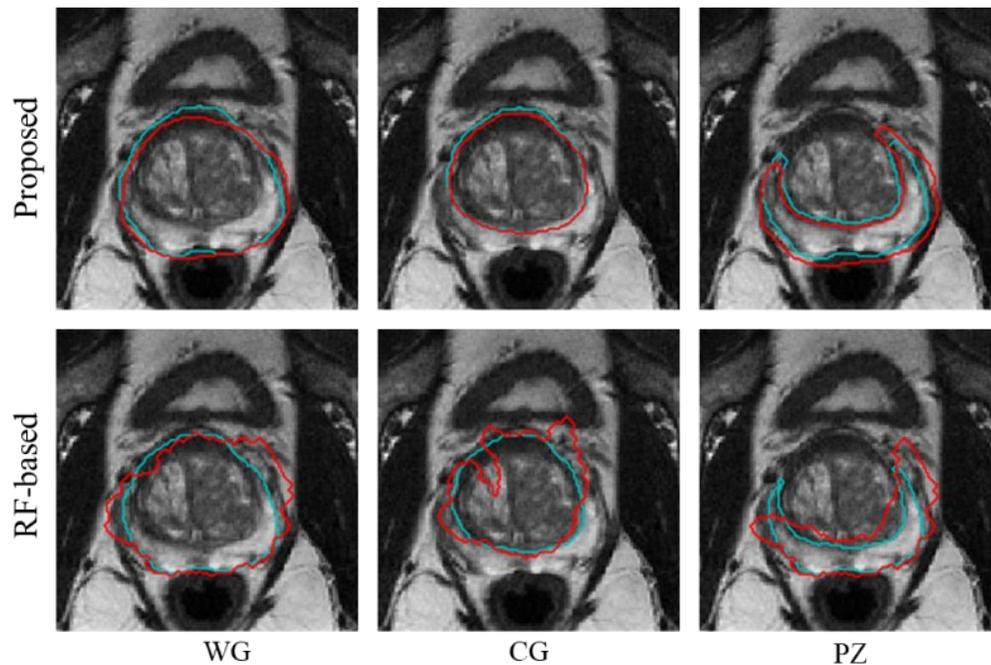


Figure 6-6. Exemplary results of prostate regional segmentation methods applied for an example of T2W MR image, where the contours computed from the expert manual and algorithm-generated segmentations are displayed in cyan and red, respectively. The results of each method are shown in the corresponding row. Reprinted with permission.

Table 6-9. Performance of RF-based method for prostate regional segmentation from T2W MR images as the median and mean values \pm standard deviation (SD) against expert manual segmentation. Reprinted with permission.

Segmented Region	Average	DSC (%)	Precision (%)	Recall (%)
WG	Median \pm SD	84.58 \pm 4.33	83.47 \pm 5.16	89.32 \pm 5.62
	Mean \pm SD	84.03 \pm 4.33	83.02 \pm 5.16	88.15 \pm 5.62
CG	Median \pm SD	79.23 \pm 5.73	74.9 \pm 7.17	90.26 \pm 6.03
	Mean \pm SD	78.6 \pm 5.73	74.64 \pm 7.17	89.78 \pm 6.03
PZ	Median \pm SD	70.29 \pm 6.01	73.91 \pm 7.19	69.79 \pm 6.79
	Mean \pm SD	69.84 \pm 6.01	73.07 \pm 7.19	69.76 \pm 6.79

6.4 Discussion

In this chapter, we describe a new fully automated method to segment the prostate WG, CG (central + transition zones) and PZ from T2W MR and ADC map images using a DL-based method. In this work, we made qualitative comparisons of our work to previous studies. In comparison to previous deep-learning-based methods [19-21], we had a larger testing dataset, where our method yielded high accuracy in DSC in segmentation across all parts of the prostate. We also performed a subgroup analysis to determine if segmentation accuracy varied depending on the prostate size (volume) to determine broader applicability in patients with and without prostatic enlargement from benign prostatic hyperplasia (BPH). This subgroup analysis, which to our knowledge has not been reported previously, demonstrates that the results of our segmentation method are not biased on the size of the prostate gland. Our U-Net based method is fully automated; therefore, it is not subject to inter and intra-observer variability. The proposed

algorithm is capable of segmenting prostate zonal anatomy across all craniocaudal slices at the basal, apical and mid-levels.

Our proposed algorithm is capable of segmenting prostate zonal anatomy across all craniocaudal slices at the basal, apical and mid-levels. Generally, the highest segmentation accuracy was at the mid-level, which is concordant with the study by Qiu et al., which showed that on average the highest DSC was reported at mid-level as $92.5 \pm 2.3\%$, $94.2 \pm 1.6\%$, and $83.1 \pm 4.7\%$ for prostate WG, CG, and PZ segmentation, respectively [18]. In addition, we also demonstrated that the segmentation accuracy of our method for both T2W and ADC map images is independent of the size of the prostate WG.

It is important to separate the PZ from the CG since cancers occur more commonly in the PZ and BPH nodules which are commonly encountered in the CG may mimic the imaging appearance of tumors [30-31]. Our method yielded similar segmentation accuracy in DSC for the PZ border for both T2W and ADC maps, whereas it yielded higher segmentation accuracy in DSC for CG for T2W images. This result is important because the CG uses T2W as the most important determining sequence for cancer diagnosis whereas the PZ uses the ADC map.

We made several choices in designing our U-Nets. For instance, we used batch-normalization and Dropout techniques to avoid the risk of network overfitting and to reduce generalization error, a measure of how accurately our algorithm performs prostate zonal segmentation for unseen images [27], [32].

Finally, we compared the performance of our automated algorithm to that of an RF-based classifier. In comparison to this classifier, our method consistently reported superior accuracy for the DSC, precision, recall, and accuracy. Since the RF-based algorithm utilizes a patch-based

segmentation approach, delineating the boundaries of prostate WG and its regions from the T2W MR image of one patient takes approximately 6 minutes, which is considerably slower than our algorithm (less than 5 sec). Additionally, this method suffers from redundant computation due to the overlap that exists between patches extracted from adjacent pixels.

Our study has several limitations. All images used in this study were from a single-institution retrospective cohort that may introduce bias to the results. Additionally, each image was annotated by only one expert radiologist. Further study is needed to investigate the inter-observer variability. As we use the segmentation maps of the two different networks to discover the boundaries of PZ, the errors of the networks are added up, which is inevitable using the proposed method. Further study is required to compare the performance of the proposed algorithm with the state-of-the-art methods for PZ segmentation to investigate whether those methods improve the accuracy.

6.5 Conclusion

A fully automated DL-based method was developed and evaluated in which two U-Nets were trained to determine prostate WG and CG boundaries from T2W and ADC map MR images. Our technique is robust and reproducible across a relatively large sample of patients with variable prostate sizes with and without clinically significant tumors. Our results confirm that automated segmentation of the prostate zonal anatomy is possible on T2W MRI, are the first to describe accurate segmentation on ADC map images, which are crucial for tumor detection in clinical practice. Our technique can be incorporated into CAD workflows for automated detection and characterization of prostate tumors.

References

1. F. Zabihollahy, N. Schieda, S. Krishna Jeyaraj, and E. Ukwatta, "Automated segmentation of prostate zonal anatomy on T2-weighted (T2W) and apparent diffusion coefficient (ADC) map MR images using U-Nets," *Med. Phys.*, 2019.
2. S. Ghose, A. Oliver, R. Martí, X. Lladó, J. C. Vilanova, J. Freixenet, "A survey of prostate segmentation methodologies in ultrasound, magnetic resonance and computed tomography images," *Comput. Methods Programs Biomed.*, 2012.
3. W. Qiu, J. Yuan, E. Ukwatta, Y. Sun, M. Rajchl, and A. Fenster, "Prostate segmentation: An efficient convex optimization approach with axial symmetry using 3-D tRUS and MR images," *IEEE Trans. Med. Imaging*, 2014.
4. W. Qiu, J. Yuan, E. Ukwatta, D. Tessier, and A. Fenster, "Three-dimensional prostate segmentation using level set with shape constraint based on rotational slices for 3D end-firing TRUS guided biopsy," *Med. Phys.*, 2013.
5. W. Qiu, J. Yuan, E. Ukwatta, D. Tessier, and A. Fenster, "Rotational-slice-based prostate segmentation using level set with shape constraint for 3D end-firing TRUS guided biopsy," in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 2012.
6. W. Qiu, J. Yuan, E. Ukwatta, and A. Fenster, "Rotationally resliced 3D prostate TRUS segmentation using convex optimization with shape priors," *Med. Phys.*, 2015.
7. J. Yuan, W. Qiu, M. Rajchl, E. Ukwatta, X. C. Tai, and A. Fenster, "Efficient 3D endfiring TRUS prostate segmentation with globally optimized rotational symmetry," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2013.
8. W. Qiu, M. Rajchl, F. Guo, Y. Sun, E. Ukwatta, A. Fenster, *et al.*, "3D prostate TRUS segmentation using globally optimized volume-preserving prior," in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 2014.
9. W. Qiu, J. Yuan, E. Ukwatta, D. Tessier, and A. Fenster, "Prostate segmentation in 3D TRUS using convex optimization with shape constraint," *Medical Imaging: Image Processing*, 2013.
10. J. Yuan, E. Ukwatta, W. Qiu, M. Rajchl, Y. Sun, X. C. Tai, *et al.*, "Jointly segmenting prostate zones in 3D MRIs by globally optimized coupled level-sets," in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 2013.
11. W. Qiu, J. Yuan, E. Ukwatta, Y. Sun, M. Rajchl, and A. Fenster, "Efficient 3D multi-region prostate MRI segmentation using dual optimization," in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 2013.

12. W. Qiu, J. Yuan, E. Ukwatta, Y. Sun, M. Rajchl, and A. Fenster, "Fast globally optimal segmentation of 3D prostate MRI with axial symmetry prior," in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 2013.
13. G. Litjens, R. Toth, W. van de Ven, C. Hoeks, S. Kerkstra, B. van Ginneken, *et al.*, "Evaluation of prostate segmentation algorithms for MRI: The PROMISE12 challenge," *Med. Image Anal.*, 2014.
14. N. Schieda, C.S. Lim, F. Zabihollahy, J. Abreu-Gomez, S. Krishna, S. Woo, *et al.*, "Quantitative Prostate MRI.," *JMRI*, 2020.
15. H. Kalinic, "Atlas-based image segmentation: A Survey," *Croat. Sci. Bibliogr.*, 2009.
16. Y. Ou and J. Doshi, "Multi-atlas segmentation of the prostate: A zooming process with robust registration and atlas selection," *MICCAI Gd. Chall. Prostate MR Image Segmentation*, 2012.
17. M. Kirschner, F. Jung, and S. Wesarg, "Automatic Prostate Segmentation in MR Images with a Probabilistic Active Shape Model," *Miccai*, 2012.
18. W. Qiu, J. Yuan, E. Ukwatta, Y. Sun, M. Rajchl, and A. Fenster, "Dual optimization based prostate zonal segmentation in 3D MR images," *Med. Image Anal.*, 2014.
19. M. N. N. To, D. Q. Vu, B. Turkbey, P. L. Choyke, and J. T. Kwak, "Deep dense multi-path neural network for prostate segmentation in magnetic resonance imaging," *Int. J. Comput. Assist. Radiol. Surg.*, 2018.
20. R. Cheng, H. R. Roth, N. Lay, L. Lu, B. Turkbey, W. Gandler, *et al.*, "Automatic magnetic resonance prostate segmentation by deep learning with holistically nested networks," *J. Med. Imaging*, 2017.
21. T. Clark, J. Zhang, S. Baig, A. Wong, M. A. Haider, and F. Khalvati, "Fully automated segmentation of prostate whole gland and transition zone in diffusion-weighted MRI using convolutional neural networks," *J. Med. Imaging*, 2017.
22. D. Le Bihan, "Apparent diffusion coefficient and beyond : What diffusion mr imaging can tell us about tissue structure," *Radiology*. 2013.
23. J. C. Weinreb, J. O. Barentsz, P. L. Choyke, F. Cornud, M. A. Haider, K. J. Macura, *et al.*, "PI-RADS Prostate Imaging - Reporting and Data System: 2015, Version 2," *Eur. Urol.*, 2016.
24. J. O. Barentsz, J. C. Weinreb, S. Verma, H. C. Thoeny, C. M. Tempany, F. Shtern, *et al.*, "Synopsis of the PI-RADS v2 Guidelines for Multiparametric Prostate Magnetic Resonance Imaging and Recommendations for Use," *European Urology*. 2016.
25. T. Clark, J. Zhang, S. Baig, A. Wong, M. A. Haider, and F. Khalvati, "Fully automated segmentation of prostate whole gland and transition zone in diffusion-weighted MRI using convolutional neural networks," *J. Med. Imaging*, 2017.
26. A. S. Puryško, A. B. Rosenkrantz, J. O. Barentsz, J. C. Weinreb, and K. J. Macura, "PI-RADS version 2: A pictorial update," *Radiographics*, 2016.
27. T.V. Laarhoven, "L2 Regularization versus Batch and Weight Normalization." ArXiv e-prints, 2017.
28. F. Milletari, N. Navab, and S. A. Ahmadi, "V-Net: Fully convolutional neural networks for volumetric medical image segmentation," in *Proceedings - 2016 4th International Conference on 3D Vision*, 2016.

29. M.D. Zeiler, "ADADELTA: an adaptive learning rate method.", arXiv:1212.5701v1, 2012.
30. S. Krishna, N. Schieda, M. D. McInnes, T. A. Flood, and R. E. Thornhill, "Diagnosis of transition zone prostate cancer using T2-weighted (T2W) MRI: comparison of subjective features and quantitative shape analysis," *Eur. Radiol.*, 2019.
31. J. S. Quon, B. Moosavi, M. Khanna, T. A. Flood, C. S. Lim, and N. Schieda, "False-positive and false negative diagnoses of prostate cancer at multi-parametric prostate MRI in active surveillance," *Insights into Imaging*. 2015.
32. N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, "Dropout: A simple way to prevent neural networks from overfitting," *J. Mach. Learn. Res.*, 2014.

Chapter 7: Fully Automated Localization of Prostate PZ Tumors on ADC Map MR Images Using an Ensemble Learning Method

In the previous chapter fully automated segmentation of prostate zones on the ADC map was described. In this chapter, we use the description of prostate PZ as an ROI to search for the prostate tumor on ADC map MRI as this sequence is the most important sequence for detecting PCa that appear in the PZ [1].

7.1 Introduction

Accurate detection and localization of PCa in men undergoing prostate MRI is a fundamental step for future targeted prostate biopsies and treatment planning. Fully automated localization of PZ PCa using the ADC map might be clinically useful. The main goal of this retrospective study is to describe the automated localization of PCa in the PZ on ADC map MR images using an ensemble U-Net-based model. We used the same dataset used for prostate zonal segmentation, where model training and testing was performed using images of 146 and 80 patients, respectively. The ground truth was established by manual delineation of the prostate and prostate PZ tumors on ADC maps by dedicated radiologists using MRI-radical prostatectomy maps as a reference standard.

PCa is the most common non-cutaneous malignancy in men [2]. Prostate MRI has been shown to be accurate for the diagnosis of PCa, particularly in the PZ [3]. Trials have shown that prostate MRI before a targeted biopsy improves the detection rate for significant PCa (which requires treatment) and lowers false-negative rates [4-5].

There are, however, limitations to prostate MRI in clinical practice. While MRI is accurate for diagnosis of many tumors, MRI is not a perfect test; for example, in the PROMIS trial, it was shown that a proportion of men with a negative MRI had an undetected significant tumor which was later diagnosed through saturation (e.g. oversampling) biopsy of the prostate [4]. False-positive cases on MRI also occur [6]. Most importantly, the inter-observer agreement for subjective evaluation of prostate multi-parametric MRI (mp-MRI) using the PI-RADS version 2, which is the current clinical standard, is at the best moderate and has been acknowledged as an area of potential improvement in future iterations of PI-RADS [7-8].

The use of more quantitative or automated methods for PCa detection from MRI is therefore clinically desirable. The addition of quantitative metrics to subjective PI-RADS improves accuracy [9] with the improved inter-observer agreement [10]. Comprehensive reviews of CAD developed for PCa detection using prostate MRI are available [11-12]. Generally, CAD methods developed for PCa detection require prostate zonal segmentation and/or feature representation within the ROI (PZ or transition zone [TZ]). Typically, in previous works, a combination of MRI sequences has been investigated and the majority of presented CADs are semi-automated, where manually segmentation of prostate zones is needed. These techniques, therefore, do not reflect clinical practice where the greater emphasis has recently been placed on DWI and derived ADC maps for PZ tumor detection [13]. Moreover, manual regional segmentation is time-consuming and subject to large intra- and inter-observer variations [14].

Recently several studies have evaluated machine-learning-based techniques for automated PCa detection using mp-MRI. Song et al. developed a computer-aided tool for PCa diagnosis from mp-MRI using a CNN [15]. They aligned and resampled low-resolution DWI/ADC images to

high-resolution T2W images using rigid and nonrigid transformation to create a registered image with three channels. The images were then artificially augmented via random rotation, shift, flip and stretching strategies to train a deep CNN. The trained network reported sensitivity and specificity of 87.0% and 90.6% on 19 cases. As they used the patch-based approach, the method suffers from redundant computation due to extracting patches from adjacent pixels. Tsehay et al. adopted a network architecture from a state-of-the-art edge detector that takes an mp-MR image as an input and localizes prostate peripheral and transition zones tumors [16]. They trained and tested their network using a dataset consisting of 52 patients with T2W, ADC, and b 2000 mm²/s DWI MR images. Although the reported results were accurate (86% detection rate at 20% false-positive rate), the studied dataset was very small. Yang et al. introduced an ML-based method using mp-MRI including T2W and ADC map sequences to first, automatically determine whether each slice of the mp-MRI contains PCa or not, and second, for slices classified as positive, to localize the position of the PCa for further targeted prostate biopsies [17]. As a first step, T2W and ADC map MR images were registered. The position of the prostate WG was then found using a network trained on T2W MR images. Both ADC and T2W were then cropped around prostate WG and normalized. A CNN was trained using patches extracted from cropped T2W and ADC map MR images to localize PCa on ADC map MRI. They trained and tested their model on 8301 and 120 images respectively and achieved a sensitivity of 0.46, 0.92 and 0.97 at 0.1, 1 and 10 false positives per normal/benign patient.

Compared to previous studies, the main advantage of our study is that our algorithm needs only ADC map MR images for PCa prediction. Additionally, our fully automated algorithm has been tested on a larger number of test images that demonstrates the robustness of the technique for PCa localization. We propose an ensemble learning technique, which is a process by which

multiple classifiers are generated and strategically combined to obtain a better predictive performance compared to any of the constituent classifiers alone [18]. The purpose of the present study was, therefore, to develop and evaluate a fully automated ensemble learning-based method for the detection and segmentation of clinically significant cancers in the prostate PZ using only ADC map MR images [19].

7.2 Materials and Methods

7.2.1 Study Subjects and Data Acquisition

This study was approved by the joint institutional review board, who waived the need for informed consent and also approved the exchange of patient anonymized data between institutions (The University of Ottawa and Carleton) through a data-sharing agreement. Between the dates of July 2012 and December 2017, we identified 154 patients who underwent mp-MRI of the prostate at 3-Tesla and radical prostatectomy (RP) at a single institution referral center for PCa. A summary of patient inclusion and exclusion criteria is provided in Figure 7-1. Age and PSA were retrieved from the electronic medical record and recorded. At RP, complete removal of the prostate gland is performed, and the removed prostate is sectioned in total and subsequently analyzed by dedicated genitourinary pathologists who outline each focus of tumor in the prostatectomy specimen. Every patient in this cohort, which formed the case group, had an mp-MRI-RP map subsequently created where the dominant significant tumor foci (those measuring $\geq 0.5 \text{ cm}^3$ in size) were identified at RP and mapped onto T2W and ADC map images by a dedicated genitourinary Pathologist and Radiologist both with 13 years of experience in PCa. For the purposes of this study, only the dominant PZ tumors in each patient were

considered such that for 154 patients there were 154 dominant PZ tumors. The Gleason score and stage of disease for each tumor foci were assessed by the pathologist and recorded. A summary of the mp-MRI protocol utilized at our institution is provided in Table 6-1 in previous chapter. Briefly, during the study period, three b values were always used including b 0, 500 and 1000 mm²/sec with an automatic derivation of the ADC map based upon an automated mono-exponential model of DWI data. A b 1500 mm²/sec image was added in 2015 after the recommendations of PI-RADS version 2; however, it is not included in the calculation of the ADC as it may add unwanted diffusion kurtosis effects into the ADC value as indicated in PI-RADS version 2.1 [8].

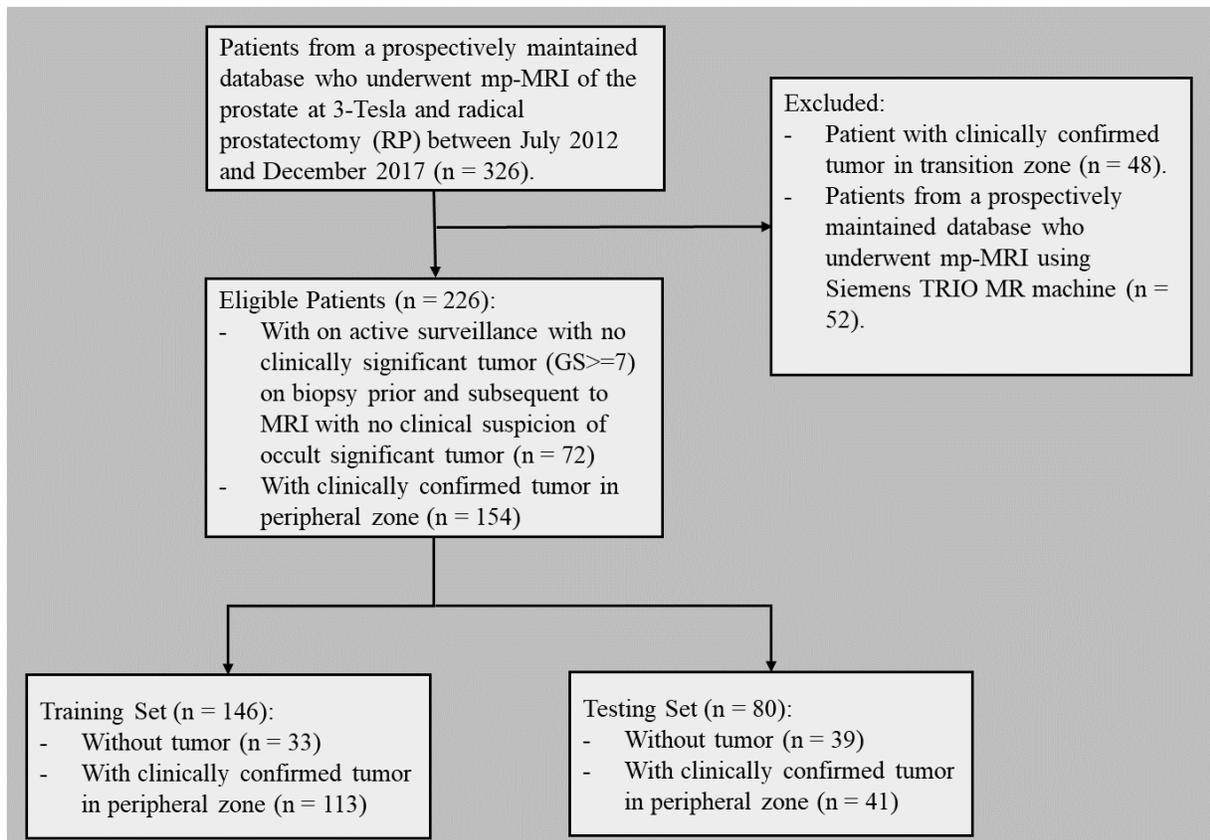


Figure 7-1. Flowchart shows the strategy for patient selection and overall experimental design for localizing PCa on ADC map MR images. Reprinted with permission.

There were 72 patients in the control group. Mean patient age and PSA for the control group were 63.9 ± 6.1 years and 9.2 ± 6.2 ng/mL. The control group consisted of patients who underwent mp-MRI with a history of not clinically significant Gleason score 3+3=6 PCa diagnosed at a prior random template biopsy undergoing active surveillance. To be included, the patients mp-MRI should not reveal any potential significant tumors (as discussed in the original MRI report and verified by the dedicated genitourinary Radiologist) and all patients underwent a repeat random template biopsy performed after MRI showing no significant tumor also with at least one year of clinical follow up showing stable clinical parameters (e.g. normal DRE and PSA metrics) with no clinical suspicion of an occult significant tumor that may have been missed on mp-MRI or biopsy. This formed a robust reference standard for the control group.

Every patient underwent mp-MRI using a single clinical 3-Tesla system, Discovery 750W (GE Healthcare, Milwaukee WI) using multi-array surface coils. A summary of the distribution of patients and tumors in the case group is summarized in Table 7-1.

Table 7-1. Summary of clinical features of 154 patients who underwent mp-MRI prior to radical prostatectomy in this study. Reprinted with permission.

Clinical and Pathological Features	Number of patients (N=154)
Age (years)	62.9 ± 6.4
PSA (ng/mL)	11.2 ± 11.1
Gleason score	
3+3=6	0
3+4=7	41.7% (65/154)
4+3=7	39.1% (61/154)
4+4=8	2.6% (4/154)
9 or 10	16.7% (26/154)
Pathological Stage	
T2	60.3% (94/154)
T3a	25.6% (40/154)
T3b	7.7% (12/154)

7.2.2 Manual Segmentation of Prostate Boundaries and Tumors

For each patient, anonymized images in Digital Imaging and Communication in Medicine (DICOM) format were exported for axial T2W and ADC map images from our PACS, Horizon Medical Imaging version 13.1, McKesson Corporation, San Francisco CA). Images were evaluated by a team of four fellowship-trained abdominal and pelvic Radiologists with between 5-13 years of experience in prostate mp-MRI using ITK-SNAP version 3.2 (Philadelphia, PA). The patients were randomly divided into four sets, where images for a particular patient in a given set were independently segmented by a single radiologist. Prior to segmentation, the four Radiologists met in person for a training and practice session and together segmented two

example patients to achieve a similar methodology for the remainder of the dataset. For each patient, Radiologists manually segmented the whole prostate gland, the PZ and the CG (TZ + central zone), first on axial T2W and then on axial ADC map images. The T2W images were segmented first to provide a better delineation of the zonal anatomy for later segmentation on ADC map images where the zonal anatomy is subjectively less well depicted than on T2W images [20]. In addition to segmentation of the PZ, Radiologists also divided and labeled the segmentations into the PZ base, middle and apical regions, as described previously [21-22]. Lastly, for patients with significant tumors identified at RP, Radiologists segmented tumors using existing mp-MRI-RP maps on both T2W and ADC map images. Examples of manual segmentation on the T2W and ADC map MR image along with their manual segmentation are shown in Figure 7-2. In this study, only the ADC maps were evaluated since T2W is not a formal component of the decision tree rules for PZ tumor detection in PI-RADS [8].

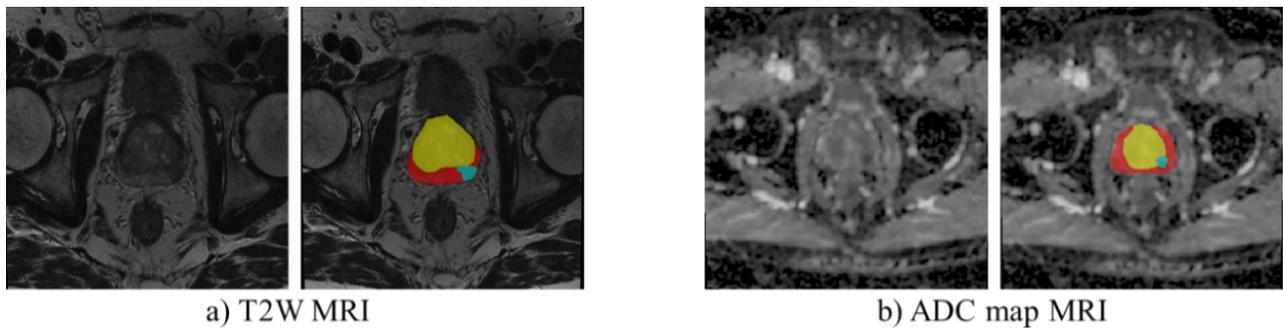


Figure 7-2. An example of axial (a) T2W and (b) ADC map prostate MR images and manual segmentation from left to right, where the surface of the prostate peripheral zone, prostate transition zone, and PCa are shown in red, yellow, and blue, respectively. Reprinted with permission.

7.2.3 Ensemble Model of U-Nets for PCa localization

Our proposed methodology is summarized in Figure 7-3. Our model is made up of three U-Nets (ADC-Net1, ADC-Net2, ADC-Net3) with different capacities and weights to effectively learn characteristics of PCa in the PZ from ADC map MR images and accomplish the distinction between normal and cancerous tissues. Figure 7-4 illustrates the architecture of ADC-Nets employed in our experiments for PCa localization and Table 7-2 indicates the structure of each network.

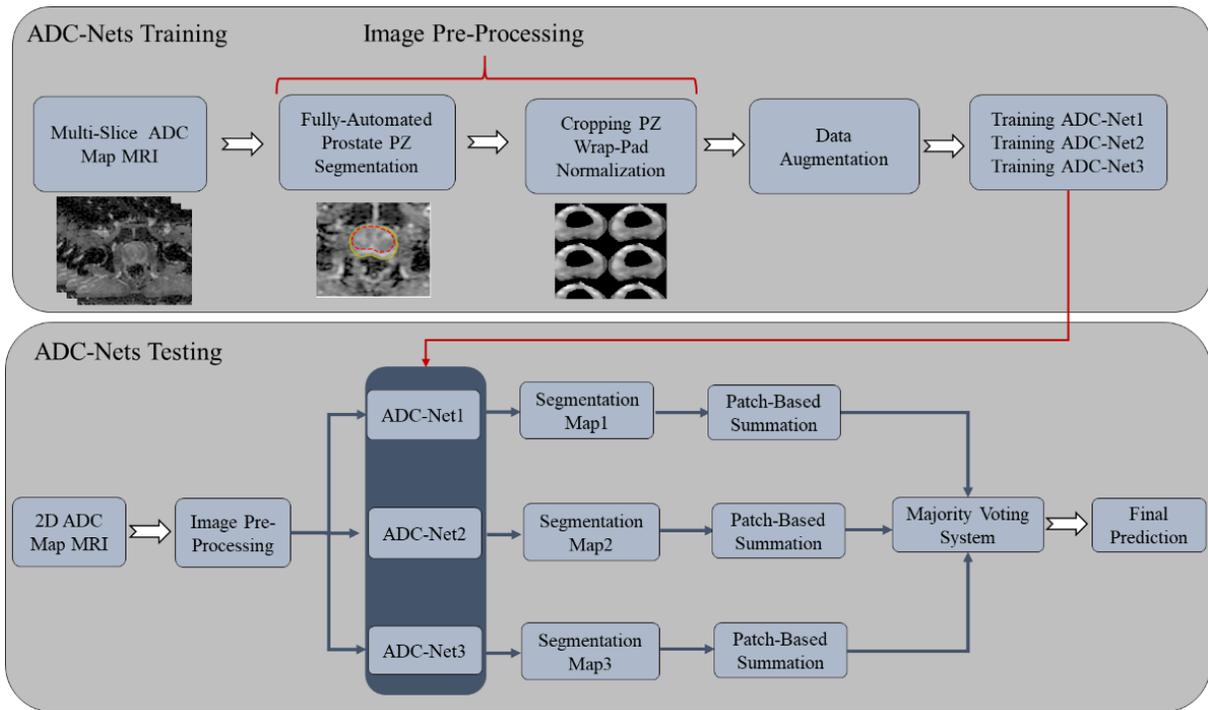


Figure 7-3. The block diagram of our proposed methodology for PCa localization of ADC map MR images. Reprinted with permission.

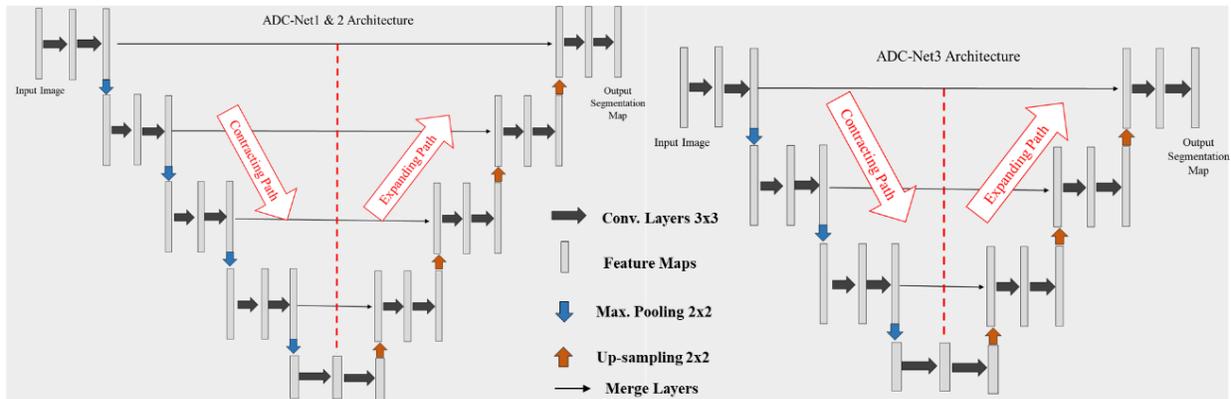


Figure 7-4. The architecture of ADC-Net1, 2 and 3 (acronyms are defined in the text). Reprinted with permission.

Table 7-2. The configuration of ADC-Nets used in our experiment for PCa localization. Reprinted with permission.

Network	No. Of Stages	No. of Filters in Each Stage
ADC-Net1	5	8,16,32,64,128
ADC-Net2	5	32, 64,128,256,512
ADC-Net3	4	16,32,64,128

As a first step, zones of the prostate were segmented using the methodology described in the previous chapter. Once prostate zones were identified, the training images were cropped automatically to a patch of size 128×128 pixels that sufficiently enclose the prostate WG. The patch was padded with the wrap of the vector along the axis to repeat the PZ pattern inside each patch while the input size remains the same. The padded images were then normalized into the range of $[0,1]$ to ensure the intensity range is consistent.

The dataset was divided into training (N=146 including 1817 slices after artificially image augmentation) and testing (N=80 including 660 slices) sets. The training dataset comprised of images from patients with and without clinically significant tumor with a ratio of approximately 1 to 2. Moreover, since tumors of the prostate are relatively small (i.e. most commonly under 2 cm in size [23-24]) they are contained within only a few axial slices extracted from base to apex of prostate WG using existing spatial resolution limitations of prostate DWI which is set at a minimum of 3-4 mm to maintain adequate signal to noise ratio [22]. Therefore, there was a large class imbalance in the dataset where the number of slices which contained tumor represented only 25% of the whole dataset. To address this problem in the training phase, we artificially augmented the data by flipping slices with tumor left-right, up-down, and rotating them by 180° and added them up to the training dataset that resulted in a total of 1817 slices for training.

We defined pixel-wise true positive of PCa as the metric to be evaluated by the network during training. The negative value of DSC was considered as a loss function to be minimized by the model during training. DSC evaluates the spatial fidelity of the network-generated segmentation to the ground truth. It has been shown that using DSC as a loss function leads to fast and accurate results on prostate volume segmentation from MRI [25]. Adadelta was employed as an optimizer during training, which is robust confronting various data modalities, different network architectures, and noisy gradient information [26]. We implemented the algorithm in Python using Keras library, on top of Tensorflow. The networks were trained for 100 epochs on Intel core i7, 2.8 GHz on a GPU-accelerated computing platform.

In the testing phase, the image was similarly pre-processed and presented to the trained networks to generate a segmentation map. As the input image was padded in wrap mode, the output

segmentation map was divided into smaller patches of the same size such that each patch encloses prostate PZ. Those patches were summed up to find the location of algorithm proposed regions as PCa. The majority voting of the 3 ADC-Nets was then applied per-pixel basis to combine generated segmentation maps for final PCa prediction. The algorithm-generated segmentation map was compared to the manual segmentations to evaluate the accuracy of our method for prostate PZ tumor detection.

All parameters of the developed model were adjusted experimentally based on the network performance for PCa detection in the PZ. To estimate the influence of each element to the final results, we conducted a causality study in which the width of the constituent networks was altered by changing the number of filters in each stage. The new network was trained for 30 epochs using the same number of training samples and its performance was re-evaluated.

7.2.4 Statistical Analysis

We used sensitivity and specificity to assess the performance of our proposed method for PCa detection in the PZ from ADC map MR images. To compute those values, a confusion matrix for each patient was created in the test dataset with multiple slices across prostate WG as follows: For each slice, the prediction mask was compared with the corresponding manual segmentation. If for a given slice, the ground truth included PCa and the algorithm predicted PCa with non-zero overlap with the manual segmentation, the prediction was counted as true positive (TP), whereas for slices with PCa, if the segmentation map did not have overlap with manual segmentation, the prediction was counted as false negative (FN). For slices without PCa, if the network predicted PCa somewhere in the prediction mask, it was counted as false-positive (FP), otherwise, it was considered as true negative (TN). Having TP, FN, FP, and TN, the sensitivity and specificity

could be calculated for each patient in the test set. DSC was computed slice-wise as a region-based metric to measure the spatial overlap between algorithm-generated and expert manual segmentations. The average (mean \pm standard deviation) of sensitivity, specificity, and DSC were reported as the performance of the developed model for PCa detection. The first and third quartiles were estimated as well to indicate the median of the lower and upper half of the estimated metrics. 25% and 75% of the values in the reported metrics lie below and above the first quartile, respectively, while this percentage is approximately 75% and 25% for the third quartile. Additionally, false-positive rate is reported. To evaluate the performance of our method, we also generated the ROC curve, which graphs the true positive rate (sensitivity) against the false-positive rate (1 - specificity) at various thresholds on the decision function. The AUC was reported as a measure of the performance of our method for PCa detection in the PZ for all possible classification thresholds. The 95% confidence interval and p-value for AUC were reported as well. To calculate the 95% confidence interval for AUC value, the bootstrapping and randomization testing was used. We bootstrapped the AUC computations (by sampling predicted and true labels for each test image across the prostate with replacement) for 10000 times. The lower and upper bounds at a 95% level of the histogram of the bootstrapped AUC values yielded the confidence intervals. Given 0.5 as the null hypothesis value of AUC, the p-value was a matter of counting the proportion of time AUC was less than or equal to 0.5.

7.3 Results

Using the method described in the previous chapter, prostate WG, CG, and PZ were delineated from ADC map MR images with the mean DSC of $88.27\% \pm 8.13\%$, $85.42\% \pm 7.72\%$, and

83.16% \pm 8.83% on 80 test images. The presented methodology was used to detect and segment PCa in the PZ from the test dataset which was comprised of 80 cases (39 normal and 41 patients with clinically significant PZ tumors; a total of 660 slices). Figure 7-5 displays the results of detected PCa in the PZ for three example patients. Table 7-3 shows the summary of results using our algorithm. For most of the test cases (29 out of 41), the tumor was detected in all slices that included tumor. Case-by-case study of sensitivity over the test population revealed that the minimum of sensitivity was 25% (for two test cases) that for all patients with a clinically significant tumor, the developed model identified PCa in at least one slice that contained tumor. The false-positive rate was 16.57% on the testing set. The average time required for PCa localization from a typical ADC map MR image including prostate zonal segmentation, image-preprocessing, and majority voting of outputs of the ADC-Nets for final prediction is 1.40 \pm 0.28 seconds on Intel Core i7, 2.8 GHz on a GPU-accelerated computing platform.

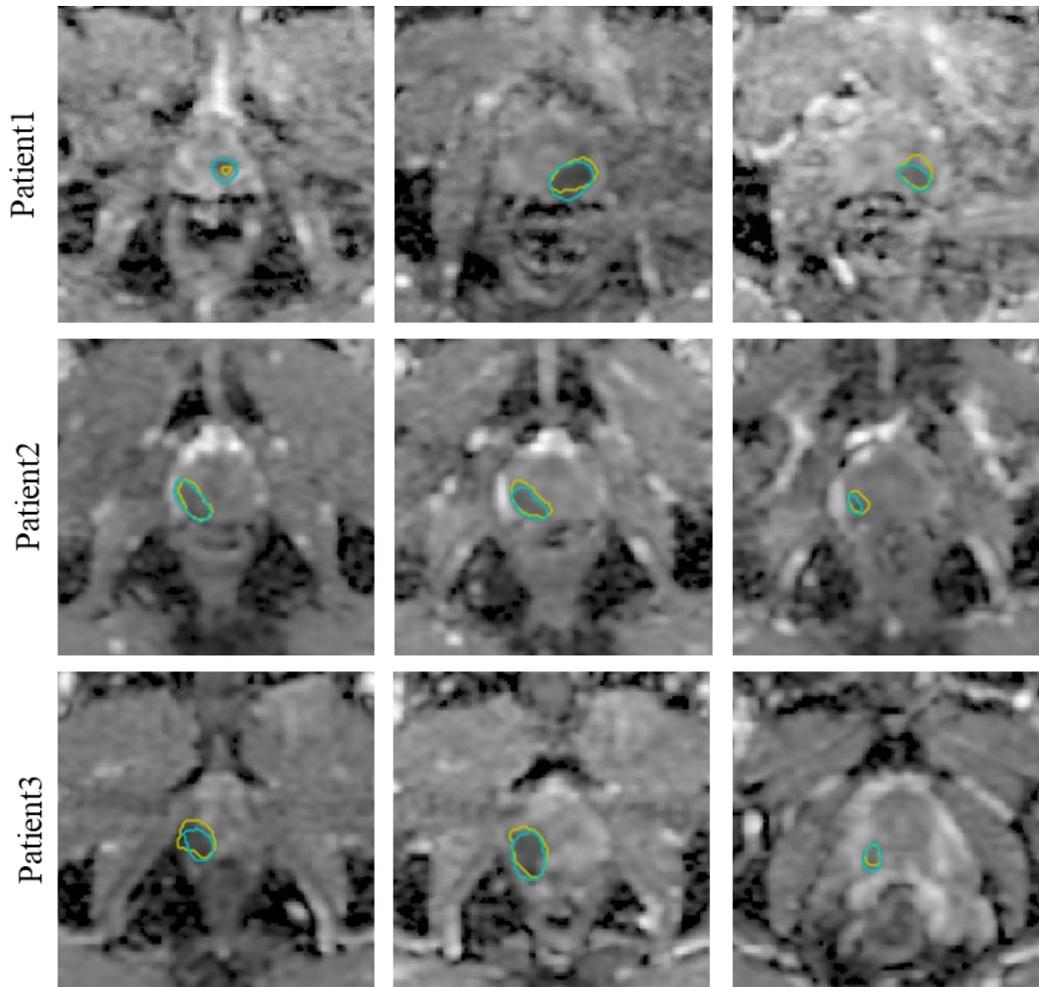


Figure 7-5. Example results from localized PCa on ADC map MR images in three patients from top to bottom. In each row, the 2D slices compiled over prostate WG from base to apex are shown from left to right. In all images, the PCa contours computed from manual and algorithm-generated segmentations are shown in cyan and yellow, respectively. Reprinted with permission.

Table 7-3. Summary of results obtained from the proposed model for fully automated PCa localization.

Reprinted with permission.

	DSC (%)	Sensitivity (%)	Specificity (%)
Mean	86.72	85.76	76.44
Standard Deviation	9.93	23.33	23.70
First quartile	80.06	66.66	65.44
Third quartile	94.96	100.00	100.00

Table 7-4 shows a summary of results of causality study on the same test dataset for each observation, where bold fonts indicate the highest sensitivity and specificity results. According to the results of the causality study, in a U-Net model with 4 stages, the sensitivity of the network for PCa detection is raised by increasing the number of filters from 8 to 16 and from 16 to 32 at the cost of specificity reduction. When we increased the depth of the network by adding another stage, the same pattern for estimated metrics was observed by increasing the number of filters from 8 to 16 whereas by enhancing the number of filters from 16 to 32 the sensitivity dropped dramatically. Therefore, the optimal solution for this problem was obtained by combining several models aiming to improve both sensitivity and specificity. A combination of 3 and 5 networks was examined through the majority voting system and sensitivity and specificity (per slice) were re-computed to find the best combination for the final ensemble model. The optimal point for sensitivity and specificity was yielded by the combination of ADC-Net1, ADC-Net2, and ADC-Net3. The paired t-test results demonstrated that compared to the ADC-Net2, the ensemble model improves sensitivity significantly ($p < 0.0001$). Additionally, the paired t-test yielded a $p < 0.0001$ in comparing the means of specificity of ADC-Net1 and ADC-Net3 with that of the ensemble model, which indicates that the ensemble model significantly ameliorates the specificity for PCa detection compared to ADC-Net1 and ADC-Net3. As an alternative, we provided an option so that based on a given sensitivity or specificity requirement in a clinical setup, the system will automatically choose and apply a suitable network configuration to the problem, given the fact that three candidate networks of the ensemble have performances evidenced by decreasing sensitivity increases specificity.

Table 7-4. Causality Experiment Results for the Proposed Model for Fully Automated PCa Localization.

Reprinted with permission.

Network	No. of Stages	No. of Filters in the First Stage	Number of Trainable Parameters	Mean of Sensitivity per Slice (%)	Mean of Specificity per Slice (%)
ADC-Net1	5	8	492,541	87.78 ± 20.85	64.58 ± 25.87
ADC-Net2	5	32	7,852,261	77.48 ± 29.09	77.10 ± 22.61
ADC-Net3	4	16	488,693	84.75 ± 25.83	68.54 ± 23.91
ADC-Net4	5	16	1,965,429	91.23 ± 16.55	54.39 ± 31.98
ADC-Net5	5	64	31,390,149	79.82 ± 28.33	72.91 ± 28.12
ADC-Net6	4	8	122,813	69.19 ± 36.06	76.72 ± 24.18
ADC-Net7	4	32	1,949,669	85.56 ± 21.05	55.68 ± 26.56
Ensemble Network	5,5,4	8,32,16	8,833,495	85.76 ± 23.33	76.44 ± 23.70

Figure 7-6 illustrates the ROC curve for the detection of clinically significant PZ PCa using our method. The AUC for the ensemble model is reported as 0.779. Figure 7-7 displays the histogram of the ROC AUC values obtained from 10000 bootstrapping of the AUC calculation. The confidence intervals at 95% level was [0.740 - 0.818] with a p-value < 0.0001.

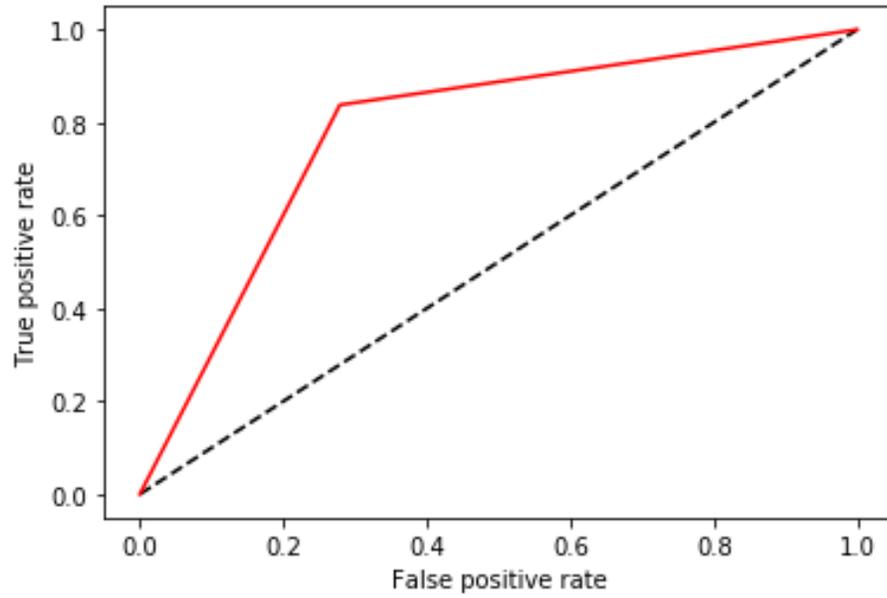


Figure 7-6. ROC curve for ensemble U-Nets, where the true positive rate (sensitivity) against false-positive rate (1 – specificity) has been plotted. Reprinted with permission.

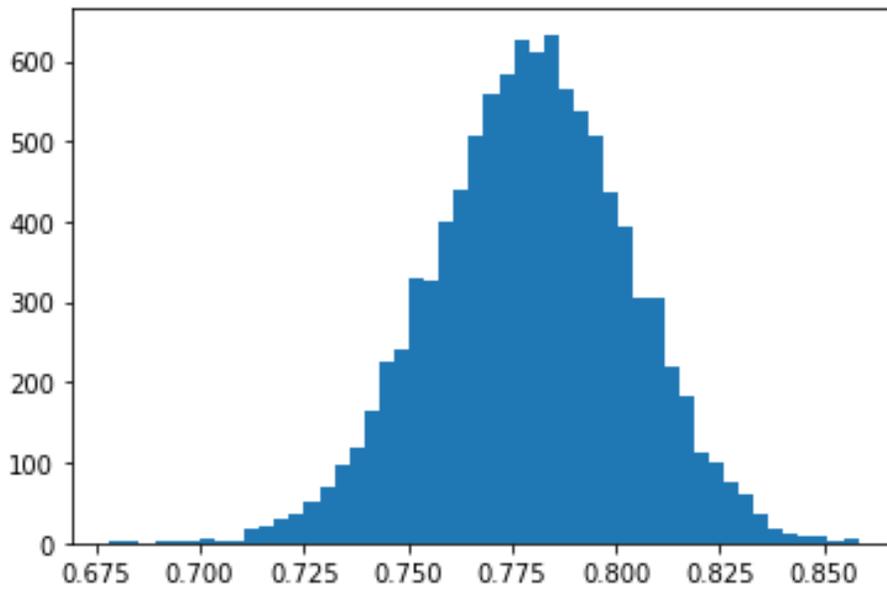


Figure 7-7. Histogram of the ROC AUC values obtained from 10,000 bootstrapping of the AUC calculation. Reprinted with permission.

Figure 7-8 shows examples of cases in which our developed model failed to detect PCa (false negative) or where PZ was marked as PCa (false-positive).

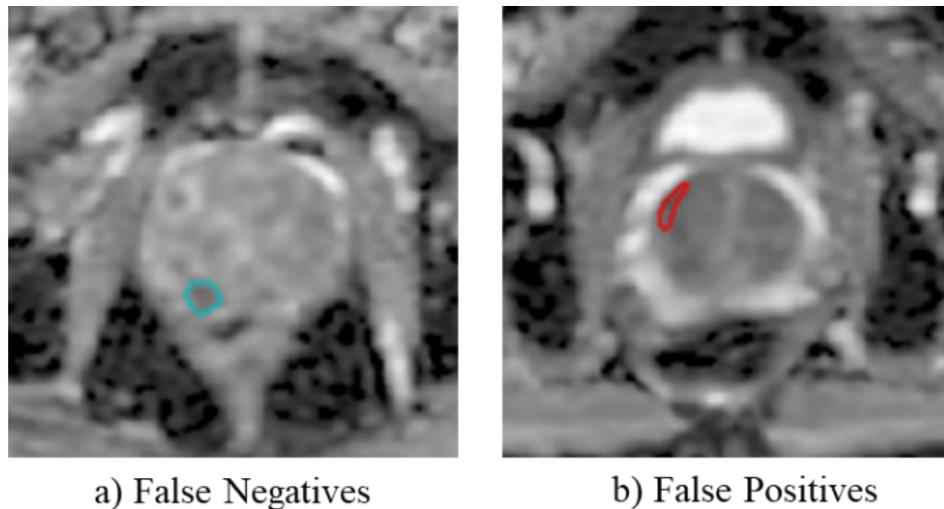


Figure 7-8. Example results of (a) undetected PCa of one patient with a significant tumor. The tumor may have been missed due to the system incorrectly attributing the tumor as representing a portion of the adjacent transition zone; (b) falsely labeled as PCa of one patient without significant tumor. The area marked as tumor represents the adjacent transition zone. The PCa contours computed from manual and algorithm-generated segmentations are shown in cyan and red, respectively. Reprinted with permission.

7.4 Discussion

This study presents an ensemble learning-based model for fully automated PCa localization in the PZ from ADC map MR images. An advantage of the proposed framework is that it does not require a full mp-MRI examination and identifies the position of potential areas of PCa using only ADC map MRI derived from diffusion-weighted images, which is the most important sequence for identification of tumors in PZ [22], [27]. The ADC map (which is a visual pixel-based rendering of individual ADC pixel values derived from the logarithmic decay curve

generated from trace low to high b-value DWI) is considered the most important image sequence for detection of PZ tumors. ADC has been shown repeatedly to be the most superior metric in prostate mp-MRI for the detection of PZ cancers and has been correlated inversely with Gleason score, stage of disease and risk of biochemical recurrence [24], [28-29]. An advantage of using only ADC is that the increased incidence of false-positive results encountered using T2W-MRI are potentially avoided [30-31] and Gadolinium is not required as it is for dynamic contrast-enhanced (DCE)-MRI, which is currently being carefully re-evaluated as a potentially unnecessary sequence in prostate mp-MRI in clinical practice. There is mounting evidence that DCE-MRI may provide no incremental advantage compared to DWI/ADC alone, in a majority of patients, and that it may be potentially removed from prostate mp-MRI, to shorten mp-MRI exam times, reduce the cost associated with MRI and address concerns over Gadolinium retention [13] and [32].

Our fully automated method was highly sensitive for PCa detection in the PZ with the added benefit of low interpretation times which only took a matter of seconds to perform.

Comparatively, semi-automated methods require human segmentation of prostatic zonal anatomy (which in our experience can take up to 20 minutes to perform accurately) and show low intra- and inter-observer agreement [33]. Since the suggested strategy is fully automated, in which manual segmentation of prostate zones is not required, this technique is entirely reproducible.

Another contribution of our study is that we evaluated prostate ADC map MR images in a large sample of men with pathologically confirmed clinically significant cancers with MRI-radical prostatectomy mapping of tumors and included a well-designed control group of patients without clinically significant tumors. This distinction is important because if automated systems are to be

used in clinical practice, methods for PCa localization should be tested in patients with and without significant tumors.

Three main reasons for using ensemble-based model are 1) statistical, which is related to the lack of adequate training data for properly representing the data distribution; 2) computational, where among multiple models that can fit a given problem, selecting the proper one is challenging, and 3) representational in which the decision boundary that separates data from different classes is too complex and cannot be implemented by a chosen classifier [34]. In designing a model for PCa detection in the PZ, we observed that for parameter adjustment there should be a trade-off between sensitivity and specificity, thus using an ensemble-based model make computational and representational sense, in our scenario. Selecting diverse elements that producing different errors is a key concept for an ensemble model to succeed in correcting the errors of its members. In other words, it is essential that constituent classifiers are negatively correlated. The most popular methods to achieve diversity are i) training a classifier using a different subset of training data. k-fold cross-validation and bagging are the most well-known algorithms of this type. In k-fold cross-validation, the training dataset is randomly divided into k subsets, where k-1 subsets are used for train and one subset for test a classifier [35]. In bagging, k samples are selected randomly from a training set contains k instances with replacement to train a classifier [36]. The outcomes of classifiers are then averaged for the final prediction. ii) training a model using a subset of features, where various features are extracted from all training data and a model is trained using different subsets of those features. Several approaches have been introduced in the literature on how to select the best subset to achieve high diversity and overall low ensemble error [37-38]. iii) training different types of classifiers, in which several different classifiers are trained on the same data and their outcomes are combined for final prediction. iv) using

different parameters for a learning algorithm [39]. In summary, in methods described in (i) and (ii), training dataset or the features derived from training samples are manipulated to train a learner whereas, in techniques explained in (iii) and (iv), multiple classifiers (same or different types) are trained using the same training samples. We employed the last approach in which three different U-Nets were designed and combined for a better PCa prediction as U-Net have recently been utilized in the medical image analysis domain for the segmentation tasks and achieved breakthrough results [40-41]. We integrated those classifiers via a voting-based system that not only predicts the location of PCa more confidently but also improves final model specificity through suppressing false-positive instances. Since the majority voting system was employed for the final prediction, the minimum of three networks was required in this study. We could have used more networks (e.g. 5) however, we found out that combining more than three networks with different width and depth deteriorate the performance of the model for PCa detection. To select proper individual networks, we referenced to the diversity of the network outputs and chose those with different detection errors to enable ensemble model for more accurate PCa localization. Also, wherever it was needed, we elected to trade off increased sensitivity for lower specificity, to preferentially reduce the risk of undetected tumors. A limitation of this strategy could be overdiagnosis of low risk Gleason score $3+3=6$ tumors; however, our study which included primarily clinically significant Gleason score $3+4=7$ or higher tumors could not evaluate this effect. Future studies, which evaluate automated tumor detection and include low risk Gleason score $3+3=6$ tumors among clinically significant cancer are needed. Moreover, we suggested using true positive as a metric and negative value of DSC as a loss function for training networks to enhance true positive rates during network training while

decreasing false-positive rates that were crucial in this research, ameliorated the performance of the suggested method.

Notably, the number of slices that encompass the prostate in ADC map MR images in the testing set varied between 5 and 14 with the median of 8, due to differences in the size of the prostate gland between patients. The number of slices with PCa also differed between 2 and 6, due to differences in the size of tumors between patients. Thus, sensitivity and specificity could decrease dramatically even for a single miss in patients with a smaller number of slices across prostate WG or a smaller number of slices containing the tumor.

Case-by-case investigation of errors revealed that in most cases, PCa detection errors happened due to the error in PZ segmentation, where segmented PZ shrinks inward such that it does not include PCa or leaks outward that causes detecting PCa in transition zone or falsely labeling non-cancerous tissue with the similar intensity profile with that of PCa structure as tumor.

We used DSC as a region-based metric to assess the performance of the developed method for PCa delineation. This metric can be quite sensitive to small errors occurring in small segmented objects (e.g. prostate gland and tumors) and conversely less sensitive to errors in larger segmented objects. As the size of PCa on an ADC map MR image can be relatively small, DSC penalizes largely on small errors.

Our study has several limitations. We only used the ADC map for the present study and not T2W-MRI or DCE-MRI, which could be considered a limitation of our study; however, registering the mp-MRI may introduce error to the algorithm. Moreover, we only evaluated PZ tumors. This was because TZ tumors are relatively less common than PZ tumors and are more difficult to differentiate from adjacent benign prostatic hyperplasia (BPH) which may overlap in

imaging features [42]. Future studies are required to determine the accuracy of automated methods of TZ tumor detection using prostate MRI. Additional research is needed to compare the performance of the proposed algorithm with the state-of-the-art methods for PCa localization using ADC map MRI.

7.5 Conclusion

We described a fully automated ensemble learning-based method in which three U-Nets were designed and incorporated to localize PCa in the PZ using only ADC map MR images. Our technique was robust and reproducible across a relatively large sample of patients with variable prostate sizes with and without clinically significant tumors. Our results confirmed that automated PCa detection and segmentation using ADC map MR image is feasible, highly sensitive and can be performed rapidly. Our results are important since the ADC map is the most important sequence for PCa detection in the PZ and there is a desire to improve the accuracy of tumor detection and inter-observer agreement in clinical practice compared to subjective interpretation.

References

1. F. Zabihollahy, E. Ukwatta, S. Krishna, and N. Schieda, "Fully automated localization of prostate peripheral zone tumors on apparent diffusion coefficient map MR images using an ensemble learning method," *J. Magn. Reson. Imaging*, 2019.
2. Canadian Cancer Society, "Canadian Cancer Society's Advisory Committee on Cancer Statistics," *Can. Cancer Stat. 2017*, 2017.
3. M. De Rooij, E. H. J. Hamoen, J. J. Fütterer, J. O. Barentsz, and M. M. Rovers, "Accuracy of multiparametric MRI for prostate cancer detection: A meta-analysis," *Am. J. Roentgenol.*, 2014.
4. H. U. Ahmed, A. El-Shater Bosaily, L. C. Brown, R. Gabe, R. Kaplan, M. K. Parmar, *et al.*, "Diagnostic accuracy of multi-parametric MRI and TRUS biopsy in prostate cancer (PROMIS): a paired validating confirmatory study," *Lancet*, 2017.
5. V. Kasivisvanathan, A. S. Rannikko, M. Borghi, V. Panebianco, L. A. Mynderse, M. H. Vaarala, *et al.*, "MRI-targeted or standard biopsy for prostate-cancer diagnosis," *N. Engl. J. Med.*, 2018.

6. J. S. Quon, B. Moosavi, M. Khanna, T. A. Flood, C. S. Lim, and N. Schieda, "False-positive and false negative diagnoses of prostate cancer at multi-parametric prostate MRI in active surveillance," *Insights into Imaging*. 2015.
7. T. C. Mussi, F. I. Yamauchi, C. F. Tridente, A. Tachibana, V. M. Tonso, D. R. Recchimuzzi, *et al.*, "Interobserver Agreement and Positivity of PI-RADS Version 2 Among Radiologists with Different Levels of Experience," *Acad. Radiol.*, 2019.
8. B. Turkbey, A. B. Rosenkrantz, M. A. Haider, A. R. Padhani, G. Villeirs, K. J. Macura, *et al.*, "Prostate Imaging Reporting and Data System Version 2.1: 2019 Update of Prostate Imaging Reporting and Data System Version 2," *European Urology*. 2019.
9. S. Gaur, S. Harmon, L. Rosenblum, M. D. Greer, S. Mehralivand, M. Coskun, *et al.*, "Can apparent diffusion coefficient values assist pi-rads version 2 dwi scoring? A correlation study using the pi-radsv2 and international society of urological pathology systems," *Am. J. Roentgenol.*, 2018.
10. S. Gaur, N. Lay, S. A. Harmon, S. Doddakashi, S. Mehralivand, B. Argun, *et al.*, "Can computer-aided diagnosis assist in the identification of prostate cancer on prostate MRI? A multi-center, multi-reader investigation," *Oncotarget*, 2018.
11. G. Lemaître, R. Martí, J. Freixenet, J. C. Vilanova, P. M. Walker, and F. Meriaudeau, "Computer-Aided Detection and diagnosis for prostate cancer based on mono and multi-parametric MRI: A review," *Computers in Biology and Medicine*. 2015.
12. S. Wang, K. Burt, B. Turkbey, P. Choyke, and R. M. Summers, "Computer aided-diagnosis of prostate cancer on multiparametric MRI: A technical review of current research," *BioMed Research International*. 2014.
13. X. ke Niu, X. hui Chen, Z. fan Chen, L. Chen, J. Li, and T. Peng, "Diagnostic performance of biparametric MRI for detection of prostate cancer: A systematic review and meta-analysis," *Am. J. Roentgenol.*, 2018.
14. M. L. Giger, H. P. Chan, and J. Boone, "Anniversary paper: History and status of CAD and quantitative image analysis: The role of Medical Physics and AAPM," *Medical Physics*. 2008.
15. Y. Song, Y. D. Zhang, X. Yan, H. Liu, M. Zhou, B. Hu, G. Yang, "Computer-aided diagnosis of prostate cancer using a deep convolutional neural network from multiparametric MRI," *J. Magn. Reson. Imaging*, 2018.
16. Y. K. Tsehay, N. S. Lay, H. R. Roth, X. Wang, J. T. Kwak, B. Turkbey, *et al.*, "Convolutional neural network based deep-learning architecture for prostate cancer detection on multiparametric magnetic resonance images," in *Medical Imaging 2017: Computer-Aided Diagnosis*, 2017.
17. X. Yang, C. Liu, Z. Wang, J. Yang, H. L. Min, L. Wang, *et al.*, "Co-trained convolutional neural networks for automated detection of prostate cancer in multi-parametric MRI," *Med. Image Anal.*, 2017.
18. L. Rokach, "Ensemble-based classifiers," *Artif. Intell. Rev.*, 2010.
19. O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 2015.
20. A. S. Puryso, A. B. Rosenkrantz, J. O. Barentsz, J. C. Weinreb, and K. J. Macura, "PI-RADS version 2: A pictorial update," *Radiographics*, 2016.

21. A. Bhavsar and S. Verma, "Anatomic Imaging of the Prostate," *BioMed Research International*. 2014.
22. J. C. Weinreb *et al.*, "PI-RADS Prostate Imaging - Reporting and Data System: 2015, Version 2," *Eur. Urol.*, 2016.
23. F. Bratan, C. Melodelima, R. Souchon, A. Hoang Dinh, F. Mège-Lechevallier, S. Crouzet, *et al.*, "How accurate is multiparametric MR imaging in evaluation of prostate cancer volume?," *Radiology*, 2015.
24. S. Krishna, C. S. Lim, M. D. F. McInnes, T. A. Flood, W. M. Shabana, R. S. Lim, *et al.*, "Evaluation of MRI for diagnosis of extraprostatic extension in prostate cancer," *J. Magn. Reson. Imaging*, 2018.
25. F. Milletari, N. Navab, and S. A. Ahmadi, "V-Net: Fully convolutional neural networks for volumetric medical image segmentation," in *Proceedings - 2016 4th International Conference on 3D Vision*, 2016.
26. M. D. Zeiler, "ADADELTA: An Adaptive Learning Rate Method.," *arXiv:1212.5701*, 2012.
27. J. O. Barentsz, J. C. Weinreb, S. Verma, H. C. Thoeny, C. M. Tempany, F. Shtern, *et al.*, "Synopsis of the PI-RADS v2 Guidelines for Multiparametric Prostate Magnetic Resonance Imaging and Recommendations for Use," *European Urology*. 2016.
28. A. B. Rosenkrantz, J. M. Ream, P. Nolan, H. Rusinek, F. M. Deng, and S. S. Taneja, "Prostate cancer: Utility of whole-lesion apparent diffusion coefficient metrics for prediction of biochemical recurrence after radical prostatectomy," *Am. J. Roentgenol.*, 2015.
29. T. H. Kim, C. K. Kim, B. K. Park, H. G. Jeon, B. C. Jeong, S. I. Seo, *et al.*, "Relationship between Gleason score and apparent diffusion coefficients of diffusion-weighted magnetic resonance imaging in prostate cancer patients," *Can. Urol. Assoc. J.*, 2016.
30. M. Chen, H. D. Dang, J. Y. Wang, C. Zhou, S. Y. Li, W. C. Wang, *et al.*, "Prostate cancer detection: Comparison of T2-weighted imaging, diffusion-weighted imaging, proton magnetic resonance spectroscopic imaging, and the three techniques combined," *Acta radiol.*, 2008.
31. H. Miao, H. Fukatsu, and T. Ishigaki, "Prostate cancer detection with 3-T MRI: Comparison of diffusion-weighted and T2-weighted imaging," *Eur. J. Radiol.*, 2007.
32. M. Scialpi, V. Rondoni, M. C. Aisa, E. Martorana, A. D'Andrea, C. M. Malaspina, *et al.*, "Is contrast enhancement needed for diagnostic prostate MRI?," *Translational Andrology and Urology*. 2017.
33. F. Khalvati, A. Salmanpour, S. Rahnamayan, M. A. Haider, and H. R. Tizhoosh, "Sequential Registration-Based Segmentation of the Prostate Gland in MR Image Volumes," *J. Digit. Imaging*, 2016.
34. T. G. Dietterich, "Ensemble methods in machine learning," in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 2000.
35. A. Krogh and J. Vedelsby, "Neural Network Ensembles, Cross Validation, and Active Learning," *Adv. Neural Inf. Process. Syst.* 7, 1995.
36. L. Breiman, "Bagging predictors," *Mach. Learn.*, 1996.
37. G. Zenobi and P. Cunningham, "Using diversity in preparing ensembles of classifiers based on different feature subsets to minimize generalization error," in *Lecture Notes in*

Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), 2001.

38. P. Melville and R. J. Mooney, "Constructing diverse classifier ensembles using artificial training examples," in *IJCAI International Joint Conference on Artificial Intelligence*, 2003.
39. G. Brown, J. Wyatt, R. Harris, and X. Yao, "Diversity creation methods: A survey and categorisation," *Inf. Fusion*, 2005.
40. G. Tong, Y. Li, H. Chen, Q. Zhang, and H. Jiang, "Improved U-NET network for pulmonary nodules segmentation," *Optik (Stuttg.)*, 2018.
41. A. Fabijańska, "Segmentation of corneal endothelium images using a U-Net-based convolutional neural network," *Artif. Intell. Med.*, 2018.
42. A. B. Rosenkrantz, S. Kim, N. Campbell, B. Gaing, F. M. Deng, and S. S. Taneja, "Transition zone prostate cancer: Revisiting the role of multiparametric MRI at 3 T," in *American Journal of Roentgenology*, 2015.

Chapter 8: Conclusion and Future Works

The following chapter concludes the thesis and provides a summary of the results. The limitations and some suggestions for future research are presented as well.

8.1 Summary

Medical imaging is a non-invasive tool that can be used for the detection of medical abnormalities. Advances in technology lead to larger image sizes, cause interpretation and manual delineation in 3D challenging and subject to inter-operator variability. Therefore, automated diagnosis and prognosis of abnormalities and disease is extremely beneficial. With the availability of large annotated datasets, DL-based methods are ideally suited for detecting abnormalities in medical images. Having access to some unique datasets of cardiac, prostate and kidney CT and MR images through our collaboration with the Ottawa Hospital and the Libin Cardiovascular Institute of Alberta at the University of Calgary inspired us to analyze abnormalities (i.e., detection, segmentation, characterization, and classification) in those medical images.

In this dissertation, I have described several algorithms: 1) to differentiate benign cyst from solid renal mass on CECT images; 2) to detect and segment renal mass on CECT scans and from there classify them into benign and malignant; 3) to segment and characterize LV scar in 3D LGE MRI; 4) to segment prostate WG, CG, and PZ on T2W and ADC map MRI, and 5) to localize PCa in the PZ using ADC map MRI. The designed algorithms evolved from a CNN with several layers to U-Net, cascaded multi-planar U-Net, and the ensemble learning system. The abnormality detection and segmentation algorithms described in this thesis are aimed toward

alleviating the measurement burden of the cardiologists and radiologists. Algorithms for kidney tumor evaluation problems were developed by integrating information extracted from images with domain knowledge and user interactions. Other methods described for LV scar segmentation and PCa localization are fully automated, thus completely reproducible without any user interactions.

An ideal algorithm is expected to be highly accurate, precise, fast, and reproducible. While it is challenging to develop methodologies that yield all those criteria for abnormality detection and segmentation in medical images, some of those criteria were achieved as described below.

8.1.1 Differentiation of Benign Cyst from Solid Renal Mass using a Patch-Based CNN (Chapter 3)

Automated classification of renal masses detected in CT examinations into benign cyst versus solid mass is clinically valuable. This distinction may be challenging at single-phase CE-CT examinations, where cysts may simulate solid masses and where renal masses are most commonly incidentally detected. This may lead to unnecessary and costly follow-up imaging for accurate characterization. In Chapter 3, we described a patch-based CNN method to differentiate benign cysts from solid renal masses using single-phase CECT images. The predictions of the network for patches extracted from a manually segmented lesion were combined through the majority voting system for renal diagnosis. Since radiologists differentiate benign cysts from solid renal masses based on the intensity and texture, we selected CNN architecture that allows learning a combination of intensity and texture features simultaneously. We used a dataset comprised of single-phase CECT images of 315 patients with 77 benign (oncocytomas, and fat

poor renal angiomyolipoma) and 238 malignant (RCC including clear cell, papillary, and chromophobe subtypes) tumors. We trained our proposed network using patches extracted and artificially augmented from 40 CECT scans. The presented algorithm was evaluated using 275 unseen CECT test images consisting of 327 renal masses by comparing algorithm-generated labels to those labeled by experts and achieved mean accuracy, precision, and recall of 88.96%, 95.64%, and 91.64%. Our method yielded an accuracy of $91.21\% \pm 25.88\%$ as a mean \pm standard deviation at the patient level. The AUC was reported as 0.804. The results indicate that our algorithm may accurately characterize benign cysts from solid masses with a high degree of accuracy and may be clinically valuable to prevent unnecessary imaging follow-up for characterization in a proportion of patients.

8.1.2 Classification of Solid Renal Masses into Benign and RCC (Chapter 4)

The objective of this study was to develop a CNN-based method for automated classification of benign from RCC solid renal masses using CECT images. We used a Dataset of CECT scans of 315 patients with 77 benign (57 oncocytomas, and 20 fp-AML) and 238 malignant (RCC: 123 clear cell, 69 papillary, and 46 chromophobe subtypes) tumors. We employed a decision fusion-based model to aggregate slice level predictions determined by CNN via a majority voting system to evaluate renal masses on CECT. As the shape of the tumor is an important feature for the differentiation of benign from RCC, we used an image-based approach instead of a path-based pipeline that enables the network to capture global features as well as the local features. The CNN-based model was trained using 7023 slices with renal masses manually extracted from CECT images of 155 patients, cropped automatically around kidneys, and augmented artificially.

We also examined the fully automated approach for renal mass evaluation on CECT. Moreover, a 3D CNN was trained and tested using the same datasets and the obtained results were compared to those acquired from slice-wise algorithms. For differentiation of RCC versus benign solid masses, the semi-automated majority voting-based CNN algorithm achieved accuracy, precision, and recall of 83.75%, 89.05%, and 91.73% using 160 test cases, respectively. Fully automated pipeline yielded accuracy, precision, and recall of 77.36%, 85.92%, and 87.22% on the same test cases, respectively. 3D CNN reported accuracy, precision, and recall of 79.24%, 90.32%, and 84.21% using 160 test cases, respectively.

For implementing a fully automated method, we used an ensemble of U-Net models to first localize renal mass on CECT images of institutional datasets. We further evaluated our algorithm using the KiTS19 images, which is a public dataset. The developed algorithm for renal mass localization yielded a mean DSC of $88.65\% \pm 7.31\%$ and $87.91\% \pm 6.82\%$ on the institutional and KiTS19 test datasets, respectively.

8.1.3 Automated Identification of LV scar from 3D LGE MRI (Chapter 5)

3D LGE MR imaging enables the quantification of myocardial scar at high resolution with unprecedented volumetric visualization. Automated segmentation of myocardial scar is critical for potential clinical translation of this technique given the number of tomographic images acquired. We describe the development of cascaded multi-planar U-Net (CMPU-Net) to efficiently segment the boundary of the LV myocardium and scar from 3D LGE-MR images. In this approach, two subnets, each containing three U-Nets, were cascaded to first segment the LV myocardium and then segment the scar within the pre-segmented LV myocardium. This method

was developed to take advantage of isotropic voxel properties in 3D MRI that allows for multi-planar reconstruction. The U-Nets were trained separately using 2D slices extracted from axial, sagittal, and coronal slices of 3D LGE-MR images. We used 3D LGE-MR images from 34 subjects with chronic ischemic cardiomyopathy. The U-Nets were trained using 8430 slices, extracted in three orthogonal directions from 18 images. In the testing phase, the outputs of U-Nets of each subnet were combined using the majority voting system for final label prediction of each voxel in the image. The developed method was tested for accuracy by comparing its results to manual segmentations of LV myocardium and LV scar from 7250 slices extracted from 16 3D LGE-MR images. Our method was also compared to numerous alternative methods based on ML, energy minimization, and intensity-thresholds. Our algorithm reported a mean DSC, AVD, and HD of $85.14\% \pm 3.36\%$, $43.72 \pm 27.18 \text{ cm}^3$, and $19.21 \pm 4.74 \text{ mm}$ for determining the boundaries of LV myocardium from LGE-MR images. Our method also yielded a mean DSC, AVD, and HD of $88.61\% \pm 2.54\%$, $9.33 \pm 7.24 \text{ cm}^3$, and $17.04 \pm 9.93 \text{ mm}$ for LV scar segmentation on the unobserved test dataset. Our method significantly outperformed the alternative techniques in segmentation accuracy ($p\text{-value} < 0.05$).

8.1.4 Prostate Zonal Segmentation using T2W and ADC MR Images (Chapter 6)

Accurate regional segmentation of the prostate boundaries on MR images is a fundamental requirement before automated PCa diagnosis can be achieved. We described a novel methodology to segment prostate WG, CG, and PZ, where $PZ + CG = WG$, from T2W and ADC map prostate MR images. We designed two similar models each made up of two U-Nets to delineate the WG, CG, and PZ from T2W and ADC map MR images, separately. The U-Net,

which is a modified version of a fully CNN, includes contracting and expanding paths with convolutional, pooling, and up-sampling layers. Pooling and up-sampling layers help to capture and localize image features with high spatial consistency. We used a dataset consisting of 225 patients (combining 153 and 72 patients with and without clinically significant prostate cancer) imaged with multiparametric MRI at 3 Tesla. Our proposed model for prostate zonal segmentation from T2W was trained and tested using 1154 and 1587 slices of 100 and 125 patients, respectively. Median of DSCs on test dataset for prostate WG, CG, and PZ were $95.33 \pm 7.77\%$, $93.75 \pm 8.91\%$, and $86.78 \pm 3.72\%$, respectively. The designed model for regional prostate delineation from ADC map images was trained and validated using 812 and 917 slices from 100 and 125 patients. This model yielded a median DSC of $92.09 \pm 8.89\%$, $89.89 \pm 10.69\%$, and $86.1 \pm 9.56\%$ for prostate WG, CG, and PZ on test samples, respectively. Further investigation indicated that the proposed algorithm reported high DSC for prostate WG segmentation from both T2W and ADC map MR images irrespective of WG size. In addition, segmentation accuracy in terms of DSC does not significantly vary among patients with or without significant tumors.

8.1.5 Automated Localization of PCa in the PZ on ADC Map MRI (Chapter 7)

Accurate detection and localization of PCa in men undergoing prostate MRI is a fundamental step for future targeted prostate biopsies and treatment planning. Fully automated localization of PZ PCa using the ADC map might be clinically useful. We described an automated localization of PCa in the PZ on ADC map MR images using an ensemble U-Net-based model. A dataset of ADC map MR images of 226 patients (154 and 72 patients with and without clinically

significant PZ PCa, respectively) was split into 146 and 80 images for train and test respectively. The ground truth was established by manual delineation of the prostate and prostate PZ tumors on ADC maps by dedicated radiologists using MRI-radical prostatectomy maps as a reference standard. Our developed algorithm yielded DSC, sensitivity, and specificity of $86.72\% \pm 9.93\%$, $85.76\% \pm 23.33\%$, and $76.44\% \pm 23.70\%$, respectively (mean \pm standard deviation) on 80 test cases consisting of 41 and 39 instances from patients with and without clinically significant tumors including 660 extracted 2D slices. AUC was reported as 0.779.

8.2 Limitations and Future Works

The main limitation of this dissertation is having a limited number of images in our datasets due to financial constraints, which is a common issue for all medical image processing problems. We employed different regularization techniques to avoid overfitting issues. Furthermore, as the datasets used in our studies were a representative of real data, in most cases we had imbalanced data in our datasets as usually, the number of cases with abnormalities is much less than the normal instances. Imbalanced data causes the classifier to pick up the patterns in the most popular classes and ignore the least popular ones (abnormalities). We addressed this problem by artificially augmenting the data from underrepresented classes in the training set. Also, metrics such as sensitivity, specificity, ROC, and PR curves rather than accuracy were chosen to evaluate the performance of the trained model in the test stage to assure the good performance of the network, facing unseen imbalanced data. Another limitation of our algorithm evaluation is the lack of ground truth for image segmentation, which is a common limitation for medical image processing. We used expert manual segmentation as a surrogate to evaluate the performance of

developed algorithms for abnormalities detection. However, it is not clear that the error is due to the algorithm or manual segmentation. Although phantoms can be used as ground truth, this is not realistic as it is challenging to make a phantom that emulates the characteristics of patients' tissue. The limitations specific to each algorithm described in this dissertation are described as follows.

It is notable that normal distribution cannot be assumed for the measured metrics such as accuracy, DSC, precision, sensitivity, and specificity across the test datasets. For this reason, mean value does not properly reflect the average and the measured mean value plus standard deviation might exceed 100% in some cases. To address this problem, median value along with first and third quartiles are reported as well.

8.2.1 Differentiation of Benign Cyst from Solid Renal Mass using a Patch-Based CNN (Chapter 3)

One limitation of our technique is that it requires expert manual identification of renal masses on CECT images. In practice, this is not a major limitation since renal lesion detection is rarely a clinical problem and renal lesion segmentation is by far less time consuming than whole kidney segmentation. Future studies are required to develop a fully automated approach including renal mass detection and segmentation. In this research, we used patch-based CNN that inherits all limitations of patch-based approach including redundant computation due to the overlap exists between adjacent patches, being slow, and incapability of capturing both local and global features simultaneously [1]. Although the sample size in this study was quite large and images were acquired from several CT scanners, images were taken from a single institution

retrospective cohort that may create the potential for population bias. Evaluation of our technique in other patient populations is required. The number of reported cysts measuring ≥ 20 HU and ≥ 30 HU at enhanced CT is higher than expected and this most likely relates to the automated HU calculation derived from the whole cyst rather than the single slice HU measurement using ROI placement in clinical practice. Though we attempted to simulate manual ROI placement in practice by trimming the volume of HU measurement in our study, it is likely that the whole cyst analysis resulted in an increased mean HU value per cyst compared to what would be expected by using single slice ROI measurements. Nevertheless, the practice of ROI placement in clinical practice is not without its own limitations subject to observer differences in the size of ROI and placement of ROI which has been reported to vary reported HU measurements and accuracy of classification [2].

8.2.2 Classification of Solid Renal Mass into Benign and RCC (Chapter 4)

The retrospective nature of our study introduces potential population bias and validation of our technique in other populations is required. To minimize inter-observer variability, only one experienced radiologist performed manual segmentation in this study. Further studies are required to examine the inter-reader variability of human and automated segmentations between centers and platforms. We did not evaluate the subjective performance of radiologists for solid renal mass characterization compared to our semi-automated method. Though it is possible to differentiate between some RCC subtypes and diagnose fp-AML using conventional CT analysis [3, 4], it is generally accepted that the differentiation of RCC from benign neoplasms is currently not accurate in clinical practice and not routinely performed [5]. Future studies may compare the

performance of automated models to radiologists' subjective interpretation and also assess how the results might alter the management of renal masses in clinical practice. We only studied renal masses on single-phase enhanced CT images, future studies might determine what benefit there is, if any, for summing information from a renal mass protocol CT which includes both an unenhanced, corticomedullary and nephrographic phase enhanced CT images. This may also be considered a limitation of our work; however, a majority of renal masses are incidentally discovered on single-phase enhanced CT images which provide more practical value to our analysis. The accuracy of our proposed system could potentially be improved by using data from multiphase CT in future studies.

8.2.3 Automated Identification of LV scar from 3D LGE MRI (Chapter 5)

While our sample dataset of 3D LGE MR images was sufficiently variable in terms of myocardial scar volume, architecture, location, and signal characteristics, all images were from a single institution retrospective cohort, which may introduce bias to our results. Validation within a multi-center cohort is desirable. Additionally, the manual segmentation was done by a single expert cardiac imager, so a remaining step of this work would be to study the interobserver variability of manual segmentation to assure this measure as an appropriate gold standard for algorithm training. Although multi-planar slice extraction increases LV scar delineation accuracy, it inherently slows segmentation efficiency compared to other methods where conventional feature learning approaches are employed. Furthermore, as we used cascaded U-Nets to segment scar within segmented LV myocardium, incorrect segmentation of LV myocardium will adversely influence the segmentation accuracy of scar.

8.2.4 Prostate Zonal Segmentation using T2W and ADC MR Images (Chapter 6)

Although the sample size in this study was larger than in previous studies, all images were from a single institution retrospective cohort. Additionally, the images were segmented by four expert Radiologists where each Radiologist performed a single unified segmentation for a part of the images in the dataset. This was necessary given the long time required to segment prostate T2W and ADC map images (approximately 30 minutes) and the large sample size. To minimize inter-observer variability, only experienced Radiologists performed manual segmentations after participating in a training session prior to the segmentation of cases used in this study. We examined different approaches for the delineation of PZ borders (i.e. segmenting PZ directly from MR images or within prostate WG as an ROI). Although subtracting WG and CG segmentation maps for generating the PZ segmentation map, yielded the highest segmentation accuracy, errors inserted by networks used for WG and CG delineation are added up, which in turn degrades PZ segmentation correctness.

8.2.5 Automated Localization of PCa in the PZ on ADC Map MRI (Chapter 7)

Although the sample size in this study was relatively large, all images were from a single institution retrospective cohort and this creates the potential for population bias. The sample is biased to overestimate the accuracy of an automated model given that it is formed from men with operated clinically significant tumors which are likely more readily detectable on ADC map images compared to lower volume and lower grade tumors which may undergo focal ablative techniques, radiotherapy or active surveillance. Additionally, the manual segmentation was done by four expert Radiologists where each Radiologist proposed a single unified segmentation for a

part of the images in the dataset which introduced inter-observer variability of segmentation into this study. This was necessary given the long time required to segment prostate T2W and ADC map images (approximately 20 minutes) and the large sample size. To minimize inter-observer variability, only experienced Radiologists performed segmentation and a training session was performed prior to segmentation of cases used in the present study. All Radiologists relied on existing MRI-Radical Prostatectomy maps showing areas of clinically significant cancers when segmenting tumors in the present study, which eliminated errors related to interpretation differences. Since our method applies several cascaded networks to delineate the PZ boundary and search within that ROI for PCa, the segmentation error accumulates to the PCa localization error, which is inevitable employing our proposed framework. We only used the ADC map for the present study and not T2W-MRI or DCE-MRI, which could be considered a limitation of our study; however, the use of DCE-MRI is now controversial with validated bi-parametric MRI protocols showing comparable accuracy for tumor detection compared to mp- MRI assessment [6]. Although bi-parametric MRI includes T2W-MRI, T2W-MRI imaging findings are not formally incorporated into PI-RADS scoring for PZ observations. It is possible that by incorporating information from other pulse sequences (e.g. T2W, trace DWI, and DCE) accuracies could be improved; however, this requires accurate registration of data between pulse sequences (which are generally performed at different spatial resolutions in clinical practice) and could introduce compounded errors related to segmentation. For example, including high b-value (e.g. $\geq 1400 \text{ mm}^2/\text{s}$) trace DWI could improve specificity for tumor detection but at the cost of expected poor automated segmentation of the zonal anatomy since the normal prostatic tissues are intentionally suppressed at high b-value DWI to better highlight areas of tumor [7]. Future studies may evaluate the incremental yield of combined sequences for fully automated tumor

detection. Future studies may also evaluate the ability of ML to not only segment the prostate and zonal anatomy, detect and segment cancers but also provide information on the extraprostatic extension of disease including seminal vesicle invasion. A final limitation of the current study is that we only evaluated PZ tumors. This was because TZ tumors are relatively less common than PZ tumors and are more difficult to differentiate from adjacent benign prostatic hyperplasia (BPH) which may overlap in imaging features [8]. Future studies are required to determine the accuracy of automated methods of TZ tumor detection using prostate MRI.

References

1. O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 2015.
2. A. T. Taner, N. Schieda, and E. S. Siegelman, "Pitfalls in Adrenal Imaging," *Semin. Roentgenol.*, 2015.
3. N. Schieda, R. S. Lim, M. D. F. McInnes, I. Thomassin, R. Renard-Penna, S. Tavolaro, *et al.*, "Characterization of small (<4 cm) solid renal masses by computed tomography and magnetic resonance imaging: Current evidence and further development," *Diagn. Interv. Imaging* 2018.
4. R. S. Lim, T. A. Flood, M. D. F. McInnes, L. T. Lavalley, and N. Schieda, "Renal angiomyolipoma without visible fat: Can we make the diagnosis using CT and MRI?," *Eur Radiol.* 2018.
5. S. K. Kang, W. C. Huang, P. V. Pandharipande, and H. Chandarana, "Solid renal masses: What the numbers tell us." *Am. J. Roentgenol* 2014.
6. J. Barentsz, "The Role of (fast) Bi-Parametric MRI versus Multi-Parametric MRI and TRUS-biopsy for Detecting Clinically Significant Prostate Cancer in Biopsy Naïve Men with Elevated PSA.," *EAU, Barcelona, Spain*; 2019.
7. S. Woo, C. H. Suh, S. Y. Kim, J. Y. Cho, and S. H. Kim, "Head-to-head comparison between high- and standard-b-value DWI for detecting prostate cancer: A systematic review and meta-analysis," *Am. J. Roentgenol.*, 2018.
8. A. B. Rosenkrantz, S. Kim, N. Campbell, B. Gaing, F. M. Deng, and S. S. Taneja, "Transition zone prostate cancer: Revisiting the role of multiparametric MRI at 3 T," in *American Journal of Roentgenology*, 2015.

Appendix A

Copyright reprint permissions

- 1) Journal of IEEE Access (Classification of Renal mass into cyst and solid)

Reuse of an author's own thesis or dissertation is allowed, as long as it is cited and has not been published elsewhere.

<https://ieeaccess.ieee.org/submitting-an-article/>

- 2) Book Chapter

 mohamed nazeeh <eng.mohamednazeeh@hotmail.com>
Wed 2020-02-19 12:39 PM
Fatemeh Zabihollahy; Shehata, Mohamed Nazih Mohamed <mohamed.shehata@louisville.edu>; ayman.elbaz@louisville.edu +1 other

[External Email]

Dear Dr. Fatma,

On behalf of Dr. El-Baz, you can consider this email as a written **permission** to reuse any material of the **book chapter** in your Dissertation for your PhD.
[Congrats](#) in advance.

Best regards,
Mohamed Shehata

...

- 3) Journal of Medical Physics (Prostate Zonal Segmentation)

JOHN WILEY AND SONS LICENSE

TERMS AND CONDITIONS

Feb 19, 2020

This Agreement between Fatemeh Zabihollahy ("You") and John Wiley and Sons ("John Wiley and Sons") consists of your license details and the terms and conditions provided by John Wiley and Sons and Copyright Clearance Center.

License Number 4772520762064

License date	Feb 19, 2020
Licensed Content Publisher	John Wiley and Sons
Licensed Content Publication	Medical Physics
Licensed Content Title	Automated segmentation of prostate zonal anatomy on T2-weighted (T2W) and apparent diffusion coefficient (ADC) map MR images using U-Nets
Licensed Content Author	Fatemeh Zabihollahy, Nicola Schieda, Satheesh Krishna Jeyaraj, et al
Licensed Content Date	May 11, 2019
Licensed Content Volume	46
Licensed Content Issue	7
Licensed Content Pages	13

Type of use Dissertation/Thesis

Requestor type Author of this Wiley article

Format Print and electronic

Portion Full article

Will you be
translating? No

Title of your
thesis /
dissertation Deep Learning Methods for Abnormality Detection and Segmentation in
Computed Tomography and Magnetic Resonance Images

Expected
completion date Mar 2020

Expected size
(number of pages)200

Requestor
Location Fatemeh Zabihollahy
339 Riversedge Crescent

Ottawa, ON K1V0Y6
Canada Attn:

Publisher Tax
ID EU826007151

Total 0.00 USD
Terms and Conditions

TERMS AND CONDITIONS

This copyrighted material is owned by or exclusively licensed to John Wiley & Sons, Inc. or one of its group companies (each a "Wiley Company") or handled on behalf of a society with which a Wiley Company has exclusive publishing rights in relation to a particular work (collectively "WILEY"). By clicking "accept" in connection with completing this licensing transaction, you agree that the following terms and conditions apply to this transaction (along with the billing and payment terms and conditions established by the Copyright Clearance Center Inc., ("CCC's Billing and Payment terms and conditions"), at the time that you opened your RightsLink account (these are available at any time at <http://myaccount.copyright.com>).

Terms and Conditions

- The materials you have requested permission to reproduce or reuse (the "Wiley Materials") are protected by copyright.
- You are hereby granted a personal, non-exclusive, non-sub licensable (on a standalone basis), non-transferable, worldwide, limited license to reproduce the Wiley Materials for the purpose specified in the licensing process. This license, **and any CONTENT (PDF or image file) purchased as part of your order**, is for a one-time use only and limited to any maximum distribution number specified in the license. The first instance of republication or reuse granted by this license must be completed within two years of the date of the grant of this license (although copies prepared before the end date may be distributed thereafter). The Wiley Materials shall not be used in any other manner or for any other purpose, beyond what is granted in the license. Permission is granted subject to an appropriate acknowledgement given to the author, title of the material/book/journal and the publisher. You shall also duplicate the copyright notice that appears in the Wiley publication in your use of the Wiley Material. Permission is also granted on the understanding that nowhere in the text is a previously published source acknowledged for all or part of this Wiley Material. Any third party content is expressly excluded from this permission.
- With respect to the Wiley Materials, all rights are reserved. Except as expressly granted by the terms of the license, no part of the Wiley Materials may be copied, modified, adapted (except for minor reformatting required by the new Publication), translated, reproduced, transferred or distributed, in any form or by any means, and no derivative works may be made based on the Wiley Materials without the prior permission of the respective copyright owner. **For STM Signatory Publishers clearing permission under the terms of the STM Permissions Guidelines only, the terms of the license are extended to include subsequent editions and for editions in other languages, provided such editions are for the work as a whole in situ and does not involve the separate exploitation of the permitted figures or extracts**, You may not alter, remove or suppress in any manner any copyright, trademark or other notices displayed by the Wiley Materials. You may not license, rent, sell, loan, lease, pledge, offer as security, transfer or assign the Wiley Materials on a stand-alone basis, or any of the rights granted to you hereunder to any other person.

- The Wiley Materials and all of the intellectual property rights therein shall at all times remain the exclusive property of John Wiley & Sons Inc, the Wiley Companies, or their respective licensors, and your interest therein is only that of having possession of and the right to reproduce the Wiley Materials pursuant to Section 2 herein during the continuance of this Agreement. You agree that you own no right, title or interest in or to the Wiley Materials or any of the intellectual property rights therein. You shall have no rights hereunder other than the license as provided for above in Section 2. No right, license or interest to any trademark, trade name, service mark or other branding ("Marks") of WILEY or its licensors is granted hereunder, and you agree that you shall not assert any such right, license or interest with respect thereto
- NEITHER WILEY NOR ITS LICENSORS MAKES ANY WARRANTY OR REPRESENTATION OF ANY KIND TO YOU OR ANY THIRD PARTY, EXPRESS, IMPLIED OR STATUTORY, WITH RESPECT TO THE MATERIALS OR THE ACCURACY OF ANY INFORMATION CONTAINED IN THE MATERIALS, INCLUDING, WITHOUT LIMITATION, ANY IMPLIED WARRANTY OF MERCHANTABILITY, ACCURACY, SATISFACTORY QUALITY, FITNESS FOR A PARTICULAR PURPOSE, USABILITY, INTEGRATION OR NON-INFRINGEMENT AND ALL SUCH WARRANTIES ARE HEREBY EXCLUDED BY WILEY AND ITS LICENSORS AND WAIVED BY YOU.
- WILEY shall have the right to terminate this Agreement immediately upon breach of this Agreement by you.
- You shall indemnify, defend and hold harmless WILEY, its Licensors and their respective directors, officers, agents and employees, from and against any actual or threatened claims, demands, causes of action or proceedings arising from any breach of this Agreement by you.
- IN NO EVENT SHALL WILEY OR ITS LICENSORS BE LIABLE TO YOU OR ANY OTHER PARTY OR ANY OTHER PERSON OR ENTITY FOR ANY SPECIAL, CONSEQUENTIAL, INCIDENTAL, INDIRECT, EXEMPLARY OR PUNITIVE DAMAGES, HOWEVER CAUSED, ARISING OUT OF OR IN CONNECTION WITH THE DOWNLOADING, PROVISIONING, VIEWING OR USE OF THE MATERIALS REGARDLESS OF THE FORM OF ACTION, WHETHER FOR BREACH OF CONTRACT, BREACH OF WARRANTY, TORT, NEGLIGENCE, INFRINGEMENT OR OTHERWISE (INCLUDING, WITHOUT LIMITATION, DAMAGES BASED ON LOSS OF PROFITS, DATA, FILES, USE, BUSINESS OPPORTUNITY OR CLAIMS OF THIRD PARTIES), AND WHETHER

OR NOT THE PARTY HAS BEEN ADVISED OF THE POSSIBILITY OF SUCH DAMAGES. THIS LIMITATION SHALL APPLY NOTWITHSTANDING ANY FAILURE OF ESSENTIAL PURPOSE OF ANY LIMITED REMEDY PROVIDED HEREIN.

- Should any provision of this Agreement be held by a court of competent jurisdiction to be illegal, invalid, or unenforceable, that provision shall be deemed amended to achieve as nearly as possible the same economic effect as the original provision, and the legality, validity and enforceability of the remaining provisions of this Agreement shall not be affected or impaired thereby.
- The failure of either party to enforce any term or condition of this Agreement shall not constitute a waiver of either party's right to enforce each and every term and condition of this Agreement. No breach under this agreement shall be deemed waived or excused by either party unless such waiver or consent is in writing signed by the party granting such waiver or consent. The waiver by or consent of a party to a breach of any provision of this Agreement shall not operate or be construed as a waiver of or consent to any other or subsequent breach by such other party.
- This Agreement may not be assigned (including by operation of law or otherwise) by you without WILEY's prior written consent.
- Any fee required for this permission shall be non-refundable after thirty (30) days from receipt by the CCC.
- These terms and conditions together with CCC's Billing and Payment terms and conditions (which are incorporated herein) form the entire agreement between you and WILEY concerning this licensing transaction and (in the absence of fraud) supersedes all prior agreements and representations of the parties, oral or written. This Agreement may not be amended except in writing signed by both parties. This Agreement shall be binding upon and inure to the benefit of the parties' successors, legal representatives, and authorized assigns.
- In the event of any conflict between your obligations established by these terms and conditions and those established by CCC's Billing and Payment terms and conditions, these terms and conditions shall prevail.
- WILEY expressly reserves all rights not specifically granted in the combination of (i) the license details provided by you and accepted in the course of this licensing transaction, (ii) these terms and conditions and (iii) CCC's Billing and Payment terms and conditions.
- This Agreement will be void if the Type of Use, Format, Circulation, or Requestor Type was misrepresented during the licensing process.
- This Agreement shall be governed by and construed in accordance with the laws of the State of New York, USA, without regards to such state's conflict of law rules. Any legal action, suit or proceeding arising out of or relating to these Terms and Conditions or the breach thereof shall be instituted in a court of competent jurisdiction in New York County in the State of New York in the United States of America and each party hereby consents and submits to the personal jurisdiction

of such court, waives any objection to venue in such court and consents to service of process by registered or certified mail, return receipt requested, at the last known address of such party.

WILEY OPEN ACCESS TERMS AND CONDITIONS

Wiley Publishes Open Access Articles in fully Open Access Journals and in Subscription journals offering Online Open. Although most of the fully Open Access journals publish open access articles under the terms of the Creative Commons Attribution (CC BY) License only, the subscription journals and a few of the Open Access Journals offer a choice of Creative Commons Licenses. The license type is clearly identified on the article.

The Creative Commons Attribution License

The [Creative Commons Attribution License \(CC-BY\)](#) allows users to copy, distribute and transmit an article, adapt the article and make commercial use of the article. The CC-BY license permits commercial and non-

Creative Commons Attribution Non-Commercial License

The [Creative Commons Attribution Non-Commercial \(CC-BY-NC\)License](#) permits use, distribution and reproduction in any medium, provided the original work is properly cited and is not used for commercial purposes.(see below)

Creative Commons Attribution-Non-Commercial-NoDerivs License

The [Creative Commons Attribution Non-Commercial-NoDerivs License \(CC-BY-NC-ND\)](#) permits use, distribution and reproduction in any medium, provided the original work is properly cited, is not used for commercial purposes and no modifications or adaptations are made. (see below)

Use by commercial "for-profit" organizations

Use of Wiley Open Access articles for commercial, promotional, or marketing purposes requires further explicit permission from Wiley and will be subject to a fee.

Further details can be found on Wiley Online Library

<http://olabout.wiley.com/WileyCDA/Section/id-410895.html> **Other Terms**

and Conditions:

v1.10 Last updated September 2015

Questions? customercare@copyright.com or +1-855-239-3415 (toll free in the US) or +1-978-646-2777.

4) Journal of MRI (Automated PCa Localization)

JOHN WILEY AND SONS LICENSE

TERMS AND CONDITIONS

Feb 19, 2020

This Agreement between Fatemeh Zabihollahy ("You") and John Wiley and Sons ("John Wiley and Sons") consists of your license details and the terms and conditions provided by John Wiley and Sons and Copyright Clearance Center.

License Number 4772530819150

License date Feb 19, 2020

Licensed Content Publisher John Wiley and Sons

Licensed Content Publication Journal of Magnetic Resonance Imaging

Licensed Content Title Fully automated localization of prostate peripheral zone tumors on apparent diffusion coefficient map MR images using an ensemble learning method

Licensed Fatemeh Zabihollahy, Eranga Ukwatta, Satheesh Krishna, et al Content Author

Licensed Content Date Aug 28, 2019

Licensed Content Volume 0

Licensed Content Issue 0

Licensed Content Pages 12

Type of use Dissertation/Thesis

Requestor type Author of this Wiley article

Format Print and electronic

Portion Full article

Will you be translating? No

Title of your thesis / dissertation Deep Learning Methods for Abnormality Detection and Segmentation in Computed Tomography and

Expected

Magnetic Resonance
Images

Mar 2020

completion date

Expected size

(number of pages)200

Requestor
Location
Fatemeh Zabihollahy
339 Riversedge Crescent
Ottawa, ON K1V0Y6
Canada Attn:

Publisher Tax
ID
EU826007151

Total
Terms and Conditions
0.00 USD

TERMS AND CONDITIONS

This copyrighted material is owned by or exclusively licensed to John Wiley & Sons, Inc. or one of its group companies (each a "Wiley Company") or handled on behalf of a society with which a Wiley Company has exclusive publishing rights in relation to a particular work (collectively "WILEY"). By clicking "accept" in connection with completing this licensing transaction, you agree that the following terms and conditions apply to this transaction (along with the billing and payment terms and conditions established by the Copyright Clearance Center Inc., ("CCC's Billing and Payment terms and conditions"), at the time that you opened your RightsLink account (these are available at any time at <http://myaccount.copyright.com>).

Terms and Conditions

- The materials you have requested permission to reproduce or reuse (the "Wiley Materials") are protected by copyright.
- You are hereby granted a personal, non-exclusive, non-sub licensable (on a standalone basis), non-transferable, worldwide, limited license to reproduce the Wiley Materials for the purpose specified in the licensing process. This license, **and any CONTENT (PDF or image file) purchased as part of your order**, is for a one-time use only and limited to any maximum distribution number specified in the license. The first instance of republication or reuse granted by this license must be completed within two years of the date of the grant of this license (although copies prepared before the end date may be distributed thereafter). The Wiley Materials shall not be used in any other manner or for any other purpose, beyond what is granted in the license. Permission is granted subject to an appropriate acknowledgement given to the author, title of the material/book/journal and the publisher. You shall also duplicate the copyright notice that appears in the Wiley publication in your use of the Wiley Material. Permission is also granted on the understanding that nowhere in the text is a previously published source acknowledged for all or part of this Wiley Material. Any third party content is expressly excluded from this permission.
- With respect to the Wiley Materials, all rights are reserved. Except as expressly granted by the terms of the license, no part of the Wiley Materials may be copied, modified, adapted (except for minor reformatting required by the new Publication), translated, reproduced, transferred or distributed, in any form or by any means, and no derivative works may be made based on the Wiley Materials without the prior permission of the respective copyright owner. **For STM Signatory Publishers clearing permission under the terms of the STM Permissions Guidelines only, the terms of the license are extended to include subsequent editions and for editions in other languages, provided such editions are for the work as a whole in situ and does not involve the separate exploitation of the permitted figures or extracts**, You may not alter, remove or suppress in any manner any copyright, trademark or other notices displayed by the Wiley Materials. You may not license, rent, sell, loan, lease, pledge, offer as security, transfer or assign the Wiley Materials on a stand-alone basis, or any of the rights granted to you hereunder to any other person.
- The Wiley Materials and all of the intellectual property rights therein shall at all times remain the exclusive property of John Wiley & Sons Inc, the Wiley Companies, or their respective licensors, and your interest therein is only that of having possession of and the right to reproduce the Wiley Materials pursuant to Section 2 herein during the continuance of this Agreement. You agree that you own no right, title or interest in or to the Wiley Materials or any of the intellectual property rights therein. You shall have no rights hereunder other than the license as provided for above in Section 2. No right, license or interest to any trademark, trade name, service mark or other branding ("Marks") of WILEY or its licensors is granted hereunder, and you agree that you shall not assert any such right, license or interest with respect thereto
- NEITHER WILEY NOR ITS LICENSORS MAKES ANY WARRANTY OR REPRESENTATION OF ANY KIND TO YOU OR ANY THIRD PARTY,

EXPRESS, IMPLIED OR STATUTORY, WITH RESPECT TO THE MATERIALS OR THE ACCURACY OF ANY INFORMATION CONTAINED IN THE MATERIALS, INCLUDING, WITHOUT LIMITATION, ANY IMPLIED WARRANTY OF MERCHANTABILITY, ACCURACY, SATISFACTORY QUALITY, FITNESS FOR A PARTICULAR PURPOSE, USABILITY, INTEGRATION OR NON-INFRINGEMENT AND ALL SUCH WARRANTIES ARE HEREBY EXCLUDED BY WILEY AND ITS LICENSORS AND WAIVED BY YOU.

- WILEY shall have the right to terminate this Agreement immediately upon breach of this Agreement by you.
- You shall indemnify, defend and hold harmless WILEY, its Licensors and their respective directors, officers, agents and employees, from and against any actual or threatened claims, demands, causes of action or proceedings arising from any breach of this Agreement by you.
- IN NO EVENT SHALL WILEY OR ITS LICENSORS BE LIABLE TO YOU OR ANY OTHER PARTY OR ANY OTHER PERSON OR ENTITY FOR ANY SPECIAL, CONSEQUENTIAL, INCIDENTAL, INDIRECT, EXEMPLARY OR PUNITIVE DAMAGES, HOWEVER CAUSED, ARISING OUT OF OR IN CONNECTION WITH THE DOWNLOADING, PROVISIONING, VIEWING OR USE OF THE MATERIALS REGARDLESS OF THE FORM OF ACTION, WHETHER FOR BREACH OF CONTRACT, BREACH OF WARRANTY, TORT, NEGLIGENCE, INFRINGEMENT OR OTHERWISE (INCLUDING, WITHOUT LIMITATION, DAMAGES BASED ON LOSS OF PROFITS, DATA, FILES, USE, BUSINESS OPPORTUNITY OR CLAIMS OF THIRD PARTIES), AND WHETHER OR NOT THE PARTY HAS BEEN ADVISED OF THE POSSIBILITY OF SUCH DAMAGES. THIS LIMITATION SHALL APPLY NOTWITHSTANDING ANY FAILURE OF ESSENTIAL PURPOSE OF ANY LIMITED REMEDY PROVIDED HEREIN.
- Should any provision of this Agreement be held by a court of competent jurisdiction to be illegal, invalid, or unenforceable, that provision shall be deemed amended to achieve as nearly as possible the same economic effect as the original provision, and the legality, validity and enforceability of the remaining provisions of this Agreement shall not be affected or impaired thereby.
- The failure of either party to enforce any term or condition of this Agreement shall not constitute a waiver of either party's right to enforce each and every term and condition of this Agreement. No breach under this agreement shall be deemed waived or excused by either party unless such waiver

or consent is in writing signed by the party granting such waiver or consent. The waiver by or consent of a party to a breach of any provision of this Agreement shall not operate or be construed as a waiver of or consent to any other or subsequent breach by such other party.

- This Agreement may not be assigned (including by operation of law or otherwise) by you without WILEY's prior written consent.
- Any fee required for this permission shall be non-refundable after thirty (30) days from receipt by the CCC.
- These terms and conditions together with CCC's Billing and Payment terms and conditions (which are incorporated herein) form the entire agreement between you and WILEY concerning this licensing transaction and (in the absence of fraud) supersedes all prior agreements and representations of the parties, oral or written. This Agreement may not be amended except in writing signed by both parties. This Agreement shall be binding upon and inure to the benefit of the parties' successors, legal representatives, and authorized assigns.
- In the event of any conflict between your obligations established by these terms and conditions and those established by CCC's Billing and Payment terms and conditions, these terms and conditions shall prevail.
- WILEY expressly reserves all rights not specifically granted in the combination of (i) the license details provided by you and accepted in the course of this licensing transaction, (ii) these terms and conditions and (iii) CCC's Billing and Payment terms and conditions.
- This Agreement will be void if the Type of Use, Format, Circulation, or Requestor Type was misrepresented during the licensing process.
- This Agreement shall be governed by and construed in accordance with the laws of the State of New York, USA, without regards to such state's conflict of law rules. Any legal action, suit or proceeding arising out of or relating to these Terms and Conditions or the breach thereof shall be instituted in a court of competent jurisdiction in New York County in the State of New York in the United States of America and each party hereby consents and submits to the personal jurisdiction of such court, waives any objection to venue in such court and consents to service of process by registered or certified mail, return receipt requested, at the last known address of such party.

WILEY OPEN ACCESS TERMS AND CONDITIONS

Wiley Publishes Open Access Articles in fully Open Access Journals and in Subscription journals offering Online Open. Although most of the fully Open Access journals publish open access articles under the terms of the Creative Commons Attribution (CC BY) License only, the subscription journals and a few of the Open Access Journals offer a choice of Creative Commons Licenses. The license type is clearly identified on the article.

The Creative Commons Attribution License

The [Creative Commons Attribution License \(CC-BY\)](#) allows users to copy, distribute and transmit an article, adapt the article and make commercial use of the article. The CC-BY license permits commercial and non-

Creative Commons Attribution Non-Commercial License

The [Creative Commons Attribution Non-Commercial \(CC-BY-NC\) License](#) permits use, distribution and reproduction in any medium, provided the original work is properly cited and is not used for commercial purposes.(see below)

Creative Commons Attribution-Non-Commercial-NoDerivs License

The [Creative Commons Attribution Non-Commercial-NoDerivs License \(CC-BY-NC-ND\)](#) permits use, distribution and reproduction in any medium, provided the original work is properly cited, is not used for commercial purposes and no modifications or adaptations are made. (see below)

Use by commercial "for-profit" organizations

Use of Wiley Open Access articles for commercial, promotional, or marketing purposes requires further explicit permission from Wiley and will be subject to a fee.

Further details can be found on Wiley Online Library

<http://olabout.wiley.com/WileyCDA/Section/id-410895.html> **Other Terms**

and Conditions:

v1.10 Last updated September 2015

Questions? customercare@copyright.com or +1-855-239-3415 (toll free in the US) or +1-978-646-2777.

5) Journal of Medical Physics (Semi-automated LV scar segmentation)

JOHN WILEY AND SONS LICENSE

TERMS AND CONDITIONS

Feb 19, 2020

This Agreement between Fatemeh Zabihollahy ("You") and John Wiley and Sons ("John Wiley and Sons") consists of your license details and the terms and conditions provided by John Wiley and Sons and Copyright Clearance Center.

License Number 4772530533343

License date Feb 19, 2020

Licensed Content Publisher John Wiley and Sons

Licensed Content Publication Medical Physics

Licensed Content Title Convolutional neural network-based approach for segmentation of left ventricle myocardial scar from 3D late gadolinium enhancement MR images

Licensed Content Author Fatemeh Zabihollahy, James A. White, Eranga Ukwatta

Licensed Content Date Feb 28, 2019

Licensed

46

Content Volume

Licensed

4

Content Issue

Licensed 12

Content Pages

Type of use Dissertation/Thesis

Requestor type Author of this Wiley article

Format Print and electronic

Portion Full article

Will you be translating? No

Title of your thesis / dissertation
Deep Learning Methods for Abnormality Detection and Segmentation in Computed Tomography and Magnetic Resonance Images

Expected completion date
Mar 2020

completion date

Expected size
(number of pages) 200

Requestor
Location

Fatemeh Zabihollahy
339 Riversedge Crescent

Ottawa, ON K1V0Y6
Canada Attn:

Publisher Tax
ID

EU826007151

Total
Terms and Conditions

0.00 USD

TERMS AND CONDITIONS

This copyrighted material is owned by or exclusively licensed to John Wiley & Sons, Inc. or one of its group companies (each a "Wiley Company") or handled on behalf of a society with which a Wiley Company has exclusive publishing rights in relation to a particular work (collectively "WILEY"). By clicking "accept" in connection with completing this licensing transaction, you agree that the following terms and conditions apply to this transaction (along with the billing and payment terms and conditions established by the Copyright Clearance Center Inc., ("CCC's Billing and Payment terms and conditions"), at the time that you opened your RightsLink account (these are available at any time at <http://myaccount.copyright.com>).

Terms and Conditions

- The materials you have requested permission to reproduce or reuse (the "Wiley Materials") are protected by copyright.
- You are hereby granted a personal, non-exclusive, non-sub licensable (on a standalone basis), non-transferable, worldwide, limited license to reproduce the Wiley Materials for the purpose specified in the licensing process. This license, **and any CONTENT (PDF or image file) purchased as part of your order**, is for a one-time use only and limited to any maximum distribution number specified in the license. The first instance of republication or reuse granted by this license must be completed within two years of the date of the grant of this license (although copies prepared before the end date may be distributed thereafter). The Wiley Materials shall not be used in any other manner or for any other purpose, beyond what is granted in the license. Permission is granted subject to an appropriate acknowledgement given to the author, title of the material/book/journal and the publisher. You shall also duplicate the copyright notice that appears in the Wiley publication in your use of the Wiley Material. Permission is also granted on the understanding that

nowhere in the text is a previously published source acknowledged for all or part of this Wiley Material. Any third party content is expressly excluded from this permission.

- With respect to the Wiley Materials, all rights are reserved. Except as expressly granted by the terms of the license, no part of the Wiley Materials may be copied, modified, adapted (except for minor reformatting required by the new Publication), translated, reproduced, transferred or distributed, in any form or by any means, and no derivative works may be made based on the Wiley Materials without the prior permission of the respective copyright owner. **For STM Signatory Publishers clearing permission under the terms of the STM Permissions Guidelines only, the terms of the license are extended to include subsequent editions and for editions in other languages, provided such editions are for the work as a whole in situ and does not involve the separate exploitation of the permitted figures or extracts,** You may not alter, remove or suppress in any manner any copyright, trademark or other notices displayed by the Wiley Materials. You may not license, rent, sell, loan, lease, pledge, offer as security, transfer or assign the Wiley Materials on a stand-alone basis, or any of the rights granted to you hereunder to any other person.
- The Wiley Materials and all of the intellectual property rights therein shall at all times remain the exclusive property of John Wiley & Sons Inc, the Wiley Companies, or their respective licensors, and your interest therein is only that of having possession of and the right to reproduce the Wiley Materials pursuant to Section 2 herein during the continuance of this Agreement. You agree that you own no right, title or interest in or to the Wiley Materials or any of the intellectual property rights therein. You shall have no rights hereunder other than the license as provided for above in Section 2. No right, license or interest to any trademark, trade name, service mark or other branding ("Marks") of WILEY or its licensors is granted hereunder, and you agree that you shall not assert any such right, license or interest with respect thereto
- NEITHER WILEY NOR ITS LICENSORS MAKES ANY WARRANTY OR REPRESENTATION OF ANY KIND TO YOU OR ANY THIRD PARTY, EXPRESS, IMPLIED OR STATUTORY, WITH RESPECT TO THE MATERIALS OR THE ACCURACY OF ANY INFORMATION CONTAINED IN THE MATERIALS, INCLUDING, WITHOUT LIMITATION, ANY IMPLIED WARRANTY OF MERCHANTABILITY, ACCURACY, SATISFACTORY QUALITY, FITNESS FOR A PARTICULAR PURPOSE, USABILITY, INTEGRATION OR NON-INFRINGEMENT AND ALL SUCH WARRANTIES ARE HEREBY EXCLUDED BY WILEY AND ITS LICENSORS AND WAIVED BY YOU.
- WILEY shall have the right to terminate this Agreement immediately upon breach of this Agreement by you.

- You shall indemnify, defend and hold harmless WILEY, its Licensors and their respective directors, officers, agents and employees, from and against any actual or threatened claims, demands, causes of action or proceedings arising from any breach of this Agreement by you.
- IN NO EVENT SHALL WILEY OR ITS LICENSORS BE LIABLE TO YOU OR ANY OTHER PARTY OR ANY OTHER PERSON OR ENTITY FOR ANY SPECIAL, CONSEQUENTIAL, INCIDENTAL, INDIRECT, EXEMPLARY OR PUNITIVE DAMAGES, HOWEVER CAUSED, ARISING OUT OF OR IN CONNECTION WITH THE DOWNLOADING, PROVISIONING, VIEWING OR USE OF THE MATERIALS REGARDLESS OF THE FORM OF ACTION, WHETHER FOR BREACH OF CONTRACT, BREACH OF WARRANTY, TORT, NEGLIGENCE, INFRINGEMENT OR OTHERWISE (INCLUDING, WITHOUT LIMITATION, DAMAGES BASED ON LOSS OF PROFITS, DATA, FILES, USE, BUSINESS OPPORTUNITY OR CLAIMS OF THIRD PARTIES), AND WHETHER OR NOT THE PARTY HAS BEEN ADVISED OF THE POSSIBILITY OF SUCH DAMAGES. THIS LIMITATION SHALL APPLY NOTWITHSTANDING ANY FAILURE OF ESSENTIAL PURPOSE OF ANY LIMITED REMEDY PROVIDED HEREIN.
- Should any provision of this Agreement be held by a court of competent jurisdiction to be illegal, invalid, or unenforceable, that provision shall be deemed amended to achieve as nearly as possible the same economic effect as the original provision, and the legality, validity and enforceability of the remaining provisions of this Agreement shall not be affected or impaired thereby.
- The failure of either party to enforce any term or condition of this Agreement shall not constitute a waiver of either party's right to enforce each and every term and condition of this Agreement. No breach under this agreement shall be deemed waived or excused by either party unless such waiver or consent is in writing signed by the party granting such waiver or consent. The waiver by or consent of a party to a breach of any provision of this Agreement shall not operate or be construed as a waiver of or consent to any other or subsequent breach by such other party.
- This Agreement may not be assigned (including by operation of law or otherwise) by you without WILEY's prior written consent.
- Any fee required for this permission shall be non-refundable after thirty (30) days from receipt by the CCC.
- These terms and conditions together with CCC's Billing and Payment terms and conditions (which are incorporated herein) form the entire agreement between you and WILEY concerning this licensing transaction and (in the absence of fraud) supersedes all prior agreements and representations of the parties, oral or written. This Agreement may not be amended except in

writing signed by both parties. This Agreement shall be binding upon and inure to the benefit of the parties' successors, legal representatives, and authorized assigns.

- In the event of any conflict between your obligations established by these terms and conditions and those established by CCC's Billing and Payment terms and conditions, these terms and conditions shall prevail.
- WILEY expressly reserves all rights not specifically granted in the combination of (i) the license details provided by you and accepted in the course of this licensing transaction, (ii) these terms and conditions and (iii) CCC's Billing and Payment terms and conditions.
- This Agreement will be void if the Type of Use, Format, Circulation, or Requestor Type was misrepresented during the licensing process.
- This Agreement shall be governed by and construed in accordance with the laws of the State of New York, USA, without regards to such state's conflict of law rules. Any legal action, suit or proceeding arising out of or relating to these Terms and Conditions or the breach thereof shall be instituted in a court of competent jurisdiction in New York County in the State of New York in the United States of America and each party hereby consents and submits to the personal jurisdiction of such court, waives any objection to venue in such court and consents to service of process by registered or certified mail, return receipt requested, at the last known address of such party.

WILEY OPEN ACCESS TERMS AND CONDITIONS

Wiley Publishes Open Access Articles in fully Open Access Journals and in Subscription journals offering Online Open. Although most of the fully Open Access journals publish open access articles under the terms of the Creative Commons Attribution (CC BY) License only, the subscription journals and a few of the Open Access Journals offer a choice of Creative Commons Licenses. The license type is clearly identified on the article.

The Creative Commons Attribution License

The [Creative Commons Attribution License \(CC-BY\)](#) allows users to copy, distribute and transmit an article, adapt the article and make commercial use of the article. The CC-BY license permits commercial and non-

Creative Commons Attribution Non-Commercial License

The [Creative Commons Attribution Non-Commercial \(CC-BY-NC\) License](#) permits use, distribution and reproduction in any medium, provided the original work is properly cited and is not used for commercial purposes.(see below)

Creative Commons Attribution-Non-Commercial-NoDerivs License

The [Creative Commons Attribution Non-Commercial-NoDerivs License](#) (CC-BY-NC-ND) permits use, distribution and reproduction in any medium, provided the original work is properly cited, is not used for commercial purposes and no modifications or adaptations are made. (see below)

Use by commercial "for-profit" organizations

Use of Wiley Open Access articles for commercial, promotional, or marketing purposes requires further explicit permission from Wiley and will be subject to a fee.

Further details can be found on Wiley Online Library

<http://olabout.wiley.com/WileyCDA/Section/id-410895.html> **Other Terms**

and Conditions:

v1.10 Last updated September 2015

Questions? customercare@copyright.com or +1-855-239-3415 (toll free in the US) or +1-978-646-2777.

6) Journal of Medical Physics (Fully automated LV scar segmentation)

JOHN WILEY AND SONS LICENSE

TERMS AND CONDITIONS

Feb 19, 2020

This Agreement between Fatemeh Zabihollahy ("You") and John Wiley and Sons ("John Wiley and Sons") consists of your license details and the terms and conditions provided by John Wiley and Sons and Copyright Clearance Center.

License Number 4772530684502

License date Feb 19, 2020

Licensed Content Publisher John Wiley and Sons

Licensed Content Publication Medical Physics

Licensed Content Title Fully automated segmentation of left ventricular scar from 3D late gadolinium enhancement magnetic resonance imaging using a cascaded multi-planar U-Net (CMPU-Net)

Licensed Content Author Fatemeh Zabihollahy, Martin Rajchl, James A. White, et al

Licensed Content Date Feb 10, 2020

Licensed Content Volume 0

Licensed Content Issue 0

Licensed Content Page 11

Content Pages

Type of use Dissertation/Thesis

Requestor type Author of this Wiley article

Format Print and electronic

Portion Full article

Will you be
translating? No

Title of your

Deep Learning Methods for Abnormality Detection and Segmentation in

thesis /

Computed Tomography and Magnetic Resonance Images

dissertation

Expected completion Mar
2020

date

Expected size

(number of 200

pages)

Fatemeh Zabihollahy

339 Riversedge Crescent

Requestor

Location

Ottawa, ON K1V0Y6

Canada Attn:

Publisher Tax

EU826007151

ID

Total 0.00 USD

Terms and Conditions

TERMS AND CONDITIONS

This copyrighted material is owned by or exclusively licensed to John Wiley & Sons, Inc. or one of its group companies (each a "Wiley Company") or handled on behalf of a society with which a Wiley Company has exclusive publishing rights in relation to a particular work (collectively "WILEY"). By clicking "accept" in connection with completing this licensing transaction, you agree that the following terms and conditions apply to this transaction (along with the billing and payment terms and conditions established by the Copyright Clearance Center Inc., ("CCC's Billing and Payment terms and conditions"), at the time that you opened your RightsLink account (these are available at any time at <http://myaccount.copyright.com>).

Terms and Conditions

- The materials you have requested permission to reproduce or reuse (the "Wiley Materials") are protected by copyright.
- You are hereby granted a personal, non-exclusive, non-sub licensable (on a standalone basis), non-transferable, worldwide, limited license to reproduce the Wiley Materials for the purpose specified in the licensing process. This license, **and any CONTENT (PDF or image file) purchased as part of your order**, is for a one-time use only and limited to any maximum distribution number specified in the license. The first instance of republication or reuse granted by this license must be completed within two years of the date of the grant of this license (although copies prepared before the end date may be distributed thereafter). The Wiley Materials shall not be used in any other manner or for any other purpose, beyond what is granted in the license. Permission is granted subject to an appropriate acknowledgement given to the author, title of the material/book/journal and the publisher. You shall also duplicate the copyright notice that appears in the Wiley publication in your use of the Wiley Material. Permission is also granted on the understanding that nowhere in the text is a previously published source acknowledged for all or part of this Wiley Material. Any third party content is expressly excluded from this permission.

- With respect to the Wiley Materials, all rights are reserved. Except as expressly granted by the terms of the license, no part of the Wiley Materials may be copied, modified, adapted (except for minor reformatting required by the new Publication), translated, reproduced, transferred or distributed, in any form or by any means, and no derivative works may be made based on the Wiley Materials without the prior permission of the respective copyright owner. **For STM Signatory Publishers clearing permission under the terms of the STM Permissions Guidelines only, the terms of the license are extended to include subsequent editions and for editions in other languages, provided such editions are for the work as a whole in situ and does not involve the separate exploitation of the permitted figures or extracts,** You may not alter, remove or suppress in any manner any copyright, trademark or other notices displayed by the Wiley Materials. You may not license, rent, sell, loan, lease, pledge, offer as security, transfer or assign the Wiley Materials on a stand-alone basis, or any of the rights granted to you hereunder to any other person.
- The Wiley Materials and all of the intellectual property rights therein shall at all times remain the exclusive property of John Wiley & Sons Inc, the Wiley Companies, or their respective licensors, and your interest therein is only that of having possession of and the right to reproduce the Wiley Materials pursuant to Section 2 herein during the continuance of this Agreement. You agree that you own no right, title or interest in or to the Wiley Materials or any of the intellectual property rights therein. You shall have no rights hereunder other than the license as provided for above in Section 2. No right, license or interest to any trademark, trade name, service mark or other branding ("Marks") of WILEY or its licensors is granted hereunder, and you agree that you shall not assert any such right, license or interest with respect thereto
- NEITHER WILEY NOR ITS LICENSORS MAKES ANY WARRANTY OR REPRESENTATION OF ANY KIND TO YOU OR ANY THIRD PARTY, EXPRESS, IMPLIED OR STATUTORY, WITH RESPECT TO THE MATERIALS OR THE ACCURACY OF ANY INFORMATION CONTAINED IN THE MATERIALS, INCLUDING, WITHOUT LIMITATION, ANY IMPLIED WARRANTY OF MERCHANTABILITY, ACCURACY, SATISFACTORY QUALITY, FITNESS FOR A PARTICULAR PURPOSE, USABILITY, INTEGRATION OR NON-INFRINGEMENT AND ALL SUCH WARRANTIES ARE HEREBY EXCLUDED BY WILEY AND ITS LICENSORS AND WAIVED BY YOU.
- WILEY shall have the right to terminate this Agreement immediately upon breach of this Agreement by you.
- You shall indemnify, defend and hold harmless WILEY, its Licensors and their respective directors, officers, agents and employees, from and against any actual or threatened claims, demands, causes of action or proceedings arising from any breach of this Agreement by you.
- IN NO EVENT SHALL WILEY OR ITS LICENSORS BE LIABLE TO YOU OR

ANY OTHER PARTY OR ANY OTHER PERSON OR ENTITY FOR ANY SPECIAL, CONSEQUENTIAL, INCIDENTAL, INDIRECT, EXEMPLARY OR PUNITIVE DAMAGES, HOWEVER CAUSED, ARISING OUT OF OR IN CONNECTION WITH THE DOWNLOADING, PROVISIONING, VIEWING OR USE OF THE MATERIALS REGARDLESS OF THE FORM OF ACTION, WHETHER FOR BREACH OF CONTRACT, BREACH OF WARRANTY, TORT, NEGLIGENCE, INFRINGEMENT OR OTHERWISE (INCLUDING, WITHOUT LIMITATION, DAMAGES BASED ON LOSS OF PROFITS, DATA, FILES, USE, BUSINESS OPPORTUNITY OR CLAIMS OF THIRD PARTIES), AND WHETHER OR NOT THE PARTY HAS BEEN ADVISED OF THE POSSIBILITY OF SUCH DAMAGES. THIS LIMITATION SHALL APPLY NOTWITHSTANDING ANY FAILURE OF ESSENTIAL PURPOSE OF ANY LIMITED REMEDY PROVIDED HEREIN.

- Should any provision of this Agreement be held by a court of competent jurisdiction to be illegal, invalid, or unenforceable, that provision shall be deemed amended to achieve as nearly as possible the same economic effect as the original provision, and the legality, validity and enforceability of the remaining provisions of this Agreement shall not be affected or impaired thereby.
- The failure of either party to enforce any term or condition of this Agreement shall not constitute a waiver of either party's right to enforce each and every term and condition of this Agreement. No breach under this agreement shall be deemed waived or excused by either party unless such waiver or consent is in writing signed by the party granting such waiver or consent. The waiver by or consent of a party to a breach of any provision of this Agreement shall not operate or be construed as a waiver of or consent to any other or subsequent breach by such other party.
- This Agreement may not be assigned (including by operation of law or otherwise) by you without WILEY's prior written consent.
- Any fee required for this permission shall be non-refundable after thirty (30) days from receipt by the CCC.
- These terms and conditions together with CCC's Billing and Payment terms and conditions (which are incorporated herein) form the entire agreement between you and WILEY concerning this licensing transaction and (in the absence of fraud) supersedes all prior agreements and representations of the parties, oral or written. This Agreement may not be amended except in writing signed by both parties. This Agreement shall be binding upon and inure to the benefit of the parties' successors, legal representatives, and authorized assigns.

- In the event of any conflict between your obligations established by these terms and conditions and those established by CCC's Billing and Payment terms and conditions, these terms and conditions shall prevail.
- WILEY expressly reserves all rights not specifically granted in the combination of (i) the license details provided by you and accepted in the course of this licensing transaction, (ii) these terms and conditions and (iii) CCC's Billing and Payment terms and conditions.
- This Agreement will be void if the Type of Use, Format, Circulation, or Requestor Type was misrepresented during the licensing process.
- This Agreement shall be governed by and construed in accordance with the laws of the State of New York, USA, without regards to such state's conflict of law rules. Any legal action, suit or proceeding arising out of or relating to these Terms and Conditions or the breach thereof shall be instituted in a court of competent jurisdiction in New York County in the State of New York in the United States of America and each party hereby consents and submits to the personal jurisdiction of such court, waives any objection to venue in such court and consents to service of process by registered or certified mail, return receipt requested, at the last known address of such party.

WILEY OPEN ACCESS TERMS AND CONDITIONS

Wiley Publishes Open Access Articles in fully Open Access Journals and in Subscription journals offering Online Open. Although most of the fully Open Access journals publish open access articles under the terms of the Creative Commons Attribution (CC BY) License only, the subscription journals and a few of the Open Access Journals offer a choice of Creative Commons Licenses. The license type is clearly identified on the article.

The Creative Commons Attribution License

The [Creative Commons Attribution License \(CC-BY\)](#) allows users to copy, distribute and transmit an article, adapt the article and make commercial use of the article. The CC-BY license permits commercial and non-

Creative Commons Attribution Non-Commercial License

The [Creative Commons Attribution Non-Commercial \(CC-BY-NC\) License](#) permits use, distribution and reproduction in any medium, provided the original work is properly cited and is not used for commercial purposes.(see below)

Creative Commons Attribution-Non-Commercial-No Derivs License

The [Creative Commons Attribution Non-Commercial-NoDerivs License](#) (CC-BY-NC-ND) permits use, distribution and reproduction in any medium, provided the original work is properly cited, is not used for commercial purposes and no modifications or adaptations are made. (see below)

Use by commercial "for-profit" organizations

Use of Wiley Open Access articles for commercial, promotional, or marketing purposes requires further explicit permission from Wiley and will be subject to a fee.

Further details can be found on Wiley Online Library

<http://olabout.wiley.com/WileyCDA/Section/id-410895.html> **Other Terms**

and Conditions:v1.10 Last updated September 2015

Questions? customercare@copyright.com or +1-855-239-3415 (toll free in the US) or +1-978-646-2777.
