Volumetric Image-based Supervised Learning Approaches for Kidney Cancer Detection and Analysis

by

Mohammad Arafat Hussain

M.A.Sc. in Engineering, University of British Columbia, Vancouver, 2015

M.Sc. in Engineering, Bangladesh University of Engineering & Technology, 2013

B.Sc. in Engineering, Bangladesh University of Engineering & Technology, 2011

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

Doctor of Philosophy

in

THE FACULTY OF GRADUATE AND POSTDOCTORAL STUDIES

(Biomedical Engineering)

The University of British Columbia

(Vancouver)

April 2020

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

Volumetric Image-based Supervised Learning Approaches for Kidney Cancer Detection and Analysis

submitted by **Mohammad Arafat Hussain** in partial fulfillment of the requirements for the degree of **Doctor of Philosophy** in **Biomedical Engineering**.

Examining Committee:

Prof. Rafeef Garbi, Electrical and Computer Engineering
Supervisor

Prof. Antony J. Hodgson, Mechanical Engineering
Supervisory Committee Member

Prof. Piotr Kozlowski, Radiology and Urological Sciences
University Examiner

Prof. Shahriar Mirabbasi, Electrical and Computer Engineering
University Examiner

Prof. Marleen de Bruijne, Erasmus MC Rotterdam, The Netherlands
External Examiner

Abstract

Kidney cancers account for an estimated 140,000 global deaths annually. According to the Canadian Cancer Society, an estimated 6,600 Canadians were diagnosed with kidney cancer, and 1,900 Canadians died from it in 2017. Computed tomography (CT) imaging plays a vital role in kidney cancer detection, prognosis, and treatment response assessment. Automated CT-based cancer analysis is benefiting from unprecedented advancements in machine learning techniques and wide availability of high-performance computers.

Typically, kidney cancer analysis requires a challenging pipeline of (a) kidney localization in the CT scan and general assessment of kidney functionality, (b) tumor detection within the kidney, and (c) cancer analysis.

In this thesis, we developed deep learning techniques for automatic kidney localization, segmentation-free volume estimation, cancer detection, as well as CT features-based gene mutation detection, renal cell carcinoma (RCC) grading, and staging. Our convolutional neural network (CNN)-based kidney localization approach produces a kidney bounding box in CT, while our CNN-based direct kidney volume estimation approach skips the intermediate segmentation step that is often used for volume estimation at the cost of additional computational overhead. We also proposed a novel collage CNN technique to detect pathological kidneys, where we introduced a unique image augmentation procedure within a multiple instance learning framework. We further proposed a multiple instance decision aggregated CNN approach for automatic detection of gene mutations and a learnable image histogram-based deep neural network (ImHistNet) approach for RCC grading and staging. These approaches could be alternatives to renal biopsy-based whole-genome sequencing, RCC grading, and staging, respectively.

Our automatic kidney localization approach reduced the mean kidney boundary localization error to 2.19 mm, which is 23% better than that of recent literature. We also achieved a mean total kidney volume estimation accuracy of 95.2%. Further, we showed a pathological *vs.* healthy kidney classification accuracy of 98% using our novel collage CNN approach. In our kidney cancer analysis works, our multiple-instance CNN demonstrated an approximately 94% accuracy in kidney-wise mutation detection. Also, our novel ImHistNet demonstrated 80% and 83% accuracies in RCC grading and staging, respectively.

Lay Summary

In this thesis, we developed several supervised learning-based techniques for kidney cancer analysis from the computed tomography (CT) images. Our methods are mostly convolutional neural network-based. These methods are capable of

1. localizing kidneys in the 3D CT volume,
2. assessing the kidney functionality via estimating the total kidney volume in a segmentation-free fashion,
3. detecting the presence of tumor in kidneys using novel collage image representation that allows using sparsely annotated volume data in the multiple instance learning framework,
4. detecting mutated genes by learning the CT-image features, and
5. determining renal cell carcinoma grades and stages by learning the CT textural features.

Preface

A part of the research presented herein involves using human data, accessed from the Vancouver General Hospital, which was approved by the UBC Clinical Research Ethics Board (CREB), certificate numbers: H15-00237. This thesis is primarily based on the following articles, resulting from the collaboration of multiple researchers.

We published a part of the studies described in chapter 3 in:
[P1] **Hussain M.A.**, Amir-Khalili A., Hamarneh G., and Abugharbieh R., Segmentation-free Kidney Localization and Volume Estimation Using Aggregated Orthogonal Decision CNN, In: Medical Image Computing and Computer-Assisted Intervention (MICCAI), pp. 612–620, Quebec-city, Canada, 2017. [1].

We are preparing a part of the studies described in chapter 3 as a journal:
[UP1] **Hussain M.A.**, Hamarneh G., and Garbi R., Segmentation-free Kidney Localization and Volume Estimation Using Mask-RCNN and FCN, IEEE Transaction on Medical Imaging. [Under Preparation]

We published some part of the studies described in chapter 4 in:
[P2] **Hussain M.A.**, Hamarneh G., O'Connell T.W., Mohammed M.F., and Abugharbieh R., Segmentation-Free Estimation of Kidney Volumes in CT with Dual Regression Forests, In: 7th International Workshop on Machine Learning in Medical Imaging (MLMI), pp. 156–163, 2016. [2], and

[P1] **Hussain M.A.**, Amir-Khalili A., Hamarneh G., and Abugharbieh R., Segmentation-

free Kidney Localization and Volume Estimation Using Aggregated Orthogonal Decision CNN, In: Medical Image Computing and Computer-Assisted Intervention (MICCAI), pp. 612–620, Quebec-city, Canada, 2017. [1]

A of the part of the studies described in chapter 4 is under preparation to submit in UP1.

We published the studies described in chapter 5 in:
[P3] **Hussain M.A.**, Amir-Khalili A., Hamarneh G., and Abugharbieh R., Collage CNN for Renal Cell Carcinoma Detection from CT, In: 8th International Workshop on Machine Learning in Medical Imaging (MLMI), pp. 229–237, Quebec-city, Canada, 2017. [3]

We published the studies described in chapter 6 in:
[P4] **Hussain M.A.**, Hamarneh G., and Garbi R., Noninvasive Determination of Gene Mutations in Clear Cell Renal Cell Carcinoma using Multiple Instance Decisions Aggregated CNN, In: Medical Image Computing and Computer Assisted Intervention (MICCAI), pp. 657–665, 2018. [4]

We published the studies described in chapter 7 in:
[P5] **Hussain M.A.**, Hamarneh G., and Garbi R., ImHistNet: Learnable Image Histogram Based DNN with Application to Noninvasive Determination of Carcinoma Grades in CT Scans, In: Medical Image Computing and Computer Assisted Intervention (MICCAI), pp. 130–138, Shenzhen-China, 2019. [5] **(Received the MICCAI Graduate Student Travel Award)**

[P6] **Hussain M.A.**, Hamarneh G., and Garbi R., Renal Cell Carcinoma Staging with Learnable Image Histogram-based Deep Neural Network, In: 10th International Workshop on Machine Learning in Medical Imaging (MLMI), pp. 533–540, Shenzhen-China, 2019. [6]

An extended version of P5 and P6 (chapter 7) is under preparation as a journal draft:

[UP2] **Hussain M.A.**, Hamarneh G., and Garbi R., ImHistNet: Deep Radiomics with Learnable Image Histograms for Renal Carcinoma Staging and Grading, Medical Image Analysis. [Under Preparation]

All published articles were revised and edited by all co-authors. In P1-P6, as the primary author, I was the main contributor to the majority of writing effort, ideation, design, implementation, and testing of the proposed methodology under the supervision of Prof. Rafeef Garbi. I also presented the oral presentation for P3 and the poster presentations P1, P2, P4, P5, P6 conference papers. In all papers, Prof. Ghassan Hamarneh helped immensely with his valuable input on developing the original idea, improving the methodology, experimental design, and writing the paper. In P1 and P2, Dr. Timothy W. O'Connell and Dr. Mohammed F. Mohammed provided the ground truth delineation of the kidney.

Table of Contents

List of Tables

List of Figures

List of Acronyms

BAP1 BRCA1-associated protein 1

CKD Chronic kidney disease

CNN Convolutional neural network

CT Computed tomography

DNN Deep neural network

ESRD End stage renal disease

FCN Fully convolutional networks

FGS Fuhrman grading system

HU Hounsfield Unit

KDM5C Lysine (K)-specific demethylase 5C

MIL Multiple-instance learning

ML Machine learning

MRI Magnetic resonance imaging

PACS Picture archiving and communication system

PBRM1 Polybromo 1

PET Positron emission tomography

RCC Renal cell carcinomas

RCNN Region convolutional neural network

ROI Region-of-interest

SETD2 SET domain containing 2

VHL von Hippel-Lindau

Acknowledgments

After the Almighty, it is my great pleasure to express gratitude to the people who made this thesis possible. Foremost, I would like to express my deepest gratitude to my supervisor, Prof. Rafeef Garbi, Ph.D., and my mentor, Prof. Ghassan Hamarneh, Ph.D., whose expertise, understanding and patience, added significantly to my graduate experience. This study could have never been done without their motivation, guidance, and inspiration. I would also like to thank Dr. Timothy W. O'Connell and Dr. Mohammed F. Mohammed at the Vancouver General Hospital, who delineated the kidney in the 3D CT images that ultimately are considered as the ground truth in some of our clinical data-based validation.

I am greatly indebted to Dr. Alborz Amir-Khalili, who was a senior graduate student in my lab when I joined. He is a mathematics and engineering genius from whom I learned so many things.

I also like to remember all my teachers from the elementary school to here at UBC, who put knowledge and value in me that helped me reach this level of my life.

I want to thank the Institute for Computing, Information, and Cognitive Systems for program support.

I would also like to thank my parents, my wife, and my two daughters, for their support. Without them, I would not be the person I am today. Especially, I feel great in debt to my wife.

Last but not least, I would like to thank NVIDIA Corporation for supporting our research through their GPU Grant Program by donating the GeForce Titan Xp.

Finally, I must say that working with brilliant researchers in the Biomedical Signal and Image Computing Laboratory (BiSICL) at the University of British

Columbia was my great honor. They provided a friendly environment for me to learn and grow.

Dedication

To my *parents*,
my supervisor and mentor *Prof. Rafeef Garbi*,
my wife *Easha*,
and my daughters *Ayesha* and *Arwaa*.

Chapter 1

Introduction

1.1 Background

Worldwide Cancer Statistics: Cancer is the second leading cause of death worldwide behind cardiovascular diseases (Fig. 1.1) [8]. In 2018, there were 18.1 million new cases of cancer worldwide that resulted in 9.6 million deaths [9]. About 40% of cancer cases occur in the abdominal organs, e.g., kidneys, liver, prostate, stomach, etc [9]. The Canadian Cancer Society projected that 220,400 Canadians would develop cancer, and 82,100 would die of it in 2019 [10]. This society also projected that about 1 in 2 Canadians would develop cancer in their lifetime, and about 1 in 4 Canadians will die of it [10].

Kidney Cancer Statistics: Kidney cancer is the 16th most common cancer worldwide. About 403,262 new kidney cancer cases were recorded and about 175,098 patients died of it globally in 2018 [9]. The Canadian Cancer Society [10] has projected about 1,900 deaths among Canadians due to kidney cancer in 2019.

1.1.1 Thesis Motivation

Socioeconomic Burden of Kidney Cancer: Kidney cancer often turns into the End stage renal disease (ESRD) [11] that typically results in kidney failure. Even though kidney cancer is the 16th most common, its economic burden goes well beyond the incident rate [12]. Kidney failure is a significant concern to patients, their

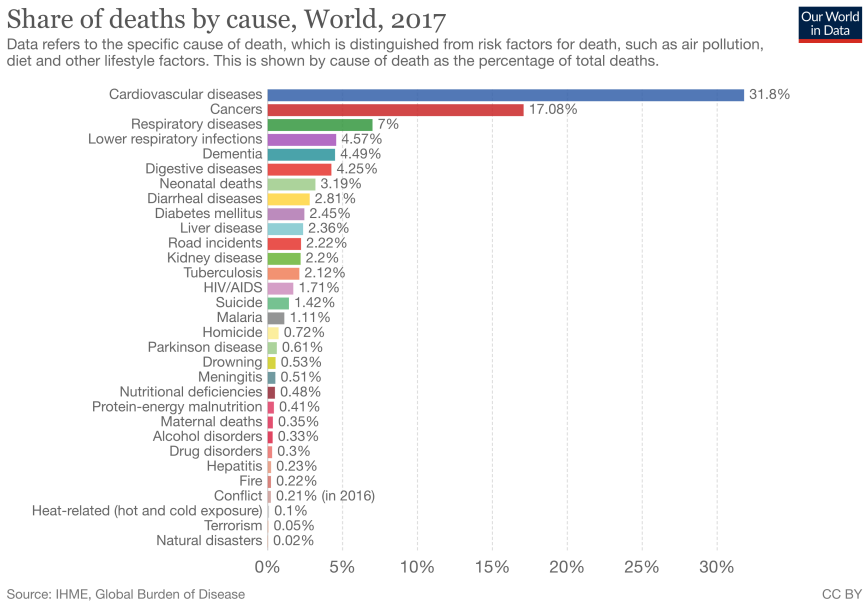

Figure 1.1: A chart showing share of deaths by cause.[1]

caregivers, and payers. It incurs high health care costs annually to manage the clinical complexities of patients with kidney diseases, including costs associated with the detection and management of kidney cancer treatment, and simultaneous management of comorbid conditions (e.g., diabetes, congestive heart failure, and hypertension). In the United States, annual medical costs per patient with ESRD often varies up to $180,000 [12]. Besides, kidney diseases result in significant productivity losses for both patients and their caregivers in terms of absenteeism, presenteeism (i.e., attending work while ill), and premature death of patients [12].

Importance of Early Tumor Detection: Despite a rapid increase in the number of patients with kidney cancer worldwide, mostly due to incidental diagnosis [13], recent developments in early detection, personalized medicine, novel treatment approaches, active surveillance, robot-assisted nephron-sparing surgical techniques, and minimally invasive procedures, such as thermal ablation, have raised hope of significantly improving kidney cancer survival [14]. For example, in the United

[1] Reproduced from [8] under the creative commons license for free use.

States, 5-year relative survival rates at diagnosis increased from 50% in 1975-77 to 57% in 1987-89 and reached 73% in 2003-09 [15]. Therefore, to improve the current kidney patient survival rate, revolutionizing the early cancer detection and prognosis system may be critical in the heterogeneous clinical setting [15].

Importance of Medical Imaging in Oncology: Recently, clinical oncology management including screening, diagnosis, treatment planning, and therapy monitoring has been revolutionized and accelerated by the explosive growth of medical imaging technologies [16]. Noninvasive medical images contain valuable information that can be extracted and utilized through computer-assisted interpretation. This process is referred to as 'radiomics,' which is a rapidly-emerging field of a study aiming to extract quantitative disease-specific data from medical images for use in clinical decision support [17]. In the context of clinical oncology, obtained information from the standard imaging modalities such as Computed tomography (CT), Magnetic resonance imaging (MRI), and Positron emission tomography (PET) scans are much more abundant, and radiomics aim to extract these high throughput quantitative features, covering the fields of texture, advanced shape modeling, and heterogeneity.

The Aim of this Thesis: Even though radiomics has immense potential to improve knowledge in tumor biology and guide the management of patients at bedside [16], kidney cancer prognosis, and prediction from volumetric medical images remains a labor-intensive and challenging pipeline of works as:

- Kidney localization in the 3D medical images and general assessment (e.g., volumetric analysis) of the kidney health,
- Detection of the malignant tumors (if any) in the kidney,
- Cancerous tumor analysis via determining mutated genes, tumor grading, tumor staging, etc.

Very often, one or more of the above-stated steps are performed manually via visual inspections [18], though several model-based image analysis techniques (i.e., level-set, active contour, graph cut, etc.) have also been traditionally used alongside [19]. However, these algorithms often require user interaction, are sen-

sitive to parameter settings, and involves heavy computation during clinical application. In contrast, a variety of Machine learning (ML) techniques [20] have also been widely applied to medical images throughout the last two decades to make the computer-aided kidney localization, segmentation and analysis tasks more feasible in the clinical environment. ML approaches enabled a shift from systems that are entirely designed by humans to systems that are trained by computers using example data from which feature vectors are extracted. Then computer algorithms determine the optimal decision boundary in the high-dimensional feature space. However, this process is still done by human researchers, and these human-engineered features are referred to as 'hand-crafted features,' which often does not optimally represent the task-specific discriminating image features. Therefore, a logical next step is to let computers learn the features that optimally represent the data for a specific task. This concept forms the basis of deep learning models, which contain many layers that transform input data to outputs while learning increasingly higher-level features [20]. The most successful type of model for image analysis to date is Convolutional neural network (CNN).

However, there are still a large number of challenges remaining in the context of kidney cancer research using different machine learning techniques, which require task-specific model designing and application. Therefore, the objective of this thesis is to develop novel machine learning approaches for volumetric medical images (e.g., CT, MRI, PET, etc.) that would make kidney cancer analysis more rapid and reproducible in the clinical environment. Although we would test the developed methods in this thesis on volumetric CT data, their applications are extendable to other modalities too.

Potential Clinical Impact: The primary aim of this thesis is to address several technical challenges associated with the current clinical image-based kidney cancer detection and analysis procedures. The scope of this thesis did not allow for extensive clinical studies, though we validated each component of this thesis on clinical data. Therefore, we mainly focus on discussing the technical contributions in this thesis. However, our works have great potential to impact patient care in clinical settings. This thesis presents a comprehensive working pipeline for kidney health analysis, starting from kidney localization in volumetric medical images to kidney cancer

analysis for treatment planning. We discuss in detail of this thesis' contributions in section 1.4. Briefly, our accurate and automatic kidney localization approaches may accelerate rapid kidney health analysis in clinical settings via presenting the background removed region-of-interest around kidneys on the point-of-care computer monitor, saving the time of clinicians in searching for kidneys in the image volume. Also, our segmentation-free and fast total kidney volume estimation approach may provide surrogate renal information, which can help clinicians to identify kidneys with reduced functionality. As we mentioned earlier that clinicians diagnose most kidney tumors incidentally nowadays, our novel collage CNN approach can be beneficial in identifying pathological kidneys in a patient, who might have primarily concerned with other diseases. Also, our image features-based non-invasive gene mutation detection, and RCC grading and staging approaches may significantly reduce the laboratory test-based diagnosis time and expenses. These methods may also help physicians in rapid treatment planning, which might be a crucial lifesaver for a patient. Besides, although we validated our methods in this thesis on kidneys, we expect these procedures to be easily transferable and practical for other human abdominal organs, e.g., liver, prostate, heart.

1.1.2 Renal Imaging

Different cross-sectional medical imaging such as contrast-enhanced ultrasound, CT, MRI, single-photon emission computed tomography, and PET have revolutionized the way of renal mass characterization (i.e., virtual biopsy) as well as the detection of metastatic disease, prognostication, and response assessment in patients with advanced kidney cancer [21]. Despite the availability of several imaging modalities, CT is often a popular choice because of its discriminatory contrast variability in the kidney based on the cancer types, grades, and stages [21]. CT is considered the gold standard for the characterization of renal tumors [22]. Although ultrasound imaging is more easily accessible than CT, imaging the retroperitoneum with ultrasound is a challenging task due to anatomic hurdles to sound wave transmission. Positioning an ultrasound probe in a suitable place to image a kidney is critical as ribs may cause posterior acoustic shadowing and often prevent accurate visualization of the renal unit beneath that rib [23]. Besides, kidney stones of 5

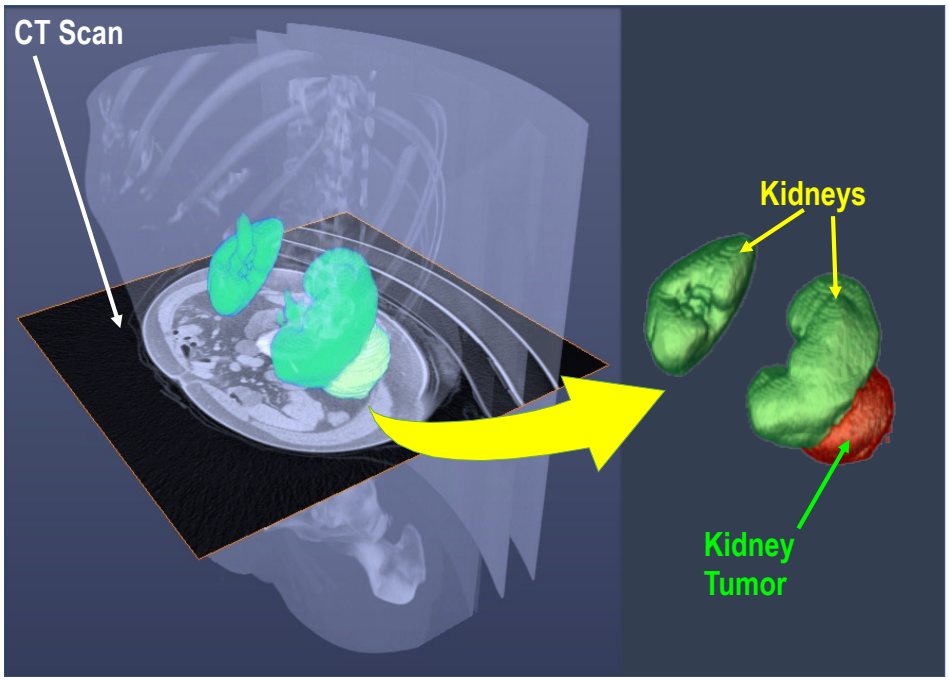

Figure 1.2: An image showing the kidney and its tumor in the abdominal CT volume.

mm or larger usually produce a posterior acoustic shadow [23]. On the other hand, MRI machines are available to a lesser extent in clinical settings than CT [24]. Thus, CT is often the first choice of imaging for the evaluation of a renal tumor. For this reason, we validated the proposed methods in this thesis on CT data, but these methods are also applicable to MRI data.

A CT scan helps to assess the tumors and other lesions (see Fig. 1.2) in a kidney. It also helps to detect and analyze obstructions such as calcification, abscesses, polycystic kidney disease as well as congenital anomalies, mainly when another type of examination like X-rays or physical exams is not conclusive. CT scans of the kidney may be used to evaluate the retroperitoneum (i.e., the back portion of the abdomen behind the peritoneal membrane) and also may be used to assist in needle placement in kidney biopsies. After the radical nephrectomy (i.e., removal of a kidney by surgery), CT scans may be used to locate abnormal masses in the space where the organ used to be. CT scans of the kidneys may also be performed

after kidney transplantation to evaluate the size and location of the new kidney to the bladder.

Often CT scanning is performed with "contrast" agent. Contrast refers to a chemical substance taken by mouth or injected into an intravenous (IV) line that causes the particular organ or tissue under the study to be seen more clearly. Contrast examinations often require the patient to fast for a certain period just before the scanning procedure. CT scans of the kidneys provide more detailed information about the organs than standard X-rays, resulting in the availability of more information related to the injuries and diseases of the kidneys.

1.1.3 Challenges in Kidney Cancer Detection and Analysis in CT

Kidney Localization: Accurate localization of kidneys in the 3D CT images is crucial as it is a start point for many automatic kidney analysis tasks, such as kidney segmentation and lesion detection [25]. Appropriate initial estimation of kidney position and extent can largely improve the performance of the subsequent treatment procedures such as thermal ablation [14]. Kidney localization allows discarding most of the non-relevant information and focuses on regions that are more likely to contain the tumor cells. Moreover, kidney localization is also important for efficient data retrieval and visual navigation of CT scans [25]. ML-based approaches show better kidney Region-of-interest (ROI) boundary wall localization accuracy compared to the traditional model-based approaches. However, the localization errors by these ML-based state-of-the-art methods are still greater than 7 mm, which may hinder the precise application of minimally invasive treatment procedures. For example, the precise application of thermal ablation like radiofrequency ablation (RFA) and cryoablation (CA) is suitable for patients who cannot undergo surgery because of comorbid illnesses [26]. These procedures are also suitable for those who have contralateral recurrences or a hereditary precancerous condition [26]. Since thermal ablation also targets to destroy at least 5 mm (up to 10 mm) thick seemingly normal tissue by the boundary of the tumor [27], inaccurate localization of kidney may lead to destroying normal tissues in the adjacent organs. Therefore, the first challenge we would address in this thesis is:

- **Challenge 1**: Reducing kidney localization error in the range of conventional

spatial resolution of modern CT scanners (0.5∼0.6 mm) [28] by using an image-based deep supervised learning approach.

Segmentation-free Kidney Volume Estimation: 'Total kidney volume' is an important biomarker in the clinical diagnosis of various renal diseases [29]. For example, it plays an essential role in the follow-up evaluation of kidney transplants [30]. Most existing methods for volume estimation rely on error-prone kidney segmentation as a prerequisite step, which has various limitations such as initialization-sensitivity and computationally-expensive optimization. Therefore, the second challenge we would address in this thesis is:

- **Challenge 2**: Developing a computationally inexpensive and robust kidney volume estimation approach, leveraging the image-based supervised learning that would bypass the segmentation procedure.

Pathological Kidney Detection: Image-based supervised learning requires a large number of annotated image data. Thus, the annotation burden to generate enough training data directly affects efficient supervised learning. The lack of sufficient annotation leads to sparse-labeled 3D datasets. 'Image labeling' is the process of recognizing different entities in an image. When a database has multi-dimensional image data having several objects of interest in those, however, labels are not present for an object for all the dimensions, we call it sparsely labeled data. Although Multiple-instance learning (MIL) [31] has shown promise in natural images for sparsely labeled data, and there are a few usages of this approach on medical data [32–34], it is yet to investigate from the image-based kidney cancer investigation perspective fully. Often kidneys are labeled as pathological, but there is no delineation of the tumor in the kidney. This scenario makes it quite difficult to use 2D slices to train a 2D CNN. On the other hand, the obvious solution to this problem is using 3D CNN. However, it poses two-fold challenges: (1) it drastically reduces the number of training samples. And (2) 3D CNNs are considerably more difficult to train as they contain significantly more parameters and necessitate the use of expensive GPUs with larger memory and require a lot more time to converge. Further, the inference time of a 3D CNN is high when running in a conventional point-of-care computer. Therefore, the third challenge we would address in this thesis is:

- **Challenge 3**: Tackling the sparse annotation problem in image-based supervised learning for pathological kidney detection.

Noninvasive Determination of Gene Mutation: Knowledge of the genetic make-up of a patient's kidney clear cell renal cell carcinoma (ccRCC) has a great prognostic value, which is helpful for treatment planning [18, 35]. Recent works [36, 37] have shown correlations between mutations in genes and different ccRCC features seen in CT images. Robust image feature identification is typically performed by expert radiologists, relying on human visual inspection. However, this process is difficult, time-consuming, and suffers from high intra/inter-observer variability. Therefore, the fourth challenge we would address in this thesis is:

- **Challenge 4**: Resolving the gene mutation detection problem in an automatic and noninvasive way via leveraging the CT-based tumor features.

Noninvasive Determination of RCC Grade: For RCC treatment planning, both RCC 'grade' and 'stage' provide critical information on the severity of renal cancer. Cancer grading is the way of classifying the cancer cells in the histopathologic images. The pathologist provides cancer a grade based on (a) how different they look from healthy cells, (b) how quickly they are growing and dividing, and (c) how likely they are to spread. Low-grade cancer cells are usually well-differentiated, and the tumors are slow-growing. In contrast, high-grade cancer cells are usually poorly differentiated or undifferentiated, and the tumors are faster growing.

The 'grade' of a ccRCC is one of the important prognostic predictors of 5-year survival of a patient. Radiologists use invasive percutaneous renal biopsy for ccRCC grading; however, inter-observer reproducibility of grades assigned by pathologists ranges from 31.3% to 97% [38]. Recent studies [39–42] proposed several ML approaches for automatic noninvasive ccRCC grading, but using hand-engineered CT 'textural' features. These features include histogram, gray-level co-occurrence matrices (GLCM), gray level run length matrix (GLRLM), gray level size zone matrix (GLSZM), etc., which are known as statistical context features [43]. On the other hand, classical CNN approaches learn features automatically and tend to outperform hand-engineered features-based ML approaches. However, classical CNN focuses on non-statistical context features like object

edges and shapes [44], and tend to put less emphasis on the statistical textural features [45], thus fails to determine the ccRCC grade. Therefore, the fifth challenge we would address in this thesis is:

- **Challenge 5**: Designing a deep neural network framework to learn powerful and discriminatory statistical radiological image features for accurate RCC grading.

Noninvasive Determination of RCC Stage: Staging is the way of classifying cancer based on the extent of cancer in the body. The stage is often based on the (a) size of the tumor, (b) whether cancer has spread (metastasized) to other parts of the body, and (c) where it has spread. Clinicians use the grade and stage of cancer, as well as other factors, to help plan treatment, estimate how cancer might respond to treatment, and give a prognosis. Knowledge of RCC 'stage' is vital for proper treatment planning and considered one of the important prognostic predictors of cancer-specific survival [46]. Clinical guidelines require clinicians to assign cancer stages before initiating any treatment [47]. For accurate staging of RCC before treatment planning, contrast-enhanced abdominal CT is considered essential [48]. Although tumor staging is believed to be dependent on the tumor size, Bradley et al. [49] suggested using CT image-based textural features to improve tumor staging. However, to our knowledge, there is no automatic CT image-based RCC staging approach present in the literature. Therefore, the sixth challenge we would address in this thesis is:

- **Challenge 6**: Investigating the potential of using CT textural features in a deep neural network for automatic and accurate RCC staging.

1.2 State-of-the-art

As we already discussed in section 1.1.1 that kidney cancer prognosis and prediction from volumetric medical images typically require a pipeline of works. Those are (a) kidney localization in the 3D medical images and general assessment of the kidney health, (b) detection of a malignant tumor (if any) in the kidney, and (c) kidney tumor analysis via determining mutated genes, tumor grading, tumor staging, etc. For years, medical imaging scientists have developed a large number of

model-based as well as ML-based approaches to tackle the above-stated working pipeline. In the following sections, we note some of the new methods from the literature related to the above steps.

1.2.1 Kidney Localization

Kidneys are often manually localized in the 3D CT data in the clinical settings [50–52]. Recently, several automatic kidney localization approaches have been proposed in the literature [53–57]. In Table 1.1, we list the most recent and relevant kidney localization-related literature. First, we discuss several conventional model-based approaches for automatic kidney localization. Yan et al. [53] proposed an improved connected component labeling algorithm based on intensity value to extract estimated kidney position. Li et al. [54] used nonlinear diffusion filtering and statistical shape model for kidney localization. Chen et al. [55] used the oriented active appearance model to localize the kidney in the 3D CT data. Xiang et al. [56] used a strategic combination of the Generalized Hough Transform and Active Appearance Model for kidney localization. Jin et al. [57] used a combination of 3D Generalized Hough Transform and 3D Active Appearance Models for kidney localization. Although these traditional methods show promising results, building a realistic model of kidney shape variability and balancing the influence of the model on the resulting segmentation are non-trivial tasks.

To overcome the limitations of the traditional methods mentioned above, several methods have been proposed based on supervised learning [58–65]. For example, Criminisi et al. [58, 59] proposed regression-forest-based anatomy localization methods that predict the boundary wall locations of a tight ROI encompassing a particular organ. Cuingnet et al. [60] fine-tuned the technique in [59] by using an additional regression forest, which improved the kidney localization accuracy by $\sim$60%. Gauriau et al. [61] used an extended cascade of regression-forests to estimate the confidence map of an organ, and the prediction was thresholded to get the final organ bounding box. Recently, Samarakoon et al. [63] proposed a light regression forest that uses fewer nodes than regular regression forests to localize different organs in the CT scans. Zhou et al. [62] used ensemble learning-based multiple 2D detectors, and their outputs are combined using a collaborative major-

ity voting in 3D to accomplish the robust kidney localization. In their subsequent works [64, 65], they localized kidney in CT images using template matching, handcrafted features, and local binary patterns.

Recently, deep learning using CNN has become a popular choice as it directly learns from the raw image data, while reducing the semantic gap created by handcrafted features and time required on designing features [66]. Several kidney localization methods using deep learning have also been proposed in the literature for the CT images. For example. Humpire et al. [66, 67] proposed a CNN-based approach to detect six organs, including kidneys. They trained three separate CNNs for classification of images taken from three orthogonal directions, where the rating of a slice is performed based on the presence or absence of a particular organ cross-section in that slice. The 3D organ bounding box is then generated by combining the classified-labels of orthogonal images. Similarly, Lu et al. [68] proposed a right-kidney localization method using a cross-sectional fusion of CNN and Fully convolutional networks (FCN). Xu et al. [25] proposed a 3D region proposal network for eleven body organs localization, including the kidney.

However, as we discussed in section 1.1.3 that the mean kidney ROI boundary localization errors by the state-of-the-art methods are still greater than 7 mm. Higher kidney localization error often makes it difficult to apply different minimally invasive treatment procedures. For example, higher kidney localization error may hinder the more precise application of minimally invasive thermal ablation like radio-frequency ablation (RFA) and cryoablation (CA). RFA and CA are suitable for patients who cannot undergo surgery because of comorbid illnesses [26]. These procedures are also useful for patients who have contralateral recurrences or a hereditary precancerous condition [26].

1.2.2 Kidney Functionality Analysis

Chronic kidney disease (CKD) refers to the reduced or absent functionality of kidneys for more than three months, which is a major risk factor for death worldwide [12]. In 2011, about 620,000 patients in the United States received treatment for ESRD either by receiving dialysis or by receiving kidney transplantation [29]. ESRD is the final stage of different CKDs, e.g., Autosomal dominant polycystic

Table 1.1: List of some recent kidney localization approaches. Here, conventional model-based approaches are denoted by category-1 (C1), classical machine learning approaches are denoted by C2, and (3) deep learning approaches are denoted by C3.

Authors	Methodology	C1	C2	C3
Yan et al. [53]	Connected component labeling algorithm	✓		
Li et al. [54]	Kidney statistical shape model	✓		
Chen et al. [55]	Oriented active appearance model	✓		
Xiang et al. [56]	2D Hough transform + 2D active appearance model	✓		
Jin et al. [57]	3D Hough transform + 3D active appearance model	✓		
Criminisi et al. [58]	Regression forest		✓	
Criminisi et al. [59]	Regression forest		✓	
Cuingnet et al. [60]	Regression forest		✓	
Gauriau et al. [61]	Cascaded regression forests		✓	
Samarakoon et al. [63]	Light regression forest		✓	
Zhou et al. [62]	Ensemble learning + majority voting		✓	
Zhou et al. [65]	3D GrabCut + contect-based image retrieval	✓	✓	
Zhou et al. [64]	ML-based template machine + Hough transform	✓	✓	
Humpire et al. [67]	3 CNN for each orthogonal direction			✓
Humpire et al. [66]	3 CNN for each orthogonal direction			✓
Lu et al. [68]	CNN + FCN			✓
Hussain et al. [1]	Orthogonal decision aggregated CNN			✓
Xu et al. [25]	3D region proposal network			✓

kidney disease (ADPKD), renal artery atherosclerosis (RAS), which are associated with the change of kidney volume. However, detection of CKDs are complicated; multiple tests such as the estimated glomerular filtration rate (eGFR) and serum albumin-to-creatinine ratio may not detect early disease and be unreliable in detecting disease and tracking its progression [29]. Also, it is known that serum albumin-to-creatinine ratio and eGFR values typically do not change until the fourth or fifth decade of life [69]. Recent works [29, 30] have suggested kidney volume as the potential surrogate marker for renal function and is thus useful for predicting and tracking the progression of different CKDs. The total kidney volume (TKV) has become the gold-standard image biomarker for the ADPKD and RAS progression at the early stages of this disease [30]. Besides, the renal volumetry has recently emerged as the most suitable alternative for evaluating the split renal function in kidney donors as well as the best biomarker in follow-up evaluation of kidney transplants [29]. The Canadian Society of Nephrology held a symposium on the topic

of TKV as a biomarker for disease severity and progression in ADPKD in April 2015. It is reported in [69] that in the majority of ADPKD patients, kidney volume increases over time, and this is attributed to an increase in cyst volume. They supported their argument by reporting a study on 241 ADPKD patients, where the estimated mean TKV of the patients was 1076mL, and the total cyst volume was 534mL. In contrast, the mean volume of normal kidneys is 196mL. Consequently, the estimation of the 'volume' of a kidney has become the primary objective in various clinical analyses of the kidney. The frequency of TKV measurements usually depends on the intended use of the information, and it has been reported that intervals of 6 months between TKV measurements may be sufficient to determine a more than 50% reduction in volume progression following drug treatment [69]. A patient is classified as having the rapidly progressive disease if the TKV increase more than 5% per year [69].

Kidney volume from 3D CT data is typically estimated using different segmentation methods. We can broadly categorize these segmentation methods into two groups based on the use of any prior kidney localization step. Some methods use manual/(semi)automatic kidney localization before segmentation [51, 53–55, 60], while some methods directly perform segmentation without using a prior localization step [50, 52, 70–83]. Although both types of methods are available in the literature, often, methods of the first category are preferred in the clinical environment. Because kidney localization not only facilitates better segmentation/volume-estimation but also can improve and speed up other algorithms such as lesion detection and registration [66].

Similar to the kidney localization approaches, we describe the kidney segmentation approaches by splitting into traditional approaches, classical machine learning approaches, and deep learning approaches. We list some of those methods in Table 1.2. As an instance of a traditional approach, Yan et al. [53] proposed a region growing approach based on a multi-scale mathematical morphology and labeling algorithm to extract the fine kidney regions. Li et al. [54] employed an optimal surface search algorithm for kidney segmentation. Chen et al. [55] used shape constrained Graph-cut methods for renal cortex segmentation. Dai et al. [50] proposed a fast GrowCut algorithm to segment the kidney in the 3D CT data. Khalifa et al. [51] used a geometric deformable model guided by a special stochastic speed

relationship for kidney segmentation. Skalski et al. [52] proposed a kidney segmentation method based on the active contour in the level set framework. Wieclawek et al. [83] presented a 3D marker-controlled watershed transform for fully automated CT kidney segmentation. Wolz et al. [71, 72] used a hierarchical atlas registration and target specific priors from an atlas database for kidney segmentation.

Recently, several hand-crafted features-based ML approaches have also been proposed for kidney segmentation in the CT data [60, 62, 64, 65, 74, 77, 82, 84]. Zhou et al. [62, 64, 65] used template matching, hand-crafted features, and local binary patterns for kidney segmentation. Cuingnet et al. [60] used a combination of regression forest and template deformation to segment kidneys. Glocker et al. [84] used a joint classification-regression forest scheme to segment different abdominal organs, including kidneys. Khalifa et al. [77] developed a 3D kidney segmentation framework integrating CT appearance features, higher-order appearance models, and adaptive shape model features into a random forest classification model. Hristova et al. [74] used a pipeline of intensity thresholding, nearest neighbor search, k-means tree, and median filtering for kidney segmentation. Zhao et al. [82] used CT features like intensity, texture, and context from the image and subsequently used regression forest for voxel-level classification to segment kidney.

Several deep learning-based approaches for kidney segmentation have also been proposed in the literature [73, 75, 76, 78, 80, 85, 86]. Chen et al. [73] proposed a 3D FCN based method for automatic multi-organ segmentation in dual-energy CT. Using dense V-network FCN, Gibson et al. [75] introduced a multi-organ segmentation approach on abdominal CT images. Valindria et al. [76] investigated the effectiveness of learning from multiple modalities for organ segmentation and shown effectiveness on kidney segmentation. Thong et al. [85] showed promising kidney segmentation performance using CNN. Similarly, using a multi-task 3D CNN, Keshwani et al. [80] proposed an ADPK segmentation approach. Sharma et al. [86] used the automated segmentation of ADPKs using FCN. Groza et al. [78] demonstrated a comparison of several CNN-based approaches to perform the segmentation of kidneys and shown that the foveal fully convolutional network is the most suitable deep architecture.

However, as we discussed in section 1.1.3 that most of the existing methods for kidney volume estimation rely on kidney segmentation as an intermediate step,

Table 1.2: List of some recent kidney segmentation approaches. Here also, conventional model-based approaches are denoted by C1, classical machine learning approaches are denoted by C2, and (3) deep learning approaches are denoted by C3.

Authors	Methodology	C1	C2	C3
Yan et al. [53]	Region growing	✓		
Li et al. [54]	Optimal surface search	✓		
Chen et al. [55]	Shape constrained Graph-cut	✓		
Dai et al. [50]	Fast GrowCut	✓		
Khalifa et al. [51]	Geometric deformable model	✓		
Skalski et al. [52]	Active contour	✓		
Wieclawek et al. [83]	3D marker-controlled watershed transform	✓		
Wolz et al. [71]	Hierarchical atlas registration + target prior	✓		
Wolz et al. [72]	Hierarchical atlas registration + target prior	✓		
Zhou et al. [62]	Ensemble learning + majority voting		✓	
Zhou et al. [65]	3D GrabCut + contect-based image retrieval	✓	✓	
Zhou et al. [64]	ML-based template machine + Hough transform	✓	✓	
Cuingnet et al. [60]	Regression forest + template deformation		✓	
Glocker et al. [84]	Joint classification-regression forest		✓	
Khalifa et al. [77]	Appearance models + shape model + random forests	✓	✓	
Hristova et al. [74]	Nearest neighbour + k-means	✓	✓	
Zhao et al. [82]	Regression forest		✓	
Chen et al. [73]	3D FCN			✓
Gibson et al. [75]	Dense V-network FCN			✓
Valindria et al. [76]	CNN			✓
Thong et al. [85]	CNN			✓
Keshwani et al. [80]	Multi-task 3D CNN			✓
Sharma et al. [86]	FCN			✓
Groza et al. [78]	Foveal FCN			✓

which has various limitations such as initialization-sensitivity and computationally-expensive optimization. Moreover, producing a more accurate kidney volume in the clinical environment would require using 2D/3D deep models that necessitate computers with larger memory. However, the point-of-care machines in a typical clinical setting are often not capable of running such a heavy computation. Besides, the inference run-time for the deep 3D models (e.g., 3D UNet [87], VNet [88]) for segmentation is often very long and significantly dependent on the machine's computation capability.

1.2.3 Pathological Kidney Detection

Generally, patients with renal tumors present clinical symptoms like flank pain, gross haematuria, or a palpable abdominal mass. Nevertheless, the detection rate of renal tumors has significantly increased due to the widespread use of various types of abdominal imaging, including ultrasonography, CT, and MRI. Typically, tumor analysis is accomplished with CT, which allows for assessment of local invasiveness, lymph node involvement, or other metastases. Nonetheless, more than 50% of kidney tumors are currently detected incidentally [13]. This tumor detection is typically carried out by radiologists through manual observation of abdominal image data. Although a good number of studies have been carried out on kidney localization as discussed in Sections 1.2.1 and 1.2.2, to the best of our knowledge, there has been no study to date that focused on automatic discrimination between healthy vs. renal cell carcinoma kidneys. A few studies [80, 86] performed ADPK segmentation, while very recently, several approaches using a different variant of 3D UNet [87] and VNet [88] have been proposed for kidney and kidney tumor segmentation using KiTS challenge database [89]. However, often these methods fail to detect and segment smaller kidney tumors. Besides, these methods require substantial computation resources, e.g., a graphics processing unit (GPU) with larger memory, which is not always available in the clinical environment.

On the other hand, medical image analysis has enjoyed significant performance improvements through the use of various ML algorithms over the past few years. Most of these algorithms are fully supervised, requiring a large number of annotated datasets for model learning and prediction accuracy analysis. Unlike 2D single- or three-channel data (e.g., gray-scale or color images), which are most commonly used in computer vision tasks, 3D medical data presents different sets of challenges for ML approaches. For example, tissue abnormalities such as tumors, cancers, nodules, stones, etc. are most often localized within a small region of anatomy and do not span the whole image volume. Localization and analysis of abnormal tissue are thus typically carried out on the 2D image slices. For example, the staging of kidney tumors is done through slice-based tumor analysis and manual boundary tracing. However, image tags or labels (e.g., healthy, cancerous, etc.) are mostly assigned per image volume or per-patient basis. Therefore, all slices of

an image are by default labeled with a single tag, though not all slices may contain the abnormal tissue. This scenario makes 'single-instance' ML approaches, especially deep learning ones such as CNN, challenging to train on the 2D slices, as the input slice often does not correspond to the assigned volume-based label. A typical solution for this problem is to use the full 3D image volume as a single-instance for learning. However, 3D CNNs are considerably more challenging to train as they contain significantly more parameters and consequently require many more training samples, necessitate the use of expensive GPUs with an extensive memory, and require a lot more time to converge. Moreover, in a typical clinical setting, point-of-care computers are often not equipped with enough computation power in terms of GPU, CPU and memory to run state-of-the-art 3D CNN models [87, 88] for inference.

An alternative approach to single-instance learning is MIL [31]. MIL is a variation of weakly supervised learning wherein the learner receives a set of labeled bags, or ensembles, each containing multiple instances. Learner assigns a class to each bag even if some of the cases are not members of that class. Using this MIL approach, the objective of our kidney tumor detection application can be formulated such that a labeled bag corresponds to a labeled CT volume, and the constituting instances within the bag correspond to the CT's 2D slices, some of which may contain tumors while many may not. This reformulation allows us to correctly incorporate volume-based labels within an easy to train 2D slice-based CNN framework. In the context of deep learning on medical images, the mutual benefits of MIL combined with the classification power of 2D CNNs have been recently demonstrated in a few applications. For example, mammogram classification for breast cancer detection [32], identifying anatomical body parts [34], colon cancer classification based on histopathology images [90], and classification of sizeable 2D microscopy images [33]. To the best of our knowledge, such an approach has not been explicitly implemented on 3D kidney data, and a novel representation of volumetric CT data is necessary to extend such techniques for the detection of kidney tumors in CT data.

1.2.4 Gene Mutation Detection

Renal cell carcinomas (RCC) are a common group of chemotherapy-resistant diseases among kidney cancer that accounted for an estimated 62,000 new patients and 14,000 deaths in the United States in 2015 alone [91]. North America and Europe have recently reported the highest numbers of new cases of RCC in the world [92]. The most common histologic subtype of RCC is clear cell RCC (ccRCC), which is known to be a genetically heterogeneous disease [36]. Recent studies [18, 93] have identified several mutations in genes associated with ccRCC. For example, the von Hippel-Lindau (VHL) tumor suppressor gene, BRCA1-associated protein 1 (BAP1) gene, Polybromo 1 (PBRM1) gene, and SET domain containing 2 (SETD2) gene have been identified as the most commonly mutated genes in ccRCC [18].

Traditionally, ccRCC underlying gene mutations are identified by genome sequencing of the ccRCC of the kidney samples after invasive nephrectomy or kidney biopsy [18]. This identification of genetic mutations is clinically important because advanced stages of ccRCC and poor patient survival are associated with the VHL, PBRM1, BAP1, SETD2, and Lysine (K)-specific demethylase 5C (KDM5C) gene mutations [18, 35]. Therefore, knowledge of the genetic make-up of a patient's kidney ccRCC has a great prognostic value that is helpful for treatment planning [18, 35]. Recent work [36, 37] shown correlations between mutations in genes and different ccRCC features seen in CT images. For example, an association between well-defined tumor margin, nodular enhancement, and intratumoral vascularity with the VHL mutation has been reported [37]. Ill-defined tumor margin and renal vein invasion were also reported to be associated with the BAP1 mutation [36], whereas PBRM1 and SETD2 mutations are mostly seen in solid (non-cystic) ccRCC cases [37]. This correlation opens a new field of study, 'Radiogenomics' [36, 37]. In Radiogenomics, radiological imaging data is used as a noninvasive determinant of the mutational status. It integrates genetic and radiomic information. From a methodological point of view, radiogenomics takes advantage of non-conventional data analysis techniques that reveal meaningful information for decision-support in cancer diagnosis and treatment [94]. Radiogenomics requires robust image feature identification, which is typically performed by expert radiologists. However, relying on human visual inspection is laborious, time-

consuming, and suffers from high intra/inter-observer variability. Recently, Kocak et al. [95] determined PBRM1 mutation using CT texture features in neural network and regression forests.

However, an image-based comprehensive deep supervised learning approach is yet to be investigated and designed for more accurate and automatic 'multiple' gene mutation detection.

1.2.5 Renal Cell Carcinoma Grading

The biological aggressiveness of ccRCC affects the prognosis and treatment planning [96]. The 'grade' of a ccRCC is one of the important prognostic predictors of 5-year survival where higher-grade tumors have an elevated risk of postoperative recurrence [41]. Although the 4-tiered Fuhrman grading system (FGS) [97] is used for ccRCC grading, in current clinical practice, pathologists prefer a simplified 2-tiered FGS that reduces variability and improves the reproducibility of the tumor grade [38, 41, 96]. The 2-tier FGS, which divides grades to low grade (Fuhrman I/II) and high grade (Fuhrman III/IV), was shown to be as effective as 4-tiered FGS in predicting cancer-specific mortality in a study population of 2,415 ccRCC patients [98].

Clinicians use invasive percutaneous renal biopsy for ccRCC FGS [38]. However, inter-observer reproducibility of grades assigned by pathologists ranges from 31.3% to 97% [38]. Oh et al. [39] tried to assess the correlation between the CT features and Fuhrman grade of ccRCC, where ccRCCs were retrospectively reviewed in consensus by two radiologists. Using logistic regression, they showed a threshold tumor size of 36 mm to predict (AUC: 70%) the high Fuhrman grade. Recently, Sasaguri et al. [40] suggested that RCCs can be characterized and graded based on CT textural features. Ding et al. [38] employed logistic regression on both non-textural features, e.g., pseudo capsule, round mass, as well as textural ones, e.g., histogram, gray-level co-occurrence matrices (GLCM), gray level run length matrix (GLRLM), and reported that textural features better discriminated high from low-grade ccRCC. Shu et al. [41] also employed logistic regression on CT textural features, e.g., GLCM, GLRLM, gray level size zone matrix (GLSZM), and achieved an FGS accuracy of 77%. Huhdanpaa et al. [42] used histogram

analysis of the peak tumor enhancement, tumor heterogeneity, and percent contrast washout in CT. They reported these parameters to be statistically different between low and high-grade ccRCC.

Current textural feature identification and quantification nonetheless faces two main challenges: it requires (1) ccRCC segmentation in CT, and (2) manual feature engineering. To our knowledge, there is no automatic ccRCC segmentation method present for CT. On the other hand, manual tumor segmentation relying on human visual inspection for feature identification is difficult, time-consuming, and suffers from high intra/inter-observer variability [4].

Avoiding sophisticated manual feature engineering, supervised deep learning using CNN has exploded in popularity for automatic feature learning, classification, as well as localization and dense labeling. In a classical CNN, the learned features in the first layer typically capture low-level features such as edges. The second layer detects motifs by spotting particular arrangements of edges. The third layer assembles motifs into larger combinations representing parts of objects, and subsequent layers detect objects as combinations of these parts [44]. These features are non-statistical context features [43] and the classical CNN tends to put less emphasis on the diffused statistical textural features that are often important for medical imaging applications, e.g., tumor characterization and analysis. This is also evident from [38, 41, 42] that CT intensity-based statistical features are important for ccRCC grading. In an attempt to learn statistical textural features via CNNs, Andrearczyk et al. [45] proposed deploying a global average pooling over each feature map of the last convolution layer of a conventional CNN to make the model object-shape unaware. However, the pooling still operates on the learned object-edge/motifs that do not capture complex and subtle textural variation in the input image. Therefore, it is still a research question to designing a novel image-based deep neural network architecture, which would be able to learn image 'textural' features for automatic ccRCC grading from CT images.

1.2.6 Renal Cell Carcinoma Staging

Clinical RCC staging is vital for proper treatment planning and thus considered one of the important prognostic predictors of cancer-specific survival [46]. The

American Joint Committee on Cancer (AJCC)/Union for International Cancer Control (UICC) specifies the criteria for tumor-node-metastasis (TNM) staging of each cancer depending on the primary tumor size (TX, T0-4); number and location of lymph node involvement (NX, N1-2); and metastatic nature, i.e., tumor spreading to other organs (M0-1) [47, 48]. Clinical guidelines require clinicians to assign TNM stages before initiating any treatment [47].

Table 1.3: Staging of RCC (AJCC TNM classification of tumors).

Anatomical Stages	**TNM Stages**		
Stage I	T1 (Tumor $\leq$7 cm)	N0	M0
Stage II	T2 (Tumor $>$7 cm but limited to kidney)	N0	M0
Stage III	T1-2, T3 (Tumour extends up to Gerota's fascia)	N1, Any	M0
Stage IV	T4, Any (Tumour invades beyond Gerota's fascia)	Any	M0-1

AJCC TNM is currently a manual process that includes two separate staging processes, performed before treatment planning and during/after surgery, to reflect the time-sensitive staging mechanism [48]. 'Clinical' staging is performed before treatment by expert radiologists via physical examination, CT image measurements, and tumor biopsies. Clinically determined TNM stages (e.g., T or M) are designated with prefix 'c' (i.e., cT and cM). 'Pathological' staging, on the other hand, is based on the resected tumor pathology results either during or after surgery [47] and designated with prefix 'p' (i.e., pT and pM). Accurate clinical staging (i.e., cT, cM) of RCC is vital for appropriate management decisions [49]. Partial nephrectomy (PN), also known as nephron-sparing surgery, is typically preferred for T1 and T2 tumors [48]. After studying 7,138 patients with T1 kidney cancer, Tan et al. [99] suggested that treatment with PN was associated with improved survival. In a similar study on pT2 tumor patients, Janssen et al. [46] showed that patients having PN had a significantly longer overall survival. Radical nephrectomy (RN), which refers to complete removal of the kidney with/without the removal of the adrenal gland and neighboring lymph node, is generally reserved for T3 and T4 tumors [49].

However, the pre-surgery clinical tumor staging often suffers from misclassification errors. For example, in a recent study, Bradley et al. [49] reported 23

disagreement cases between cT and pT stages of 90 patients. The study further indicated that five patients were misclassified with cT3 but later down-staged to pT2, while six patients were misclassified with cT2 but later up-staged to pT3 for the same patient cohort ($\sim$12%). In another study on 1,250 patients who underwent nephrectomy, Shah et al. [100] reported 11% (140 patients) upstaging of tumors from cT1 to pT3. Besides, there was tumor recurrence in 44 patients (31.4% of the pT3 promoted cases), where most of these patients initially had PN. These alarming findings suggest that PN is associated with better survival in low stage tumors (T1 and T2), while RN is associated with reduced recurrence in high stage (T3 and T4) tumors. However, high stage tumors (T3-4) are often misclassified as the low stage (T1-2) in the clinical staging phase. Also, we see in the rows 1-3 of Table 1.3 that the tumor classifying criterion is not well defined for stages T1, T2, and T3. Therefore, radiologists often use the TNM description to assign an overall 'Anatomical stage' from 1 to 4 using the Roman numerals I, II, III, and IV [48], see Table 1.3.

For accurate staging of RCC before treatment planning, contrast-enhanced abdominal CT is considered essential [48]. Although tumor staging is believed to be dependent on the tumor size, by studying the pT stages of 94 kidney samples, Bradley et al. [49] argued that stages $>$ T3 do not always correlate with tumor size. This study further suggested using CT image-based textural features to improve tumor staging, like in Furhman tumor grading. However, to our knowledge, there is no automatic CT image-based RCC staging approach present in the literature.

1.3 Thesis Contributions

Our main technical contributions in this thesis are summarized as follows:

- We address the challenge of reducing the kidney localization error in the CT volumes by using novel DNN approaches. First, we discuss a practical deep CNN-based approach (localization approach 1) for tight kidney ROI localization. Here, we aggregated orthogonal 2D slice-based probabilities of containing kidney cross-sections into a voxel-based decision that ultimately predicts whether an interrogated voxel sits inside or outside of a kidney ROI. Also, we discuss a second deep learning approach that further improves the

automatic kidney localization performance than that by the localization approach 1. This method uses an effective CNN-guided Region convolutional neural network (RCNN) approach for efficient kidney localization in CT images.

- We propose three segmentation free ML-based approaches for total kidney volume estimation. First, we present a novel kidney volume estimation approach for 3D CT images with dual regression forests that skipped the segmentation step. Also, we present two methods that use a CNN and FCN, respectively, to predict slice-based cross-sectional kidney areas followed by integration over these values across axial kidney span to produce the volume.

- We propose a novel DNN approach in the MIL framework for efficiently learning from the sparsely labeled 2D data, which is, at the same time, computationally inexpensive both in training and inference. In this work, we propose a novel collage image representation for the CNN framework for pathological kidney classification.

- We propose a novel DNN approach that can efficiently learn CT-based ccRCC features for automatic determination of mutated genes. We develop a CNN approach that automatically determines the ccRCC image features. Then the binary decisions (i.e., presence/absence of a mutation) for all the ccRCC slices in a particular kidney sample are aggregated into a robust singular decision that ultimately determines whether an interrogated kidney sample has undergone a specific mutation or not.

- We propose a novel DNN architecture that specifically learns image 'textural' features for automatic RCC grading and staging. We present a learnable image histogram (LIH) layer within a DNN framework capable of learning complex and subtle task-specific textural features from raw images directly, adhering to the classical input-output mapping of a CNN.

1.4 Thesis Organization

In addition to this introductory chapter, this thesis includes seven chapters. The final chapter discusses the conclusions and directions for future work.

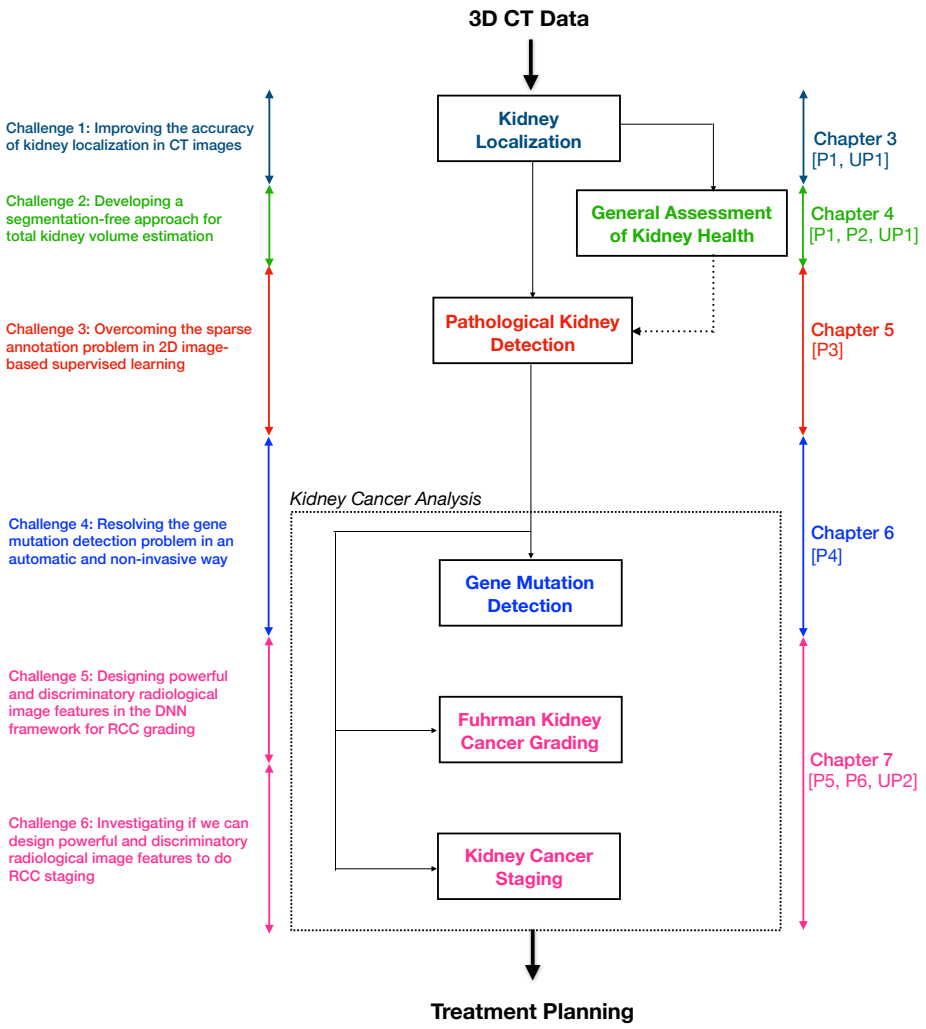

Figure 1.3: A flowchart of our kidney cancer analysis working pipeline with component-wise associated challenges, publications and chapter numbers that discuss the technical contributions of this thesis.

We show an overview of chapters 3 to 7 in the flowchart in Fig. 1.3. It also shows relevant publications associated with each of the research questions.

Chapter 3 presents our kidney localization approaches that address the first

research challenge of this thesis.

In chapter 4, we present our segmentation free ML-based approaches for total kidney volume estimation that address the second research challenge.

Later, in chapter 5, we present our pathological kidney detecting DNN approach in the MIL framework that efficiently learns from the sparsely annotated data. This approach addresses the third research challenge in this thesis.

Chapter 6 presents our novel DNN approach that efficiently learns CT-based ccRCC features for automatic determination of mutated genes. This approach addresses the fourth research challenge in this thesis.

Finally, in chapter 7, we discuss our learnable image histogram-based DNN approach that specifically learns CT 'textural' features for automatic RCC grading and staging. This approach addresses our fifth and sixth research challenges in this thesis.

Chapter 2

Data and Experimental Setup

2.1 Private Database

We accessed abdominal CT scans from 100 patients from the Picture archiving and communication system (PACS) of the Vancouver General Hospital (VGH), Vancouver, BC, Canada, with all ethics review board approvals in place. Appendix **??** shows the ethics certificate of this study. We show a few examples of CT slices from the VGH patient pool in Fig. 2.1. These data were collected using the CT scanner Siemens SOMATOM Definition Flash (Siemens Healthcare GmbH, Erlangen, Germany). Two expert radiologists at VGH performed the kidney delineation to produce the ground truth using medical image viewing software OsiriX MD. We show a summary of these data in Table 2.1:

2.2 Public Database 1

We also obtained access to 267 patients' CT scans from The Cancer Genome Atlas Kidney Renal Clear Cell Carcinoma (TCGA-KIRC) database [101]. These data were collected in multiple institutions across the United States by different types of CT scanners. We show a few examples of CT slices from the TCGA-KIRC patient pool in Fig. 2.2. We also collected clinical information about these patients from the same database. We collected the corresponding gene mutation information from the *cBioPortal for Cancer Genomics* [35] database. We show a summary of

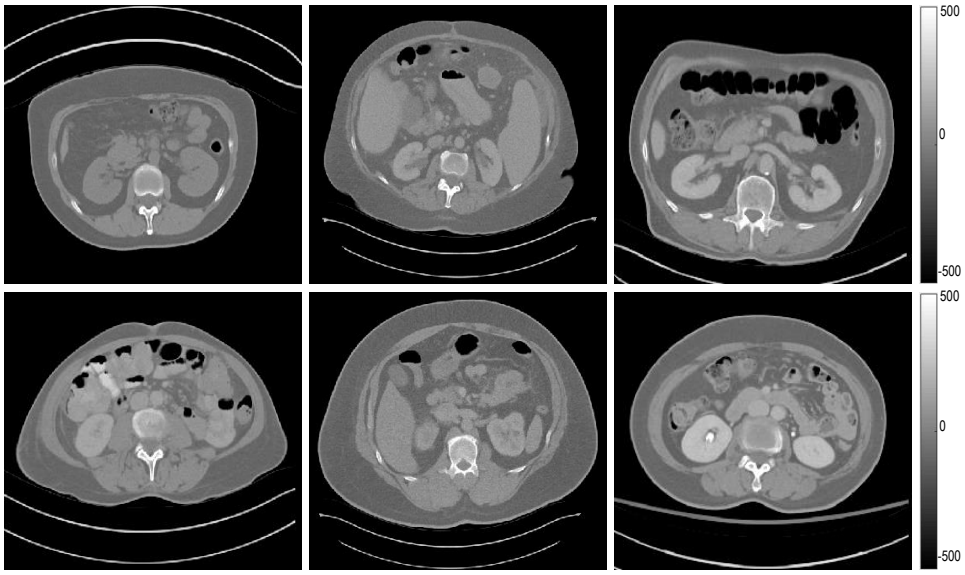

Figure 2.1: Examples of kidney data from our VGH patient pool.

these data in Table 2.2:

2.3 Public Database 2

We also obtained access to 210 patients' CT scans from the 2019 Kidney Tumor Segmentation (KiTS) Challenge database [102]. Patients who underwent partial or radical nephrectomy for one or more kidney tumors at the University of Minnesota Medical Center between 2010 and 2018 were candidates for inclusion in this database. We show a few examples of CT slices from the KiTS patient pool in Fig. 2.3. We also collected the kidney segmentation data from the same database. A summary of these data is shown in Table 2.3:

2.4 Strategies to Overcome the Effects of Limited Data

Typically, the performance of a supervised deep model increases logarithmically with the volume of training data size [103]. However, while both computation power and model capacity has continued to grow, datasets to train these models have remained stagnant [103]. This scenario is especially true for medical imaging

Table 2.1: Summary of relevant and available information of the CT data from VGH.

Items	**Descriptions**
Modality	CT
Pixel Dimensions	Axial: 1.5 $\sim$ 3 mm
	Coronal: 0.5820 $\sim$ 0.9766 mm
	Sagittal: 0.5820 $\sim$ 0.9766 mm
Contrast Agent Used	45 cases
Total Patients	100
Number of Males	50
Number of Females	50
Age	Mean: 56.71$\pm$15.81 Y
	Minimum Age: 19 Y
	Maximum Age: 89 Y
Number of Pathological Kidneys	12 kidneys in 12 patients

data and gets worse by the annotation burden of large 3D datasets. In this thesis, we also faced a similar problem, where our datasets are reasonably smaller and often lack appropriate annotations. To tackle this problem, we adopted the following procedures to generate effective and scalable performance by our proposed methods:

- We always pre-trained our deep models, wherever useful, using the ImageNet challenge database [104].
- In a seminal work, Yosinski et al. [105] reported that the first three layers in a CNN contain generic and reusable features. Beyond the third layer, the features gradually become more specific to the source data set. Therefore, during the fine-tuning of the pre-trained models on our medical images, we chose local learning rates very carefully. We chose small learning rates for the first few layers, while comparatively larger for the deep layers.
- We increased the volume of our medical imaging data via various types of augmentations.
- To make our deep model generalized to avoid over-fitting, we adopted regularization in the form of Dropouts [106]. We also kept an eye on the training

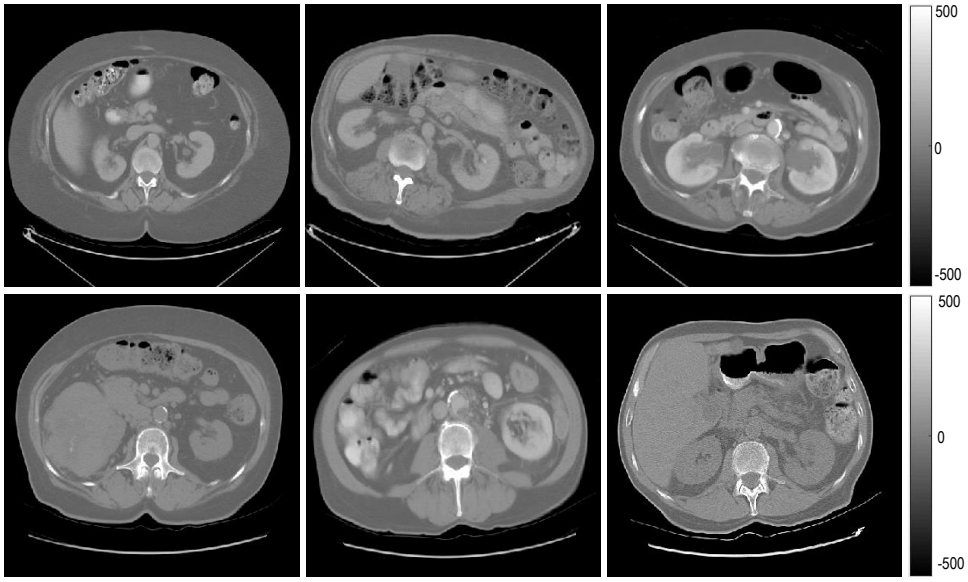

Figure 2.2: Examples of kidney data from our TCIA-KIRC patient pool.

and validation losses during model training to prevent over-fitting.

- To avoid the problem of class-imbalance in our datasets, we often adopted class-wise different but carefully calculated overlaps among patches.

Table 2.2: Summary of relevant and available information of the CT data from the TCGA-KIRC.

Items	**Descriptions**
Modality	CT/MR
Pixel Dimensions	Axial: 1.5 $\sim$ 7.5 mm
	Coronal: 0.29 $\sim$ 1.87 mm
	Sagittal: 0.29 $\sim$ 1.87 mm
Total Patients	267
Number of Males	176
Number of Females	91
Age	Mean: 60.28$\pm$12.04 Y
	Minimum Age: 27 Y
	Maximum Age: 89 Y
Race	White: 241
	Black or African American: 22
	Asian: 3
Pathological Kidneys	Either one or both kidneys in all patients

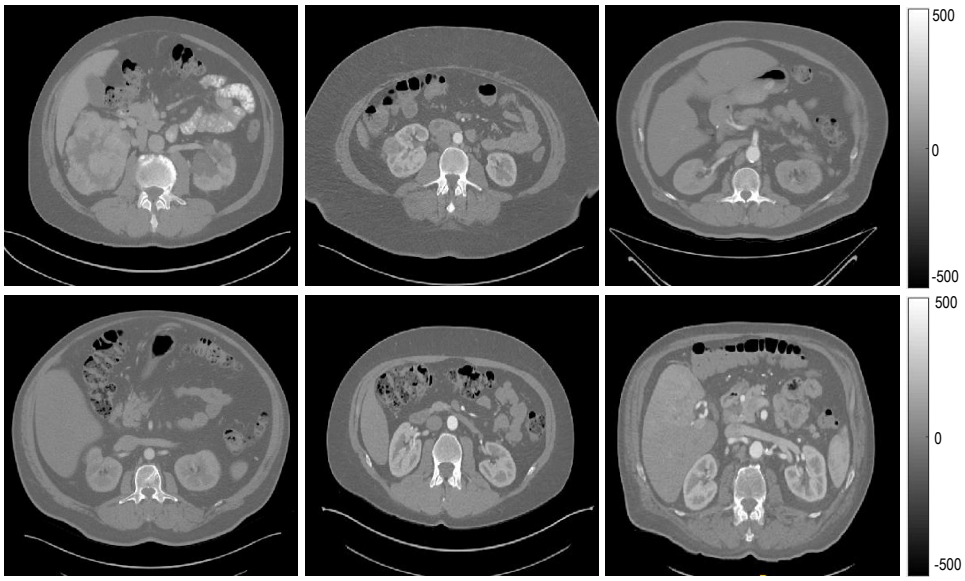

Figure 2.3: Examples of kidney data from our KiTS patient pool.

Table 2.3: Summary of relevant and available information of the CT data from the KiTS.

Items	**Descriptions**
Modality	CT
Pixel Dimensions	Axial: 3 mm (spacing was uniformed across cases)
	Coronal: 0.7816 mm (spacing was uniformed across cases)
	Sagittal: 0.7816 mm (spacing was uniformed across cases)
Total Patients	210
Contrast Agent Used	in all cases
Number of Males	Not available
Number of Females	Not available
Age	Mean: Not available
	Minimum Age: Not available
	Maximum Age: Not available
Pathological Kidneys	Either one or both kidneys in all patients

Chapter 3

Kidney ROI Localization

3.1 Aggregated Orthogonal Decision CNN for Kidney Localization

In this section, we present a CNN-based approach that addressed the challenge of automatic kidney localization in the CT image. We originally published the methodology presented in this section (3.1) in Hussain et al. [1]. This method used an effective deep CNN-based approach for tight kidney ROI localization. Here, we aggregated the probabilities of containing 2D kidney cross-sections into a voxel-based decision that ultimately predicts whether an interrogated voxel sits inside or outside of a kidney ROI.

3.1.1 Orthogonal Decision CNN for Kidney Localization

We use a deep CNN to predict the locations of six walls of the tight ROI boundary around a kidney by aggregating individual probabilities associated with three intersecting orthogonal (axial, coronal, and sagittal) image slices (Fig. 3.1). The CNN has eleven layers excluding the input. It has five convolutional layers, three fully connected layers, two softmax layers, and one additive layer. All but the last three layers contain trainable weights. The input is a 256×256 pixel image slice, either from the axial, coronal or sagittal directions, sampled from the initially generated local kidney-containing volumes. At first, we pre-train this single CNN (from layer

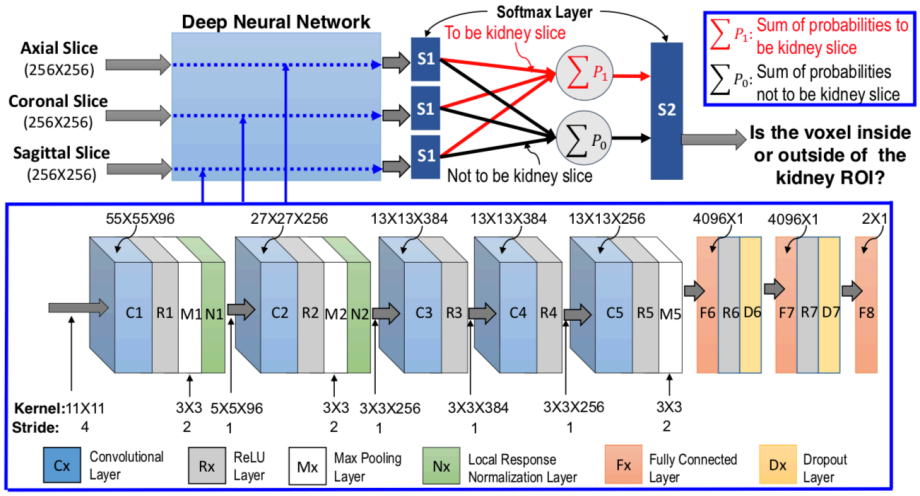

Figure 3.1: Orthogonal decision aggregated CNN for kidney localization.

1 to layer 8) using the ImageNet Challenge dataset [104]. Then we fine-tune this pre-trained model using our dataset containing a mix of equal numbers of 3D orthogonal image slices. CNN uses convolutional layers for sequentially learning the high-level non-linear spatial image features (e.g., object edges, intensity variations, orientations of objects, etc.). Subsequent fully connected layers prepare these features for optimal classification of the object (e.g., kidney cross-section) present in the image. In our case, five convolutional layers followed by three fully connected layers make a reasonable decision on orthogonal image slices if they include kidney cross-sections or not. During testing, we feed three different orthogonal image slices parallelly to this CNN, and acquired the probabilities for each slice of being a kidney slice or not at the softmax layer, $S1$ (Fig. 3.1). These individual probabilities from the three orthogonal image slices are added class-wise (i.e., containing kidney cross-section [class: 1] or not [class: 0]) in the additive layer, shown as $\sum P_1$ and $\sum P_0$ in Fig. 3.1. We then use a second softmax layer $S2$ with $\sum P_1$ and $\sum P_0$ as inputs. This layer decides whether the voxel where the three orthogonal input slices intersect is inside or outside the tight kidney ROI. The second softmax layer $S2$ is included to remove any potential miss-classification by the first softmax layer $S1$. We consider the voxels having probability ($\in [0,1]$) > 0.5 at $S2$ to be inside the kidney ROI. Finally, we record the locations of six boundary walls (two in each

orthogonal direction) of this ROI from the maximum span of the distribution of these voxels (with probability > 0.5) along with three orthogonal directions.

3.1.2 Experimental Setup and Data Acquisition

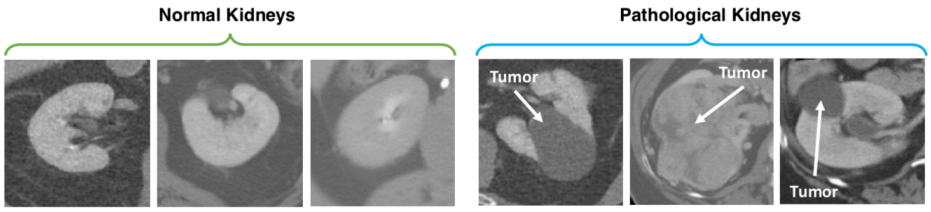

Figure 3.2: Example kidney data from our patient pool demonstrating data variability, ranging from normal to pathological.

We used 100 patients' CT scans from our VGH patient pool (discussed in section 2.1). We were able to use a total of 200 kidney samples (both left and right kidneys) from which we used 120 samples (from 60 randomly chosen patients) for training, 20 samples from 10 randomly chosen patients for validation, and the rest is unseen. Our dataset included 12 pathological kidney samples (with endo- and exophytic tumors), and our training and test data contained 6 cases, each. The CNN was implemented using *Caffe* [107]. The base learning rate for CNN pre-training was set to 0.01 and was decreased by a factor of 0.1 to 0.0001 over 25,000 iterations. During fine-tuning, the base learning rate was set to 0.001 and was decreased by a factor of 0.1 to 0.0001 over 20,000 iterations.

3.1.3 Data Pre-processing

Before training the CNN, we did some basic pre-processing of the data. We programmed an automatic routine for 'abdominal' CT that separates the left and right kidneys. Since the left and right kidneys always fall in the separate half volumes, the routine simply divided the abdominal CT volume medially along the left-right direction. The routine also discarded a few slices in the pelvic region from an image (where applicable). However, this step was optional and only carried out on slices beyond ~ 52 cm (4 times the typical kidney length ~ 13 cm) from the chest

Table 3.1: Comparison of mean kidney ROI boundary localization error (mm) and mean kidney ROI centroid localization error (mm) in terms of Euclidean distance. Not reported values are shown with (-).

Methods	**Boundary Error (mm)**		**Centroid Error (mm)**	
	Left	Right	Left	Right
Cascaded RF [60]	7.00 ± 10.0	7.00 ± 6.00	11.0 ± 18.0	10.0 ± 12.0
RF1 [58]	17.3 ± 16.5	18.5 ± 18.0	-	-
RF2 [59]	13.6 ± 12.5	16.1 ± 15.5	-	-
CS-(CNN+FCN) [68]	-	-	-	7.80 ± 9.40
Proposed Method	**6.23±6.06**	**5.92±6.55**	**7.91±4.99**	**7.69±4.23**

side of the image. Finally, our pre-processing routine re-sized the medially separated CT volumes to generate fixed resolution cubic volumes (e.g., $256 \times 256 \times 256$ voxel in our case) using either interpolation or decimation, as needed.

3.1.4 Validation on Kidney Data

We provide results of our proposed kidney localization on 3D kidney data to enable direct comparisons with those obtained by recently reported kidney localization [58–60, 68]. Since the recently published methods we use for comparison are mostly either RF-based or deep CNN-based, reproduce their results are impossible without access to the code and data on which they tested. However, the type of data these methods used is similar to ours in terms of resolution and imaging modality. Therefore, we conservatively use their reported accuracy values for comparison, rather than using our implementation of their models.

In Table 3.1, we present kidney boundary and centroid localization performance comparisons of cascaded RF-based [60], single RF-based (RF1 [58] and RF2 [59]), cross-sectional (CS) fusion of CNN and FCN-based [68], and our proposed method. The centroid of a kidney is the approximate mid-point of the kidney, which can be inferred directly by a supervised learning model or can be estimated from the estimated six kidney boundary walls. We estimate the mean kidney wall

error E_{mean} for all the test samples as:

$$E_{mean} = \frac{1}{6N}\sum_{1}^{N}|G_L - P_L| + |G_R - P_R| + |G_A - P_A| + |G_P - P_P| + |G_S - P_S| + |G_I - P_I|, \tag{3.1}$$

where N is the total number of test kidney samples, G represents ground truth and P represents the predicted wall location in *mm*, and the subscripts L, R, A, P, S, and I denote the left, right, anterior, posterior, superior and inferior directions, respectively.

The cascaded RF method used the RF1 [58] for coarse localization of both left and right kidneys, then fine-tuned these locations using an additional RF per left/right kidney. Even then, its centroid localization errors and boundary localization errors were higher than those of the proposed method. The RF2 [59] was an incremental work over the RF1 [58], and both use regression-forests for different anatomy localization. Both methods exhibited higher boundary localization errors than those of the cascaded RF and proposed methods and did not report any centroid localization accuracy. The recently proposed CS-(CNN+FCN) [68] method reported significantly better kidney centroid localization performance than the cascaded RF [60]. However, this method was only validated on the right kidneys and did not report the kidney boundary localization accuracy. As evident from the quantitative results, compared to all these recent methods, the proposed method demonstrates better performance in both kidney boundary and centroid localization by producing the lowest localization errors in all categories.

3.1.5 Discussion

In this section, we discussed a deep learning approach for human kidney localization. This method enabled a clinical approach for kidney localization in the raw abdominal CT data. We formulated an effective deep CNN-based method for kidney ROI localization, which aggregates 2D orthogonal slice-based kidney candidacy decisions. Our deep CNN better captured the rich and complex variability in kidney anatomy and outperformed the hand-engineered feature representations used in [59, 60]. Our experimental results demonstrated the best kidney ROI localization performance compared to that of recent literature [58–60, 68].

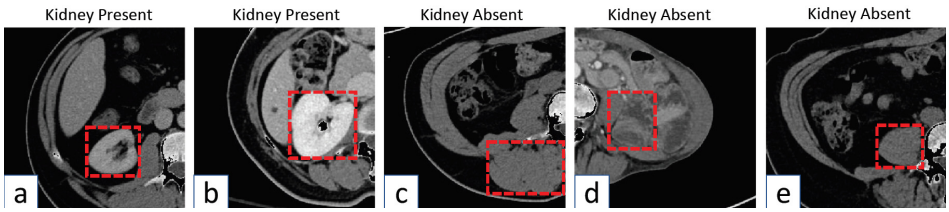

Figure 3.3: Example CT data from our patient pool demonstrating RCNN performance on kidney bounding box localization. The RCNN correctly localized the kidney in (a) and (b), but produced false-positive kidney bounding boxes in (c)-(e), where the kidney was absent.

3.2 CNN Guided Mask-RCNN for Kidney Localization

Typically, medical imaging scientists treat organ localization in an image as a detection problem [108], and most of the kidney localization approaches discussed in section 1.2.1 achieved organ detection by performing an initial regression or classification step. Region-based convolutional neural network (RCNN) [109] has recently emerged as a promising approach in computer vision that solves the localization problem by operating within the "recognition using regions" paradigm. RCNN is quite capable of directly predicting 2D bounding boxes around the objects of interest in an image. However, the effectiveness of using RCNN for organ localization in 3D medical images are often error-prone. In our own experience, we found that RCNN often produces false-positive bounding boxes around image objects that are similar looking to the organ of interest. For example, Fig. 3.3 shows localization obtained using RCNN, which demonstrates several false-positive kidney bounding box localization results. We believe a possible cause of such false positives could be the fact that a typical RCNN network gets optimized on local region proposals. Thus it always looks for an object or a set of objects in a test image via analyzing local regions only. Hence, an RCNN network lacks the global context, which is especially important in the medical image data.

In this section, we discuss a deep learning approach that further improves the automatic kidney localization performance than that mentioned in section 3.1. The method presented in this chapter uses an effective CNN-guided RCNN approach for efficient kidney localization in volumetric CT images. Briefly, we construct a deep learning architecture comprising three steps: Firstly, we use ResNet-50 [110],

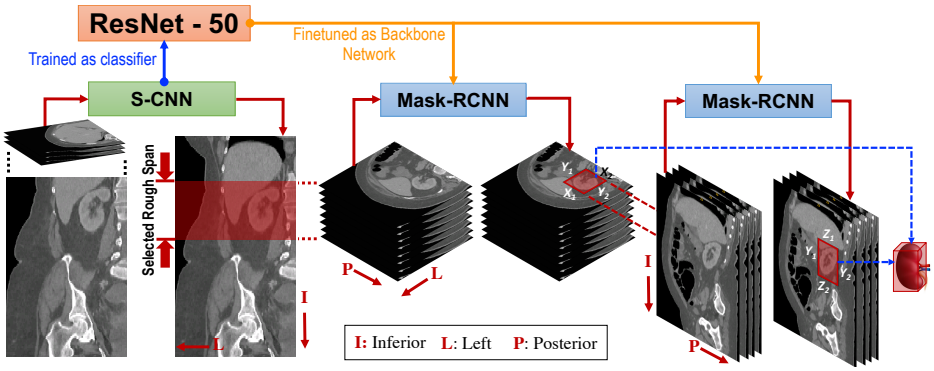

Figure 3.4: Schematic diagram of the proposed selection CNN guided Mask-RCNN for efficient kidney localization in the volumetric CT images.

which call S-CNN, detects an approximate span of 2D axial slices encompassing the target kidney. Secondly, we feed axial slices from the S-CNN identified span region to a Mask-RCNN [7] to get the 2D kidney bounding box lying on the axial plane. Finally, we use the same Mask-RCNN to detect the 2D kidney bounding box in the 2D sagittal slices, which we take strictly inside the organ span estimated by the Mask-RCNN in the previous step. Since RCNN typically produces false-positive kidney bounding boxes in those slices that do not contain the kidney, the CNN pipeline of our method controls the choice of slices (fed to the RCNNs) by extracting those from the kidney containing region only. Thus, the main novelties of the proposed localization method lie (1) in operating the localization problem within the "recognition using regions" paradigm, and (2) in the cascaded CNN architecture that extracts better 3D bounding box estimates with almost no false positives. Note that we use the same ResNet-50 as S-CNN as well as the backbone network in Mask-RCNN. Also, note that we extract the kidney region from the masks as it covers the kidney cross-section more tightly than the predicted object boundary.

3.2.1 Kidney Span Detection using S-CNN

We use the S-CNN (ResNet-50) to classify 2D axial slices that enable a rough detection of the kidney span along the axial direction (see Fig. 3.4). The initial slice classification labels (i.e., 0: kidney absent, 1: kidney present) may contain a few

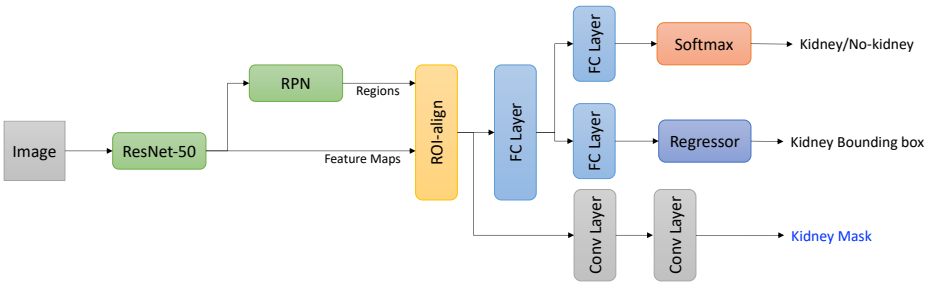

Figure 3.5: Block diagram of the Mask-RCNN [7] used in the proposed method.

false positives and false negatives. To remove those, we perform a moving average over the label values along the axial direction with a moving window size of 12 cm as a typical kidney length is approximately 12 cm [111]. Then we normalize the average values and estimate the organ span from the range of values ≥ 0.75. Since this is a rough estimation of a kidney span, we empirically choose the threshold value of 0.75. This approximate span could be bigger than the actual kidney span. If the estimated span comes out smaller than a typical kidney length, we take extra slices into the span in the superior and inferior directions. This S-CNN (ResNet-50) network was pre-trained on the ImageNet dataset, and we fine-tune the network weights on our kidney dataset. Although the S-CNN aims to detect approximate kidney span in the axial direction, we fine-tune this network using cross-sectional 2D slices from all three orthogonal directions. For details of this network structure and function, we refer readers to [110].

3.2.2 Bounding Box Detection in the Coronal-Sagittal Direction

In this stage, we use a Mask-RCNN [7] to detect the 2D kidney bounding box along the coronal and sagittal directions (see Fig. 3.4 and 3.5). The input to the Mask-RCNN is the 2D axial slices strictly taken from the inside of the selected span by S-CNN. The Mask-RCNN is capable of classification, bounding box generation, and instance segmentation of an object in an image. It comprises two stages: (1) the first stage generates proposals about the regions where there might be an object based on the input image. (2) the second stage predicts the class of the object, refines the bounding box, and generates a mask in the pixel level of the

object based on the first stage proposal. Both stages are connected to the backbone structure, which is a feature pyramid network (FPN) style deep neural network. It consists of a bottom-up pathway, a top-bottom pathway, and lateral connections. The bottom-up pathway can be any CNN, which extracts features from raw images. The top-bottom pathway generates a feature pyramid map that is similar in size to the bottom-up pathway. Lateral connections are convolution and adding operations between two corresponding levels of the two pathways. A region proposal network (RPN) works on the FPN feature map to propose an object region. A pooling layer (ROI-align) then works on the proposed regions to extract a fixed-length feature vector. Then each feature vector is fed into a sequence of fully connected layers or convolution layers that finally branch into three sibling output layers: an object bounding box layer, an object classification layer, and an object masking layer. This entails the use of a multitask loss function $L = L_{cls} + L_{bbox} + L_{mask}$, where L_{cls}, L_{bbox} and L_{mask} are the class loss, bounding box loss and mask loss, respectively. As a 'backbone' network of the Mask-RCNN, we use the ResNet-50 [110], which we fine-tune from S-CNN. For fine-tuning, we use a kidney containing 2D slices from all three orthogonal directions. During inference, we restrict the Mask-RCNN to produce a single bounding box and kidney mask per slice. Although the Mask-RCNN produces a 2D bounding box around a kidney cross-section, in most of the cases, it does not tightly encompass a kidney cross-section. Rather gaps are seen between the predicted boundary line and the actual kidney boundary. Therefore, we use the predicted kidney mask to generate the rectangular kidney bounding box. Finally, we find the sagittal and coronal edges of a bounding box, which is the Union set of all the axial bounding box, by $X_1 = min(x_1)$, $X_2 = max(x_2)$, $Y_1 = min(y_1)$ and $Y_2 = max(y_2)$, where *min* and *max* are the minimum and maximum operators, respectively, and X_1, X_2 and Y_1, Y_2 are the Union box edges along the coronal and sagittal directions, respectively (see Mask-RCNN output in Fig. 3.4). Note that finding the rough kidney span along the axial direction by the initial S-CNN is important for this stage, as false-positive bounding boxes may corrupt these estimates.

3.2.3 Bounding Box Detection in the Axial-Sagittal Direction

In this final detection stage, we use the same Mask-RCNN. The input to the Mask-RCNN in this stage is the 2D sagittal slices strictly taken from the inside of the selected span X_1 and X_2 in the previous step (see Fig. 3.4). In this stage, the Mask-RCNN detects the kidney bounding box along with the sagittal and coronal directions. This stage updates the estimated axial kidney span in the previous step. Finally, we find the axial edges of a Union bounding box, which is the Union of all the sagittal bounding boxes, by $Z_1 = min(z_1)$ and $Z_2 = max(z_2)$, where z_1 and z_2 are the edges along the axial directions (see Fig. 3.4). Lastly, we combine the final predicted spans in the second and third stages by the Mask-RCNN to produce the 3D bounding box around the kidney (see Fig. 3.4).

3.2.4 Experimental Setup and Data Acquisition

We used 100 patients' CT scans from our VGH patient pool (discussed in section 2.1). Our data provided a total of 200 kidney samples (both left and right kidneys) among which we used 130 samples (from 65 randomly chosen patients) for training, 20 samples (from 10 randomly chosen patients) for validation, and the remaining 50 samples for testing. Our dataset included 12 pathological kidney samples (with endo- and exophytic tumors), and our training and test data contained six pathological cases, each. We also used 210 patients' CT scans from our KiTS patient pool (discussed in section 2.3). We used 160 randomly chosen patients' data for training, 15 randomly chosen patients' data for validation, and the remaining 35 patients data (70 kidney samples) for testing. We implemented the S-CNN with Caffe [107], and Mask-RCNN with the TensorFlow codes provided in [112]. Our training was performed on a workstation with Intel 4.0 GHz i7 processor, an Nvidia Titan Xp GPU with 12 GB of VRAM, and 32 GB of host memory.

3.2.5 Data Pre-processing

Before training the CNN, we did some basic pre-processing of the data. We programmed an automatic routine for 'abdominal' CT that separates the left and right kidneys. Since the left and right kidneys always fall in the separate half volumes,

Table 3.2: Comparison of mean kidney bounding wall localization error (mm).

Methods	**Short Name**	**Wall Error (mm)**	
		Left Kidney	Right Kidney
Criminisi et al. 2010 [58]	M-1	17.3 ± 16.5	18.5 ± 18.0
Cuingnet et al. 2012 [60]	M-2	7.00 ± 10.0	7.00 ± 6.00
Criminisi et al. 2013 [59]	M-3	13.6 ± 12.5	16.1 ± 15.5
Gauriau et al. 2015 [61]	M-4	5.5 ± 4.0	5.6 ± 3.0
Hussain et al. 2017 [1]	M-5	6.19 ± 6.02	5.86 ± 6.40
Samarakoon et al. 2017 [63]	M-6	11.52 ± 9.60	10.98 ± 9.60
Humpire et al. 2018 [66]	M-7	2.67 ± 7.18	3.03 ± 9.30
Xu et al. 2019 [25]	M-8	4.31 ± 4.18	3.89 ± 3.47
Proposed Method (on KiTS data)	-	2.06 ± 4.39	3.18 ± 14.02
Proposed Method (on VGH data)	-	$\mathbf{1.93 \pm 1.21}$	$\mathbf{2.45 \pm 1.75}$

the routine simply divided the abdominal CT volume medially along the left-right direction.

3.2.6 Validation on Kidney Data

We quantitatively compared the performance of our proposed kidney localization method with those reported in recent kidney localization approaches [1, 25, 58–61, 63, 66] in Table 3.2. Our bounding box has six walls, and for a particular kidney sample, we used the mean of the Euclidean distance errors between the estimated and ground-truth locations for all six walls. Please note that each method was independently implemented and tested on different CT databases (none of previous methods implementations were available as publicly shared code). However, the type of data these methods used is very similar to ours in terms of resolution, area scanned, and scan quality. Therefore, our comparisons are conservative, and rather than using our implementation of the other contrasting methods, we compare to each authors' best self-reported accuracy values. Here also, we use Eq. 3.1 to estimate mean kidney ROI boundary wall localization error.

Table 3.2 compares our kidney boundary localization performance to those of the methods proposed by Cuingnet et al. 2012 [60] (M-2), Criminisi et al. 2010 [58] (M-1) and Criminisi et al. 2013 [59] (M-3), Gauriau et al. 2015 [61] (M-

4), Hussain et al. 2017 [1] (M-5), Samarakoon et al. 2017 [63] (M-6), Humpire et al. 2018 [66] (M-7), and Xu et al. 2019 [25] (M-8). The M-2 method used the M-1 method for coarse localization of both left and right kidneys, then fine-tuned these locations using an additional RF per left/right kidney. Nonetheless, their resultant boundary localization errors remained higher than those of our proposed method. The M-3 method was an incremental work over the M-1, and both used RFs for various organ localization tasks. Both M-1 and M-2 methods exhibited higher boundary localization errors than those of the M-2 and proposed methods. The M-4 method used an extended cascade of RFs to estimate the confidence map of an organ, and the prediction was thresholded to obtain a final organ bounding box. The mean error by this method is worse than that of the proposed method. The M-5 method (our previous method discussed in section 3.1) used a deep CNN-based method for kidney 3D bounding box localization, which aggregates 2D orthogonal slice-based kidney candidacy decisions. This method also showed worse boundary localization performance than that of the proposed method. The M-6 method proposed a light RF consisting of less number of nodes than regular RF to localize different organs in the CT scans. However, its kidney bounding wall localization errors are too high. The closest performer to our proposed method was M-7, where the wall localization errors were comparably lower than those by other techniques. However, the standard deviation of the bounding wall estimation by this method is higher than that of the proposed method tested on the VGH data for both the left and right kidney. The M-8 method used a 3D region proposal network to detect eleven abdominal organs, including the left and right kidneys. This method showed the fourth-best performance after the M-7 method. Then we show the results by our proposed method on the KiTS data. Although KiTS datasets contain tumors in the kidney, our method performs better in mean kidney boundary wall localization than the other techniques. However, we observe a higher standard deviation for the right kidneys. It happened because some of the right kidneys in this dataset have large tumors in the upper pole, thus confusing boundary estimation. Finally, we can see in the Table 3.2 that the mean wall localization error for both the left and right kidneys are the lowest for the proposed method on the VGH data. Note that we also tested the performance of the proposed method by changing the image orientation and found almost no difference in the 3D bounding box localization

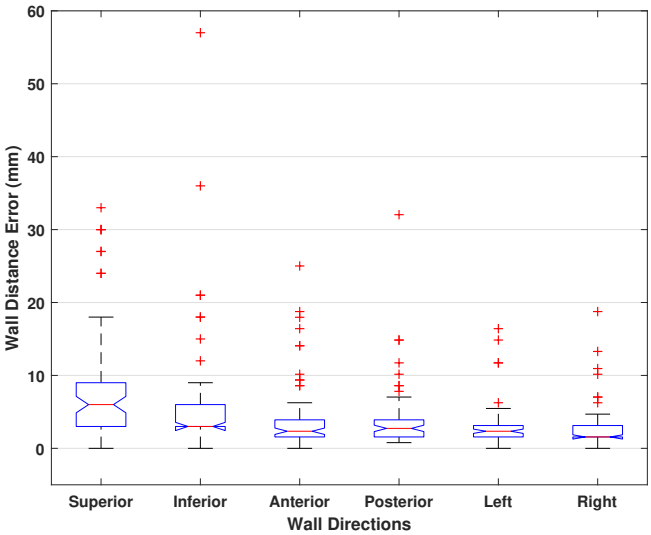

Figure 3.6: Box-plot of wall distance error (mm) per wall side of the kidney by the proposed method on the KiTS data.

accuracy.

In Fig. 3.6, we show the box plot of the wall distance errors (mm) by the proposed method in the superior, inferior, anterior, posterior, left, and right directions of a kidney in the KiTS dataset. This figure further supports the mean error reported in Table 3.2. We also see in this figure that the errors in the superior-inferior direction are comparatively higher than those in the anterior-posterior, and left-right directions. As we explained above, it happened because some of the right kidneys in this dataset have large tumors in the upper pole.

We also show the box plot of the wall distance errors (mm) for kidneys in the VGH data in Fig. 3.7. This figure also supports the mean error reported in Table 3.2. Here also, we see that the errors in the superior-inferior direction are comparatively higher than those in the anterior-posterior and left-right directions. In this case, the possible explanation could be that the slice thickness is higher in the axial direction than in the coronal and sagittal directions.

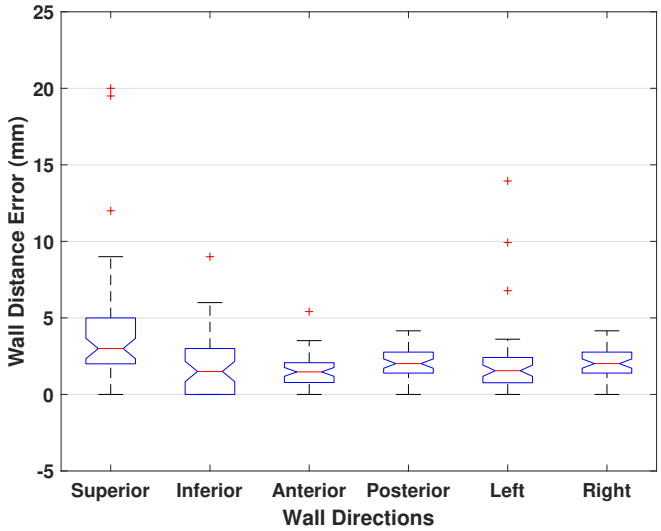

Figure 3.7: Box-plot of wall distance error (mm) per wall side of the kidney by the proposed method on the VGH data.

3.2.7 Discussion

In this section, we discussed a deep learning approach for kidney localization. This method enabled a clinical approach for kidney localization in the raw abdominal CT data. Our contribution comprises a novel 3-step CNN-based architecture that reduces false positives in organ bounding boxes of the targeted organ. Our Mask-RCNN in the second stage operates strictly on our S-CNN selected slices. Similarly, the Mask-RCNN in the third stage operates strictly on the sagittal slices falling inside the span estimated in the second stage. As a result, the mean bounding box error proved to be very low compared to current methods. Our experimental results demonstrated a 23% increase in kidney boundary localization accuracy compared to those of recent literature.

3.3 Summary

In this chapter, we discussed two deep learning approaches for human kidney localization in the volumetric medical images. The first method, discussed in section 3.1, was one of the successful deep learning approaches in the literature for

both kidney localization. That method aggregates 2D orthogonal slice-based kidney candidacy decisions. Later, we developed another deep learning approach that further improves the automatic kidney localization performance than that discussed in section 3.1. We discussed this method in section 3.2 that uses an effective CNN-guided Mask-RCNN approach for efficient kidney localization in the volumetric CT images. On the VGH patient data, the CNN-guided Mask-RCNN approach showed the lowest mean kidney ROI boundary localization error. This error is also the lowest among other comparing methods, thought the state-of-the-art methods were validated on different datasets. Overall our proposed methods showed robust kidney ROI localization performance; however, it sometimes produces higher boundary localization error in the kidney superior wall (see Fig. 3.6). We empirically observed that this higher boundary localization error occurs when there is a large tumor in the upper pole of the kidney.

Chapter 4

Segmentation-free Kidney Volume Estimation

4.1 Dual-regression Forests for Volume Estimation

Accurate estimation of kidney volume is essential for clinical diagnosis and therapeutic decisions related to different chronic kidney diseases (CKD). The abnormal volume of a kidney often related to the presence of a tumor or cancer in it [30]. Existing kidney volume estimation methods rely on an intermediate computationally expensive segmentation step. In this chapter, we discuss a novel method for direct estimation of kidney volumes for 3D CT images with dual regression forests that skipped the segmentation step. We originally published this work in Hussain et al. [2]. After the determination of kidney locations by using any of the methods in chapter 3 within the 3D abdominal CT images, our method used dual regression forests, one for predicting the anatomical area in a particular image plane, and another one for boosting the results by removing outliers from the initially estimated areas. We adopted a smaller subpatch-based approach to increase the number of observations, which ultimately improve the results.

This novel segmentation-free kidney volume estimation technique is divided into three steps as shown in Fig. 4.1. In section 4.1.1, we discuss the 2D image patch representation. Then, in section 4.1.2, we discuss the training of regression forests and the subsequent prediction of kidney areas. Finally, in section 4.1.3,

we discuss the estimation of kidney volumes based on the predictions by the dual regression forests.

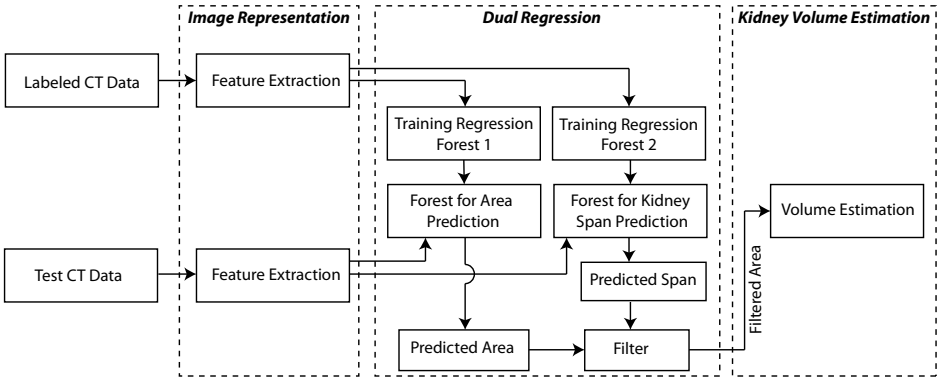

Figure 4.1: Flowchart showing different components of the proposed method.

4.1.1 Subpatch Image Representation

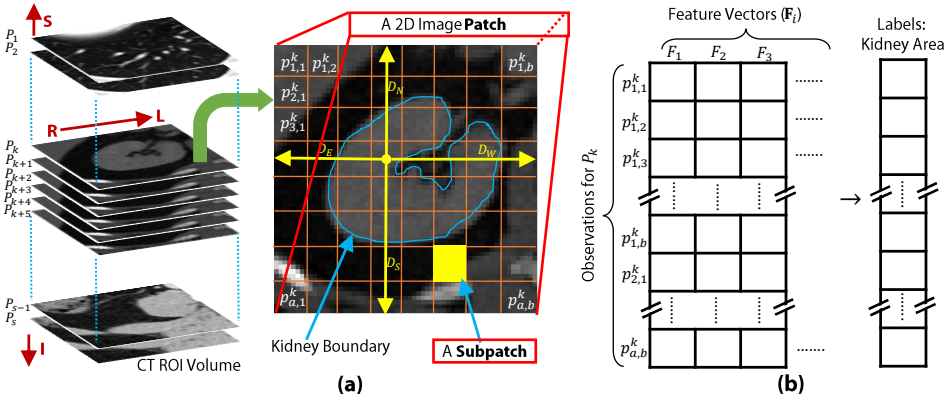

Figure 4.2: Illustration of (a) the representation of our 2D image patch containing kidney, and (b) the formation of feature vectors from its sub-patches.

We divide each image patch into square subpatches (Fig. 4.2(a)). Then, to obtain the prediction of the kidney area for each of the sub-patches, we train a regression forest with these sub-patches as observations. We use various features

$\mathbf{F}_i$ for each subpatch p: (1) the sum of image intensities $\sum_p I$; (2) sum of non-overlapped binned intensities $\sum_p I_b$, where b stands for different bin numbers and $\min(I) \leq I_b \leq \max(I)$; (3) entropy $E = -\sum h \times log_2(h)$, where h are the histogram counts of I; (4) sum of image intensity ranges $\sum_p R$, which is (max value - min value) in a 3×3 pixel neighborhood around the corresponding pixel; (5) sum of standard deviations $\sum_p SD$, where SD is estimated in a 3×3 neighborhood pixels around the corresponding pixel; and (6) axially aligned distances D_E, D_W, D_N and D_S of the interrogated subpatch center from the east, west, north and south boundaries of the 2D image patch, respectively (Fig. 4.2(a)). Features (3)-(5) capture the texture information in a subpatch.

4.1.2 Dual-regression Forests for Kidney Area Prediction

Regression forest 1 (Fig. 4.1) learns the correspondence between input features and kidney areas for training subpatches, and then predicts organ areas in unseen subpatches. For feature matrix $v = (\mathbf{F}_1, \mathbf{F}_2, \ldots, \mathbf{F}_d)$, where $\mathbf{F}_i$ is a feature vector (Fig. 4.2(b)) and d is the total number of features, forest 1 learns to associate observations $\mathbf{F}_i(r,s)$, $(i = 1,..,d)$ with a continuous scalar value $y^k(r,s)$ which is the estimated organ area in the corresponding subpatch $p^k_{r,s}$. Here, k is the patch index, and r and s are the subpatch indices along the posterior-anterior (P-A) and right-left (R-L) directions, respectively. The distribution of estimated kidney area values $D(\tilde{y})$ *vs.* subpatches for an kidney sample (kidney) is shown in Fig. 4.3(b). However, due to extensive variation in kidney shapes, sizes and orientations across subjects, we observed that non-zero volumes are predicted for areas devoid of kidney tissue (Figs. 4.3(a) and (b)). These false positives are removed using a spatial filter (Fig. 4.3(c)) having an extent (or bandwidth) equal to a spatial kidney span measure (along superior-inferior direction). This important span parameter is learned by forest 2. For training forest 2, we rearrange (i.e., negligible extra computations) the feature vectors as $\hat{\mathbf{F}}^k_i = \sum_{m=1}^{a \times b} \mathbf{F}^k_i(m)$, where a and b are the total number of subpatches along the P-A and R-L directions, respectively. We define a unit step function $U(\tilde{u})$ whose spatial bandwidth is equivalent to the span $\tilde{u}$ predicted by forest 2 for a particular kidney sample (Fig. 4.3(c)). We approximate the most probable kidney span in the false positives-corrupted D by calculating the

cross-correlation between D and U defined as $\rho(l) = \sum_{q=1}^{Q} D(q) \cdot U(q+l)$, where Q is the total number of subpatches in an investigated ROI containing kidney. The lag corresponding to the maximum of $\rho(l)$, $l_{max} = \text{argmax}_l\{\rho(l)\}$ is then used to align U with D. Finally, an element-wise multiplication $D \cdot U$ generates the filtered area distribution (D_f), where almost all of the false positives are removed (Fig. 4.3(d)). Note that although we use subpatches, we are not labeling every pixel, as done for classification-based segmentation in [60, 84, 113], but rather inferring a scalar area for every subpatch.

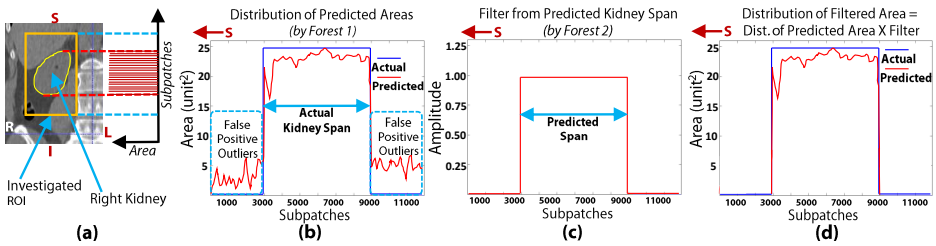

Figure 4.3: (a) A schematic diagram showing an example investigated ROI and its most likely kidney-area *vs.* subpatches distribution. (b) A typical distribution of predicted kidney-area *vs.* subpatches (red), overlaid on the actual kidney-area *vs.* subpatches (deep blue). Predicted areas include false-positive outliers as shown with the light-blue dashed-boxes. (c) An example plot of a predicted kidney span. (d) The final distribution of the filtered kidney-area *vs.* subpatches, overlaid on the actual kidney-area *vs.* subpatches where most of the outliers are removed.

4.1.3 3D Volume Estimation from 2D Area Estimates

Some subpatches completely lie inside the organ cross-section, and we expect the predicted kidney areas for those subpatches to be the maximum, S_A (area of a subpatch). However, we observed that almost no predicted-subpatch-area (by forest 1) reaches this obvious maximum value of S_A. On the other hand, there are few false positives still left inside the filtered area distribution D_f. So, we choose an empirical threshold g and fine-tune D_f as: $D_f(p) = 0$, if $D_f(p) < g$, and $D_f(p) = S_A$, if $S_A - D_f(p) < g$. Finally, we estimate the volume of a kidney by integrating the areas in D_f in the axial direction.

Table 4.1: A comparison of volume estimation accuracies, estimation speeds, and requirements of extra-time for parameter optimization during the execution for different types of methods. Execution time is the MATLAB run-time on Intel Xeon CPU E3 @ 3.20GHz with 16 GB RAM.

	Method Types	**Methods**	**Per Sample Run-time**	**Mean Volume Error (%)**
1	Seg	Intensity Threshold	1.10±0.06	67.57±114.10
2		3D Active Contour [114]	70.10±10.19	55.91±98.17
3	Manual	Ellipsoid Fit [115]	∼180	14.20±13.56
4	Seg-free	Single Reg. + 2D Patch [116]	1.57±0.07	36.14±20.86
5		Single Reg. + 2D Subpatch	1.88±0.12	16.88±10.82
6		Single Reg. + 3D Subpatch	1.02±0.04	26.88±20.11
7		**Dual Regression**	3.75±0.23	**9.97±8.69**

4.1.4 Experimental Setup and Data Acquisition

We used abdominal CT images of 45 patients from our VGH patient pool (discussed in section 2.1). We used a total of 90 kidney samples (both left and right kidneys), among which we used 46 samples (from 23 randomly chosen patients) for training, and the rest is unseen. We calculated the ground truth kidney volumes (referred to as 'actual volumes') from kidney delineations performed by expert radiologists. We used a leave-one-kidney-sample-out cross-validation approach on the training set only to choose suitable tree and leaf sizes. We use a small size of 5×5 pixels for a subpatch so that we have more training samples. Also, the smaller subpatch helps to incur smaller-valued false positives in D_f, which can be easily removed by simple thresholding with g. We also use $g = S_A/5$ throughout the paper. Nowadays, the available public CT data are acquired with intensity-standardized CT machines, where intensity for the same organ does not vary significantly across CT machines manufactured by different companies. Also, kidney intensity typically falls within the range of [-150, 300]HU. Therefore, we expect that our choice of g value would work for other CT datasets.

4.1.5 Validation on Kidney Data

We provide comparative results of our proposed method with those obtained by four generic approaches: two segmentation algorithms, a naïve manual method, and three forests-based approaches. But first, we show the performance of the

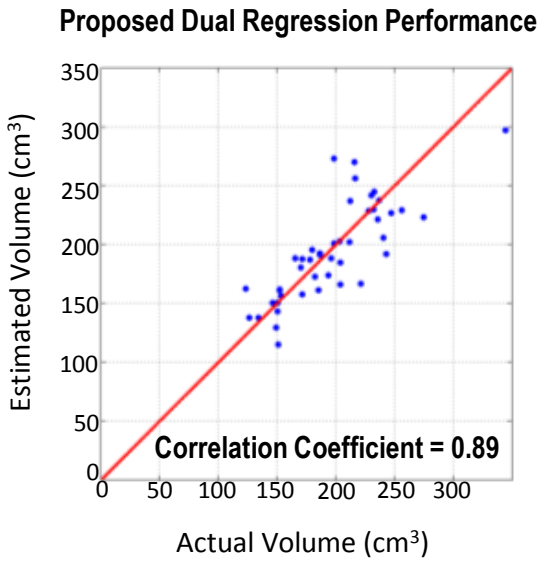

Figure 4.4: Scatter plot showing the volume correlations between the actual and proposed dual regression-based estimates.

proposed method visually in Fig. 4.4, where we illustrate the correlation between the actual and estimated kidney volumes. This figure shows that, aside from few exceptions, almost all of the estimates are close to their corresponding ground truth measurements.

We also show the performance comparison of the execution time, volume estimation accuracy, and extra-time requirements for parameter optimization for different methods in Table 4.1. We see in Table 4.1: rows 1 & 2 that we had to use extra-time for kidney-sample-wise parameter optimization for both segmentation-based approaches. It is sometimes possible to find optimal settings of parameters for energy-minimizing segmentation methods via cross-validation. However, the pursuit of optimal parameters is computationally expensive and near infeasible. We also see in row 1 that the estimated mean volume error for the intensity thresholding-based method is the highest. It cannot differentiate between two different organs if the intensities associated with these organs fall inside the same user-defined or automatically chosen range. On the other hand, the 3D active contours-based method [114] produces a kidney surface that leaks through the

weaker boundaries, even with the best empirical parameter configuration. As a result, the mean volume error performance is poor (Table 4.1: row 2). Moreover, it is time inefficient, as well.

Then we consider a manual approach that is typically used by the radiologists in the clinical settings. They obtain three principal axes on a kidney, which correspond to a 3D ellipsoid that approximates that particular kidney. In Table 4.1: row 3, we see that the estimated mean volume error (computed by expert radiologists) for this approach is approximately 15% with high standard deviations. Besides, it takes around 3 minutes per kidney sample.

Finally, we consider four segmentation-free approaches using regression forests. The first approach [116] uses a single forest + 2D patch, and the corresponding mean volume error performance is poor, as seen in row 4. This approach works well for cardiac bi-ventricles but fails for a kidney since kidney sizes, shapes, and orientations vary more extensively across subjects. Subsequently, we adopt an efficient approach to learning using image subpatches (5×5 pixels). This subpatch-based approach improves the mean volume error performance than that of the patch-based method (see rows 4 & 5). However, false-positive estimates still corrupt these subpatch-based results. We also tested using 3D subpatches ($5\times5\times2$ voxels). Since CT axial resolution is lower than those of the coronal and sagittal, $5\times5\times2$ closely resembles a cube shape. However, we see in row 6 that the corresponding mean volume error is worse than those of 2D subpatches (row 5). We suspect that this poor performance may be a result of the reduced number of training samples. The proposed method (dual regression+2D subpatch) combines the 2D subpatch-based area prediction and patch-based kidney span prediction, which ultimately results in the best mean volume error performance. While the mean accuracy of the forest 2-based kidney span prediction is approximately 95.5% alone, the Table (row 7) depicts that the mean volume error by the proposed method falls below 10%, with the cost of a prediction time of $\sim$4 sec per kidney sample, which we can further accelerate via a GPU-implementation. Besides, we performed the Student t-test between the actual and estimated volumes, and the estimated p value is 0.8170, which fails to reject the Null hypothesis. Therefore, the ground truth and estimated kidney volumes do not statistically differ.

4.1.6 Discussion

In this section, we discussed an effective method for segmentation-free estimation of kidney volumes from 3D CT images. We formulated our volume estimation problem as a 2D subpatch learning-based regression problem and were able to skip the problematic segmentation step. Though kidney shapes, sizes, and orientations vary extensively across subjects, we addressed this challenge by adopting a dual regression forest formulation. We train it by making use of the same extracted image features, and their combined predictions resulted in satisfactory kidney volume estimates. Our experimental results showed that the proposed method could estimate kidney volumes with high correlations of 89% with those obtained manually by expert radiologists and reported the mean volume estimation error of 10%. However, this approach used hand-engineered features that may be difficult to design in the clinical environment optimally. An alternate solution could be using a CNN approach that would learn features automatically from subpatches. However, the features used in this dual-regression approach are statistical in nature. On the other hand, CNN relies mostly on non-statistical object appearance features [43]. Therefore, a conventional CNN would fail to learn anything from a 5×5 subpatch, and our experiment supported our hypothesis. That is why we omitted to report any CNN-based performance on subpatches.

4.2 Deep Supervised Learning for Volume Estimation

In the previous section, we discussed a dual regression forests-based segmentation-free kidney volume estimation approach using hand-engineered features. However, hand-engineered features are often difficult to design in the clinical environment optimally. Therefore, in this section, we discuss a couple of deep supervised learning approaches for segmentation-free kidney volume estimation. We published the first part of this work (discussed in section 4.2.1) in Hussain et al. [1], which uses a deep CNN to predict slice-based cross-sectional kidney areas followed by integration over these values across axial kidney span to produce the volume estimate. The second approach discussed in section 4.2.2 uses a deep FCN instead of deep CNN to predict more accurate slice-based cross-sectional kidney areas than that in section 4.2.1. In our both approaches, discussed in section 4.2.1 and section 4.2.2, we tried to choose the simplest networks possible to predict accurate kidney cross-sectional areas. We increased the depth of the networks gradually as well as chose the filter sizes and strides iteratively to improve the training and validation accuracy during network training.

4.2.1 Regression CNN for Volume Estimation

In chapter 3, we estimated the kidney encompassing tight ROI. Typically, kidney shape and appearance vary across patients (Fig. 3.2). Training our CNN requires 2D image patches of consistent size. Also, the patch size needs to be universal so that it always contains the kidney cross-sections. Therefore, to generate training data, we choose a patch size of 120×120 pixel, making sure that the cross-section of the initially estimated kidney ROI is at the center of it. We also ensure that there is enough free space around a kidney cross-section. The ratio between the number of pixels fall inside a kidney cross-section to the total number of pixels in the image patch (120×120 pixel) is considered as the output variable (label) for that particular image patch.

We estimate the cross-sectional area of a kidney in each slice using a deep CNN shown in Fig. 4.5. The CNN performs regression and has seven layers, excluding the input. It has four convolutional layers, three fully connected layers, and one Euclidean loss layer. We also use dropout layers along with the first two fully

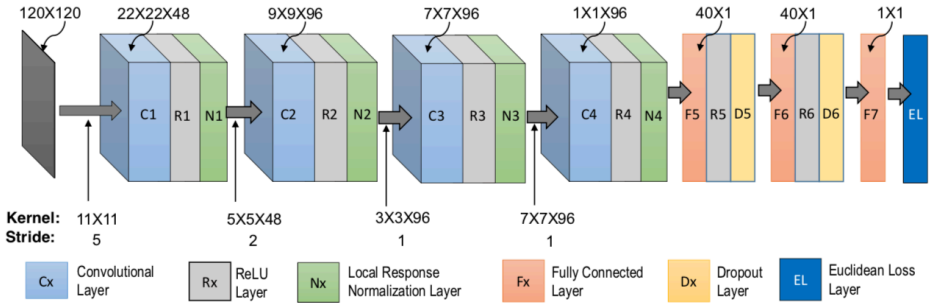

Figure 4.5: Segmentation-free kidney volume estimation using deep CNN.

connected layers to avoid over-fitting. As mentioned earlier, the input is a 120×120 pixel image patch, and the output is the ratio of kidney pixels to the total image size. We train the CNN by minimizing the Euclidean loss between the desired and predicted values. After training the CNN model, we deploy the model to predict the kidney area in a particular image patch. Finally, we estimate the volume of a specific kidney by integrating the predicted areas in all of its image patches in the axial direction.

4.2.2 Regression FCN for Volume Estimation

To generate training data for the FCN shown in Fig. 4.6, we choose a patch size of 128×128 pixel. The ratio between the number of pixels fall inside a kidney cross-section to the total number of pixels in the image patch (128×128 pixel) is considered as the output variable (label) for that particular image patch.

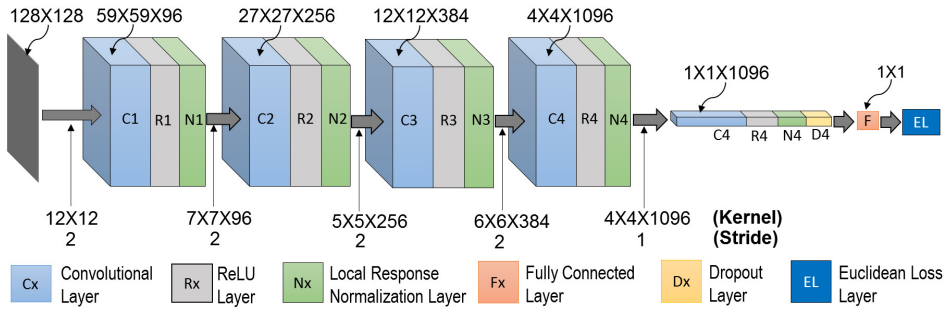

Figure 4.6: Segmentation-free kidney area estimation using deep FCN.

We estimate the cross-sectional area of a kidney in each slice using an FCN shown in Fig. 4.6. The FCN performs regression and has six layers, excluding the input. It has five convolutional layers, one fully connected layer (only to generate single activation), and one Euclidean loss layer. We also use the dropout layer with the last convolution layer (containing flatten activation) to avoid over-fitting. As mentioned earlier, the input is a 128×128 pixel image patch, and the output is the ratio of kidney pixels to the total image size. We train the FCN by minimizing the Euclidean loss between the desired and predicted values. After training the FCN model, we deploy the model to predict the kidney area in a particular image patch. Finally, we estimate the volume of a specific kidney by integrating the predicted areas in all of its image patches in the axial direction.

4.2.3 Experimental Setup and Data Acquisition

We used 100 patients' CT scans from our VGH patient pool (discussed in section 2.1). We were able to use a total of 200 kidney samples (both left and right kidneys) from which we used 120 samples (from 60 randomly chosen patients) were used for training, 20 samples from 10 randomly chosen patients for validation, and the rest is unseen. Our dataset included 12 pathological kidney samples (with endo- and exophytic tumors), and our training and test data contained 6 cases, each. We also used 210 patients' CT scans from our KiTS patient pool (discussed in section 2.3). We used 160 randomly chosen patients' data for training, 15 randomly chosen patients' data for validation, and the remaining 35 patients data (70 kidney samples) for testing. We implemented both neural networks using *Caffe* [107]. We set the base learning rate for training to 0.01 and decreased by a factor of 0.1 to 0.0001 over 15,000 iterations. Since we performed experiments in sections 4.2.1 and 4.2.2 separately, we discuss their results separately in sections 4.2.4 and 4.2.5, respectively.

4.2.4 Results Comparison to CNN-based Approach

Table 4.2 shows quantitative comparative results of our direct volume estimation module (including the localization step) with those obtained by a manual ellipsoid fitting method, two segmentation-based methods, and two segmentation-free

regression-forest-based methods. We use the mean volume errors by [115, 116] reported in [2] for comparison. For the manual approach [115], we see in Table 4.2 that the estimated mean volume error for this approach is approximately 14% with high standard deviation. Then we consider two segmentation-based methods [60, 117]. These methods reported their volume estimation accuracy in terms of the Dice similarity coefficient (DSC), which does not relate linearly to the percentage of volume error. Since segmentation-free methods do not perform any voxel classification, DSC cannot be calculated for these methods. Therefore, it is difficult to directly compare DSC performance to the percentage of volume estimation error. However, [60] used 2 regression forests (RF), an ellipsoid fitting and subsequent template deformation for kidney segmentation. Even then, authors in [60] admitted that this method did not correctly detect/segment about 20% of left and 20% of right kidneys (DSC < 0.9, non-correctness criterion in [60]), and failed on about 10% left and 10% right kidneys (DSC < 0.65, failing criterion in [60]). But our method successfully estimated volumes for all our kidney samples and achieved a mean volume error of 7%. Moreover, authors in [60] mentioned that the RF-based voxel classification was uncertain and the subsequent deformation step relies on the initial kidney shape. Due to this drawback, [60] is likely to fail on pathological kidneys. Crucially, authors in [60] did not include the truncated (tumor removal during partial nephrectomy) kidneys (16% of their data) in their evaluation. For a similar reason, the multi-atlas image registration-based method [117] was evaluated only on 22 kidney samples out of 28, because 6 samples contained tumors. In addition, the test dataset size in [117] was very small (22 *vs.* 60 in our dataset). In contrast, our dataset includes 12 pathological kidney samples, where our training and test data contained 6 cases, each. As the results show, our method does not fail for any kidney and thus, suggests it is less sensitive to kidney truncation or tumors. Finally, we consider two RF-based segmentation-free kidney volume estimation approaches [2, 116]. For the single RF-based method [116], we see that the corresponding volume estimation error is worse than those of the dual RF and proposed methods as seen in Table 4.2. Using smaller 2D patches and the dual RF, [2] outperformed [116] but still shows a higher volume estimation error than the proposed method likely due to the use of non-optimal hand-engineered features. The last row of Table 4.2 reports the mean error of our proposed method

Table 4.2: Volume estimation accuracies compared to state-of-the-art methods. Not reported values are shown with (-).

Method Types	Methods	Kidney Samples	Mean Volume Error (%)	Mean Dice Index
Manual	Ellipsoid Fit [115]	44	14.20 ± 13.56	-
Seg	RF+Template [60]	358	-	0.752 ± 0.222 †
	Atlas-based [117]	22	-	0.952 ± 0.018
Seg-free	Single RF [116]	44	36.14 ± 20.86	-
	Dual RF [2]	44	9.97 ± 8.69	-
	Proposed Method	**60**	**7.16±8.91**	-

† Estimated from reported Dice quartile values (in [60]) using the method in [118].

around 7%, which is the lowest among all three segmentation-free methods.

4.2.5 Results Comparison to FCN-based Approach

We provide comparative results of our proposed method with those obtained by three generic approaches: a manual clinical method, two regression forest-based approaches, and three deep learning approaches. But first, we show the performance of the proposed method visually in Figs. 4.7 and 4.8, where we illustrate the correlation between the actual and estimated kidney volumes for the VGH and KiTS data, respectively. These figures show that almost all of the estimates are close to their corresponding ground truth measurements.

We also show the quantitative comparative results of our segmentation-free volume estimation approach in Table 4.3. Note that most of the state-of-the-art kidney volume estimation approaches are segmentation-based, and they report their accuracy in terms of the Dice similarity coefficient (DSC), which does not relate linearly to the percentage of volume error. Since segmentation-free methods do not perform any voxel classification, we cannot calculate DSC for these methods. Therefore, it is difficult to directly compare DSC performance to the percentage of volume estimation error. Therefore, we emphasize the performance comparison mostly concerning the state-of-the-art segmentation-free methods. First, we consider a manual approach [115], which is typically used by the radiologists in the clinical settings. The experts obtain three major axes on a kidney, which correspond to a 3D ellipsoid that approximates that particular kidney. In Table 4.3: row

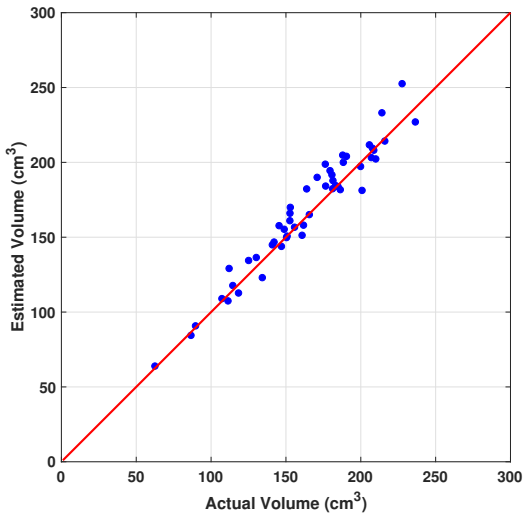

Figure 4.7: Scatter plot showing the volume correlations between the actual and proposed FCN-based estimates for the VGH data. Correlation coefficient = 0.9714.

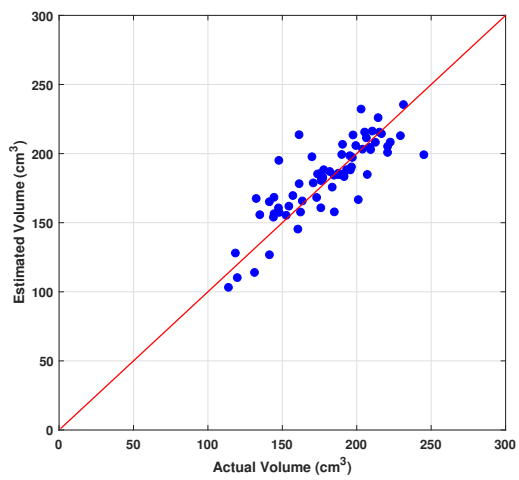

Figure 4.8: Scatter plot showing the volume correlations between the actual and proposed FCN-based estimates for the KiTS data. Correlation coefficient = 0.9645.

1, we see that the estimated mean volume error (computed by expert radiologists) for this approach is approximately 15% with high standard deviations. Also, it takes around 3 min per kidney sample. Next, we consider two regression forest-based approaches [2, 116] used for segmentation-free kidney volume estimation. The method in [116] used a single regression forest, and the corresponding volume estimation error is the worst among the comparing methods (see Table 4.3: row 2). On the other hand, using dual regression forests, our initial work [2] shows better volume estimation accuracy than that by [116] (see Table 4.3: row 3). Later, we used a CNN-based approach [1] for segmentation-free kidney volume estimation. It showed better volume estimation accuracy as CNN better captured the rich and complex variability in the kidney anatomy and outperformed the hand-engineered feature representations in [2, 116] (see Table 4.3: row 4). In this work, we further improve the volume estimation accuracy using a comparatively deeper network than that in [1]. Besides, this network utilized fully convolution layers except for the layer before the loss layer to accumulate the network activation as a single value. We see that the FCN approach shows the best volume estimation performance among all the methods on the VGH data (see Table 4.3: row 5). To check the performance of a network with a fully connected layer instead of a convolution layer of a similar dimension, we replaced the final convolution layer (of size $1 \times 1 \times 1096$) with a fully connected layer of equal size (i.e., 1096×1). This change makes the FCN a CNN, which predicts worse kidney cross-sectional area estimates than that by the FCN (see Table 4.3: row 6). We infer that the FCN performs better than CNN because of the better feature correspondence among convolution layers, i.e., preservation of the spatial context throughout the network. In contrast, a fully connecting layer typically learns a completely new set of weights based on the activation of the previous layer. Finally, we show the volume estimation performance of the proposed FCN approach on the KiTS data in Table 4.3: row 7. Since almost all the kidney samples in this dataset contain tumors of various sizes and shapes, the volume estimation error is slightly higher than that for the VGH data (row 5).

Since we cannot estimate the DSC for the segmentation-free volume estimates, we visually demonstrate the comparison of the mean distribution of the ground truth and estimated kidney cross-sectional area in Figs. 4.9 and 4.10 for the VGH and KiTS data, respectively. Our FCN predicts the ratio between the kidney cross-

Table 4.3: Volume estimation accuracy compared to state-of-the-art competing methods.

Method Type	Methods	Test Samples	Mean Volume Error (%)
Manual Ellipsoid Fitting	Zakhari et al. 2014 [115]	44	14.20 ± 13.56
Regression Forest	Zhen et al. 2014 [116]	44	36.14 ± 20.86
(Seg-free)	Hussain et al. 2016 [2]	44	9.97 ± 8.69
Deep Learning	Hussain et al. 2017 [1]	60	8.05 ± 8.91
(Seg-free)	**Proposed with FCN** (VGH Data)	60	**4.80±3.89**
	Proposed with FC Layer (VGH Data)	60	5.92 ± 4.50
	Proposed with FCN (KiTS Data)	70	7.26 ± 6.80

sectional area and the 2D ROI area. So, in Figs. 4.9 and 4.10, we plot the mean kidney area to the ROI area ratio for all the test kidney samples along the axial direction. Since kidney span along the axial direction varies across kidneys, we re-sample all the kidney spans to 25 slices to make those consistent across all the samples. We can see in Figs. 4.9 and 4.10 that the mean and standard deviations of the kidney area to ROI area ratio by the proposed method follows the same trend as the ground truth. Besides, we performed the Student t-test on both samples, and the estimated p values are 0.775 and 0.6442 for the VGH and KiTS data, respectively. These statistical tests fail to reject the Null hypothesis. Therefore, the ground truth and estimated kidney area to ROI area ratio do not statistically differ.

Our proposed FCN for segmentation-free kidney volume estimation is also very light in terms of the number of trainable parameters ($\sim$94,000). In contrast, one of the recent and popular segmentation-based organ volume estimation approach, 3D U-Net [87], has $\sim$19,070,000 trainable parameters, which is approximately 200x more than that of our proposed FCN approach.

4.2.6 Discussion

In this section, we discussed two deep learning approaches for segmentation-free kidney volume estimation from the raw abdominal CT data. We formulated our volume estimation problem as a 2D image patch-based regression problem and were able to skip the often problematic segmentation step. Our deep neural net-

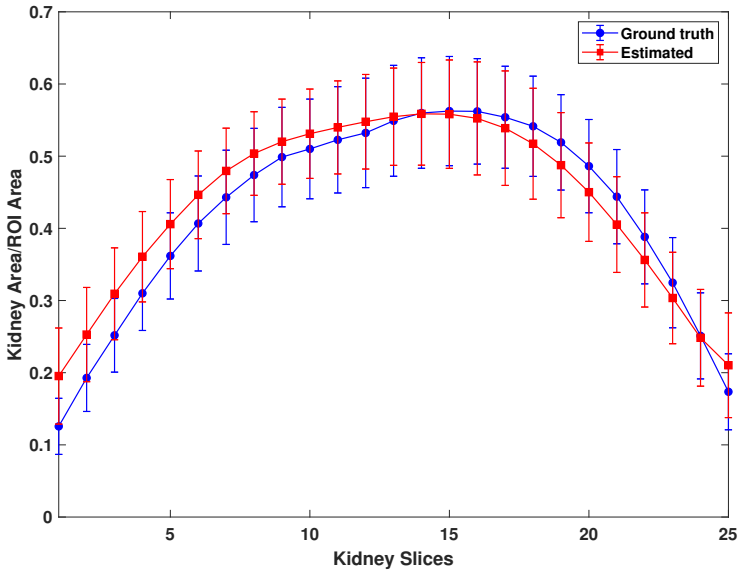

Figure 4.9: Distribution of kidney cross-sectional areas for the VGH data along the axial direction.

works better captured the rich and complex variability in kidney anatomy and outperformed the hand-engineered feature representations used in [2, 59, 60]. Further, we showed that an FCN performs better than a CNN of a similar size in a regression problem. Our experimental results demonstrated a 40% increase in volume estimation accuracy compared to those of the recent literature.

4.3 Summary

In this chapter, we discussed three ML-based methods for segmentation-free kidney volume estimation, which bypasses the intermediate segmentation approach. All of these approaches estimate the 3D kidney volume via estimating 2D slice-based kidney cross-sectional area estimation. The use of 2D slices enables using more training data, which in turn makes the supervised model more robust on the test data. Besides, all of these methods are very light in terms of the number of trainable parameters. In fact, our most successful segmentation-free volume estimation approach (discussed in section 4.2.2), i.e., FCN-based approach, has

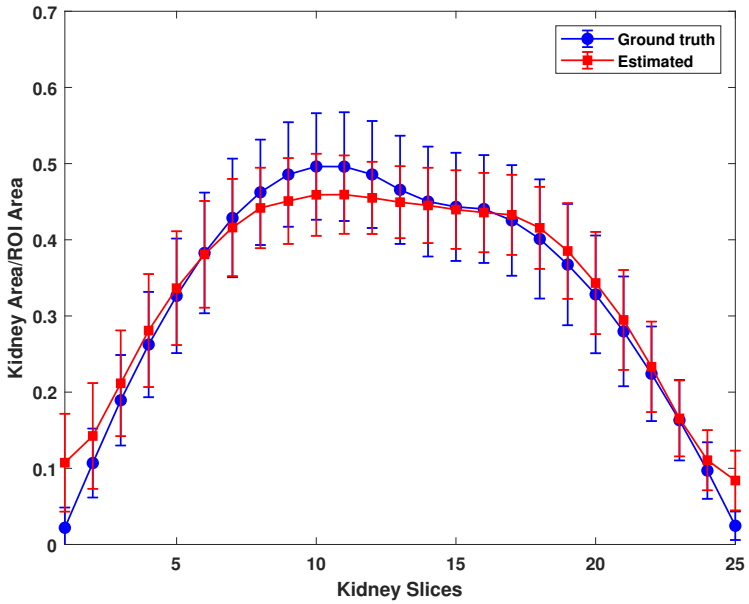

Figure 4.10: Distribution of kidney cross-sectional areas for the KiTS data along the axial direction.

around 200x less number of trainable parameters than one of the state-of-the-art segmentation approach 3D U-Net [87]. Thus our methods are more suitable for using a clinical environment where substantial computational resources are often not available. However, our segmentation-free volume estimation approach tends to perform worse when a kidney has a tumor(s). We also observe this phenomenon from the performance of the FCN approach on the KiTS data shown in Table 4.3. Therefore, more training data with pathological kidneys are required to make our model more robust.

Chapter 5

Collage CNN for Pathological Kidney Detection in CT

Tissue abnormalities such as tumors, cancers, nodules, stones, etc. are most often localized within a small region of anatomy and do not span the whole image volume. Localization and analysis of abnormal tissue are thus typically carried out on the 2D image slices. For example, the staging of kidney tumors often requires slice-based tumor analysis and manual boundary tracing. However, assignments of image tags or labels (e.g., healthy, cancerous, etc.) are mostly per image volume or per-patient basis. Therefore, all slices of an image are by default labeled with a single tag, though not all slices may contain the abnormal tissue. This scenario makes 'single-instance' ML approaches, especially deep learning ones such as CNN, challenging to train on the 2D slices, as the input slice often does not correspond to the assigned volume-based label. In this chapter, we discuss a CNN based kidney classification method that makes use of a novel collage image representation. The image slices in a 3D volume are rearranged side-by-side into a virtual extended 2D image slice, which in turn correctly corresponds to the single available label for that dataset. Our proposed collage also allowed for data augmentation by a random reshuffling of the locations of axial image slices within the collage. We originally published this work in Hussain et al. [3].

5.1 Collage Representation of 3D Image Data

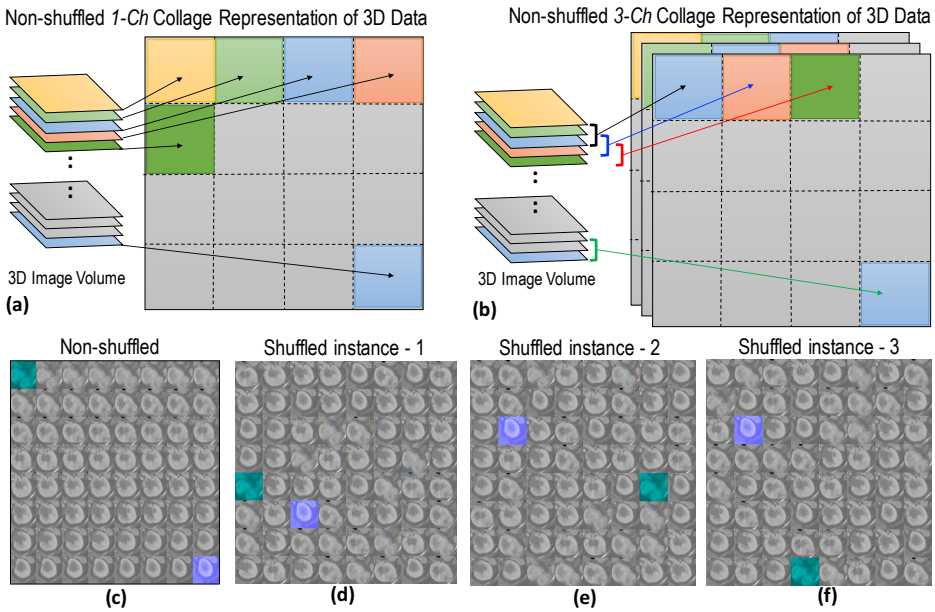

Figure 5.1: Schematic diagrams showing the non-shuffled (a) 1-channel and (b) 3-channels 2D collage representations of a 3D image volume. (c) An example 1-channel 2D collage image slice (512×512 pixels) containing 64 individual (non-shuffled) axial slices (64×64 pixels) of an actual kidney CT volume. The axially top and bottom slices (two corner slices in (c)) are colored to locate those in the randomly shuffled collages in (d)-(f).

Typically, tumors grow in different regions of the kidney and are clinically scored based on their CT slice-based image features. For example, tumor size, margin, composition, necrosis, growth pattern (endophytic or exophytic), calcification, etc., [36]. Of course, not all kidney slices necessarily contain tumors. Nonetheless, clinical labels (healthy/pathological) are normally recorded on a kidney- or a patient-basis. Therefore, it is not possible to use slice-based inputs in the training of a CNN because the volume-based label does not apply to all constituent axial slices. To address this challenge, we propose a novel approach where we rearranged the slices within the 3D image into an extended 2D image collage (Fig. 5.1). In a non-shuffled collage representation, each consecutive image slice (for 1-channel) or, a

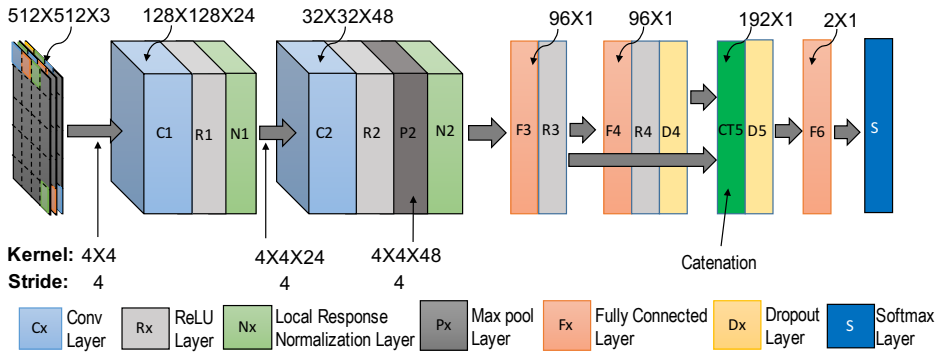

Figure 5.2: The architecture of our collage deep convolutional neural network for pathological *vs.* healthy kidney classification. See Fig. 5.1 for the input image representation.

group of n consecutive image slices (for n-channels, where $n > 1$) along a particular direction are sequentially placed on a 2D plane, which is schematically shown for a 1-channel and a 3-channels ($n = 3$) image in Fig. 5.1(a) and (b), respectively. Note that we opt to keep the collage dimension square (i.e., 512×512 pixels) in this experiment; however, it is not a necessity. This collage not only ensures meaningful correspondence to the volume's single label but also allows for invaluable data augmentation by a simple random reshuffling of image slices as well as by rotation and flipping. We show a non-shuffled 2D image collage representing an actual kidney CT data and its shuffle-based three augmented collages in Fig. 5.1(c) and (d)-(f), respectively. Our collage representation enables CNN to look at all the image slices of the 3D volume at once. At the same time, the volume-based label corresponds correctly to that particular collage. Shuffling-based augmentation further enables multiplying the number of training data and also introduces variations in the input collage. Our collage CNN can efficiently learn the discriminatory features from all the constituent shuffled and non-shuffled images in the collage without requiring any adjacency information preserved in the longitudinal direction. In a crucial case of keeping longitudinal adjacency information, we can increase the value of n in the collage.

Note that we prepare our CNN input data in a process shown in Fig. 5.1(b), where we set $n = 3$. The resulting dimension of a single CNN input data is

512×512×3 pixels, and the output was either 0 (healthy) or 1 (pathological).

5.2 Collage CNN for Kidney Classification

Our proposed CNN has seven layers excluding the input. All of these layers, except the 5th layer (concatenation layer), contained trainable weights.

Layer C1 is a convolutional layer that filters the input image with 24 kernels of size 4×4×3. Since we used collage-based image representation, we needed to carefully design our filter sizes and strides in a way that the convolutional (Cx) and max pooling (Px) filters do not overlap between two adjacent slices. To achieve this, we chose each edge size of the convolution filter to equal the stride in a particular layer. For example, the edge size of the convolution filter and the stride in the C1 layer were 4 and 4, respectively (Fig. 5.2). We chose a small convolutional filter size, which tends to achieve better classification accuracy, as demonstrated in [119]. Layer C2 is the second convolutional layer with forty-eight 4×4×24 kernels applied to the output of C1. Unlike C1, we used a max-pooling (P2) of 4×4×48 windows in this layer to reduce the image size to 8×8 from 32×32.

The output of C2 is connected to a fully connected layer (F3), which contains 96 units. Similarly, a layer F4 comprises 96 units and is fully connected to F3. We concatenated the units of F3 and F4 into CT5 to reduce possible information loss. This bypassing connection is typically known for better classification accuracy [120]. Note that the CT5 layer did not have any trainable weights.

The CT5 layer is connected to an F5 layer having two units. These units are connected to a softmax layer (S), which produces the relative probabilities for back-propagation and classification.

5.3 Experimental Setup and Data Acquisition

Our clinical dataset consisted of 160 kidney scans of 160 patients accessed from our TCGA-KIRC patient pool (discussed in section 2.2). We used 80 healthy kidney samples from 80 patients who had one healthy kidney. The 80 pathological kidney samples used were from another 80 patient scans. Of the 80 healthy and 80 carcinoma scans, we randomly chose 45 and 10 cases from each set to use for training and validation, respectively, and the remaining 25 for testing. We trained

our network by minimizing the softmax loss between the desired and predicted labels. We used an optimization method called *Adam* [121]. All the parameters for this solver were set to the suggested default values, i.e., $\beta_1 = 0.9$, $\beta_2 = 0.999$, and $\varepsilon = 10^{-8}$. We also employed a unit dropout (Dx) that drops 50% of units in both F4 and CT5 layers and used a weight decay of 0.005. The base learning rate was set to 0.01 and was decreased by a factor of 0.1 to 0.0001 over 25000 iterations with a batch of 32 images processed at each iteration.

5.4 Validation on Kidney Data

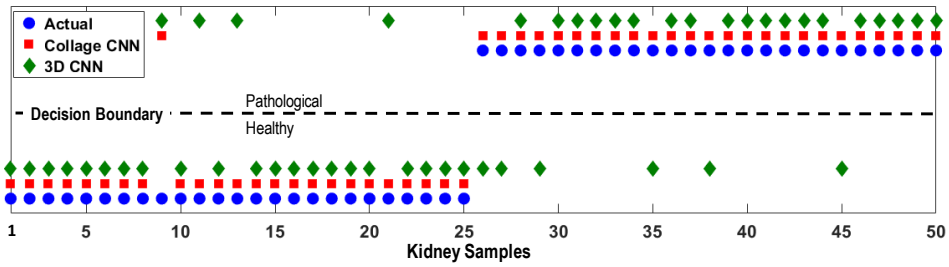

Figure 5.3: Scatter plot showing the actual *vs.* predicted labels by the collage image-based and 3D CNNs.

We provide the classification accuracy results of our proposed collage image-based CNN as a bar plot in Fig. 5.3. We also compare our performance to that of a 3D CNN on the same plot. For the 3D CNN, we replaced the collage input (512×512×3 pixels) with the full 3D volume (64×64×64 pixels) of the kidney and performed 3D convolutions with a filter size of 4×4×64 with stride 4. For a fair comparison, we chose the 3D volume dimension as 64×64×64 pixels, since each constituent axial slice in the collage was of 64×64 pixels. Other layer configurations remained the same as in Fig. 5.2. Both CNNs were implemented using *Caffe* [107]. The pre-processing of the data, visualizations, and comparisons were done in MATLAB using the *MatCaffe* interface. Before generating the collage representation of the input data, we ensured a uniform voxel spacing in the image volume of all axial, coronal, and sagittal planes using interpolation. We manually defined the kidney ROI (sub)volume within the CT data in such a way that leaves

an approximately 25% background area framing a kidney. Before training, both the training and testing datasets were standardized.

In our experiments, we augmented the number of training samples by a factor of 40 by flipping and rotating the image slices as well as by random reshuffling the slice location within the collage. This augmentation process enabled by our new image representation yielded a total of 4,400 2D image collages for training.

As demonstrated in Fig. 5.3, our proposed method succeeded in all but only one case out of 50 tested kidney samples, resulting in a classification accuracy of 98%. In comparison, the 3D convolution-based CNN failed in eight cases resulting in an accuracy of 80%.

Our preliminary results suggest that our proposed collage image representation may offer significant advantages for deep CNN-based classification tasks on 3D data. Our collage representation allows the convolution kernel to slide over all the axial 2D slices in a 3D volume, which is impossible in case of a 3D CNN. The training time of the collage CNN was approximately 5 hours (on our primary machine), while the 3D CNN took nearly 7 hours to converge. We also augmented the 3D data by using data rotation and flipping before feeding to the 3D CNN, and we expect the performance of the 3D CNN to be better than our collage CNN. But because of the better augmentation capability of the collage representation, it performed better compared to the 3D CNN in our experiment. Thus, the collage representation seems best suited in the insufficient annotated medical data scenario. It is worth noting that to improve the classification accuracy by the 3D CNN approach, one may have to increase the convolution kernel size and decrease the stride size to capture more features from the image volume. However, this would drastically increase the number of trainable weights, which would necessitate the use of expensive GPUs with large memory, and would cost more time to converge.

It is worth mentioning that there are some other potential ways of using deep learning approaches on the sparsely annotated data. One approach could be the 3D CNN that we use to compare our results within this chapter. This special type of 3D CNN uses 2D filters applied to all slices. However, our results showed that this approach is not as effective as our collage CNN approach. Another approach could be weight sharing in the axial direction for 3D CNN. However, tumors are often very small compared to the actual kidney. Therefore, sharing weights along the

axial direction may result in the loss of tumor information. Thus the network could be ineffective in detecting pathological kidneys. For the same reason, we believe that a simple MIL approach with prediction averaging would not be effecting for pathological kidney detection.

We further like to mention that our proposed collage CNN is very light in terms of the number of trainable parameters ($\sim$39,168). In contrast, one of the recent and popular 3D CNN, 3D U-Net [87], has $\sim$19,070,000 trainable parameters, which is approximately 487x more than that of our proposed collage CNN approach.

5.5 Summary

In this chapter, we discussed a novel collage image representation within a CNN based classification scheme to enable deep learning from sparsely labeled 3D datasets. We applied our proposed method on CT abdominal scans from the TCIA database to discriminate healthy from cancerous kidneys containing renal cell carcinoma. Our method enables efficient 2D slice-based learning in the absence of slice-based labels. Also, the proposed collage inherently allows for easy data augmentation through a random reshuffling of the locations of image slices within the collage, thus facilitating more effective training of the implicit relationship between bag labels and feature representation in weekly supervised ML settings. We showed our approach to be impressively effective (98% classification accuracy) on weakly labeled data on a small-sized database of 160 kidney CTs outperforming 3D CNNs. However, the latter's performance could potentially be better with a significant increase in labeled data as well as computation cost.

Chapter 6

Gene Mutations Detection in Kidney ccRCC

In this chapter, we discuss a deep CNN approach that addresses the challenge of automatic mutation detection in kidney ccRCC. Our method is a variant of the conventional MIL approach, where we use multiple instances for robust binary classification while using single instances for training CNN to facilitate a higher number and variation of training data. The CNN automatically learns the ccRCC image features, and the aggregation of binary decisions (i.e., presence/absence of a mutation) for all the ccRCC slices in a particular kidney sample ultimately determines whether an interrogated kidney sample has undergone a certain mutation or not. The frequency of occurrence of various mutations in ccRCC varies significantly, e.g., VHL, PBRM1, BAP1, SETD2, and KDM5C were found in 76%, 43%, 14%, 14% and 8% of kidney samples of our dataset, respectively. In this study, we consider the four most prevalent gene mutations (i.e., VHL, PBRM1, BAP1, and SETD2). We achieve this via four multiple instance decision aggregation CNNs. However, our approach is directly extendable to more mutation types depending on the availability of sufficient training data. We originally published this work in Hussain et al. [4].

6.1 Multiple Instance Decision Aggregation for Mutation Detection

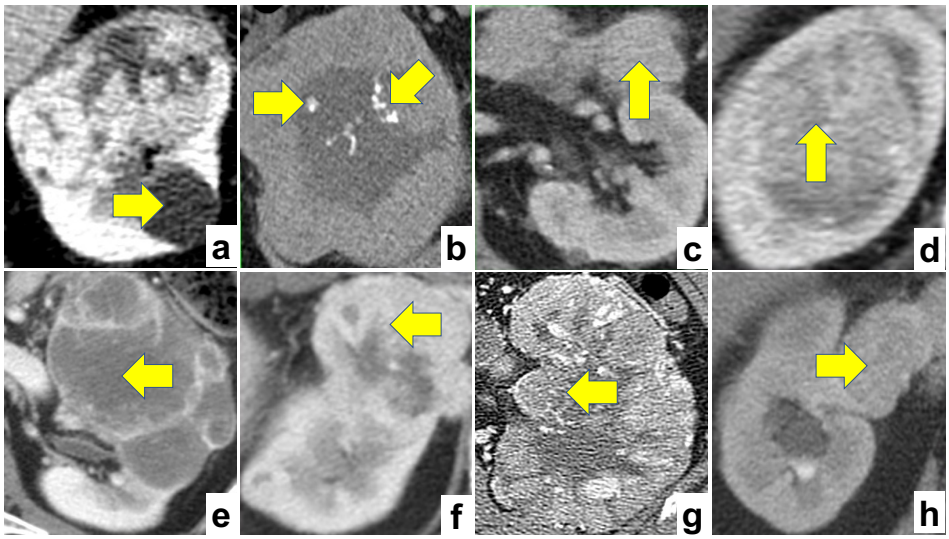

Figure 6.1: Illustration of CT features of ccRCC seen in the data of this study. (a) Cystic tumor architecture, (b) calcification, (c) exophytic tumor, (d) endophytic tumor, (e) necrosis, (f) ill-defined tumor margin, (g) nodular enhancement, and (h) renal vein invasion. Arrow indicates a feature of interest in each image.

Typically, ccRCC grows in different regions of the kidney and is clinically scored based on their CT slice-based image features. For example, size, margin (well- or ill-defined), composition (solid or cystic), necrosis, growth pattern (endophytic or exophytic), calcification, etc. [36]. We show some of these features in our dataset in Fig. 6.1. Recent work [36, 37] shown correlations between mutations in genes and different ccRCC features seen in CT images. For example, an association between well-defined tumor margin, nodular enhancement, and intratumoral vascularity with the VHL mutation has been reported [37]. Ill-defined tumor margin and renal vein invasion were also reported to be associated with the BAP1 mutation [36], whereas PBRM1 and SETD2 mutations are mostly seen in solid (non-cystic) ccRCC cases [37]. We propose to learn these features from the CT images using four different CNNs: VHL-CNN, PBRM1-CNN, SETD2-CNN,

and BAP1-CNN, each for one of the four mutations (VHL, PBRM1, SETD2, and BAP1). Using a separate CNN per-mutation alleviates the problem of data imbalance among mutation types, given that the mutations are not mutually exclusive. For a particular CNN targeting to detect a particular gene (say, 'x') mutation, we used two sets of data for training: one set with x-mutation present, and another set with x-mutation absent but may or may not have other mutations present.

6.1.1 CNN Architecture

All the CNNs in this study (i.e., VHL-CNN, PBRM1-CNN, SETD2-CNN, and BAP1-CNN) have similar configurations, but we train those separately (Fig. 6.2). Each CNN has twelve layers excluding the input: five convolutional (Conv) layers; three fully connected (FC) layers; one softmax layer; one average pooling layer; and two thresholding layers. All but the last three layers contain trainable weights. The input is the $227 \times 227 \times 3$ pixel image slice containing the kidney+ccRCC. We train these CNNs (layers 1–9) using a balanced dataset for each mutation case separately (i.e., a particular mutation-present and absent). During training, we fed images to the CNNs in a randomly shuffled single instance fashion. Typically, Conv layers are known for sequentially learning the high-level non-linear spatial image features (e.g., ccRCC size, orientation, edge variation, etc.). We used five Conv layers as the 5th Conv layer typically grabs an entire object (e.g., ccRCC shape) in an image even if there are a significant pose variation [122]. Subsequent FC layers prepare those features for the optimal classification of an interrogated image. In our case, we deployed three FC layers to decide on the learned features from the 3-ch images to decide if a particular gene mutation is probable or not. The number of FC layers plays a vital role as the overall depth of the model is important for obtaining good performance [122], and we achieve optimal performance with three FC layers. Layers 10, 11, and 12 (i.e., two thresholding and one average pooling layers) of the CNNs are used during the testing phase and do not contain any trainable weights.

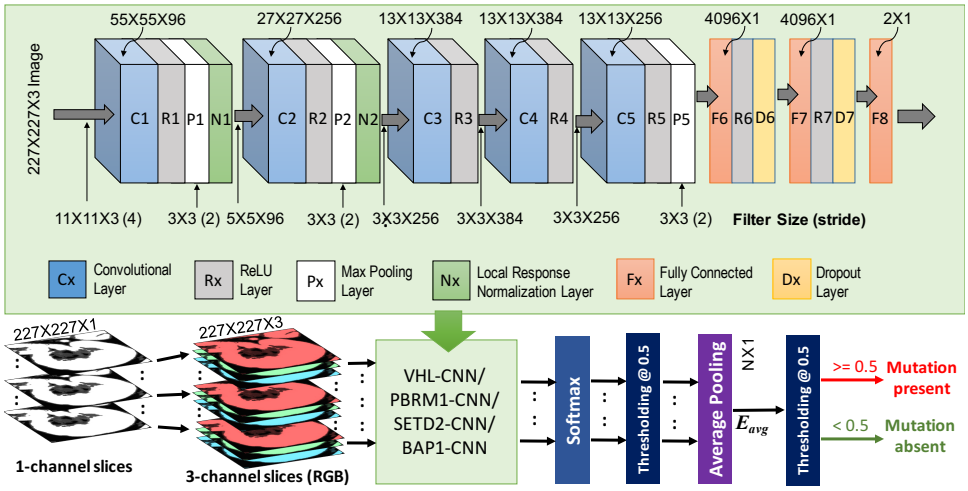

Figure 6.2: Multiple instance decisions aggregated CNN for gene mutation detection.

6.1.2 Mutation Detection

After all the CNNs are trained (from layer 1 to 9), we use the full configuration (from layer 1 to 12) in the testing phase. Although we use only ccRCC containing kidney slices during training and validation, often not all the ccRCC cross-sections contains the discriminating features for proper mutation detection. Therefore, our trained CNN (from layer 1 to 9) often miss-classifies the interrogated image slice based on the probability estimated at layer 9 (i.e., softmax layer). To address this miss-classification by our CNNs, we adopt a multiple instance decision aggregation procedure. In this procedure, we feed all the candidate image slices of a particular kidney to the trained CNN and accumulate the slice-wise binary classification labels (0 or 1) at layer 10 (the thresholding layer). We fed these labels into a $N \times 1$ average pooling layer, where N is the total number of 3-channel axial slices of an interrogated kidney. Finally, we fed the estimated average (E_{avg}) at layer 11 to the second thresholding layer (layer 12), where $E_{avg} \geq 0.5$ indicates the presence of the mutation in that kidney, and no-mutation otherwise (see Fig. 6.2).

6.2 Experimental Setup and Data Acquisition

We used 160 patients' CT scans from our TCGA-KIRC patient pool (discussed in section 2.2). In this dataset, 138 scans contained at least one mutated gene because of ccRCC. For example, 105 patients had VHL, 60 patients had PBRM1, 20 patients had SETD2, and 20 patients had BAP1 mutations. Besides, some of the patients had multiple types of mutations. However, nine patients had CT scans acquired after nephrectomy and, therefore, those patients' data were not usable for this study. We show the number of kidney samples used in the training, validation, and testing stages in Table 6.1. During training, validation, and testing, we use only those slices of the kidney that contain ccRCC as our CNNs aim to learn ccRCC features. We train all the networks by minimizing the softmax loss between the expected and detected labels (1: mutation present and 0: mutation absent). We used the *Adam* optimization method [123]. All the parameters for this solver were set to the suggested (by [123]) default values, i.e., $\beta_1 = 0.9$, $\beta_2 = 0.999$ and $\varepsilon = 10^{-8}$. We also employed a Dropout unit (Dx) that dropped 50% of units in both F6 and F7 layers (Fig. 6.2) and used a weight decay of 0.005. The base learning rate for CNN pre-training was set to 0.01 and was decreased by a factor of 0.1 to 0.0001 over 25,000 iterations. During fine-tuning, the base learning rate was set to 0.001 and was decreased by a factor of 0.1 to 0.0001 over 20,000 iterations with a batch of 256 images processed at each iteration. We performed the training on a workstation with Intel 4.0 GHz Core-i7 processor, an Nvidia GeForce Titan Xp GPU with 12 GB of VRAM, and 32 GB of RAM.

Table 6.1: The number of kidney samples used in training, validation, and testing per-mutation case. An acronym used: xM: 'x' type mutation.

Genes	# Training Samples		# Validation Samples		# Test Samples	
(x)	xM-Present	xM-Absent	xM-Present	xM-Absent	xM-Present	xM-Absent
VHL	74	74	10	10	15	15
PBRM1	35	35	6	6	10	10
SETD2	11	11	3	3	5	5
BAP1	10	10	3	3	4	4

6.3 Data Pre-processing

We form a 3-channel image from each scalar-valued CT slice by generating channel intensities [I, I-50, I-100] Hounsfield Unit (HU), where I represents the original intensities in a CT image slice tightly encompassing a kidney+ccRCC cross-section (Fig. 6.1), whereas I-50 and I-100 represent two variants of I with different HU values. We add these variations in channel intensity values as similar ccRCC features may have different X-ray attenuation properties across patients [36]. Also, contrast-shifting produces an augmented version of the original image without changing the tumor shape features. We resized all the image data by resampling into a size of $227 \times 227 \times 3$ pixels. We augmented the number of training samples by a factor of 24 by flipping and rotating the 3-ch image slices as well as by re-ordering the three channels in each image. We normalized the training and validation data before training by subtracting the image mean and dividing by the image standard-deviation.

6.4 Validation on Kidney Data

We compare the mutation detection performance by a wide range of methods. At first, we tested the performance using a single instance (SI)-based random forest (RF) approach, where hand-engineered image features were used. In a typical SI-based classification approach, the class-label is decided from the maximum among the predicted class-probabilities [124]. Similarly, in our SI-based approaches, the presence or absence of a certain mutation is decided from the maximum among the estimated probabilities associated with all the ccRCC image slices in a particular kidney. Then we demonstrate the effectiveness of automatic feature learning compared to the hand-engineered features generation using the CNN approach. Afterward, we show the effect of incorporating augmented data in the training dataset and compared the mutation detection performance for three different types of augmentation (i.e., image flipping+rotation, 3-ch re-ordering, and those combined). Finally, we demonstrated the effectiveness of using multiple instance decision aggregations in our proposed method.

In Table 6.2, we estimated the overall mutation prediction errors by the VHL-

Table 6.2: Automatic gene mutation detection performance of different methods. We use acronyms as M: mutation, x: one of VHL/PBRM1/SETD2/BAP1, Aug: augmentation, SI: single instance, MI: multiple instances, 3ch: 3-channel data with augmentation by channel reordering, F: augmentation by flipping, and R: augmentation by rotation.

Methods	Genes	# Test Samples		# Correct Detection		Overall Error	Mean Error
	x	M-present	M-absent	M-present	M-absent	OE_x (%)	ME (%)
Random Forest (SI+1ch) No Aug	VHL	15	15	5	7	60	53.75
	PBRM1	10	10	4	5	55	
	SETD2	5	5	2	3	50	
	BAP1	4	4	2	2	50	
x-CNN (SI+1ch) No Aug	VHL	15	15	7	8	50	41.88
	PBRM1	10	10	6	6	40	
	SETD2	5	5	3	3	40	
	BAP1	4	4	2	3	37.50	
x-CNN (SI+3ch)	VHL	15	15	12	9	30	29.38
	PBRM1	10	10	4	7	45	
	SETD2	5	3	4	4	30	
	BAP1	4	4	3	4	12.5	
x-CNN (SI+1ch +F+R)	VHL	15	15	11	13	20	21.88
	PBRM1	10	10	8	7	25	
	SETD2	5	5	3	4	30	
	BAP1	4	4	4	3	12.50	
x-CNN (SI+3ch +F+R)	VHL	15	15	15	11	13.33	13.96
	PBRM1	10	10	9	9	10	
	SETD2	5	5	5	3	20	
	BAP1	4	4	3	3	12.50	
Proposed (MI+3ch +F+R)	VHL	15	15	14	13	**10**	**6.25**
	PBRM1	10	10	9	10	**5**	
	SETD2	5	5	5	4	**10**	
	BAP1	4	4	4	4	**0**	

CNN, PBRM1-CNN, SETD2-CNN, and BAP1-CNN as:

$$OE_x = 100\left(1 - \frac{C_{MP} + C_{MA}}{T_{MP} + T_{MA}}\right)\%, \tag{6.1}$$

where, OE stands for overall error (shown in the second last column of Table 6.2), x represents either of the four mutations (i.e., VHL or PBRM1 or SETD2 or BAP1), C_{MP} denotes the correct number of predictions for x-mutation presence, C_{MA} denotes the correct number of predictions for x-mutation absence, T_{MP} denotes the total number of test cases for x-mutation presence, and T_{MA} denotes the total num-

ber of test cases for x-mutation absence. We also report the mean error (ME) for each of the comparing methods in the last column of Table 6.2 by combining the individual errors (i.e., OE_x) as:

$$ME = \frac{OE_{VHL} + OE_{PBRM1} + OE_{SETD2} + OE_{BAP1}}{4}. \tag{6.2}$$

In row 1 of the comparison Table 6.2, we show results of a traditional RF approach with hand-engineered image features proved to be useful in anatomy classification task [124]: histogram of the oriented gradient, Haar features, and local binary patterns. Here, we did not augment any manually transformed data to the training samples. We trained four RFs for the four different mutation cases, and as we see in Table 6.2, the resulting mean detection error was the highest ($\sim$54%) among all contrasted methods. Row 2 shows the results of a deep CNN (namely, x-CNN, where x: VHL/PBRM1/SETD2/BAP1 (see Fig. 6.2)) approach with no data augmentation. Since the CNN learns the image features automatically, it may have helped this CNN method perform better (mean error $\sim$42%) than that of the hand-engineered features-based RF approach. Row 3 shows results for x-CNN, where we used data augmentation by deploying 3-ch data and re-ordering of channels (see section 6.3). We fed these data to x-CNN, and we can see how the SI-based mutation detection performance by this approach (mean error $\sim$29%) outperformed that with no data augmentation. Thus, including channels with different intensity ranges, mimicking the tumor intensity variation across patients, have shown a positive impact on the mutation detection task. Row 4 shows results for x-CNN with a different augmentation process, which deploys the flipping and rotating of the 1-ch training samples. This approach (mean error $\sim$22%) outperformed that with 3-ch augmentation. So it is clear that the flipping+rotation-based augmentation introduced more variation in the training data than that by the 3-ch augmentation, resulting in a better generalization of the model. In the method shown in row 5, we combined the flipping+rotation augmentation with the 3-ch re-ordering augmentation. The performance of the x-CNN with these data was better in mutation detection (mean error $\sim$14%) than that of flipping+rotation or 3-ch augmentation alone (see Table 6.2). Finally, row 6 demonstrates the results of our proposed method, where we used flipping, rotation, and 3-ch re-ordering augmentations.

Also, binary classification was performed based on the multiple instance decisions aggregation. We see in the Table 6.2 that the mean mutation detection error by our method is $\sim$6%, which is the lowest tested. Also, detection errors for individual mutation cases were small and in the range of 10%. Thus, our multiple instance decision aggregation procedure made our CNN models more robust on SI-based miss-classification.

6.5 Summary

In this chapter, we discussed a multiple instance decision aggregation-based deep CNN approach for automatic mutation detection in kidney ccRCC. We have shown how our approach automatically learned discriminating ccRCC features from CT images and aggregated the binary decisions on the mutation-presence/absence for all the ccRCC slices in a particular kidney sample. This aggregation produced a robust decision on the presence of a certain mutation in an interrogated kidney sample. Also, our multiple instance decision aggregation approach achieved better accuracy in mutation detection than that of a typical single instance-based approach. On the other hand, better performance by conventional MIL approaches is subject to the availability of a sufficient number of data. At the same time, in applications such as ours, there are usually very few data samples for some of the mutation cases. Therefore, an end-to-end MIL approach will most likely fail for those mutation cases with few data samples. However, this chapter included several meaningful comparisons to highlight the effects of different augmentation, pooling schemes, etc. within the context of insufficient data. We believe that these comparisons provide more interesting findings and appear to be suitable for ccRCC Radiomics, where the learned mutations would aid in better ccRCC diagnosis, prognosis, and treatment response assessment. Our experimental results demonstrated an approximately 94% accuracy in kidney-wise mutation detection.

An image-based noninvasive gene mutation detection method has a promising clinical implication. Although the biopsy-based diagnosis is an inseparable part of the clinical workflow, it often requires considerable time in the process of performing the biopsy and subsequent radiological analysis. Our image-based noninvasive approach can be effective in such a scenario. While a patient waits for the biopsy

conduction and results, an image-based approach can help physicians to diagnose ahead and prepare the treatment plan. The biopsy results can confirm the decision.

Our proposed method showed promising results in noninvasive gene mutation detection. However, we had to train four separate CNNs because of the lack of adequate data for the PBRM1, BAP1, and SETD2 mutation cases. A multitasking CNN often leverages the complementary features of the different objects of interest, which makes the model more robust on the classification task. Therefore, we plan to investigate using a multitasking CNN for gene mutation detection if we get access to more dataset in the future.

Chapter 7

ImHistNet for RCC Grades and Stages Detection in CT

In this chapter, we discuss the ImHistNet, a Deep neural network (DNN) for an end to end texture-based image classification. Our ImHistNet approach makes the following contributions: (1) we propose a learnable image histogram (LIH) layer within a DNN framework. It is capable of learning complex and subtle task-specific textural features from raw images directly, adhering to the classical input-output mapping of a CNN. (2) We remove the requirement for fine pre-segmentation of the RCC as the proposed learnable image histogram can stratify tumor and background textures well, thus enabling the model to focus specifically on the tumor texture. And (3) we demonstrate ImHistNet's capabilities by performing automatic RCC grade classification for the 2-tiered FGS as well as automatic categorization of RCC into anatomical stage low (I/II) and high (III/IV) on an extended clinical dataset from real patients. Note that it is an important finding of our experiment that the RCC stages correlate with the CT textural features, which is, to our knowledge, not investigated to date. We originally published these works in two parts in Hussain et al. [5] and Hussain et al. [6].

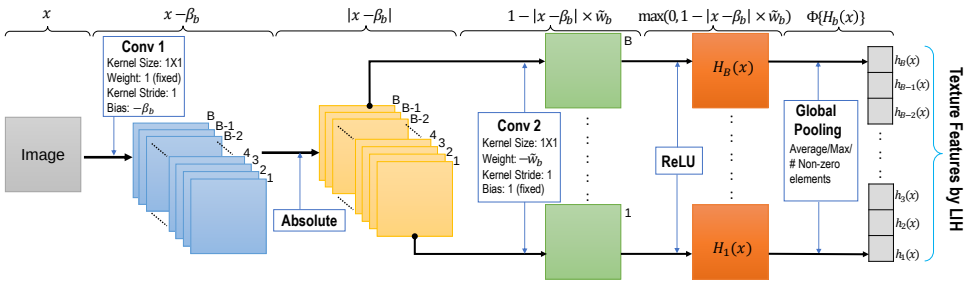

Figure 7.1: The architecture of our learnable image histogram using CNN layers.

7.1 Learnable Image Histogram

Our proposed learnable image histogram (LIH) stratifies the pixel values in an image x in different learnable and possibly overlapping intervals (bins of width w_b) with arbitrary learnable means (bin centers β_b). The feature value $h_b(x) : b \in \mathscr{B} \rightarrow \mathscr{R}$, corresponding to the pixels in x that fall in the bth bin, is estimated as:

$$h_b(x) = \Phi\{H_b(x)\} = \Phi\{\max(0, 1 - |x - \beta_b| \times \widetilde{w}_b)\}, \tag{7.1}$$

where $\mathscr{B}$ is the set of all bins, Φ is the global pooling operator, $H_b(x)$ is the piece-wise linear basis function that accumulates positive votes from the pixels in x that fall in the bth bin of interval $[\beta_b - w_b/2, \beta_b + w_b/2]$, and $\widetilde{w}_b$ is the learnable weight related to the width w_b of the bth bin: $\widetilde{w}_b = 2/w_b$. Any pixel may vote for multiple bins with different $H_b(x)$ since there could be an overlap between adjacent bins in our learnable histogram. The final $|\mathscr{B}| \times 1$ feature values from the learned image histogram are obtained using a global pooling Φ over each $H_b(x)$ separately. This pooling can be a 'non-zero elements count', which matches the convention of a traditional histogram, or can be an 'average' or 'max' pooling, depending on the task-specific requirement. The linear basis function $H_b(x)$ of the LIH is piece-wise differentiable and can back-propagate (BP) errors to update β_b and $\widetilde{w}_b$ during

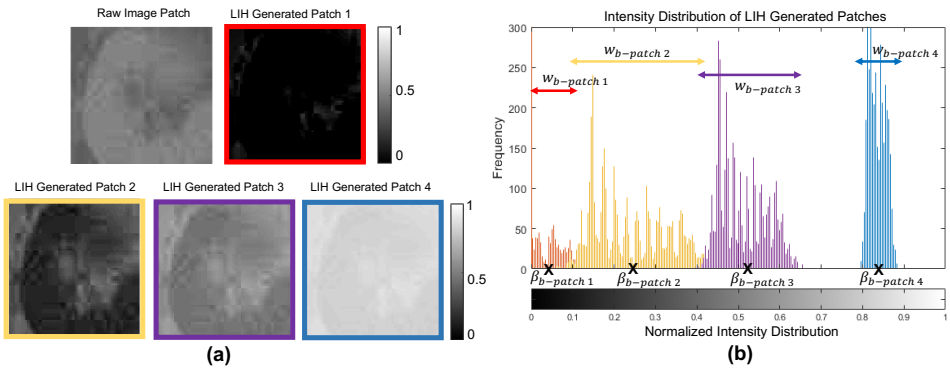

Figure 7.2: Illustration of LIH generated image patches with variable intensity distribution. (a) Raw CT image patch (x) of size 64×64 pixels and four randomly selected image patches [$H_B(x)$] before the global pooling in Fig. 7.1. (b) Corresponding intensity distributions of patches 1-4 in (a) are shown with Histogram of variable bin centers β_b and widths w_b.

training. The gradients of β_b and $\widetilde{w}_b$ for a loss $\mathscr{L}$ are estimated as:

$$\frac{\partial \mathscr{L}}{\partial \beta_b} = \begin{cases} \widetilde{w}_b & \text{if } H_b(x) > 0 \text{ and } x - \beta_b > 0, \\ -\widetilde{w}_b & \text{if } H_b(x) > 0 \text{ and } x - \beta_b < 0, \\ 0 & \text{otherwise.} \end{cases} \tag{7.2}$$

$$\frac{\partial \mathscr{L}}{\partial \widetilde{w}_b} = \begin{cases} |x - \beta_b| & \text{if } H_b(x) > 0, \\ 0 & \text{otherwise.} \end{cases} \tag{7.3}$$

7.2 Design of LIH using CNN Layers

The proposed LIH is implemented using CNN layers as illustrated in Fig. 7.1. The input of LIH can be a 2D or vectorized 1D image, and the output is a $|\mathscr{B}| \times 1$ histogram feature vector. The operation $x - \beta_b$ for a bin centered at β_b is equivalent to convolving the input by a 1×1 kernel with fixed weight of 1 (i.e., with no updating by BP) and a learnable bias term β_b ('Conv 1' in Fig. 7.1). A total of $B = |\mathscr{B}|$ number of similar convolution kernels are used for a set of $\mathscr{B}$ bins. Then an absolute value layer produces $|x - \beta_b|$. This is followed by a set of convolutions ('Conv 2'

in Fig. 7.1) with a total of B separate (non-shared across channels) learnable 1×1 kernels and a fixed bias of 1 (i.e., no updating by BP) to model the operation of $1 - |x - \beta_b| \times \widetilde{w}_b$. We use the rectified linear unit (ReLU) to model the $\max(0, \cdot)$ operator in Eq. 7.1. The final $|\mathscr{B}| \times 1$ feature values $h_b(x)$ are obtained by global pooling over each feature map $H_b(x)$ separately.

In Fig. 7.2(a), we show an example raw CT image patch x and corresponding LIH generated image patches randomly selected from the feature maps of $H_b(x)$ (see Fig. 7.1). We also show the intensity distributions of the selected patches in Fig. 7.2(a) in terms of histogram in Fig. 7.2(b), where we can observe the learned histogram of variable bin centers β_b and bin widths w_b. We also observe in Fig. 7.2(b) that the learned w_b for different feature maps in $H_b(x)$ have overlaps among those.

7.3 ImHistNet Classifier Architecture

The classification network comprises ten layers: the LIH layer, five (F1-F5) fully connected layers (FCLs), one softmax layer, one average pooling (AP) layer, and two thresholding layers (see Fig. 7.3). The first seven layers contain trainable weights. The input is a 64×64 pixel image patch extracted from the kidney+ccRCC slices. During training, we fed randomly shuffled image patches individually to the network. The LIH layer learns the variables β_b and $\widetilde{w}_b$ to extract characteristic textural features from image patches. In implementing the proposed ImHistNet, we chose $B = 128$ and 'average' pooling at $H_b(x)$. We set subsequent FCL (F1-F5) size to 4096×1. The number of FCLs plays a vital role as the overall depth of the model is important for good performance [122]. Empirically, we achieved good performance with five FCL layers. Layers 8, 9, and 10 of the ImHistNet are used during the testing phase and do not contain any trainable weights.

7.4 RCC Grade and Stage Classification

After training ImHistNet (layers 1 to 7) by estimating errors at layer 7 (i.e., Softmax layer), we used the full configuration (from layer 1 to 10) in the testing phase. Although we used patches from only RCC-containing kidney slices during training and validation, not all the RCC cross-sections contained discriminant features

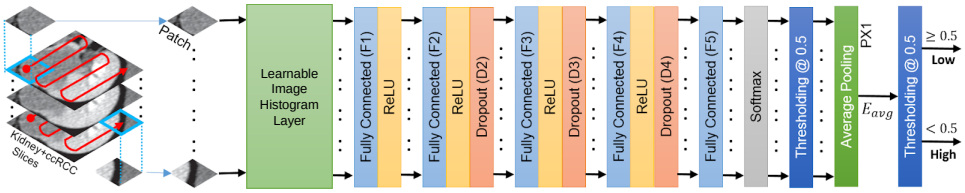

Figure 7.3: Multiple instance decisions aggregated ImHistNet for grade classification.

for proper grade identification. Thus our trained network may miss-classify the interrogated image patch. To reduce such miss-classification, we adopt a similar multiple instance decision aggregation procedure proposed by Hussain et al. [4]. In this approach, we feed randomly shuffled single image patches as inputs to the model during training. During inference, we feed all candidate image patches of a particular kidney to the trained network and accumulate the patch-wise binary classification labels (0 or 1) at layer 8 (the thresholding layer). We then feed these labels into a $P \times 1$ average pooling layer, where P is the total number of patches of an interrogated kidney. Finally, we feed the estimated average (E_{avg}) from layer 9 to the second thresholding layer (layer 10), where $E_{avg} \geq 0.5$ indicates the Fuhrman low or stage low, and $E_{avg} < 0.5$ indicates Fuhrman high or stage high (see Fig. 7.3).

7.5 Experimental Setup and Data Acquisition

We used CT scans of 159 patients from our TCGA-KIRC patient pool (discussed in section 2.2). These patients were diagnosed with clear cell RCC, of which 64 were graded Fuhrman low (I/II), and 95 were graded Fuhrman high (III/IV). Also, 99 patients were staged low (I-II), and 60 were staged high (III-IV) in the same cohort. We divided the dataset for training/validation/testing as 44/5/15 and 75/5/15 for Fuhrman low and Fuhrman high, respectively. For anatomical staging, we divided the dataset for training/validation/testing as 81/3/15 and 42/3/15 for stage low and stage high, respectively. Note that typical tumor radiomic analysis comprises [125]: (i) 3D imaging, (ii) tumor detection and/or segmentation, (iii) tumor phenotype quantification, and (iv) data integration (i.e., phenotype + genotype + clinical + proteomic) and analysis. Our approach falls under step-iii. The input data to our method are thus 2D image patches of size 64×64 pixels,

taken from kidney+RCC (i.e., both mutually inclusively present) bounding boxes. We do not require any fine pre-segmentation of the RCC rather only assume a kidney+RCC bounding box, generated in step-ii. Given data imbalance where samples for Fuhrman low are fewer than for Fuhrman high and stage high are fewer than for stage low, we allowed more overlap among adjacent patches for the Fuhrman low and stage high dataset. We calculated the amount of overlap to balance the samples from both cohorts.

We trained two separate ImHistNets for Fuhrman grading and anatomical staging of RCC. We implemented our networks in *Caffe* [107] and trained by minimizing the binary cross-entropy loss between the ground truth and predicted labels (1: Fuhrman low/stage low, and 0: Fuhrman high/stage high). We used Stochastic gradient descent for updating parameters. We employed a Dropout unit (Dx) that drops 20%, 30%, and 40% of units in F2, F3, and F4 layers, respectively (Fig. 7.3) and used a weight decay of 0.005. The base learning rate was set to 0.001 and was decreased by a factor of 0.1 to 0.0001 over 250,000 iterations with a batch of 128 patches. We performed the training on a workstation with Intel 4.0 GHz Core-i7 processor, an Nvidia GeForce Titan Xp GPU with 12 GB of VRAM, and 32 GB of RAM.

7.6 Validation on Kidney Data

7.6.1 RCC Fuhrman Grade Classification

We compared our RCC grade classification performance in terms of accuracy (%) to a wide range of methods. Note that for all our implementations, we trained models with shuffled single image patches, and used multiple instance decision aggregations per kidney during inference. We fixed our patch size to 64×64 pixels across all contrasted methods.

First, we use ResNet-50 [110] with transfer learning in order to test the performance of conventional CNN (see Table 7.1). Here, we used the full kidney+RCC slices as well as patches as inputs. As we mentioned in section 1.2.5 that a classical CNN typically fails to capture textural features, it has become evident from the results in Table 7.1 where such CNNs performed poorly in learning the textural

Table 7.1: Automatic RCC Fuhrman grade classification performance by conventional CNN. NTS: Number of test samples.

Methods	NTS	Accuracy
Full image+ResNet-50	30	53%
Patch+ResNet-50	30	50%

features of RCC.

Table 7.2: Automatic RCC Fuhrman grade classification performance by hand-engineered features-based conventional machine learning approaches. SVM: support vector machines, xFCV: x-fold cross-validation, LxOCV: leave-x-out cross-validation, '-': Not reported.

Methods	NTS	Accuracy
Patch+Histogram (128 bins)+SVM	30	56%
Patch+Histogram (256 bins)+SVM	30	63%
Shu et al. [41] (5FCV on 260 samples)	-	**77**%
Fei et al. [126] (L1OCV on 90 samples)	-	70%

Next, to evaluate the performance of hand-engineered features-based conventional machine learning approaches, we tested support vector machine (SVM) employing the conventional image histogram of 128 and 256 bins as shown in Table 7.2. We also compared two state-of-the-art methods [41, 126] where we quote the authors' best self-reported performances. These methods mostly relied on the RCC textural features and used classical predictive models, e.g., logistic regression. Here, the method by Shu et al. [41] performed the best with 77% classification accuracy (see Table 7.2).

Then, we cross-examine the performance of hand-engineered features with deep neural network (DNN) and the LIH features with SVM in Table 7.3. To contrast the performance of a SVM against a DNN, we fed the conventional histogram (128 and 256 bins) features to a DNN of 5 FCL with weight sizes (4096×1)-(4096×1)-(4096×1)-(4096×1)-(2×1). We choose this FCL configuration as our ImHistNet contains the same. Next, to evaluate the hand-engineered features against

Table 7.3: Automatic RCC Fuhrman grade classification performance by hand-engineered features with deep learning and LIH features with conventional machine learning approaches. AP: Average pooling.

Methods	NTS	Accuracy
Patch+Histogram (128 bins)+5 FCL	30	50%
Patch+Histogram (256 bins)+5 FCL	30	50%
Patch+LIH (128 bins)+AP+SVM	30	60%

LIH features, we used LIH features to train an SVM. We see in Table 7.3 that the SVM with LIH features outperformed the SVM with conventional histogram features (see Table 7.2).

Table 7.4: Automatic RCC Fuhrman grade classification performance by combined ImHistNet and conventional CNN.

Methods	NTS	Accuracy
Patch+LIH (128 bins)+AP+5 FCL $\parallel$ AlexNet	30	53%
Full Image+LIH (128 bins)+AP $\parallel$ AlexNet	30	50%

To evaluate the performance of a DNN, combining a conventional CNN and the ImHistNet, we added a CNN of AlexNet [127] equivalent configuration in parallel to the ImHistNet. The last FCLs of size 4096×1 in both networks were concatenated and the total network was trained end-to-end. We implemented two such approaches using the full kidney+RCC images, as well as the patches as inputs. To use patches as inputs to the AlexNet, we up-sampled those to a size of 227×227 pixels. We observed in Table 7.4 that the classical CNN affect the performance of the proposed ImHistNet negatively, i.e., results were worse than those by ImHistNet (see Table 7.5).

To achieve optimum results from LIH, we varied the number of bins (64/128/256) and FCLs of size 4096×1 (4/5/6), and the pooling types (AP/NZEC) with the LIH layer and present results in Table 7.5. We see that ImHistNet with 128 bins, average pooling, and 5 FCL achieved the highest accuracy (80%) among all contrasted methods shown in Tables 7.1-7.5. The closest performance to ImHistNet is shown

Table 7.5: Automatic RCC Fuhrman grade classification performance of LIH with a different number of bins, FCLs, and different types of pooling. NZEC: Non-zero elements count.

Methods	NTS	Accuracy
Patch+LIH (128 bins)+NZEC+5 FCL	30	50%
Patch+LIH (128 bins)+AP+4 FCL	30	50%
Patch+LIH (128 bins)+AP+6 FCL	30	50%
Patch+LIH (64 bins)+AP+5 FCL	30	50%
Patch+LIH (256 bins)+AP+ 5 FCL	30	43%
ImHistNet [LIH (128 bins)+AP+5 FCL]	30	**80**%

by the method of Shu et al. [41] with 77% accuracy (see Table 7.2).

7.6.2 RCC Stage Classification

We also compared our RCC stage classification performance in terms of accuracy (%) to a wide range of methods in Table 7.6. To our knowledge, there is no automatic and machine learning-based approach for RCC stage classification. Therefore, we compare the RCC staging performance of different methods by implementing those in our capacity. Similar to RCC grade classification, we trained models with shuffled single image patches and used multiple instance decision aggregations per kidney during inference. We fixed our patch size to 64×64 pixels across all contrasted methods except for ResNet-50.

Table 7.6: Automatic RCC stage classification performance by different methods.

Methods	NTS	Accuracy
Patch+Histogram (16 bins)+SVM	30	53%
Patch+Histogram (64 bins)+SVM	30	53%
Patch+Histogram (16 bins)+5 FCL	30	50%
Patch+Histogram (64 bins)+5 FCL	30	50%
Full Image+ResNet-50 [110]	30	60%
ImHistNet [LIH (128 bins)+AP+5 FCL]	30	**83**%

First, to compare the performance of ImHistNet to that of traditional hand-

engineered feature-based machine learning approaches, we evaluated an SVM employing a conventional image histogram of 16 and 64 bins and Table 7.6 shows a resulting poor performance at 53% accuracy for both the cases. Next, to contrast the performance of SVM against DNN, we fed the conventional histogram (16 and 64 bins) features to a DNN of 5 FCL with weight sizes (4096×1)-(4096×1)-(4096×1)-(4096×1)-(2×1). We chose this FCL configuration for fairer comparisons since our ImHistNet contains the same. Table 7.6 shows that the FCL with conventional histogram performed the worst achieving a 50% accuracy. Next, we used ResNet-50 [110] with transfer learning to test the performance of high performing modern CNN (see Table 7.6). We used full kidney+RCC slices of size 224×224 pixels as input. As mentioned in section 1.2.5, a classical CNN typically fails to capture textural features, which is evident from our results where ResNet-50 performed poorly in learning the textural features of RCC, resulting in 60% accuracy. Finally, we show the performance of our proposed method in Table 7.6 where ImHistNet achieved the highest accuracy (83%) among all contrasted methods.

7.7 Summary

In this chapter, we discussed a learnable image histogram-based DNN framework for end to end image classification. We demonstrated our approach to a cancer grade and stage prediction task providing automatic 2-tiered FGS (Fuhrman low and Fuhrman high) grade classification as well as stage low and stage high classification of RCC from CT scans. Our approach learns a histogram directly from the image data and deploys it to extract representative discriminant textural image features. We increased efficacy by using small image patches to increase the number and variability of training samples, as well as address class imbalances in the training data via overlap control. We also used multiple instance decision aggregations to robustify binary classification further. Our proposed ImHistNet outperformed current competing approaches for this task, including conventional ML, deep learning, as well as manual human radiology experts. ImHistNet appears well-suited for radiomic studies, where learned textural features using the learnable image histogram may aid in better diagnosis.

Similar to the previous chapter, an image-based RCC grading and staging has a promising clinical implication. Although biopsy-based RCC grading is an inseparable part of the clinical workflow, it often requires considerable time in the process of performing the biopsy and subsequent radiological analysis. Our image-based noninvasive approach can be effective in such a scenario. While a patient waits for the biopsy conduction and results, an image-based approach can help physicians to diagnose ahead and prepare the treatment plan. The biopsy results can confirm the decision.

Our proposed ImHistNet efficiently stratifies the intensity spectrum into learnable bins. However, a random perturbation of the input image would lead to the same histogram. In this process, the network loses the spatial context of the image contents. Therefore, we plan to investigate a process to incorporate spatial texture context via learning the co-occurrence statistics within the DNN framework.

Chapter 8

Conclusions

Kidney cancer is the 7th most common cancer in men and the 10th most common cancer in women [128], accounting for an estimated 140,000 global deaths annually [38]. Despite a rapid increase in the number of patients with kidney cancer worldwide, recent developments in personalized medicine and novel treatment approaches have raised hope of significantly improving kidney cancer survival [14]. Medical imaging technologies, accelerated by the advent of modern machine learning techniques, now play a central role in clinical oncology. Automated CT-based cancer analysis is also benefiting from unprecedented advancements in machine learning techniques and wide availability of high-performance computers. However, there are still many challenges remaining in kidney cancer research using different machine learning techniques.

8.1 Summary of Thesis Contributions

Typically, kidney cancer analysis requires a challenging pipeline of (a) kidney localization in the CT and primary assessment of the kidney health, (b) tumor detection within the kidney, and (c) cancer analysis. In this thesis, we developed machine learning techniques for automatic kidney localization, segmentation-free volume estimation, cancer detection, as well as CT features-based gene mutation detection, and cancer grading and staging. The main contributions of this thesis (see Fig. 8.1) are summarized as follows:

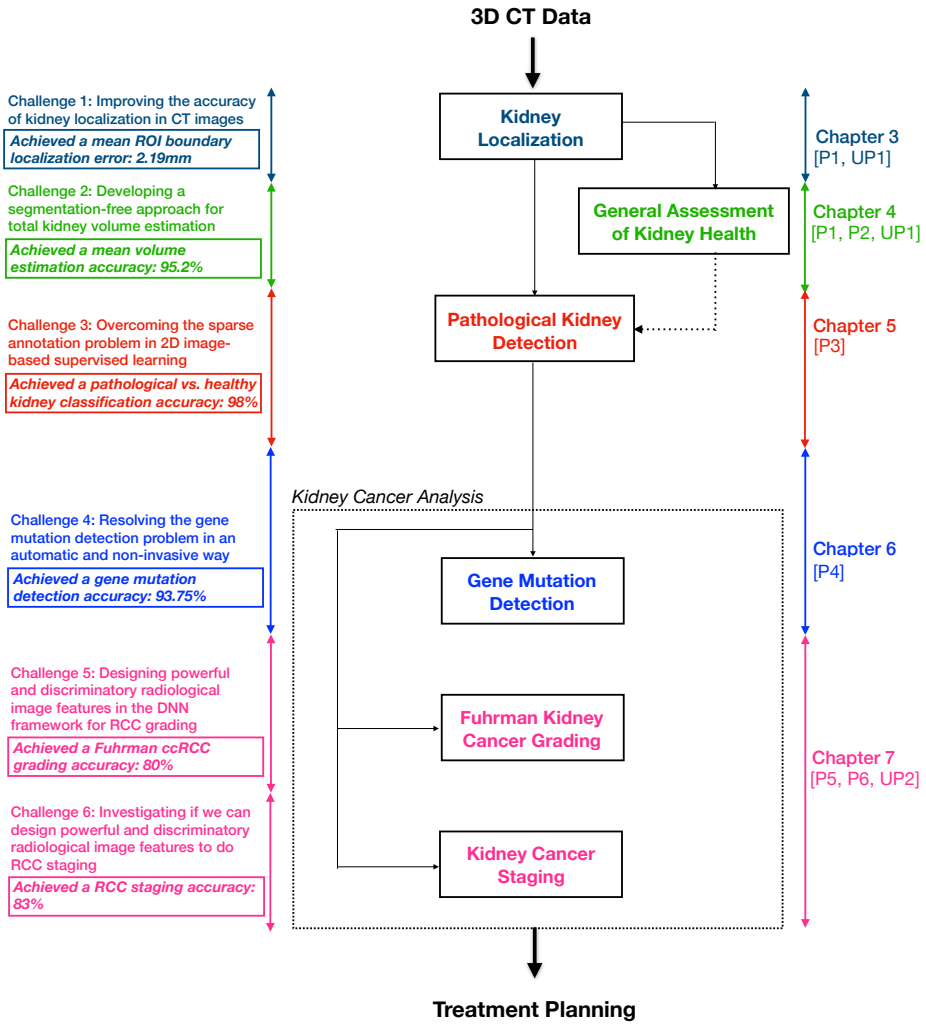

Figure 8.1: A flowchart of our kidney cancer analysis approach with component-wise associated challenges, publications, achieved accuracy and chapter numbers that discuss the technical contributions of this thesis.

8.1.1 Kidney Localization in CT Volume

Chapter 3 presented two methods for automatic kidney localization in the CT image. The first method used an effective deep CNN-based approach for tight kidney ROI localization. The second approach further improved the automatic kidney localization performance than that in the first approach, which used an effective CNN-guided Mask-RCNN approach for efficient kidney localization in the volumetric CT images. Accurate kidney ROI localization is important as it helps to achieve better performance in different kidney analysis tasks in the downstream of our working pipeline. For example, we retrained our CNN-based segmentation-free volume estimation network (discussed in section 4.2.1) with slices taken (a) within the estimated tight ROI and (b) with $\sim$5mm free space around the kidney cross-section. We observed that the volume estimation performance deteriorates by approximately 1.5% when the ROI includes $\sim$5mm free space. A list of our contributions in kidney localization are:

- We produced the 2D slice-level predictions of the presence or absence of kidney cross-section using only a single deep CNN, which was pre-trained on the ImageNet dataset and then fine-tuned on orthogonal 2D CT image slices from all three directions (P1 [1]).

- To reduce the false positives and false negatives from the initial 2D predictions, we combined the 2D CNN-produced probabilities from all three directions into voxel-level decisions, whether it is inside or outside a kidney ROI (P1 [1]).

- We presented a second and improved kidney localization approach by adopting a CNN-guided Mask-RCNN approach (UP1).

- We tackled the challenge of reducing the false positive kidney cross-section predictions by the Mask-RCNN via confining its operation only into a prospective kidney region (UP1).

- Rather than using the regression-based boundary wall predictions, which is not tight enough around a kidney cross-section, we adopted to use the predicted kidney mask to know the kidney boundary. Although the kidney mask

does not correspond to actual kidney contour, it is sufficient enough to provide tight boundary information (UP1).

- We reduced the mean kidney boundary localization error to 2.19 mm (UP1), which is 23% better than those of recent literature.

8.1.2 Segmentation-free Kidney Volume Estimation

After the determination of kidney locations by using any of the methods in chapter 3 within the 3D abdominal CT images, our methods discussed in chapters 4 estimated the kidney volume directly, without requiring the intermediate computationally expensive segmentation step. The first method in chapter 4 used dual regression forests, while other methods in chapter 4 used deep CNN and deep FCN for predicting the anatomical kidney area in a particular image plane. A list of our contributions in kidney volume estimation are:

- We devised a segmentation-free volume estimation approach that bypasses the computationally expensive segmentation step (P2 [2]).

- Our method used dual regression forests, one for predicting the anatomical area in a particular image plane, and another one for boosting the results by removing outliers from the initially estimated areas (P2 [2]).

- We adopted a smaller subpatch-based approach to increase the number of observations, which ultimately improve the results (P2 [2]).

- We also presented two direct kidney volume estimation approaches, P1 [1] and UP1, by using deep learning approaches that learns the image features automatically and demonstrated better kidney volume estimation accuracy than by P2 [2].

- The first method [1] used a deep CNN to predict slice-based cross-sectional kidney areas followed by integration over these values across axial kidney span to produce the volume estimate.

- The second approach (UP1) used a deep FCN instead of deep CNN to predict more accurate slice-based cross-sectional kidney areas than that by [1].

- We achieved a mean volume estimation accuracy of 95.2% (UP1), which is 40% better compared to those of the recent literature.

8.1.3 Pathological Kidney Detection

In chapter 5, we addressed the challenge of using sparsely labeled 2D data in 2D deep learning approaches. Since tissue abnormalities such as tumors, cancers, nodules, etc. are most often localized within a small region of anatomy, localization and analysis of abnormal tissue are typically carried out on the 2D image slices. However, image tags or labels (e.g., healthy, cancerous, etc.) are mostly assigned per image volume or per-patient basis, creating 2D data labels sparse. A typical solution of using sparely labeled data in deep learning is to use the full 3D image volume as a single-instance for learning. However, 3D CNNs are considerably more difficult to train and necessitate the use of expensive GPUs with extensive memory and require a lot more time to converge. We brought a solution to this problem in chapter 5 and our contributions are:

- We proposed a CNN based kidney classification method that makes use of a novel collage image representation (P3 [3]).
- In the collage representation, the 2D image slices in a 3D volume are rearranged side-by-side into a virtual extended 2D image slice, which in turn correctly corresponds to the single available label for that dataset (P3 [3]).
- Our proposed collage also allowed for data augmentation by a random reshuffling of the locations of axial image slices within the collage (P3 [3]).
- We achieved a pathological *vs.* healthy kidney classification accuracy of 98% (P3 [3]).

8.1.4 Detection of Mutated Genes in ccRCC from CT Features

In chapter 6, we discuss a deep multiple instance decision aggregated CNN approach for detecting mutated genes out of the four mostly mutated genes, namely VHL, PBRM1, BAP1, and SETD2, in the ccRCC cases. We have shown how our approach automatically learned discriminating ccRCC features from CT images

and aggregated the binary decisions on the mutation-presence/absence for all the ccRCC slices in a particular kidney sample. A list of our contributions in chapter 6 are:

- We proposed a multiple instance decision aggregation-based deep CNN approach for automatic mutation detection in kidney ccRCC (P4 [4]).
- Our multiple instance decision aggregation approach achieved better accuracy in mutation detection than that of a typical single instance-based approach (P4 [4]).
- Our experimental results demonstrated an approximately 94% accuracy in kidney-wise mutation detection (P4 [4]).

8.1.5 Learnable Image Histogram for RCC Grading and Staging

In chapter 7, we discuss a novel DNN approach [5, 6] that is capable of learning task-specific image-inherent textural features, unlike a conventional CNN approach. We made the bin-center and bin-width of a histogram variable. These are learned with respect to an objective function during the model training. In this way, the trained model can focus on a set of (number of bins) value ranges in the intensity spectrum. Also, bins with variable bin-width and bin-centers can have overlaps among them. A list of our contributions in chapter 7 are:

- We proposed a learnable image histogram (LIH) layer within a DNN framework capable of learning complex and subtle task-specific textural features from raw images directly, adhering to the classical input-output mapping of a CNN (P5 [5], P6 [6], UP2).
- We removed the requirement for fine pre-segmentation of the RCC as the proposed learnable image histogram can stratify tumor and background textures well, thus enabling the model to focus specifically on the tumor texture (P5 [5], P6 [6], UP2).
- We demonstrated ImHistNet's capabilities by performing automatic RCC grade classification for the 2-tiered FGS (P5 [5], UP2).

- We also demonstrated ImHistNet's capability of automatic classification of RCC into anatomical stage low (I/II) and high (III/IV) on an extended clinical dataset from real patients (P6 [6], UP2).

8.2 Potential Impact in Clinical Settings

Although the primary aim of this thesis was to address different technical challenges associated with the current clinical practices of image-based kidney cancer detection and analysis, we believe our works demonstrated considerable potentials to be transferred in the clinical settings to improve patient care. In this thesis, we presented a comprehensive working pipeline for kidney cancer detection and analysis, as visually shown in Fig. 8.1. We broke down this comprehensive working pipeline into several working steps and addressed some critical technical challenges in each step via proposing novel supervised learning approaches and data representations. We believe this whole working pipeline could be very beneficial in clinical practices. For example, our accurate and automatic kidney localization approaches may accelerate rapid kidney health analysis in clinical settings via presenting the background removed region-of-interest around kidneys on the point-of-care computer monitor, saving the time of clinicians in searching for kidneys in the image volume. Further, our segmentation-free and fast total kidney volume estimation approach may provide surrogate renal information, which can help clinicians to identify kidneys with reduced functionality. Our novel collage CNN approach can be beneficial in identifying pathological kidneys in a patient who might have primarily concerned with other diseases. Last but not least, our image features-based noninvasive gene mutation detection, and RCC grading and staging approaches may significantly reduce the laboratory test-based diagnosis time and expenses. These methods may also help physicians in rapid treatment planning, which might be a crucial lifesaver for a patient. Besides, we expect our proposed methods to be easily transferable and practical for analyzing other human abdominal organs, e.g., liver, prostate, heart, etc.

8.3 Future Work

In this thesis, we developed novel supervised learning techniques for improved kidney localization, segmentation-free kidney volume estimation, collage CNN-based kidney cancer detection, CT tumor features-based gene mutation detection, and CT textural features-based cancer grading and staging. Our investigations opened up some new challenges and some potential future works in kidney cancer analysis:

8.3.1 Multi-staging of RCC Using Deep Learning

We discussed in chapter 1 that the American Joint Committee on Cancer (AJCC) and Union for International Cancer Control (UICC) specified the criteria (see Table 8.1) for tumor-node-metastasis (TNM) staging of each cancer depending on the primary tumor size (TX, T0-4); number and location of lymph node involvement (NX, N1-2); and metastatic nature, i.e., tumor spreading to other organs (M0-1) [47, 48]. Clinical guidelines require clinicians to assign TNM stages before initiating any treatment [47], which is typically performed manually. In chapter 7, we discussed a method of automatic 'anatomical' staging of RCC into low (stages I/II) and high (stages III/IV). In the future, we plan to develop a method for the automatic anatomical staging of RCC into four individual groups. However, one of the challenges of doing so would be the lack of data for stages I and IV in the publicly available databases like TCIA [101]. Therefore, we also plan to investigate the feasibility of using 3D Generative Adversarial Network (GAN) to generate artificial data for anatomical stages I and IV.

Table 8.1: Staging of RCC (AJCC/UICC TNM classification of tumors).

Anatomical Stages	**TNM Stages**		
Stage I	T1 (Tumor $\leq$7 cm)	N0	M0
Stage II	T2 (Tumor $>$7 cm but limited to kidney)	N0	M0
Stage III	T1-2, T3 (Tumour extends up to Gerota's fascia)	N1, Any	M0
Stage IV	T4, Any (Tumour invades beyond Gerota's fascia)	Any	M0-1

8.3.2 Survival Analysis of the RCC Patients

Better characterization and understanding of RCC development and progression lead to better diagnosis and clinical outcomes. Recent studies identified a significantly increased mutation frequency of PBRM1 and KDM5C in tumors from male patients and BAP1 from female patients [129]. Mutation of BAP1 had previously been significantly associated with poorer overall survival; however, when stratified by gender, mutation of BAP1 only significantly affected overall survival in female patients [129]. In chapter 7, we showed that gene mutations could be identified from the CT image-based ccRCC features. So, there should be a direct link between the CT image-based ccRCC features to the patient survival rate. Therefore, in the future, we plan to investigate and develop deep learning-based techniques for patient survival prediction and analysis.

8.3.3 Development of a Clinical Software

As we discussed in the previous section that the presented works in this thesis have high potentials to aid in the clinical settings to improve patient care, we plan to develop a clinical software using these thesis works. The software would take a volumetric CT image as input, and would be able to extract the background removed region-of-interest around kidneys, show *total kidney volume* estimate, and identify the presence of any tumor in kidneys. We also expect this software would be able to identify the mutated genes if ccRCC is suspected as well as would be able to indicate the severity of the RCC by predicting its grade and stage from the kidney radiomic analysis. This software would be auxiliary support for physicians in rapid treatment planning.

Bibliography

[1] M. A. Hussain, A. Amir-Khalili, G. Hamarneh, and R. Abugharbieh, "Segmentation-free kidney localization and volume estimation using aggregated orthogonal decision CNN," in *International Conference on Medical Image Computation and Computer Assisted Intervention*, Springer, 2017. → pages vi, vii, 13, 33, 43, 44, 56, 62, 63, 96, 97

[2] M. A. Hussain, G. Hamarneh, T. W. O'Connell, M. F. Mohammed, and R. Abugharbieh, "Segmentation-free estimation of kidney volumes in CT with dual regression forests," in *International Workshop on Machine Learning in Medical Imaging*, pp. 156–163, Springer, 2016. → pages vi, 48, 59, 60, 62, 63, 64, 97

[3] M. A. Hussain, A. Amir-Khalili, G. Hamarneh, and R. Abugharbieh, "Collage CNN for renal cell carcinoma detection from CT," in *International Workshop on Machine Learning in Medical Imaging*, Springer, 2017. → pages vii, 66, 98

[4] M. A. Hussain, G. Hamarneh, and R. Garbi, "Noninvasive determination of gene mutations in clear cell renal cell carcinoma using multiple instance decisions aggregated CNN," in *International Conference on Medical Image Computing and Computer Assisted Intervention*, pp. 657–665, Springer, 2018. → pages vii, 21, 73, 87, 99

[5] M. A. Hussain, G. Hamarneh, and R. Garbi, "ImHistNet: Learnable image histogram based DNN with application to noninvasive determination of carcinoma grades in CT scans," in *International Conference on Medical Image Computing and Computer Assisted Intervention*, pp. 1–8, Springer, 2019. → pages vii, 83, 99

[6] M. A. Hussain, G. Hamarneh, and R. Garbi, "Renal cell carcinoma staging with learnable image histogram-based deep neural network," in

International Workshop on Machine Learning in Medical Imaging, pp. 533–540, Springer, 2019. → pages vii, 83, 99, 100

[7] K. He, G. Gkioxari, P. Dollár, and R. Girshick, "Mask R-CNN," in *Proceedings of the IEEE International Conference on Computer Vision*, pp. 2961–2969, 2017. → pages xvii, 39, 40

[8] E. Gakidou, A. Afshin, A. A. Abajobir, K. H. Abate, C. Abbafati, K. M. Abbas, F. Abd-Allah, A. M. Abdulle, S. F. Abera, V. Aboyans, *et al.*, "Global, regional, and national comparative risk assessment of 84 behavioural, environmental and occupational, and metabolic risks or clusters of risks, 1990–2016: a systematic analysis for the Global Burden of Disease Study 2016," *The Lancet*, vol. 390, no. 10100, pp. 1345–1422, 2017. → pages 1, 2

[9] F. Bray, J. Ferlay, I. Soerjomataram, R. L. Siegel, L. A. Torre, and A. Jemal, "Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries," *CA: A Cancer Journal for Clinicians*, vol. 68, no. 6, pp. 394–424, 2018. → page 1

[10] C. C. S. A. Committee *et al.*, "Canadian cancer statistics: a 2018 special report on cancer incidence by stage [internet]. Toronto: Canadian Cancer Society; 2018 [cited 2019 jul 3]." → page 1

[11] E. L. Yanik, C. A. Clarke, J. J. Snyder, R. M. Pfeiffer, and E. A. Engels, "Variation in cancer incidence among patients with ESRD during kidney function and nonfunction intervals," *Journal of the American Society of Nephrology*, vol. 27, no. 5, pp. 1495–1504, 2016. → page 1

[12] V. Wang, H. Vilme, M. L. Maciejewski, and L. E. Boulware, "The economic burden of chronic kidney disease and end-stage renal disease," in *Seminars in Nephrology*, vol. 36, pp. 319–330, Elsevier, 2016. → pages 1, 2, 12

[13] S. K. Kang, L. D. Scherer, A. J. Megibow, L. J. Higuita, N. Kim, R. S. Braithwaite, and A. Fagerlin, "A randomized study of patient risk perception for incidental renal findings on diagnostic imaging tests," *American Journal of Roentgenology*, vol. 210, no. 2, pp. 369–375, 2018. → pages 2, 17

[14] C. Fitzmaurice, C. Allen, R. M. Barber, L. Barregard, Z. A. Bhutta, H. Brenner, D. J. Dicker, O. Chimed-Orchir, R. Dandona, L. Dandona, *et al.*, "Global, regional, and national cancer incidence, mortality, years of

life lost, years lived with disability, and disability-adjusted life-years for 32 cancer groups, 1990 to 2015: a systematic analysis for the Global Burden of Disease Study," *JAMA Oncology*, vol. 3, no. 4, pp. 524–548, 2017. → pages 2, 7, 94

[15] U. Capitanio and F. Montorsi, "Renal cancer," *The Lancet*, vol. 387, no. 10021, pp. 894–906, 2016. → page 3

[16] E. Limkin, R. Sun, L. Dercle, E. Zacharaki, C. Robert, S. Reuzé, A. Schernberg, N. Paragios, E. Deutsch, and C. Ferté, "Promises and challenges for the implementation of computational medical imaging (radiomics) in oncology," *Annals of Oncology*, vol. 28, no. 6, pp. 1191–1206, 2017. → page 3

[17] R. J. Gillies, P. E. Kinahan, and H. Hricak, "Radiomics: images are more than pictures, they are data," *Radiology*, vol. 278, no. 2, pp. 563–577, 2015. → page 3

[18] C. G. A. R. Network *et al.*, "Comprehensive molecular characterization of clear cell renal cell carcinoma," *Nature*, vol. 499, no. 7456, p. 43, 2013. → pages 3, 9, 19

[19] L. K. Lee, S. C. Liew, and W. J. Thong, "A review of image segmentation methodologies in medical image," in *Advanced Computer and Communication Engineering Technology*, pp. 1069–1080, Springer, 2015. → page 3

[20] G. Litjens, T. Kooi, B. E. Bejnordi, A. A. A. Setio, F. Ciompi, M. Ghafoorian, J. A. Van Der Laak, B. Van Ginneken, and C. I. Sánchez, "A survey on deep learning in medical image analysis," *Medical Image Analysis*, vol. 42, pp. 60–88, 2017. → page 4

[21] K. M. Krajewski and I. Pedrosa, "Imaging advances in the management of kidney cancer," *Journal of Clinical Oncology*, vol. 36, no. 36, pp. 3582–3590, 2018. → page 5

[22] B. Ljungberg, K. Bensalah, S. Canfield, S. Dabestani, F. Hofmann, M. Hora, M. A. Kuczyk, T. Lam, L. Marconi, A. S. Merseburger, *et al.*, "EAU guidelines on renal cell carcinoma: 2014 update," *European Urology*, vol. 67, no. 5, pp. 913–924, 2015. → page 5

[23] D. B. Rukstalis, J. Simmons, and P. F. Fulgham, "Renal ultrasound," in *Practical Urological Ultrasound*, pp. 51–76, Springer, 2017. → pages 5, 6

[24] T. J. van Oostenbrugge, J. J. Fütterer, and P. F. Mulders, "Diagnostic imaging for solid renal tumors: A pictorial review," *Kidney Cancer*, no. Preprint, pp. 1–15, 2018. → page 6

[25] X. Xu, F. Zhou, B. Liu, D. Fu, and X. Bai, "Efficient multiple organ localization in CT image using 3D region proposal network," *IEEE Transactions on Medical Imaging*, 2019. → pages 7, 12, 13, 43, 44

[26] M. Regier and F. Chun, "Thermal ablation of renal tumors: indications, techniques and results," *Deutsches Ärzteblatt International*, vol. 112, no. 24, p. 412, 2015. → pages 7, 12

[27] C. Brace, "Thermal tumor ablation in clinical use," *IEEE Pulse*, vol. 2, no. 5, pp. 28–38, 2011. → page 7

[28] J. Wang and D. Fleischmann, "Improving spatial resolution at ct: Development, benefits, and pitfalls," 2018. → page 8

[29] A. Diez, J. Powelson, C. P. Sundaram, T. E. Taber, M. A. Mujtaba, M. S. Yaqub, D. P. Mishler, W. C. Goggins, and A. A. Sharfuddin, "Correlation between CT-based measured renal volumes and nuclear-renography-based split renal function in living kidney donors. Clinical diagnostic utility and practice patterns," *Clinical Transplantation*, vol. 28, no. 6, pp. 675–682, 2014. → pages 8, 12, 13

[30] E. Widjaja, J. Oxtoby, T. Hale, P. Jones, P. Harden, and I. McCall, "Ultrasound measured renal length versus low dose CT volume in predicting single kidney glomerular filtration rate," *The British Journal of Radiology*, vol. 77, no. 921, pp. 759–764, 2004. → pages 8, 13, 48

[31] M.-A. Carbonneau, V. Cheplygina, E. Granger, and G. Gagnon, "Multiple instance learning: A survey of problem characteristics and applications," *Pattern Recognition*, vol. 77, pp. 329–353, 2018. → pages 8, 18

[32] W. Zhu, Q. Lou, Y. S. Vang, and X. Xie, "Deep multi-instance networks with sparse label assignment for whole mammogram classification," *ArXiv Preprint arXiv:1612.05968*, 2016. → pages 8, 18

[33] O. Z. Kraus, J. L. Ba, and B. J. Frey, "Classifying and segmenting microscopy images with deep multiple instance learning," *Bioinformatics*, vol. 32, no. 12, pp. i52–i59, 2016. → page 18

[34] Z. Yan, Y. Zhan, Z. Peng, S. Liao, Y. Shinagawa, S. Zhang, D. N. Metaxas, and X. S. Zhou, "Multi-instance deep learning: Discover discriminative local anatomies for bodypart recognition," *IEEE Transactions on Medical Imaging*, vol. 35, no. 5, pp. 1332–1343, 2016. → pages 8, 18

[35] J. Gao, B. A. Aksoy, U. Dogrusoz, G. Dresdner, B. Gross, S. O. Sumer, Y. Sun, A. Jacobsen, *et al.*, "Integrative analysis of complex cancer genomics and clinical profiles using the cBioPortal," *Sci. Signal.*, vol. 6, no. 269, pp. pl1–pl1, 2013. → pages 9, 19, 27

[36] A. B. Shinagare, R. Vikram, C. Jaffe, O. Akin, J. Kirby, E. Huang, J. Freymann, N. I. Sainani, C. A. Sadow, T. K. Bathala, *et al.*, "Radiogenomics of clear cell renal cell carcinoma: preliminary findings of The Cancer Genome Atlas–Renal Cell Carcinoma (TCGA–RCC) Imaging Research Group," *Abdominal Imaging*, vol. 40, no. 6, pp. 1684–1692, 2015. → pages 9, 19, 67, 74, 78

[37] C. A. Karlo, P. L. Di Paolo, J. Chaim, A. A. Hakimi, I. Ostrovnaya, P. Russo, H. Hricak, R. Motzer, J. J. Hsieh, and O. Akin, "Radiogenomics of clear cell renal cell carcinoma: associations between CT imaging features and mutations," *Radiology*, vol. 270, no. 2, pp. 464–471, 2014. → pages 9, 19, 74

[38] J. Ding, Z. Xing, Z. Jiang, J. Chen, L. Pan, J. Qiu, and W. Xing, "CT-based radiomic model predicts high grade of clear cell renal cell carcinoma," *European Journal of Radiology*, vol. 103, pp. 51–56, 2018. → pages 9, 20, 21, 94

[39] S. Oh, D. J. Sung, K. S. Yang, K. C. Sim, N. Y. Han, B. J. Park, M. J. Kim, and S. B. Cho, "Correlation of CT imaging features and tumor size with fuhrman grade of clear cell renal cell carcinoma," *Acta Radiologica*, vol. 58, no. 3, pp. 376–384, 2017. → pages 9, 20

[40] K. Sasaguri and N. Takahashi, "CT and MR imaging for solid renal mass characterization," *European Journal of Radiology*, vol. 99, pp. 40–54, 2018. → page 20

[41] J. Shu, Y. Tang, J. Cui, R. Yang, X. Meng, Z. Cai, J. Zhang, W. Xu, D. Wen, and H. Yin, "Clear cell renal cell carcinoma: CT-based radiomics features for the prediction of Fuhrman grade," *European Journal of Radiology*, vol. 109, pp. 8–12, 2018. → pages 20, 21, 89, 91

[42] H. Huhdanpaa, D. Hwang, S. Cen, B. Quinn, M. Nayyar, X. Zhang, F. Chen, B. Desai, G. Liang, I. Gill, *et al.*, "CT prediction of the fuhrman grade of clear cell renal cell carcinoma (RCC): towards the development of computer-assisted diagnostic method," *Abdominal Imaging*, vol. 40, no. 8, pp. 3168–3174, 2015. → pages 9, 20, 21

[43] Z. Wang, H. Li, W. Ouyang, and X. Wang, "Learnable histogram: Statistical context features for deep neural networks," in *European Conference on Computer Vision*, pp. 246–262, Springer, 2016. → pages 9, 21, 55

[44] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, no. 7553, p. 436, 2015. → pages 10, 21

[45] V. Andrearczyk and P. F. Whelan, "Using filter banks in convolutional neural networks for texture classification," *Pattern Recognition Letters*, vol. 84, pp. 63–69, 2016. → pages 10, 21

[46] M. Janssen, J. Linxweiler, S. Terwey, S. Rugge, C.-H. Ohlmann, F. Becker, C. Thomas, A. Neisius, J. Thüroff, S. Siemer, *et al.*, "Survival outcomes in patients with large ($\geq$ 7cm) clear cell renal cell carcinomas treated with nephron-sparing surgery versus radical nephrectomy: Results of a multicenter cohort with long-term follow-up," *PloS One*, vol. 13, no. 5, p. e0196427, 2018. → pages 10, 21, 22

[47] A. K. AAlAbdulsalam, J. H. Garvin, A. Redd, M. E. Carter, C. Sweeny, and S. M. Meystre, "Automated extraction and classification of cancer stage mentions from unstructured text fields in a central cancer registry," *AMIA Summits on Translational Science Proceedings*, vol. 2018, p. 16, 2018. → pages 10, 22, 101

[48] B. Escudier, C. Porta, M. Schmidinger, N. Rioux-Leclercq, A. Bex, V. Khoo, V. Gruenvald, and A. Horwich, "Renal cell carcinoma: ESMO clinical practice guidelines for diagnosis, treatment and follow-up," *Annals of Oncology*, vol. 27, no. suppl_5, pp. v58–v68, 2016. → pages 10, 22, 23, 101

[49] A. Bradley, L. MacDonald, S. Whiteside, R. Johnson, and V. Ramani, "Accuracy of preoperative CT T staging of renal cell carcinoma: which features predict advanced stage?," *Clinical Radiology*, vol. 70, no. 8, pp. 822–829, 2015. → pages 10, 22, 23

[50] G. Y. Dai, Z. C. Li, J. Gu, L. Wang, X. M. Li, and Y. Q. Xie, "Segmentation of kidneys from computed tomography using 3d fast growcut algorithm," in *Applied Mechanics and Materials*, vol. 333, pp. 1145–1150, Trans Tech Publ, 2013. → pages 11, 14, 16

[51] F. Khalifa, A. Elnakib, G. M. Beache, G. Gimel'farb, M. A. El-Ghar, R. Ouseph, G. Sokhadze, S. Manning, P. McClure, and A. El-Baz, "3D kidney segmentation from CT images using a level set approach guided by a novel stochastic speed function," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pp. 587–594, Springer, 2011. → pages 14, 16

[52] A. Skalski, K. Heryan, J. Jakubowski, and T. Drewniak, "Kidney segmentation in CT data using hybrid level-set method with ellipsoidal shape constraints," *Metrology and Measurement Systems*, vol. 24, no. 1, pp. 101–112, 2017. → pages 11, 14, 15, 16

[53] G. Yan and B. Wang, "An automatic kidney segmentation from abdominal CT images," in *Intelligent Computing and Intelligent Systems (ICIS), 2010 IEEE International Conference on*, vol. 1, pp. 280–284, IEEE, 2010. → pages 11, 13, 14, 16

[54] X. Li, X. Chen, J. Yao, X. Zhang, and J. Tian, "Renal cortex segmentation using optimal surface search with novel graph construction," in *International Conference on Medical Image Computing and Computer Assisted Intervention*, pp. 387–394, Springer, 2011. → pages 11, 13, 14, 16

[55] X. Chen, R. M. Summers, M. Cho, U. Bagci, and J. Yao, "An automatic method for renal cortex segmentation on CT images: evaluation on kidney donors," *Academic Radiology*, vol. 19, no. 5, pp. 562–570, 2012. → pages 11, 13, 14, 16

[56] D. Xiang, X. Chen, and C. Jin, "Fast renal cortex localization by combining generalized Hough transform and active appearance models," in *International MICCAI Workshop on Computational and Clinical Challenges in Abdominal Imaging*, pp. 175–183, Springer, 2013. → pages 11, 13

[57] C. Jin, D. Xiang, and X. Chen, "Renal cortex localization by combining 3D generalized Hough transform and 3D active appearance models," in *Biomedical Imaging (ISBI), 2014 IEEE 11th International Symposium on*, pp. 1275–1278, IEEE, 2014. → pages 11, 13

[58] A. Criminisi, J. Shotton, D. P. Robertson, and E. Konukoglu, "Regression rorests for efficient anatomy detection and localization in CT studies," *MCV*, vol. 2010, pp. 106–117, 2010. → pages 11, 13, 36, 37, 43

[59] A. Criminisi, D. Robertson, E. Konukoglu, J. Shotton, S. Pathak, S. White, and K. Siddiqui, "Regression forests for efficient anatomy detection and localization in computed tomography scans," *Medical Image Analysis*, vol. 17, no. 8, pp. 1293–1303, 2013. → pages 11, 13, 36, 37, 43, 64

[60] R. Cuingnet, R. Prevost, D. Lesage, L. D. Cohen, B. Mory, and R. Ardon, "Automatic detection and segmentation of kidneys in 3D CT images using random forests," in *International Conference on Medical Image Computing and Computer Assisted Intervention*, pp. 66–74, Springer, 2012. → pages 11, 13, 14, 15, 16, 36, 37, 43, 51, 59, 60, 64

[61] R. Gauriau, R. Cuingnet, D. Lesage, and I. Bloch, "Multi-organ localization with cascaded global-to-local regression and shape prior," *Medical Image Analysis*, vol. 23, no. 1, pp. 70–83, 2015. → pages 11, 13, 43

[62] X. Zhou, A. Watanabe, X. Zhou, T. Hara, R. Yokoyama, M. Kanematsu, and H. Fujita, "Automatic organ segmentation on torso CT images by using content-based image retrieval," in *Medical Imaging 2012: Image Processing*, vol. 8314, p. 83143E, International Society for Optics and Photonics, 2012. → pages 11, 13, 15, 16

[63] P. N. Samarakoon, E. Promayon, and C. Fouard, "Light random regression forests for automatic multi-organ localization in CT images," in *Biomedical Imaging (ISBI 2017), 2017 IEEE 14th International Symposium on*, pp. 371–374, IEEE, 2017. → pages 11, 13, 43, 44

[64] X. Zhou, S. Morita, X. Zhou, H. Chen, T. Hara, R. Yokoyama, M. Kanematsu, H. Hoshi, and H. Fujita, "Automatic anatomy partitioning of the torso region on CT images by using multiple organ localizations with a group-wise calibration technique," in *Medical Imaging 2015: Computer-Aided Diagnosis*, vol. 9414, p. 94143K, International Society for Optics and Photonics, 2015. → pages 12, 13, 15, 16

[65] X. Zhou, T. Ito, X. Zhou, H. Chen, T. Hara, R. Yokoyama, M. Kanematsu, H. Hoshi, and H. Fujita, "A universal approach for automatic organ segmentations on 3D CT images based on organ localization and 3D GrabCut," in *Medical Imaging 2014: Computer-Aided Diagnosis*, vol. 9035, p. 90352V, International Society for Optics and Photonics, 2014. → pages 11, 12, 13, 15, 16

[66] G. E. Humpire-Mamani, A. A. A. Setio, B. van Ginneken, and C. Jacobs, "Efficient organ localization using multi-label convolutional neural networks in thorax-abdomen CT scans," *Physics in Medicine & Biology*, vol. 63, no. 8, p. 085003, 2018. → pages 12, 13, 14, 43, 44

[67] G. E. H. Mamani, A. A. A. Setio, B. van Ginneken, and C. Jacobs, "Organ detection in thorax abdomen CT using multi-label convolutional neural networks," in *Medical Imaging 2017: Computer-Aided Diagnosis*, vol. 10134, p. 1013416, International Society for Optics and Photonics, 2017. → pages 12, 13

[68] X. Lu, D. Xu, and D. Liu, "Robust 3D organ localization with dual learning architectures and fusion," in *International Workshop on Large-Scale Annotation of Biomedical Data and Expert Label Synthesis*, pp. 12–20, Springer, 2016. → pages 12, 13, 36, 37

[69] N. Tangri, I. Hougen, A. Alam, R. Perrone, P. McFarlane, and Y. Pei, "Total kidney volume as a biomarker of disease progression in autosomal dominant polycystic kidney disease," *Canadian Journal of Kidney Health and Disease*, vol. 4, p. 2054358117693355, 2017. → pages 13, 14

[70] T. Okada, M. G. Linguraru, Y. Yoshida, M. Hori, R. M. Summers, Y.-W. Chen, N. Tomiyama, and Y. Sato, "Abdominal multi-organ segmentation of CT images based on hierarchical spatial modeling of organ interrelations," in *International MICCAI Workshop on Computational and Clinical Challenges in Abdominal Imaging*, pp. 173–180, Springer, 2011. → page 14

[71] R. Wolz, C. Chu, K. Misawa, K. Mori, and D. Rueckert, "Multi-organ abdominal CT segmentation using hierarchically weighted subject-specific atlases," in *International Conference on Medical Image Computing and Computer Assisted Intervention*, pp. 10–17, Springer, 2012. → pages 15, 16

[72] R. Wolz, C. Chu, K. Misawa, M. Fujiwara, K. Mori, and D. Rueckert, "Automated abdominal multi-organ segmentation with subject-specific atlas generation," *IEEE transactions on medical imaging*, vol. 32, no. 9, pp. 1723–1730, 2013. → pages 15, 16

[73] S. Chen, H. Roth, S. Dorn, M. May, A. Cavallaro, M. M. Lell, M. Kachelrieß, H. Oda, K. Mori, and A. Maier, "Towards automatic abdominal multi-organ segmentation in dual energy CT using cascaded 3D

fully convolutional network," *ArXiv Preprint arXiv:1710.05379*, 2017. → pages 15, 16

[74] E. Hristova, H. Schulz, T. Brosch, M. P. Heinrich, and H. Nickisch, "Nearest neighbor 3D segmentation with context features," in *Medical Imaging 2018: Image Processing*, vol. 10574, p. 105740M, International Society for Optics and Photonics, 2018. → pages 15, 16

[75] E. Gibson, F. Giganti, Y. Hu, E. Bonmati, S. Bandula, K. Gurusamy, B. Davidson, S. P. Pereira, M. J. Clarkson, and D. C. Barratt, "Automatic multi-organ segmentation on abdominal CT with dense v-networks," *IEEE Transactions on Medical Imaging*, 2018. → pages 15, 16

[76] V. V. Valindria, N. Pawlowski, M. Rajchl, I. Lavdas, E. O. Aboagye, A. G. Rockall, D. Rueckert, and B. Glocker, "Multi-modal learning from unpaired images: Application to multi-organ segmentation in CT and MRI," in *Applications of Computer Vision (WACV), 2018 IEEE Winter Conference on*, pp. 547–556, IEEE, 2018. → pages 15, 16

[77] F. Khalifa, A. Soliman, A. Elmaghraby, G. Gimel'farb, and A. El-Baz, "3D kidney segmentation from abdominal images using spatial-appearance models," *Computational and Mathematical Methods in Medicine*, vol. 2017, 2017. → pages 15, 16

[78] V. Groza, T. Brosch, D. Eschweiler, H. Schulz, S. Renisch, and H. Nickisch, "Comparison of deep learning-based techniques for organ segmentation in abdominal CT images," in *1st Conference on Medical Imaging with Deep Learning (MIDL 2018), Amsterdam, The Netherlands*, pp. 1–3, 2018. → pages 15, 16

[79] W. Thong, S. Kadoury, N. Piché, and C. J. Pal, "Convolutional networks for kidney segmentation in contrast-enhanced CT scans," *Computer Methods in Biomechanics and Biomedical Engineering: Imaging & Visualization*, vol. 6, no. 3, pp. 277–282, 2018.

[80] D. Keshwani, Y. Kitamura, and Y. Li, "Computation of total kidney volume from CT images in autosomal dominant polycystic kidney disease using multi-task 3D convolutional neural networks," in *International Workshop on Machine Learning in Medical Imaging*, pp. 380–388, Springer, 2018. → pages 15, 16, 17

[81] X. Liu, S. Guo, B. Yang, S. Ma, H. Zhang, J. Li, C. Sun, L. Jin, X. Li, Q. Yang, *et al.*, "Automatic organ segmentation for CT scans based on

super-pixel and convolutional neural networks," *Journal of Digital Imaging*, pp. 1–13, 2018.

[82] F. Zhao, P. Gao, H. Hu, X. He, Y. Hou, and X. He, "Efficient kidney segmentation in micro-CT based on multi-atlas registration and random forests," *IEEE Access*, vol. 6, pp. 43712–43723, 2018. → pages 15, 16

[83] W. Wieclawek, "3D marker-controlled watershed for kidney segmentation in clinical CT exams," *Biomedical Engineering Online*, vol. 17, no. 1, p. 26, 2018. → pages 14, 15, 16

[84] B. Glocker, O. Pauly, E. Konukoglu, and A. Criminisi, "Joint classification-regression forests for spatially structured multi-object segmentation," *Computer Vision–ECCV 2012*, pp. 870–881, 2012. → pages 15, 16, 51

[85] W. Thong, S. Kadoury, N. Piché, and C. J. Pal, "Convolutional networks for kidney segmentation in contrast-enhanced CT scans," *Computer Methods in Biomechanics and Biomedical Engineering: Imaging & Visualization*, pp. 1–6, 2016. → pages 15, 16

[86] K. Sharma, C. Rupprecht, A. Caroli, M. C. Aparicio, A. Remuzzi, M. Baust, and N. Navab, "Automatic segmentation of kidneys using deep learning for total kidney volume quantification in autosomal dominant polycystic kidney disease," *Scientific Reports*, vol. 7, no. 1, p. 2049, 2017. → pages 15, 16, 17

[87] Ö. Çiçek, A. Abdulkadir, S. S. Lienkamp, T. Brox, and O. Ronneberger, "3D U-Net: learning dense volumetric segmentation from sparse annotation," in *International Conference on Medical Image Computing and Computer Assisted Intervention*, pp. 424–432, Springer, 2016. → pages 16, 17, 18, 63, 65, 72

[88] F. Milletari, N. Navab, and S.-A. Ahmadi, "V-net: Fully convolutional neural networks for volumetric medical image segmentation," in *2016 Fourth International Conference on 3D Vision (3DV)*, pp. 565–571, IEEE, 2016. → pages 16, 17, 18

[89] N. Heller, N. Sathianathen, A. Kalapara, E. Walczak, K. Moore, H. Kaluzniak, J. Rosenberg, P. Blake, Z. Rengel, M. Oestreich, J. Dean, M. Tradewell, A. Shah, R. Tejpaul, Z. Edgerton, M. Peterson, S. Raza, S. Regmi, N. Papanikolopoulos, and C. Weight, "The KiTS19 Challenge

Data: 300 kidney tumor cases with clinical context, ct semantic segmentations, and surgical outcomes," 2019. → page 17

[90] Y. Xu, T. Mo, Q. Feng, P. Zhong, M. Lai, I. Eric, and C. Chang, "Deep learning of feature representation with multiple instance learning for medical image analysis," in *Acoustics, Speech and Signal Processing (ICASSP), 2014 IEEE International Conference on*, pp. 1626–1630, IEEE, 2014. → page 18

[91] R. L. Siegel, K. D. Miller, and A. Jemal, "Cancer statistics, 2015," *CA: a cancer journal for clinicians*, vol. 65, no. 1, pp. 5–29, 2015. → page 19

[92] C. A. Ridge, B. B. Pua, and D. C. Madoff, "Epidemiology and staging of renal cell carcinoma," in *Seminars in Interventional Radiology*, vol. 31, pp. 003–008, Thieme Medical Publishers, 2014. → page 19

[93] G. Guo, Y. Gui, S. Gao, A. Tang, X. Hu, Y. Huang, W. Jia, *et al.*, "Frequent mutations of genes encoding ubiquitin-mediated proteolysis pathway components in clear cell renal cell carcinoma," *Nature Genetics*, vol. 44, no. 1, p. 17, 2012. → page 19

[94] M. Incoronato, M. Aiello, T. Infante, C. Cavaliere, A. Grimaldi, P. Mirabelli, S. Monti, and M. Salvatore, "Radiogenomic analysis of oncological data: a technical survey," *International Journal of Molecular Sciences*, vol. 18, no. 4, p. 805, 2017. → page 19

[95] B. Kocak, E. S. Durmaz, E. Ates, and M. B. Ulusan, "Radiogenomics in clear cell renal cell carcinoma: machine learning–based high-dimensional quantitative CT texture analysis in predicting PBRM1 mutation status," *American Journal of Roentgenology*, vol. 212, no. 3, pp. W55–W63, 2019. → page 20

[96] K. Ishigami, L. V. Leite, M. G. Pakalniskis, D. K. Lee, D. G. Holanda, and D. M. Kuehn, "Tumor grade of clear cell renal cell carcinoma assessed by contrast-enhanced computed tomography," *SpringerPlus*, vol. 3, no. 1, p. 694, 2014. → page 20

[97] S. A. Fuhrman, L. C. Lasky, and C. Limas, "Prognostic significance of morphologic parameters in renal cell carcinoma.," *The American Journal of Surgical Pathology*, vol. 6, no. 7, pp. 655–663, 1982. → page 20

[98] A. Becker, D. Hickmann, J. Hansen, C. Meyer, M. Rink, M. Schmid, C. Eichelberg, K. Strini, T. Chromecki, J. Jesche, *et al.*, "Critical analysis

of a simplified fuhrman grading scheme for prediction of cancer specific mortality in patients with clear cell renal cell carcinoma–impact on prognosis," *European Journal of Surgical Oncology (EJSO)*, vol. 42, no. 3, pp. 419–425, 2016. → page 20

[99] H.-J. Tan, E. C. Norton, Z. Ye, K. S. Hafez, J. L. Gore, and D. C. Miller, "Long-term survival following partial vs radical nephrectomy among older patients with early-stage kidney cancer," *Jama*, vol. 307, no. 15, pp. 1629–1635, 2012. → page 22

[100] P. H. Shah, D. M. Moreira, V. R. Patel, G. Gaunay, A. K. George, M. Alom, Z. Kozel, O. Yaskiv, S. J. Hall, M. J. Schwartz, *et al.*, "Partial nephrectomy is associated with higher risk of relapse compared with radical nephrectomy for clinical stage T1 renal cell carcinoma pathologically up staged to T3a," *The Journal of Urology*, vol. 198, no. 2, pp. 289–296, 2017. → page 23

[101] K. Clark, B. Vendt, K. Smith, J. Freymann, J. Kirby, P. Koppel, S. Moore, S. Phillips, D. Maffitt, M. Pringle, *et al.*, "The Cancer Imaging Archive (TCIA): maintaining and operating a public information repository," *Journal of Digital Imaging*, vol. 26, no. 6, pp. 1045–1057, 2013. → pages 27, 101

[102] N. Heller, N. Sathianathen, A. Kalapara, E. Walczak, K. Moore, H. Kaluzniak, J. Rosenberg, P. Blake, Z. Rengel, M. Oestreich, *et al.*, "The KiTS19 Challenge Data: 300 kidney tumor cases with clinical context, ct semantic segmentations, and surgical outcomes," *ArXiv Preprint arXiv:1904.00445*, 2019. → page 28

[103] 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*, pp. 843–852, 2017. → page 28

[104] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei, "ImageNet: A Large-Scale Hierarchical Image Database," in *CVPR09*, 2009. → pages 29, 34

[105] J. Yosinski, J. Clune, Y. Bengio, and H. Lipson, "How transferable are features in deep neural networks?," in *Advances in Neural Information Processing Systems*, pp. 3320–3328, 2014. → page 29

[106] G. E. Hinton, N. Srivastava, A. Krizhevsky, I. Sutskever, and R. R. Salakhutdinov, "Improving neural networks by preventing co-adaptation of feature detectors," *ArXiv Preprint arXiv:1207.0580*, 2012. → page 29

[107] Y. Jia, E. Shelhamer, J. Donahue, S. Karayev, J. Long, R. Girshick, S. Guadarrama, and T. Darrell, "Caffe: Convolutional architecture for fast feature embedding," in *Proceedings of the 22nd ACM international conference on Multimedia*, pp. 675–678, ACM, 2014. → pages 35, 42, 58, 70, 88

[108] S. Ren, K. He, R. Girshick, and J. Sun, "Faster R-CNN: towards real-time object detection with region proposal networks," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 39, no. 6, pp. 1137–1149, 2017. → page 38

[109] R. Girshick, J. Donahue, T. Darrell, and J. Malik, "Region-based convolutional networks for accurate object detection and segmentation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 38, no. 1, pp. 142–158, 2016. → page 38

[110] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 770–778, 2016. → pages 38, 40, 41, 88, 91, 92

[111] K.-Y. Kang, Y. J. Lee, S. C. Park, C. W. Yang, Y.-S. Kim, I. S. Moon, Y. B. Koh, B. K. Bang, and B. S. Choi, "A comparative study of methods of estimating kidney length in kidney transplantation donors," *Nephrology Dialysis Transplantation*, vol. 22, no. 8, pp. 2322–2327, 2007. → page 40

[112] W. Abdulla, "Mask R-CNN for object detection and instance segmentation on keras and tensorflow." https://github.com/matterport/Mask_RCNN, 2017. → page 42

[113] W. Thong, S. Kadoury, N. Piché, and C. J. Pal, "Convolutional networks for kidney segmentation in contrast-enhanced CT scans," *Computer Methods in Biomechanics and Biomedical Engineering: Imaging & Visualization*, pp. 1–6, 2016. → page 51

[114] Y. Zhang, B. J. Matuszewski, L.-K. Shark, and C. J. Moore, "Medical image segmentation using new hybrid level-set method," in *BioMedical Visualization, 2008. MEDIVIS'08. Fifth International Conference*, pp. 71–76, IEEE, 2008. → pages 52, 53

[115] N. Zakhari, B. Blew, and W. Shabana, "Simplified method to measure renal volume: the best correction factor for the ellipsoid formula volume calculation in pre-transplant computed tomographic live donor," *Urology*, vol. 83, no. 6, pp. 1444–e15, 2014. → pages 52, 59, 60, 63

[116] X. Zhen, Z. Wang, A. Islam, M. Bhaduri, I. Chan, and S. Li, "Direct estimation of cardiac bi-ventricular volumes with regression forests.," in *International Conference on Medical Image Computing and Computer Assisted Intervention*, pp. 586–593, 2014. → pages 52, 54, 59, 60, 62, 63

[117] G. Yang, J. Gu, Y. Chen, W. Liu, L. Tang, H. Shu, and C. Toumoulin, "Automatic kidney segmentation in CT images based on multi-atlas image registration," in *Engineering in Medicine and Biology Society (EMBC), 2014 36th Annual International Conference of the IEEE*, pp. 5538–5541, IEEE, 2014. → pages 59, 60

[118] X. Wan, W. Wang, J. Liu, and T. Tong, "Estimating the sample mean and standard deviation from the sample size, median, range and/or interquartile range," *BMC Medical Research Methodology*, vol. 14, no. 1, p. 135, 2014. → page 60

[119] P. Sermanet, D. Eigen, X. Zhang, M. Mathieu, R. Fergus, and Y. LeCun, "Overfeat: Integrated recognition, localization and detection using convolutional networks," *ArXiv Preprint arXiv:1312.6229*, 2013. → page 69

[120] Y. Sun, X. Wang, and X. Tang, "Deep learning face representation from predicting 10,000 classes," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1891–1898, 2014. → page 69

[121] D. Kingma and J. Ba, "Adam: A method for stochastic optimization," *ArXiv Preprint arXiv:1412.6980*, 2014. → page 70

[122] M. D. Zeiler and R. Fergus, "Visualizing and understanding convolutional networks," in *European Conference on Computer Vision*, pp. 818–833, Springer, 2014. → pages 75, 86

[123] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," *arXiv preprint arXiv:1412.6980*, 2014. → page 77

[124] H. Wang, M. Moradi, Y. Gur, P. Prasanna, and T. Syeda-Mahmood, "A multi-atlas approach to region of interest detection for medical image

classification," in *International Conference on Medical Image Computing and Computer Assisted Intervention*, pp. 168–176, Springer, 2017. → pages 78, 80

[125] H. J. Aerts, "The potential of radiomic-based phenotyping in precision medicine: a review," *JAMA Oncology*, vol. 2, no. 12, pp. 1636–1642, 2016. → page 87

[126] F. Meng, X. Li, G. Zhou, and Y. Wang, "Fuhrman grade classification of clear-cell renal cell carcinoma using computed tomography image analysis," *Journal of Medical Imaging and Health Informatics*, vol. 7, no. 7, pp. 1671–1676, 2017. → page 89

[127] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," in *Advances in Neural Information Processing Systems*, pp. 1097–1105, 2012. → page 90

[128] R. L. Siegel, K. D. Miller, and A. Jemal, "Cancer statistics, 2016," *CA: A Cancer Journal for Clinicians*, vol. 66, no. 1, pp. 7–30, 2016. → page 94

[129] C. J. Ricketts and W. M. Linehan, "Gender specific mutation incidence and survival associations in clear cell renal cell carcinoma (ccRCC)," *PloS One*, vol. 10, no. 10, p. e0140257, 2015. → page 102