University of Louisville

ThinkIR: The University of Louisville's Institutional Repository

Electronic Theses and Dissertations

8-2022

Role of deep learning techniques in non-invasive diagnosis of human diseases.

Hisham Abouelseoud Elsayem Abdeltawab
University of Louisville

ROLE OF DEEP LEARNING TECHNIQUES IN NON-INVASIVE DIAGNOSIS OF
HUMAN DISEASES

By

Hisham Abouelseoud Elsayem Abdeltawab
M.Sc., Department of Biomedical and Systems Engineering,
Cairo University, Cairo, Egypt, 2015

A Dissertation
Submitted to the Graduate School,
the University of Louisville,
in Partial Fulfillment of the Requirements
for the Degree of

Doctor of Philosophy
in Interdisciplinary Studies: Specialization in Translational Bioengineering

Interdisciplinary Studies
University of Louisville
Louisville, Kentucky

August 2022

ROLE OF DEEP LEARNING TECHNIQUES IN NON-INVASIVE DIAGNOSIS OF
HUMAN DISEASES

By

Hisham Abouelseoud Elsayem Abdeltawab
M.Sc., Department of Biomedical and Systems Engineering,
Cairo University, Cairo, Egypt, 2015

A Dissertation Approved on

July 11, 2022

———————————————————————

by the Following Dissertation Committee:

———————————————————————
Ayman El-Baz, Ph.D., Dissertation Director

———————————————————————
Guruprasad Giridharan, Ph.D.

———————————————————————
Hermann Frieboes, Ph.D.

———————————————————————
Tamer Mohamed, Ph.D.

———————————————————————
Dibson Gondim, M.D.

DEDICATION

This dissertation is dedicated to my father.

ACKNOWLEDGMENTS

ABSTRACT

ROLE OF DEEP LEARNING TECHNIQUES IN NON-INVASIVE DIAGNOSIS OF

HUMAN DISEASES

Hisham Abouelseoud Elsayem Abdeltawab

July 11, 2022

Machine learning, a sub-discipline in the domain of artificial intelligence, concentrates on algorithms able to learn and/or adapt their structure (e.g., parameters) based on a set of observed data. The adaptation is performed by optimizing over a cost function. Machine learning obtained a great attention in the biomedical community because it offers a promise for improving sensitivity and/or specificity of detection and diagnosis of diseases. It also can increase objectivity of the decision making, decrease the time and effort on health care professionals during the process of disease detection and diagnosis. The potential impact of machine learning is greater than ever due to the increase in medical data being acquired, the presence of novel modalities being developed and the complexity of medical data. In all of these scenarios, machine learning can come up with new tools for interpreting the complex datasets that confront clinicians. Much of the excitement for the application of machine learning to biomedical research comes from the development of deep learning which is modeled after computation in the brain. Deep learning can help in attaining insights that would be impossible to obtain through manual analysis. Deep learning algorithms and in particular convolutional neural networks are different from traditional machine learning approaches. Deep learning algorithms are known by their ability to learn complex representations to enhance pattern recognition from raw data. On the other hand,

traditional machine learning requires human engineering and domain expertise to design feature extractors and structure data. With increasing demands upon current radiologists, there are growing needs for automating the diagnosis. This is a concern that deep learning is able to address. In this dissertation, we present four different successful applications of deep learning for diseases diagnosis. All the work presented in the dissertation utilizes medical images.

In the first application, we introduce a deep-learning based computer-aided diagnostic system for the early detection of acute renal transplant rejection. The system is based on the fusion of both imaging markers (apparent diffusion coefficients derived from diffusion-weighted magnetic resonance imaging) and clinical biomarkers (creatinine clearance and serum plasma creatinine). The fused data is then used as an input to train and test a convolutional neural network based classifier. The proposed system is tested on scans collected from 56 subjects from geographically diverse populations and different scanner types/image collection protocols. The overall accuracy of the proposed system is 92.9% with 93.3% sensitivity and 92.3% specificity in distinguishing non-rejected kidney transplants from rejected ones.

In the second application, we propose a novel deep learning approach for the automated segmentation and quantification of the LV from cardiac cine MR images. We aimed at achieving lower errors for the estimated heart parameters compared to the previous studies by proposing a novel deep learning segmentation method. Using fully convolutional neural networks, we proposed novel methods for the extraction of a region of interest that contains the left ventricle, and the segmentation of the left ventricle. Following myocardial segmentation, functional and mass parameters of the left ventricle are estimated. Automated Cardiac Diagnosis Challenge dataset was used to validate our framework, which gave better segmentation, accurate estimation of cardiac parameters, and produced less error compared to other methods applied on the same dataset. Furthermore, we showed that our segmentation approach generalizes well across different datasets by testing its perfor-

mance on a locally acquired dataset.

In the third application, we propose a novel deep learning approach for automated quantification of strain from cardiac cine MR images of mice. For strain analysis, we developed a Laplace-based approach to track the LV wall points by solving the Laplace equation between the LV contours of each two successive image frames over the cardiac cycle. Following tracking, the strain estimation is performed using the Lagrangian-based approach. This new automated system for strain analysis was validated by comparing the outcome of these analysis with the tagged MR images from the same mice. There were no significant differences between the strain data obtained from our algorithm using cine compared to tagged MR imaging.

In the fourth application, we demonstrate how a deep learning approach can be utilized for the automated classification of kidney histopathological images. Our approach can classify four classes: the fat, the parenchyma, the clear cell renal cell carcinoma, and the unusual cancer which has been discovered recently, called clear cell papillary renal cell carcinoma. Our framework consists of three convolutional neural networks and the whole-slide kidney images were divided into patches with three different sizes to be inputted to the networks. Our approach can provide patch-wise and pixel-wise classification. Our approach classified the four classes accurately and surpassed other state-of-the-art methods such as ResNet (pixel accuracy: 0.89 Resnet18, 0.93 proposed).

In conclusion, the results of our proposed systems demonstrate the potential of deep learning for the efficient, reproducible, fast, and affordable disease diagnosis.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER I

INTRODUCTION

For decades, the disease diagnosis process has relied on the conventional approaches where the physician analyzes symptoms, perform lab tests, and refer to medical diagnostic guidelines. However, there are limitations in these approaches. For example, lab tests might be late indicator for some diseases such as acute renal transplant rejection. Furthermore, traditional approaches might be slow and require intensive workload and considerable time in order to reach an accurate diagnosis. Other limitations include the case where the healthcare professionals are not available in certain local areas or the case where the cost of the diagnosis is not affordable by the patient. Therefore, there is unmet need and another approach for diseases diagnosis is required to overcome the mentioned limitations in the traditional approaches. Computer-aided diagnosis and detection can offer promising solutions to the current limitations in the medical diagnosis field. By utilizing medical imaging, researchers can implement computer-aided diagnosis systems that can detect the disease early and with an accuracy that is comparable to the human operators. Usually, these systems can be implemented by the power of artificial intelligence (AI). The prospect of predictive analysis can be offered by the AI and its subfield of deep learning. Deep learning can help in attaining insights that would be impossible to obtain through manual analysis. Deep learning algorithms and in particular convolutional neural networks (CNNs) are different from traditional machine learning approaches. Deep learning algorithms are known by their ability to learn complex representations to enhance pattern recognition from raw data. On the other hand, traditional machine learning requires human engineering and domain expertise to design feature extractors and structure data. The application of deep learning in medical imaging is a promising field. The interpretation of the human radiol-

ogist for the medical images, regardless of modality, is required to obtain a diagnosis in a timely fashion. With increasing demands upon current radiologists, there are growing needs for automating the diagnosis. This is a concern that deep learning is able to address. In order to integrate deep learning into routine clinical practice, we must obtain a diagnostic accuracy that is comparable or superior to healthcare professionals. Furthermore, the proposed solutions must present other benefits such as efficiency, speed, cost, and maintenance of ethical conduct.

In this dissertation, we present four different successful applications of deep learning for diseases diagnosis. All the work presented in the dissertation utilizes medical images. Therefore, we present below an overview on the medical imaging and the used modalities in the proposed applications.

A. Medical Imaging

There is a plenty of medical imaging modalities and their use frequency is increasing. Smith-Bindman et al. [15] investigated the use of medical imaging modalities from 1996 to 2010 in six large healthcare facilities in the United States, where they looked at 30.9 million examinations. They found that over the investigation period, positron emission tomography (PET), magnetic-resonance imaging (MRI), and computed tomography (CT) usage increased 57%, 10%, and 7.8%, respectively.

Modalities of medical images include X-ray, CT, MRI, PET, single photon emission computed tomography (SPECT), histology slides, retinal photography, and and dermoscopy images. Examples of medical images are shown in Fig. 1. Some of these techniques scan multiple organs such as CT and MRI, while others examine certain organs such as retinal photography, and dermoscopy. Also, the size of the generated data from each modality varies. Therefore, there are technical implications on the method of data pre-processing and on the selection of the architecture of the algorithm, in the context of memory and processor limitations. We can classify medical images based on their modali-

ties (Fig. 1) or based on the nature of the information obtained from the technique (i.e., the structure or the function of the organ being scanned, refer to Fig. 2). **In this dissertation, all of our work deals with the analysis of MRI images and histology slides, therefore, the details of these imaging modalities are presented below.**
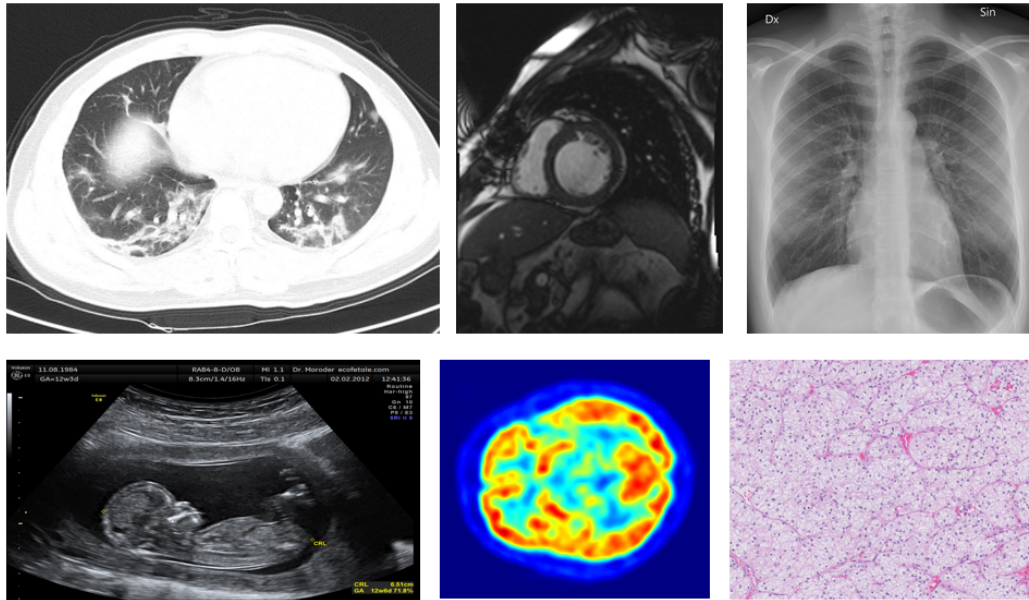
FIGURE 1: Examples of medical images, from left to right, top to bottom: an axial CT lung scan with Covid-19 lesion, a short-axis view MRI scan for the left ventricle, a normal chest X-ray scan, an ultrasound image of a fetus in the womb, PET scan for the brain, and a histology digital image for clear cell renal cell carcinoma.

1. Magnetic Resonance Imaging

MRI is a medical imaging modality that follows the same principles of nuclear magnetic resonance spectroscopy [16]. MRI is considered the most central and powerful non-invasive tool for the clinical diagnosis of diseases [17]. The basic idea of MRI relies on the usage of a strong static magnetic field. In that field, the hydrogen nuclei of the water molecules in human tissues are aligned to the magnetic field. Then, an external radio frequency (RF) wave/pulse is applied to the unpaired magnetic spins aligned in the
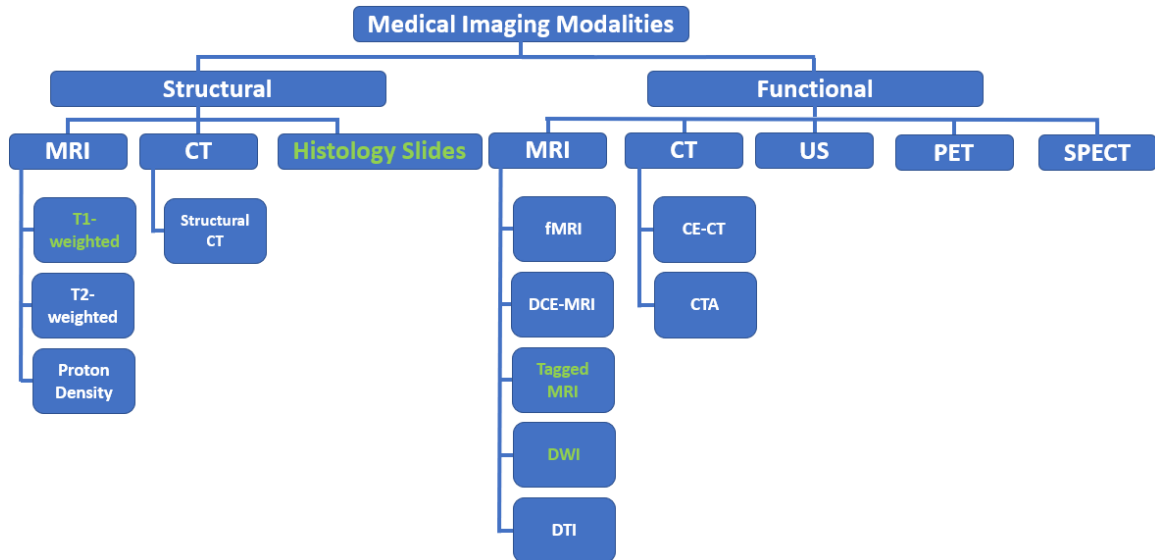
FIGURE 2: Classes of medical image modalities based on the nature of the information that they provide about the organ being scanned, i.e., structural or functional imaging. Categories written in green color are the focus of this dissertation.

magnetic field. The application of the RF pulse change the direction of spinning of the spins [18]. The interaction between the RF pulse and the spins results in periodic absorption and emission of energy. The protons tend to relax back by returning to their lower energy state. This relaxation leads to the release of detectable signals that are spatially encoded to construct the MR image. Different types of tissues (fat, muscle, cerebral spinal fluid, etc.) send back tissue specific signals after the application of the same RF wave. The way the image is acquired strongly determines the contrast of the MR image. Different parts of the scanned region can be highlighted utilizing different pulse sequences: a preselected shape, strength, and timing of the used RF and gradient pulses (external fields). In general, MRI can be used to construct 2D images, 3D volumes, or sequences of 3D volumes (i.e., 4D images). **The following sub-sections describe the structural MRI and the MRI techniques which are used throughout this dissertation.**

    a. Structural MRI   Structural MRI includes the techniques that show the structure of the body and its tissues. These techniques are T1-weighted MRI, T2-weighted MRI,

and proton density (PD). The signal strength (amount of energy) of each of these three techniques mainly relies on the magnetic relaxation properties of body atomic nuclei. The time spent by nuclei to return to their relaxation states after the application of the RF pulse is known as the time of relaxation. There are two relaxation times: longitudinal relaxation time (T1) or transverse relaxation time (T2), based on the orientation of the component with respect to the magnetic field. In the human body, every tissue has its own T1 and T2 values, that depend on the concentration of protons in the human tissue in the form of water and macromolecules. T1-weighted and T2-weighted scans are the most common MRI sequences. T1-weighted images are generated by the usage of short time to echo (TE) and repetition time (TR). The T1 properties of the tissue predominately determine the contrast and the brightness of the image. Conversely, T2-weighted images are generated by the usage of longer TE and TR times. In these images, T2 properties of the tissue predominately determine the contrast and the brightness. In practice, complementary information can be obtained from both T1- and T2-weighted images. Therefore, the two methods are important for characterizing abnormalities. On the other hand, the proton density weighted images have no T1 or T2 contrast. The signal differences come from the differences in the amount of available spins in the tissue. The prime advantage of these proton density weighted scans is the increase in the contrast between fluid and non-fluid regions. However, proton density weighted images have lower contrast than T1-weighted and T2-weighted images, because the difference in hydrogen concentration of soft tissues is relatively low.

b. Diffusion MRI MRI techniques that use contrast agent (e.g., gadolinium) might be harmful to some patients, especially, patients that suffer from kidney problems. In recent years, researchers avoided the use of contrast agents by introducing an MRI technique called diffusion imaging. Diffusion MRI is considered a functional technique that rely on the measurement of the micro-movements (Brownian, random) of extracellular water molecules in the body. An indirect information about the structure surrounding these molecules can be obtained from these micro-movements. Diffusion MRI is considered a

noninvasive method that has the advantages of rapid acquisition without the need of any contrast agent or specialized hardware. We can classify diffusion MRI into three main categories, namely, diffusion weighted imaging (DWI), diffusion tensor imaging (DTI), and diffusion spectrum imaging (DSI). In this thesis, we use DWI in the project of transplanted kidney classification.

In DWI, the contrast of the images depends on the differences in water molecule mobility. This can be performed by the addition of diffusion magnetic field gradients during data acquisition. The b-factor (in s/mm^2) determines the degree of diffusion weighting of the sequence. The b-factor relies on the amplitude of the field gradient, time of the application, and the time interval between magnetic field gradients. A typical DWI-MRI for the kidney is shown in Figure 3. Researchers used DWI as well-established MRI technique for investigation of brain disorders, such as multiple sclerosis, epilepsy, brain abscesses and tumors and hypertensive encephalopathy [19], tumor localization and diagnosis [20], and in-vivo study of tissue micro-structure [21].

c. Tagged Magnetic Resonance Imaging   Tagged MRI is a popular MRI technique for noninvasive and detailed visualization of cardiac motions [22]. This imaging method can facilitate the localization of cardiac diseases (e.g., coronary atherosclerosis). Furthermore, it can help in the evaluation of cardiac strain [23]. A pre-specified pattern of temporary markers is placed in cardiac tagged MRI. These markers are called tags and are placed inside the soft body tissues. For example, tag lines constructed by patterns of magnetic spin in the examined tissues so that the movement in the tagged tissues can be measured from the images [24].

This modality is considered a complement for traditional anatomical imaging and can obtain a detailed cardiac information over time. The tag lines can help in computing displacement, rotation, velocity, elongation, twist, and strain of the heart. While conventional MRI methods carry information only about the motion at the boundaries of an object, the tag lines allow us to investigate the displacement and the strain of the interior of the tis-
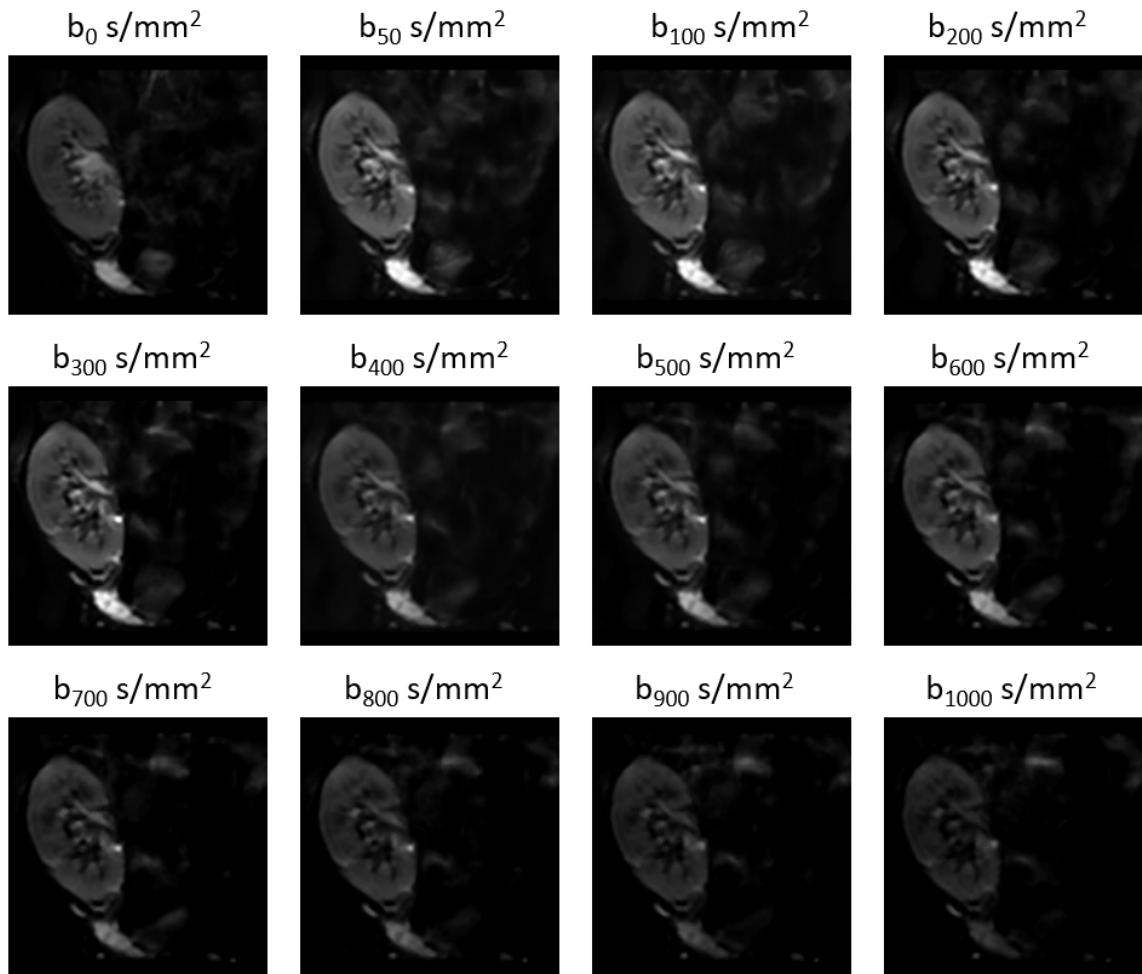
FIGURE 3: Example of diffusion weighted imaging: A coronal kidney cross-section of a stable transplanted kidney at 12 b-values.

sue in much details [25]. A typical tagged MRI time series of the heart is shown in Figure 4.

    d.   *Histopathological Images*   Valuable information for the clinicians and pathologists can be obtained from histopathological images [26]. Histopathology is a medical field where pathologists study the disease by a microscope that examines a tissue biopsy placed in glass slides. The glass slides are then digitized by a scanner to produce digital histopathological images that can be stored on a computer for processing and analysis. Stains are used as a dye for the tissue to show the different parts of the tissue under the

Time Frames During the Cardiac Cycle
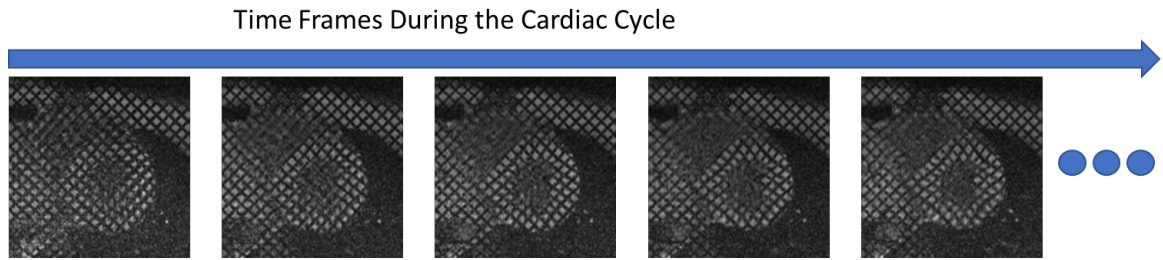
FIGURE 4: An example of cardiac tagged MRI time series.

microscope. Pathologists have used Hematoxylin-Eosin (H and E) stain for more than a hundred year. Hematoxylin dyes the nuclei by blue colour. On the other hand, Eosin dyes cytoplasm and connective tissue by pink colour, see Figure 5. The disadvantages of using H and E images for disease diagnosis are the inter-operator variability, the effort, and the time required for the examination and diagnosis.
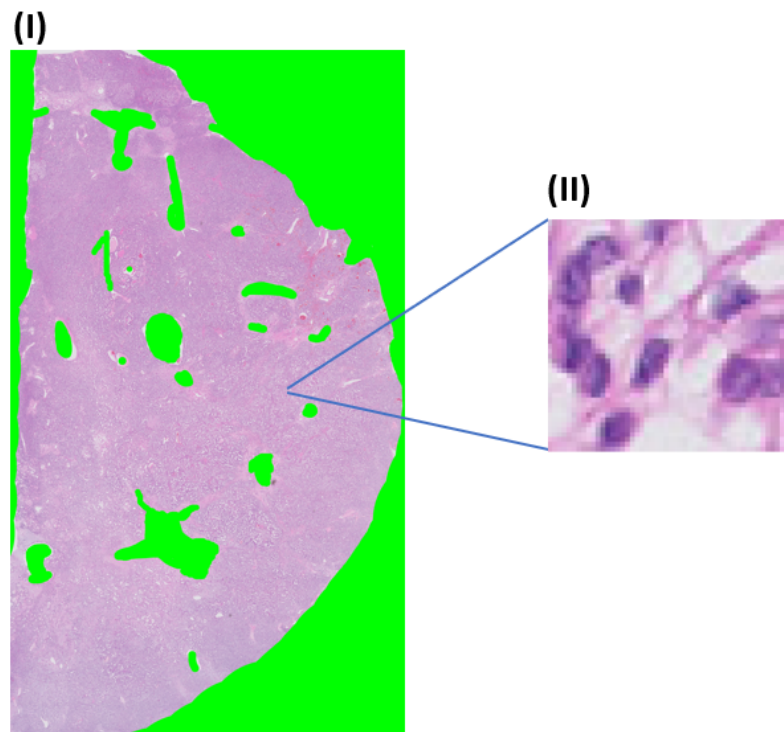
FIGURE 5: An example of a histology image that shows renal cell carcinoma. (I) is the whole digitized slide, and (II) is a zoomed region.

B. Automated Diseases Diagnosis using Deep Learning and Medical Imaging

The prospect of predictive analysis can be offered by the artificial intelligence (AI) and its subfield of deep learning [27]. Deep learning can help in attaining insights that would be impossible to obtain through manual analysis [28]. Deep learning algorithms and in particular convolutional neural networks (CNNs) are different from traditional machine learning approaches. Deep learning algorithms are known by their ability to learn complex representations to enhance pattern recognition from raw data. On the other hand, traditional machine learning requires human engineering and domain expertise to design feature extractors and structure data [11]. The application of deep learning in medical imaging is a promising field [29,30]. The interpretation of the human radiologist for the medical images, regardless of modality, is required to obtain a diagnosis in a timely fashion. With increasing demands upon current radiologists, there are growing needs for automating the diagnosis. This is a concern that deep learning is able to address [31]. In order to integrate deep learning into routine clinical practice, we must obtain a diagnostic accuracy that is comparable or superior to healthcare professionals. Furthermore, the proposed solutions must present other benefits such as efficiency, speed, cost, and maintenance of ethical conduct.

The goal of this work is to develop automated computer aided diagnostic system for disease diagnosis using deep learning and medical imaging. The dissertation proposes four different systems. The first system aimed at developing an automated tool for the early detection of acute renal transplant rejection using DWI. The second system aimed at developing an automated tool for the segmentation and quantification of the left ventricle from cardiac cine MRI. The third system aimed at developing an automated tool for the assessment of LV strain using mice cine MRI and tagged MRI. The fourth system aimed at developing an automated tool for the detection of renal cell carcinoma from histopathological kidney images.

1. Early Detection of Renal Transplant Rejection

Acute renal transplant rejection is a critical problem in the urology field. In the United States, approximately 22,393 kidney transplants were performed in 2018 [32]. Usually, the number of kidney donors is limited. Therefore, it is important to maintain the transplanted kidney in a healthy state. Acute renal rejection can occur in up to 40% in kidney recipients within 3 weeks after transplantation. Typically, high-dose steroids or antibody therapy can treat acute renal rejection. A gradual damage in the transplanted kidney function leads to chronic rejection, which starts about at 3 months after transplantation [33]. Given that providing an effective therapy for chronic rejection is an open area of research, avoiding acute renal rejection is the most promising way to prevent chronic rejection [34]. Therefore, it is of paramount importance to early detect rejection to intervene with the suitable medical and immune therapy in kidney transplant recipients [35].

Currently, initial assessment of kidney transplant function is based on blood and urine tests such as plasma creatinine and creatinine clearance. creatinine clearance is a laboratory test that can be used to evaluate kidney function, such as glomrular filtration rate (GFR). Urologists consider creatinine clearance as a practical measurement to evaluate kidney function. However, we can not rely on such index to effectively detect renal rejection due to the following reasons: (1) creatinine clearance gives an overall idea about the function of both kidneys together not about a specific kidney [36], (2) we can observe a significant change in creatinine level only after 60% loss of kidney function [37]. Biopsy is considered the gold standard to evaluate graft function, but only as the last option because it is associated with potential morbidity and high costs. Also, the invasive nature of this procedure imposes a risk of infection and bleeding to the patients. Furthermore, the sample that we get from the needle biopsy is small and might lead to underestimation or overestimation of the degree of inflammation in the entire graft [38]. Therefore, a technology that is based on accurate and noninvasive techniques is so much needed to evaluate transplanted kidney function and to early detect acute renal rejection.

Various noninvasive imaging techniques have been used in clinical settings to evaluate the function of transplanted kidney. Graft function can be evaluated quantitatively and qualitatively by an imaging modality called radionuclide imaging [39]. However, this technique suffers from limited spatial resolution, therefore, it does not show accurate anatomical information of functional abnormalities inside the kidney [40]. Additionally, radionuclide imaging is associated with exposure to radiation which limits the range of its applications. Ultrasound imaging can be used to assess the function of the graft early in the postoperative period and it is suitable for long-term evaluation of the graft. However, this modality suffers from the following drawbacks: (1) investigator's experience greatly affects the sensitivity and reliability of this technique, (2) it has low signal-to-noise ratio, (3) it suffers from shadowing artifacts and speckles which decrease the diagnostic confidence [41]. Computed tomography (CT) is successful in diagnosing post transplantation complications [42]. However, it has low specificity and its process contains nephrotoxic agent. Different magnetic resonance imaging (MRI) techniques have been introduced to perform assessment for kidney transplant function. For instance, dynamic contrast enhanced (DCE) MRI utilizes contrast agents, such as gadolinium to assess tissue perfusion, which gives an idea about kidney function [43]. Despite the fact that good anatomical and functional information can be obtained from DCE-MRI, the use of the contrast agent (gadolinium) can lead to nephrogenic systemic fibrosis [44], especially when $GFR < 30 \; ml/min$ per $1.73 \; m^2$. Blood oxygen level dependent (BOLD) MRI which assess the renal oxygenation level has been utilized for the evaluation of kidney transplant function and the detection of acute renal rejection [45]. However, BOLD-MRI has several limitations such as susceptibility to breathing motion artifact and bowel gas artifacts [46, 47].

Diffusion weighted (DW) MRI is an imaging modality that is not associated with contrast agents. DW-MRI has been successful in detecting and characterizing tumors, neuroimaging, and kidney transplant function evaluation [48]. DW-MRI measures water's motion inside the tissue, and thus it is considered a way to evaluate the diffusion charac-

teristics for the tissue. We can obtain apparent diffusion coefficient (ADC) from DW-MRI. ADCs are quantitative maps which represent the diffusion and can be measured at different magnetic field strengths and duration noted as b-values [48]. In this dissertation, we propose a machine learning framework that aims at classifying stable transplanted kidneys (non-rejection transplants) versus acute renal rejection transplants by analyzing DW-MRI data, see Figure 6. Our framework is efficient in terms of it does not require kidney segmentation, and non-invasive in terms of it requires only imaging and clinical biomarkers. Furthermore, our framework is intended to early detect acute renal transplant rejection. It is composed of the following three steps: (1) preprocessing where we perform histogram equalization and region of interest (ROI) extraction, (2) feature extraction where we estimate the 3D ADC maps and fuse the clinical biomarkers, (3) CNN based classification to non-rejection and acute renal rejection transplants.
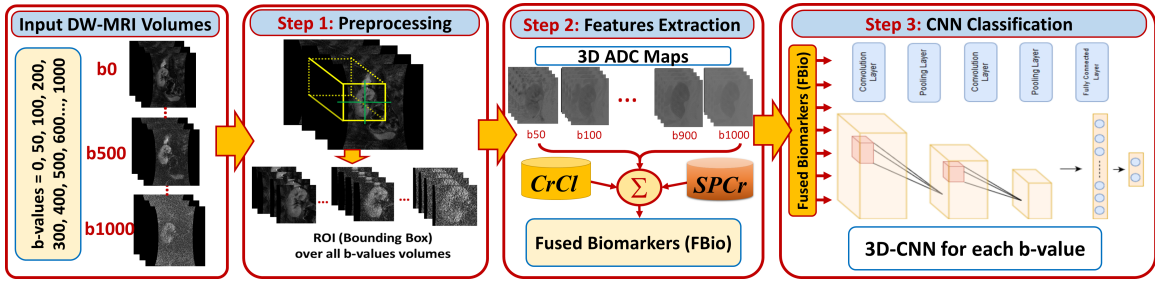
FIGURE 6: The proposed machine learning framework for early detection of acute renal rejection using DW-MRI analysis. The framework is composed of three main steps. In the first step, we equalize the histogram to reduce the inhomogeneity and noise. Then, we construct an ROI that encloses the kidney for each subject. In the second step, we estimate the 3D ADC maps for the chosen ROI and we fuse the clinical biomarkers (creatinine clearance (CrCl) and serum plasma creatinine (SPCr)). Finally, the fused markers (3D volumes of size 150x150x24) are inputted to the CNN classifier to classify non-rejection versus acute rejection allografts.

2. Automated Segmentation and Quantification of the Left Ventricle of the Human Heart

CVDs are dangerous heath issues as they are responsible for the highest rate of mortality worldwide [49]. About 836,546 patients in the United States die each year because of CVDs [50]. A non-invasive quantitative evaluation for the heart functionality can be available for the cardiologist from Cardiac magnetic resonance (CMR) which is an important imaging technique. The cardiologist can get functional heart indexes from performing segmentation for the LV in CMR short-axis view cine images [51]. These indexes are ESV, EDV, EF, wall mass, and regional indexes (e.g., wall thickening) which are essential for heart diagnosis and treatment. However, to estimate these parameters, accurate delineation of myocardial walls is required. Manual segmentation of LV in CMR short-axis view images requires significant effort and time. Furthermore, manual segmentation is prone to inter- and intra-operator variability [52]. Therefore, an alternative technique to manual segmentation is needed to perform automated segmentation for the LV cavity and myocardium. Furthermore, this technique should be accurate to obtain left ventricular functional metrics with high accuracy.

The need for accurate and automated estimation of left ventricular parameters, and the success of the CNNs motivated us to build a framework for the automated functional assessment of the heart. This framework should segment and quantify the LV with a comparable accuracy to the human operator. In this dissertation, we propose a novel fully automated framework that segment and quantify the LV from CMR cine images. Our proposed framework for the automated segmentation and quantification of the LV is shown in Figure 7. The quantification is performed by the calculation of left ventricular functional parameters and mass. Our framework consists of four main steps: (1) ROI extraction using a FCN called FCN1, where the center of the ROI is at the center of left ventricular cavity, (2) Image cropping for all CMR images using the extracted ROI, (3) Segmentation of the left ventricular cavity and myocardium using a FCN called FCN2, and (4) Calculation of left ventricular functional parameters and mass.

13
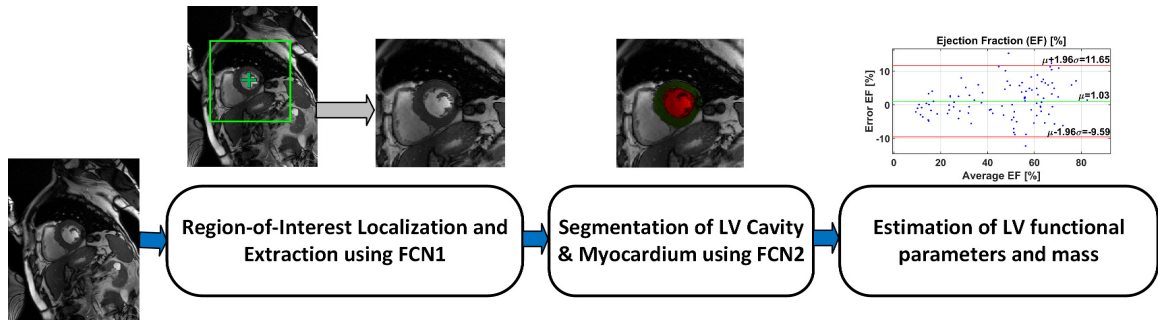
FIGURE 7: Illustration for the proposed framework for automated left ventricle segmentation and quantification (calculation of LV functional indexes and mass).

We performed segmentation to the LV cavity and myocardium to estimate the heart physiological parameters. Our main achievement is the accurate quantification of the LV indexes and reaching a lower error compared to other previous techniques applied on the same dataset. Our framework that is based on FCN has the following contributions:

- We propose an efficient method that is based of FCN to extract a region-of-interest (ROI) that contains the LV from CMR cine images. We perform this extraction process in the beginning to alleviate the class-imbalance problem and reduce the memory and computational requirements.

- After ROI extraction, we use a novel FCN model for cardiac segmentation. The architecture of this novel model is inspired from the U-net; i. e., the input is passed to a contracting path followed by an expanding path. The addition to the conventional model is the incorporation of multiple bottleneck layers that describe the input by multiple representations. The bottleneck layers are up-sampled and combined to estimate the final segmentation. Our proposed model requires less fewer number of parameters than the state-of-the-art models such as U-net, yet it showed better performance.

- We propose a novel loss function that works on minimizing the difference between the ground truth LV contours and the predicted contours. We refer to the new loss by

14

radial loss and we incorporate it with the cross-entropy loss.

- We evaluated the generalization strength of our proposed segmentation method by estimating the segmentation performance of our method when we use the ACDC 2017 data as a training set and a local dataset as a testing set. Our segmentation approach demonstrated good segmentation accuracy that is comparable to another model that uses only our local dataset.

3. Automated Assessment of the strain of the Left Ventricle in Mice

Mice are the most preferable species for cardiovascular modeling for several reasons including the low procedure costs, ease of handling, and the capability to manipulate the mouse genome [53]. By utilizing animal models, we increase our knowledge for improving the diagnosis and treatment of cardiovascular diseases. For cardiac function quantification, MRI is considered the most accurate imaging technique (gold standard) [54–58]. The major advantages of MRI are ability to rebuild 3D structure and function, clear tissue structure, and high reproducibility [59]. currently, cardiac MRI is clinically utilized to evaluate myocardial function, heart structure, perfusion, and viability in humans [60]. Multiple mouse works by our group and others [54,61–63] got benefit from the use of standard cardiac cine MRI to evaluate heart structure and functional indexes, e.g., EDV, ESV, and EF. However, tagged cardiac MRI is required for strain analysis of myocardial contractility. Tagged cardiac MRI takes considerable time and difficult to interpret, especially in mice. It is difficult to interpret because of the small size of the mice heart, which prevent the creation of multiple $1mm^2$ diamonds within the wall of the ventricle for strain analysis. Therefore, researchers abstain from performing strain analysis from tagged MRI. On the other hand, the picture of the heart function is fully completed by strain analysis which is an essential component because it is the only parameter that distinguish between active and passive movement of myocardial segments, assess longitudinal myocardial shortening, and

15

quantify intraventricular dyssynchrony. Furthermore, evaluating myocardial strain gives us a comprehensive evaluation of diastolic myocardial function, that is not inferred from EF [64].

In this dissertation, we aimed at the development of an automated framework for the accurate estimation of strain from standard cardiac cine MRI of mice. The used mice are sham and myocardial infarction (MI) cases. Our work is an alternative to tagged imaging of mice. Myocardial segmentation is an important initial step for the estimation of LV functional indexes. Therefore, a deep-learning based technique was proposed for the automated segmentation of LV borders. Subsequently, we developed a Laplace based technique to track myocardial points along the cardiac cycle to accurately evaluate the strain from the standard cardiac cine MRI. We validated the Laplace-based technique by performing a comparison between the Laplace-based technique and the analyzed strain outcome resulted from the tagged imaging which is considered the current gold standard for strain analysis. Finally, we also utilized our framework to calculate other global and local cardiac indexes. Our framework consists of two primary stages. The first stage is the segmentation of the LV cavity and myocardium using a deep learning approach that is based on a FCN. The second stage is the calculation of the functional and structural heart parameters. The parameters include global volumetric indexes (e.g., EDV, ESV, EF) and local indexes involving wall thickening and myocardium strain. Fig. 8 depicts the overall framework. The development and validation of the framework will be described in detail in the upcoming subsections.

4.   Automated Detection of Renal Cell Carcinoma

Kidney cancer is an abnormal cell growth in the kidney. Kidney cancer is ranked fourteenth among common cancers and contributes about 2.4% of the total number of diagnosed cancers [65]. A recent estimate by the American Cancer Society revealed that in the United States about 76,080 new patients will be diagnosed with kidney cancer and about 13,780 patients will die from the disease [66]. Cancers of the kidney are of several types,
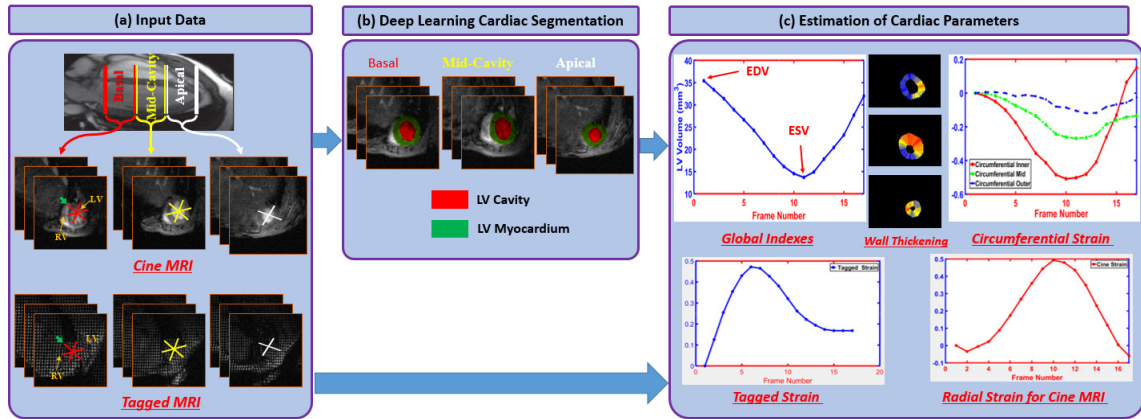
FIGURE 8: The proposed framework for the quantification of left ventricular functional and structural parameters.

renal cell carcinoma (RCC) being the most prevalent, accounting for nearly 85% of kidney cancers. RCC is itself a heterogeneous set of cancers with distinct molecular characteristics, responses to treatment, and clinical outcomes [67]. The most prevalent subtypes of RCC include clear cell (70%—80%), papillary (14%-–17%), chromophobe (4%—8%), and clear cell papillary RCC (4%) [68]. We can classify RCC subtypes by utilizing the morphological features observed on histopathological slides stained with hematoxylin and eosin. However, there is a significant morphological overlap between clear cell RCC and clear cell papillary RCC subtypes due to the presence of clear cells. It is critical to be able to classify between clear cell RCC and clear cell papillary RCC to determine the correct patient management. Clear cell RCC is associated with poor prognosis because it is likely to metastasize, while clear cell papillary RCC is considered an indolent neoplasm, which is not associated with metastatic spread [69]. When there is a significant overlap in morphological features, immunohistochemistry helps in separating the two subtypes of the tumor. However, we cannot find immunohistochemistry in all areas of the world. We can use deep learning algorithms to differentiate between the types of tumor by analyzing histopathological images.

In this dissertation, we propose a computer aided diagnostic system that uses CNNs

for the automated classification of kidney tissues and RCC, see Figure 9. Our framework can partition histology images into four classes including fat, parenchyma, clear cell RCC, and clear cell papillary RCC. Our framework has the following contributions:

- The first study to discriminate between clear cell RCC and clear cell papillary RCC. This classification has high clinical relevance.

- We propose a pyramidal deep learning model that utilizes a hierarchy of three CNNs that process different image sizes. The deep learning improves the precision of the diagnosis and decrease human error. Furthermore, it produces reproducible results and objective assessment.

- Our approach can provide both patch-wise classification and pixel-wise classification.

- We incorporate a statistical approach based on Markov-Gibbs random field (MGRF) to remove inconsistencies in the final pixelwise classification.

FIGURE 9: Schematic illustration of the proposed computer aided diagnostic system for automated classification of kidney tissues.

C. Dissertation Organization

This dissertation consists of seven chapters. The following remarks summarize the scope of each chapter:

- Chapter I presents some basic concepts about medical imaging and the used imaging modalities in this dissertation. Furthermore, the chapter provides a brief summary for the the basic contributions of the proposed research.

- Chapter II introduces what is meant by deep learning and the most popular algorithms used in this domain. Furthermore, the chapter presents the use of deep learning in the medical imaging domain.

- Chapter III presents a novel deep learning based approach for early detection of acute renal transplant rejection using diffusion weighted magnetic resonance imaging.

- Chapter IV presents a novel deep learning based approach for the automated segmentation and quantification of the left ventricle from cardiac cine magnetic resonance imaging.

- Chapter V presents a novel approach for the automated assessment of left ventricle strain using mice cine MRI and tagged MRI.

- Chapter VI presents a deep learning based approach for the automated detection of renal cell carcinoma from histopathological whole slide images.

- Chapter VII concludes the work presented in this dissertation and outlines future directions.

CHAPTER II

DEEP LEARNING

This chapter introduces deep learning and its algorithms. Deep learning is as sub-field in machine learning which in turn is a sub-field in the artificial intelligence. One of the algorithms of deep learning is artificial neural networks which are considered mathematical function approximators and are inspired from the workings of the brain and the connections of the nervous system. The basic unit of artificial neural network is the neuron. The chapter introduces the feedforward neural network, activation functions, loss function, and network optimization. Another algorithm in deep learning is convolutional neural networks which are inspired from the human visual cortex. It is the first choice in computer vision applications. Furthermore, the chapter introduces deep learning in medical imaging.

A. Introduction

Until recently, shallow-structured architectures had been adopted in most machine learning (ML) and signal processing techniques. Typically, these architectures include at most one or two nonlinear features transformations layers. Examples of shallow architectures involve linear or nonlinear dynamical systems, Gaussian mixture models (GMMs), maximum entropy (MaxEnt) models, conditional random fields (CRFs), logistic regression, support vector machines (SVMs), multilayer perceptrons (MLPs) with only one single hidden layer [70]. Various simple or well-constrained problems can be solved effectively by shallow architectures. However, complicated real-world applications can not be solved by shallow architecture due to their limited modeling and representational power. Real-world applications involve natural signals such as natural sound, human speech and language, and

natural images and medical images.

Processing mechanisms of human information (e.g., audition and vision) suggest the need for deep architectures for building internal representation and extracting complex structure from rich sensory inputs. For instance, systems of production and perception of human speech are equipped with layered hierarchical structures for converting the information from its waveform level to its linguistic level [71]. Similarly, The human visual system is hierarchical in the perception and in the generation sides [72]. Historically, from the domain of artificial neural network research originated the concept of deep learning. Therefore, one might hear the discussion of the new generation of neural networks. Deep learning is considered a subfield of ML which in turn is a subfield of artificial intelligence (AI) as shown in Figure 10. Feed forward neural networks or multi layer perceptrons (MLPs) that contain many hidden layers are usually referred to deep neural networks (DNNs). DNNs are typical examples of deep architectures models. The parameters of DNNs need to be learned during the training process. In 1980, Back-propagation (BP) algorithm became popular for learning these parameters. In practice, BP alone did not achieve good results for learning the parameters of networks with high number of hidden layers [73]. The difficulty in the learning that BP experiences is due to the presence of many local optima and the optimization difficulties in the non-convex objective function of the DNNs. BP uses local gradient information and its work starts with selecting random initial points. Therefore, it can be trapped in unsatisfactory local optima when we use batch-mode or even the algorithm of stochastic gradient descent BP. The more the network depth increases, the more the severity increases. ML researches steer away from using DNNs because of this training difficulty and they prefer to use shallow architectures that have convex loss function (e.g., CRFs, SVMs, and MaxEnt models). These shallow models can reach global optima efficiently, however, they have a reduced modeling power [70].

Different nonlinearities and better learning algorithms contributed to the success of DNNs. In the case of the training set is large, the Stochastic gradient descend (SGD) al-

21
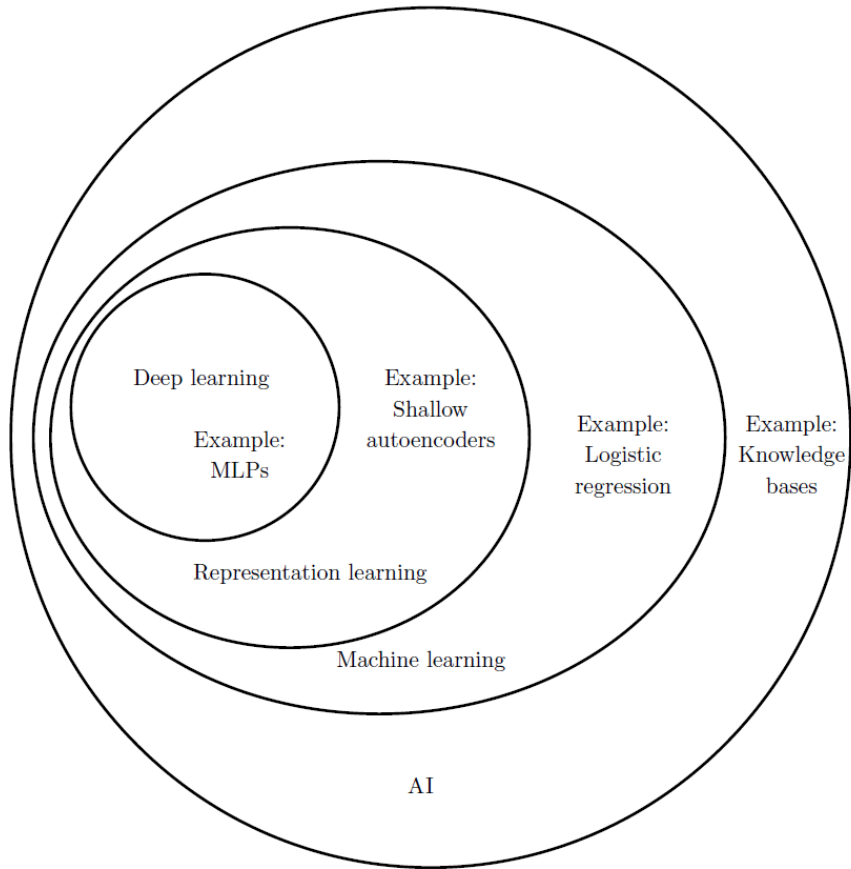
FIGURE 10: A Venn diagram that shows the relationship between artificial intelligence sub-domains [4].

gorithm is very efficient [74]. Recently, researchers used SGD for parallel computations over many machines [75] or over multiple GPUs [76]. Furthermore, SGD can jump out of local optima because it involves the estimation of noisy gradients from a small batch of samples. We can find similar ability in other learning algorithms such as Krylov subspace methods [77] or Hessian free [78]. When the optimization problem of the DNN is highly non-convex, it is obvious that good initialization methods for network parameter will result in good models. However, it is not obvious how to correctly initialize the network parameters and how big amounts of training data can mitigate the learning problem till more recently [73, 79]. Unsupervised pretraining method is an initialization method which

attracted a lot of attention [80].

A source of difficulty in real world AI applications is that there are factors of variations that affect every observed single piece of data. For example, the red pixels of a car in an image might appear black at night. Also, the viewing angle determines the shape of the silhouette of the car. Therefore, we need to disentangle these factors of variations. However, it is very challenging to extract such abstract features from the raw data. Deep learning solves this challenge of representation learning by making representations that are described in terms of other simpler representations. Deep learning allows the computer to make complex concepts from simpler concepts [4]. Figure 11 demonstrates how a deep learning model can describe complex concept such as an image of a person by integrating simpler concepts, such as the contours which are in turn described in terms of edges.

B. Artificial Neural Networks

Artificial neural networks (ANNs) are mathematical function approximators that are commonly used in the field of computational intelligence [81]. They are inspired from the workings of the brain and the connections of the nervous system [82]. For more than seventy years, researchers studied ANNs [83]. Given enough computational resources and data, non-linear and complex functional mapping can be learned using ANNs. The more training data, the better results of ANNs. In the following subsections, we will discuss the basic architecture and the widely used model of ANNs to analyze data.

1. Artificial Neuron

The basic unit of ANNs is the neuron. Figure 12(a) shows the neuron and its components. The basic components of an artificial neuron are an input x which is a vector valued where $x \in \mathbb{R}^n$ (n is the dimension of the input), bias b which is a scalar value, weight w which is a vector valued where $w \in \mathbb{R}^n$, an activation function ϕ, and an output
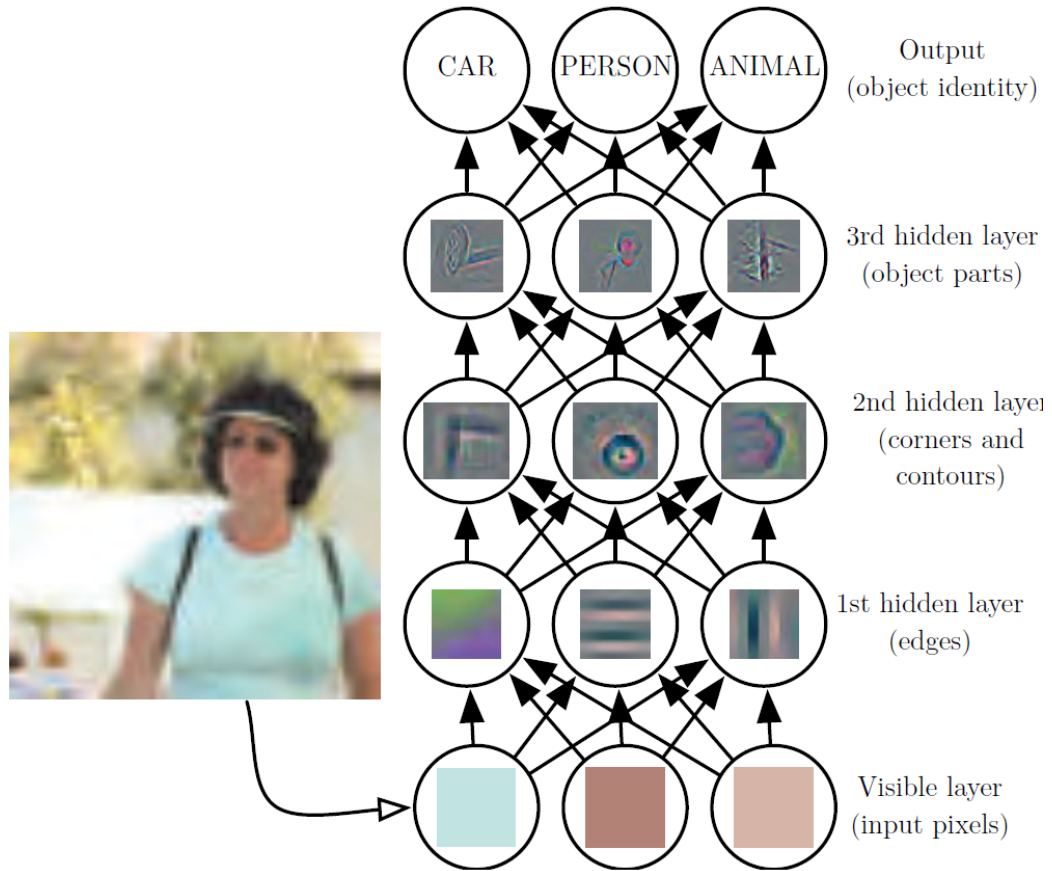
FIGURE 11: Demonstration of a deep learning model. The function that maps a set of pixels to an object identity is complicated. The learning of this mapping is very challenging if we want to evaluate it directly. Deep learning solves this difficulty by breaking the complex mapping into a series of simpler mappings. Each mapping is described by a layer in the model. That is how a deep learning model can describe complex concept such as an image of a person by integrating simpler concepts, such as the contours which are in turn described in terms of edges. [4]

y. The argument of the activation function is $z = w^T x + b$. The output of the neuron is defined as

$$y = \phi(z) = \phi(w^T x + b) \tag{1}$$

If we have only the argument z, we will get a linear discriminant function [84].

24

However, using the activation function ϕ which is also known as the transfer function, we can obtain a non-linear transformation. The non-linear transformation useful because it mimics the real world which is non-linear. There are many types of the activation functions, such as hyperbolic tangent, ReLu, sigmoid, Heaviside function, Signum function, and softmax.
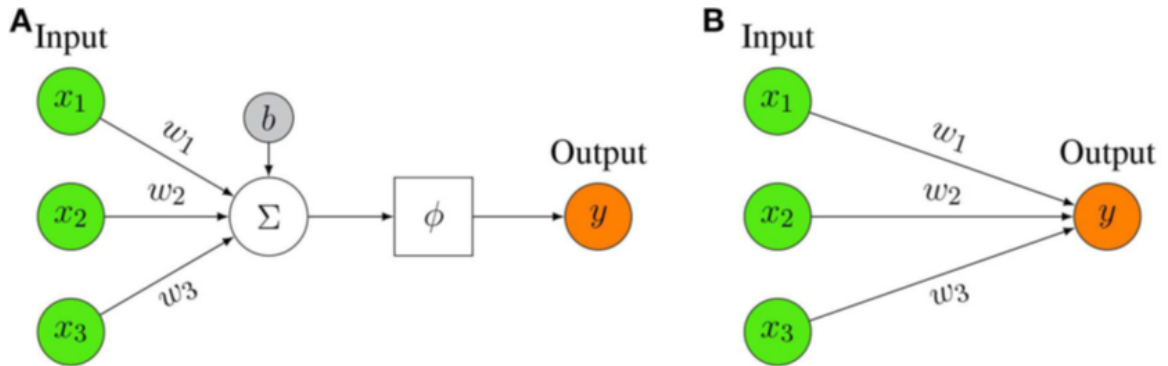
FIGURE 12: (A) Schematic illustration of mathematical artificial neuron. (B) Simplified artificial neuron [5].

ReLu is an abbreviation for rectified linear unit which is the most widely used activation function in ANNs [85]. Another popular function is the softmax activation function [86]:

$$y_i = \frac{e^{x_i}}{\sum_j^n e^{x_j}} \tag{2}$$

The softmax function transform the vector x of n-dimension into a vector y of n-dimension and y satisfies that $\sum y_i = 1$ Therefore, softmax produces a vector contains probabilities for each of the n-elements. Researchers often use softmax as the final layer of the neural network. The model of the neuron is called perceptron if the Heaviside step function is used as an activation function [87].

Usually, the artificial neuron is represented in a simplified form by reducing the focus of its important components. Figure 12(B) shows the artificial neuron in its simplified form.

2.   Feedforward Neural Networks

It is required that the neurons connect with each other, in order to construct neural networks (NN). Feedforward architecture is the simplest structure of a NN. Figure 13 shows exemplary architectures of the shallow and deep NN. The number of the non-linear transformations between the disconnected layers is referred to the network depth, while the network width is the number of the hidden layers. For example, the shallow network in Figure 13(A) has a depth of 2 and a width of 1 whereas the network in Figure 13(B) has a depth of 4 and width of 3. The researchers calls a network deep network when it has more than 2 hidden layers [73].
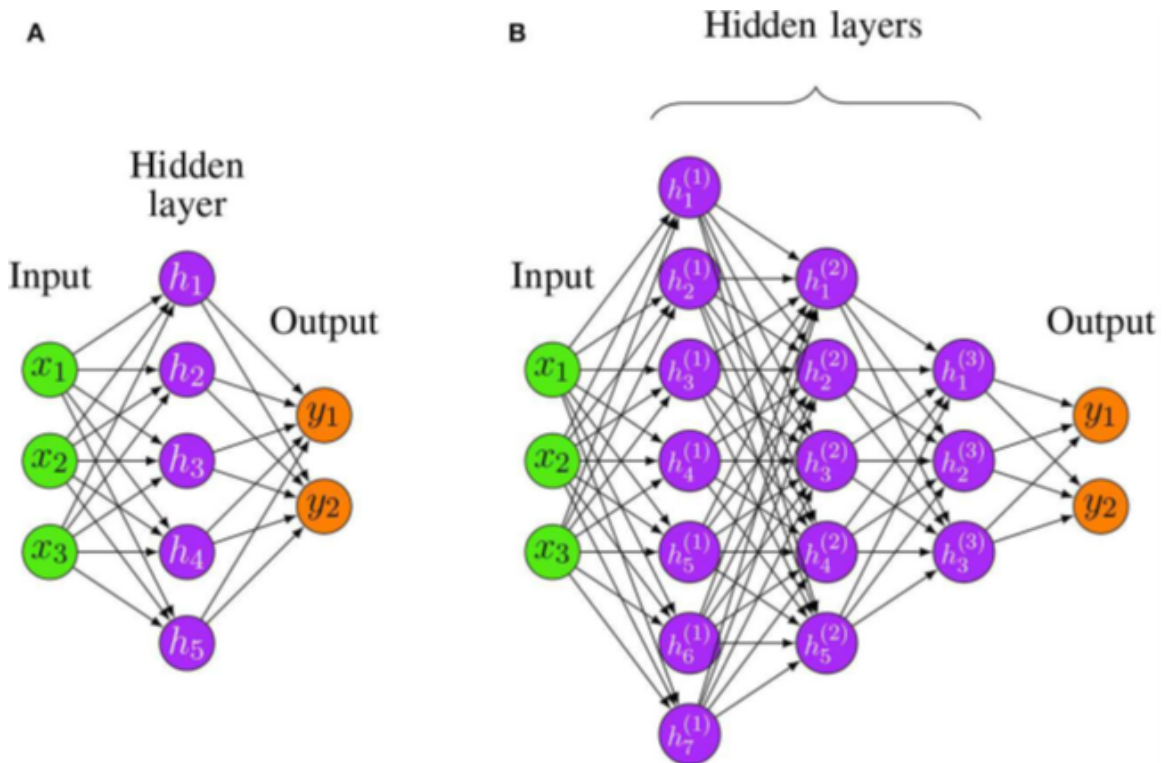
FIGURE 13: Two examples of feedforward NN. (A) Shallow feedforward NN. (B) Deep feedforward NN with three hidden layers [5].

The feedforward NN is also known as multilayer perceptron where linear or nonlinear activation functions can be used. The feedforward NN has no cycles. The output of

the multilayer perceptron is defined as follows [84]:

$$f(x) = \phi^{(2)}(W^{(2)}\phi^{(1)}(W^{(1)}x + b^{(1)}) + b^{(2)}) \tag{3}$$

Equation 3 is defined as the discriminant function of the NN. A learning rule is required to obtain the optimal parameters of the NN. Researchers state an error function (called the cost function) and an optimization algorithm to obtain the best parameters of the network. This is usually done by minimizing the cost function for the training data.

3. Activation Functions

In order to allow the network to learn complex patterns in the data, activation functions were added to the ANN. When comparing activation function with the model of a neuron in our brains, the activation function lies at the end where it decides what is to be fired to the next neuron. It receives the output from the previous cell and converts it to another form. Then this new form is given as input signal to the next cell [88]. Figure 14 summarizes the comparison between the neuron model and the activation function.
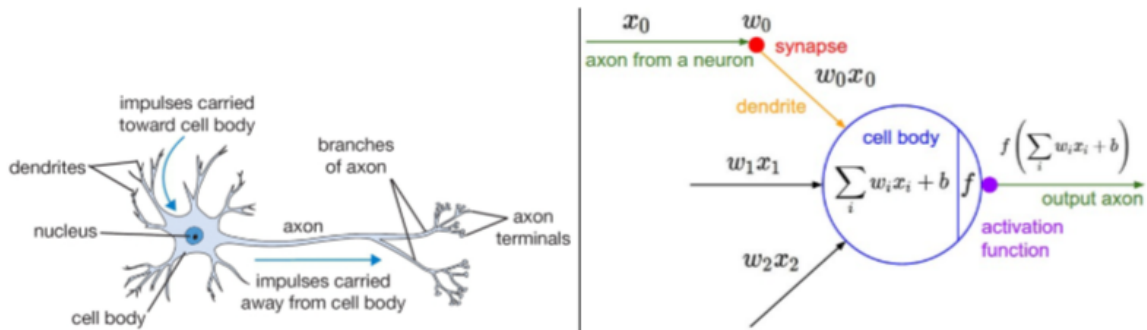
FIGURE 14: A cartoon drawing of a biological neuron (left) and its mathematical model (right) [6]

Activation functions are important elements to the network for various reasons. First, it is biologically inspired from the model of the neuron in our brains. Second, activation functions keep the value of the neuron's output restricted to a certain required limit.

Keeping the output of the neuron at a certain limit is important because the output can go very high in magnitude, particularly when the network is very deep. Third, the most important benefit of activation functions is that they add non-linearity into ANN. The ability of the ANN to learn non-linear patterns depends the addition of non-linear layers such as activation functions. Non-linearity is common in real world datasets. For example, voice recordings, images, or videos are composed of various dimensions, and thus, they can not be described with linear transformations. Activation functions have to be differentiable. This is necessary to perform the optimization strategy of back-propagation. In the learning context, the gradient descent (GD) optimization algorithm uses back-propagation. GD adjusts the weights of the neurons by estimating the gradient of the loss function.

Most of the activation functions struggle with the issue of vanishing gradient problem. As we add more layers with a certain activation function to an ANN, the gradients of the loss function start to be near zero which freezes network training. A near zero gradient dictates that the biases and weights of the first layers will not be updated efficiently during network training, and this can result in a loss of accuracy because the network is unable to recognize essential elements from the input data, which often happen at model entrance.

Figure 15 summarizes the most widely used activation functions. We will discuss them one by one. The sigmoid is an activation function which has an output range from 0 to 1. Therefore, it is used for networks where we need to predict the probability as an output. It is also centered between 0 and 1, which makes the gradient updates become too far in different directions, and thus, the optimization problem is harder than using other alternatives to the sigmoid activation function. The main issue with using the sigmoid activation function is that it is affected by the vanishing gradient problem and therefore, it is recommended to use the sigmoid activation function only with networks with low number of layers. The Tanh function looks similar to the sigmoid, however, the Tanh function is centered between -1 and 1. When we have two class classification problem, we can use the Tanh function. On the other hand, Tanh function is also affected by the

vanishing gradient problem, and thus we should consider the appropriate network size. The softamax is a generalized logistic activation function that can be used for multi-class classification, not just binary classification. It maps each output in such a way that the sum = 1. Given that its output is a probability distribution, we can use softmax as the final layer in an ANN classifier. ReLU is a very efficient and simple activation function that gained a lot of attention recently, especially in convolutional neural networks (CNN). It avoids the vanishing gradient problem. However, ReLU has a slope of 0 for negative values. Therefore, if a neuron stuck in negative values, it will always outputs zero (dead neuron). This problem is known as dying ReLU. There are other variants of ReLU where there is a non-zero slope at negative values [88].
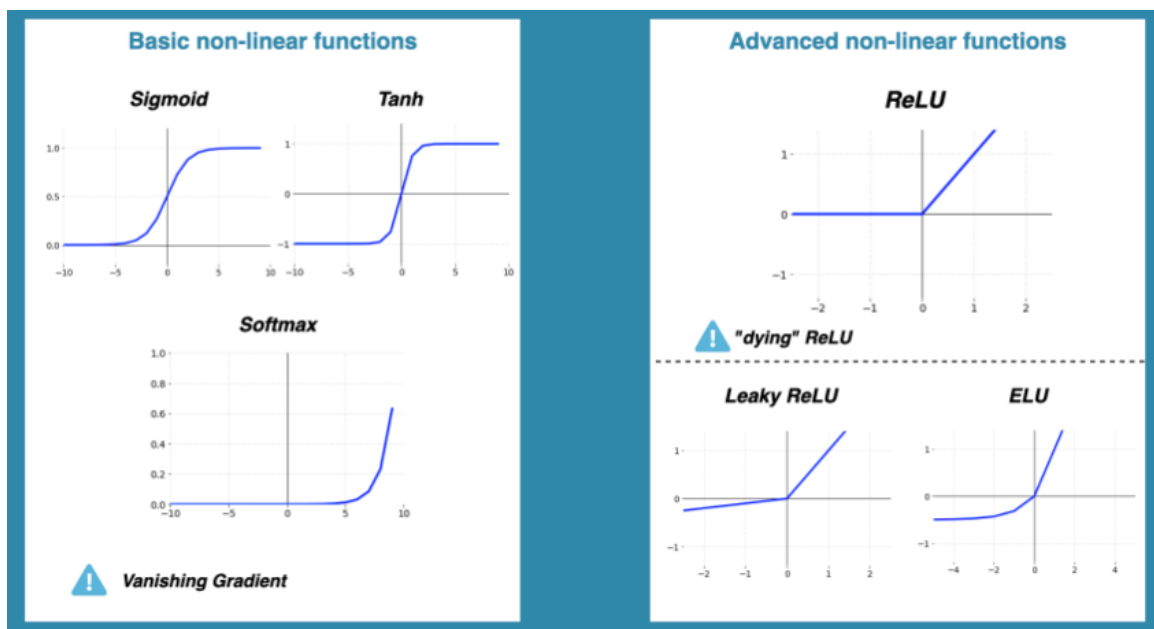
FIGURE 15: Popular activation functions [7].

4. Loss Function

In the domain of optimization, a candidate solution (i.e. set of weights) is evaluated using a function called the objective function. Our job might be minimizing or maximizing

29

the objective function, which means that we are looking for a candidate solution that has the lowest or the highest score, respectively. In the domain of ANN, we desire to minimize the objective function. Therefore, the cost/loss function term is used to refer to the objective function. We call the value estimated by the loss function as the loss. In other words, the loss function describes the prediction error that we want to minimize. The loss function distills all the aspects of the network into a single value and improvements in that value indicates an improvement in the model [4]. There are many loss functions: Mean Squared Error (MSE), Binary Crossentropy (BCE), Categorical Crossentropy (CCE), Sparse Categorical Crossentropy (SCCE), etc. MSE loss is usually used for regression problems. This loss is estimated by taking the average of squared differences between the actual values and the predicted outputs. In binary classification problems, BCE is used with a network that contains only one output neuron to classify the data into two classes. In order to limit the range of the output between 0 and 1, we should pass the output through a sigmoid activation function. CCE loss is used when we want to perform multi-class classification. The number of output neurons must equal the number of classes. The output neurons must pass through a softmax activation function to obtain a probability distribution. SCCE loss is similar to CCE loss. However, in SCCE, we do not need to encode the target vector by one hot encoding.

5.   Network Optimization

Gradient descent (GD) is an optimization algorithm that aims at finding the minimum of a function. In NN domain, the function that we want to minimize is the loss function. GD is a popular way to optimize neural NN [4]. Let us refer to the loss function we want to minimize by $J(\theta)$ which is parameterized by the model weights/parameters $\theta \in \mathbb{R}^d$. The algorithm finds the best solution by updating the weights in the opposite direction of the derivative of the loss function $(\nabla_\theta J(\theta))$ with respect to the weights. The size of the step that the algorithm takes is determined by the learning rate η. There are 3

variants of GD, where each one uses different amount of data to estimate the gradient of the loss function. Therefore, there is a trade off between the time needed to make an update and the accuracy of the weight update.

1. Batch gradient descent: it is also called vanilla GD. The entire training dataset is used to estimate the gradient of the loss function w.r.t. the weights θ.

$$\theta = \theta - \eta \cdot \nabla_\theta J(\theta) \tag{4}$$

As shown in Equation 4, we update our parameters in the opposite direction of the gradient with the learning rate controls the size of the step. To perform one update, it is required to estimate the gradients for the entire dataset. If the dataset does not fit in memory, batch GD became very slow and intractable. Updating the model online with new examples is not available in batch GD.

2. Stochastic gradient descent: compared to batch GD, SGD updates the parameters for each training example $x^{(i)}$ and label $y^{(i)}$.

$$\theta = \theta - \eta \cdot \nabla_\theta J(\theta; x^{(i)}; y^{(i)}) \tag{5}$$

For large datasets, GD can be very slow because one iteration requires a prediction for all training examples. Therefore, SGD is recommended for large datasets where training is much faster and a good set of parameters is reached after one to ten passes through the dataset. SGD is associated with frequent updates with high variance which lead to fluctuations in the loss function.

3. Mini-batch gradient descent: mini-batch GD makes an update for every min-batch of n samples.

$$\theta = \theta - \eta \cdot \nabla_\theta J(\theta; x^{(i:i+n)}; y^{(i:i+n)}) \tag{6}$$

Mini-batch GD has advantages such as (1) it has stable convergence because the variance of the parameter updates is reduced, and (2) it uses highly optimized matrix

31

computations that are found in the deep learning libraries. Therefore computing the gradients is performed in highly efficient way. Mini-batch GD is the algorithm of choice when training NN. Usually SGD term is used also when mini-batches are used.

Mini-batch GD has many problems:

1. Selecting the appropriate learning rate is challenging. The model will converge slowly if the learning rate is small. On the other hand, a high learning rate can hinder convergence and lead to fluctuations in the loss function around the minimum.

2. We should adjust the learning rate during training by using learning rate schedules. This is done by reducing the learning rate at a pre-defined schedule or when the loss function falls below a certain threshold. We have to define the schedules and the thresholds in advance. Therefore, these values do not adapt to the characteristics of the dataset.

3. NNs have non-convex error function. Therefore, mini-batch GD algorithm might be trapped in one of suboptimal local minima.

To overcome the aforementioned limitations, various optimizers have been proposed in the literature. For example, Nesterov accelerated gradient [89], Adagrad [90], Adadelta [91], and Adam [92].

C.   Convolutional Neural Networks

CNNs are inspired by the human visual cortex [93]. It is popular and the first choice in computer vision applications such as object classification and detection, and video recognition. CNN can also be used in other fields such as drug discovery [94] and natural language processing [95]. Figure 16 illustrates an example CNN that is used for the classification of hand written digits. The input to the CNN is a 2D image with size $28 \times 28 \times 1$

pixels. Generally, the CNN is composed of a series of convolution and down-sampling layers and at the end a fully connected layer followed by a normalizing layer such as softmax activation function. A progressively more refined features are created by the series of the convolution layers that are present from the input to the output layers. The fully connected layer that follows the series of the convolution and down-sampling layers performs classification. Between the convolution layers, down-sampling or pooling layers are usually inserted.
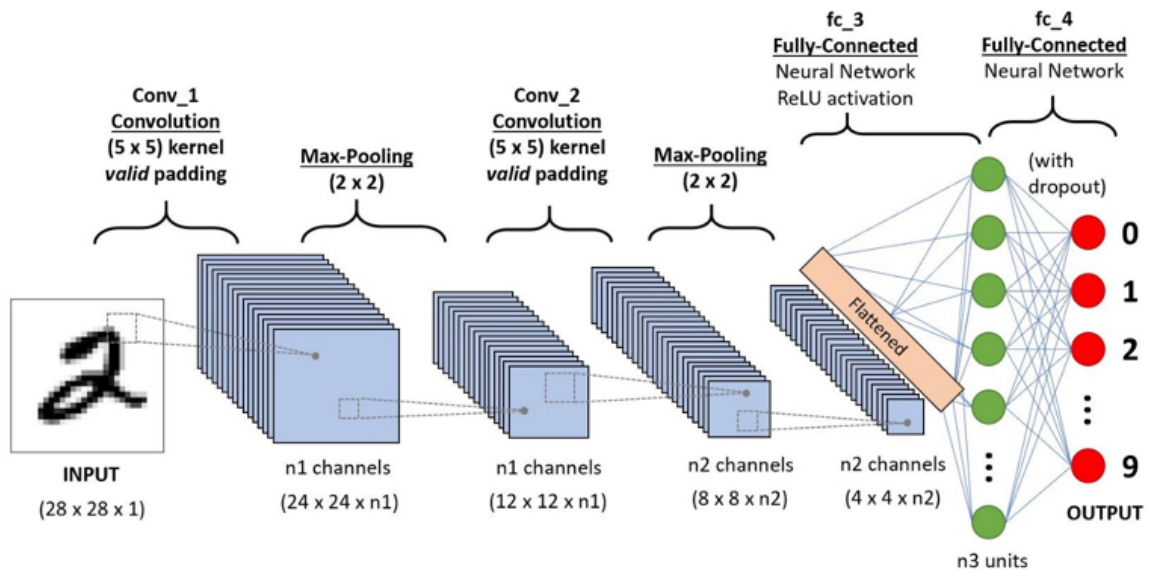
FIGURE 16: Example of a convolutional neural network for hand-written digits classification [8]

The kernels or filters are group of 2D neurons which are the basic components of each layer. Unlike traditional NN, in CNN the neurons are not connected to all neurons in the previous layer. Instead in the CNN, there is a concept of the receptive field where the neurons are only connected to a fixed number of neurons. The connection between two adjacent layers is determined by the spatial mapping. Reducing the number of connections between adjacent layers decreases chances of overfitting and training time. The neurons of the kernel are connected to the same number of neurons in the previous layer or feature

map. These facts reduce the learning time and decrease network's memory requirement. Therefore, each neuron in the kernel finds the same texture but in different areas of the image. Figure 17 shows an example of convolving a $5 \times 5 \times 1$ image with a kernel of size $3 \times 3 \times 1$ to get a $3 \times 3 \times 1$ feature map.

$$(5x1)+(8x0)+(3x1)+(7x0)+(5x1)+(4x1)+(9x1)+(2x1)+(1x0)$$
$$= 28$$
$$5 + 0 + 3 + 0 + 5 + 4 + 9 + 2 + 0$$
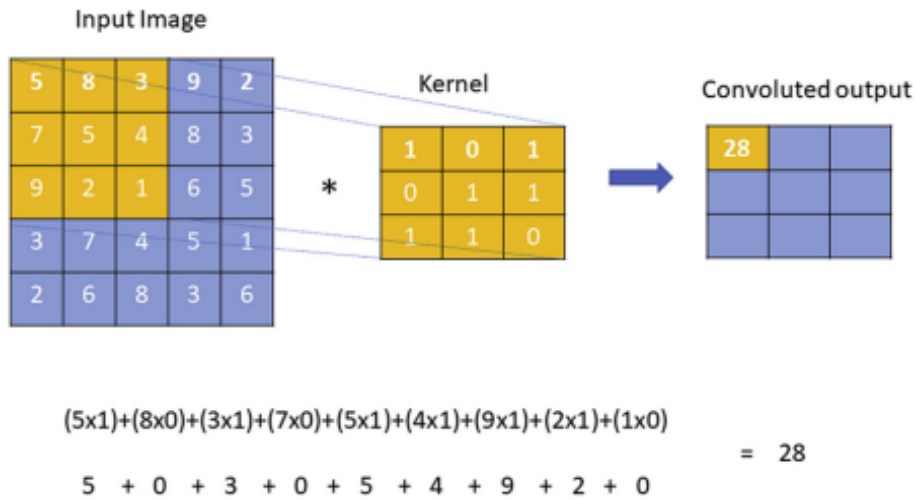
FIGURE 17: Convolution operation in the convolution layer of the CNN [9].

If we have a stride length = 1, the kernel will shift 9 times. Figure 18 shows the movement of the kernel. The kernel shifts to the right with a stride = 1 until it parses the complete width. Then, the kernel hops down to the beginning of the image with the same stride value, repeating the same operation till the whole image is traversed.
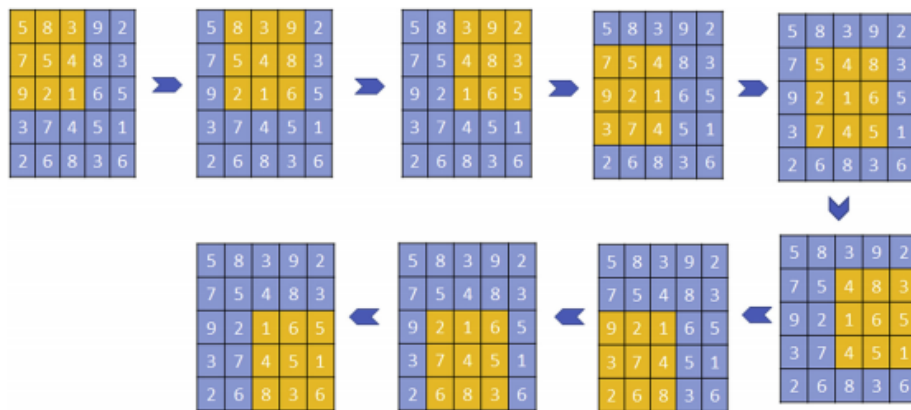
FIGURE 18: Movement of the kernel during the convolution operation [9].

Down-sampling or pooling layers decrease the size of the network. It also work on reducing the susceptibility of the network to scale, shift, and distortion of images [96]. To achieve down-sampling, max-pooling or average-pooling are often used. Figure 19 shows an example of applying max-pooling on an image of size $4 \times 4 \times 1$ with a $2 \times 2 \times 1$ kernel and stride = 2.
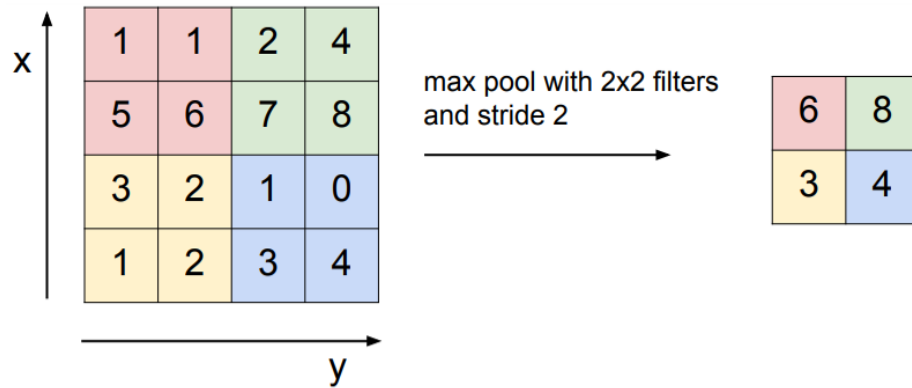
FIGURE 19: Pooling operation [6].

Actual classification is performed by the final layers in the CNN where layers are fully connected. The fact that the CNN is a deep architecture results in good quality representations and in the same time reducing parameters, maintaining locality, and invariance to minor changes in the input image [97]. Usually, backpropagation is performed to CNN training to get the optimal network parameters (weights and biases).

D.   Encoder-Decoder Architectures

Fully convolutional neural networks (FCNN) is a CNN with only convolutional and down-sampling layers [98]. In other words, FCNN is a CNN with the final fully connected layer replaced with convolutional layer. FCNN is mainly used for image semantic segmentation. Image semantic segmentation involves classification for the pixels of the image. FCNN has an encoder-decoder architecture which is a popular end-to-end model. Other architectures include U-net [10] and Deeplab [99]. In these models, image features are ex-

tracted by an encoder, while the original image size is restored from the image features by a decoder. The decoder also produces the final segmentation. Although the end-to-end architecture is useful for medical image segmentation, it works on reducing the interpretability of the model. The U-net is the first high impact encoder-decoder architecture proposed by Ronneberger et al. [10]. The U-net which is shown in Figure 20 has been widely used for performing segmentation for medical images.
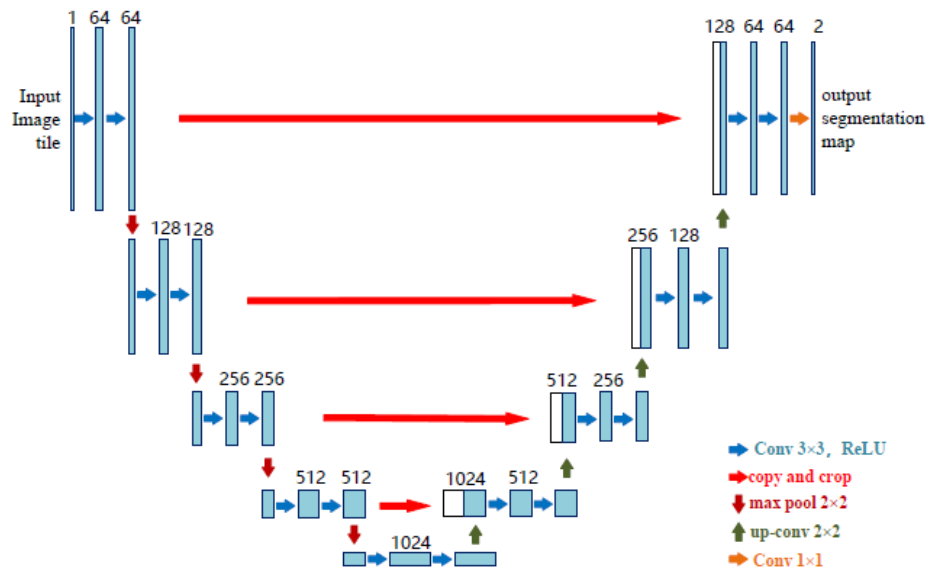
FIGURE 20: U-net architecture [10].

U-net uses a perfectly symmetric structure with skip connections. Apart from usual image segmentation tasks, the segmentation of medical images needs to tackle the noise and blurring in the image. Therefore, it is challenging to recognize objects in medical imaging by relying only on the low-level features of the image. Meanwhile, it is difficult to get accurate boundaries by relying only on semantic features of the image because there is a lack in image details. On the other hand, U-net makes a fusion between low-level and high-level features of the image. This is performed by integrating low-resolution and high-resolution feature maps through skip connections. Skip connection is a perfect solution for performing segmentation for medical images. Currently, U-net is considered the benchmark for medical image segmentation and resulted in a lot of successes.

E. Deep Learning Tools

1. TensorFlow

TensorFlow is an open-source ML platform [100]. It is used to implement deep learning applications. Google team created TensorFlow to research and develop on fantastic ideas of AI. To use TensorFlow, we program in Python programming language, and thus it is considered easy to use and interpret. TensorFlow has the following features:

- It optimizes and estimates mathematical expressions easily with the use of arrays called tensors. Tensors are the data structures in TensorFlow platform, see Figure 21.

- It features various levels of abstraction, thus you can select the right level for your task. For example, you can choose the high-level Keras API that makes building ML applications easy.

- It offers a direct path to production. Whether you intend to build your application on servers, web, or edge devices, TensorFlow allows you to train and deploy your model very easily, regardless of the programming language or platform you use. For example you can use TensorFlow Lite to run inference on edge devices and mobile, and you can use TensorFlow.js to train and deploy applications in JavaScript environments.

- You can build and train state-of-the-art architectures without sacrificing performance or speed. You can use Keras Functional API or Model Subclassing to build complex architectures.

- It harness the power of GPU to improve model performance.
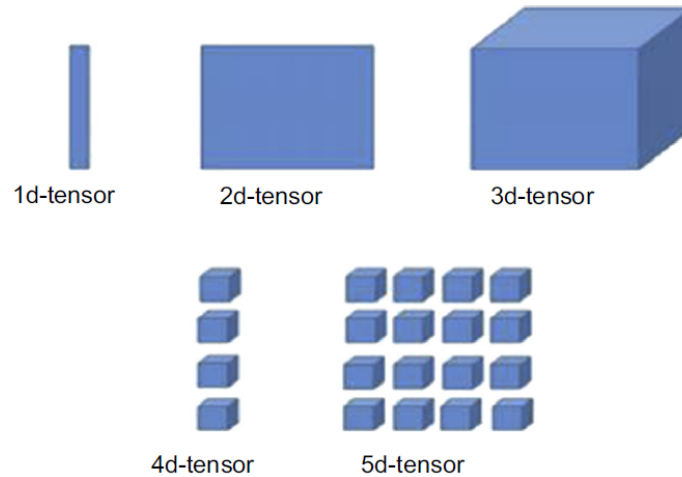
FIGURE 21: Different dimensions of the tensors in TensorFlow platform [9].

2. CAFFE

CAFFE (Convolutional Architecture for Fast Feature Embedding) is an open-source deep learning framework developed at University of California by Yangqing Jia during his Ph.D. at the university [101]. CAFFE is written in C++ programming language with a Python interface. CAFFEE has the following features:

- It supports various types of deep learning models, such as CNN, region based CNN, long short term memory (LSTM), and fully connected ANN.

- It supports CPU- and GPU-based acceleration computational libraries such as Intel MKL and NVIDIA cuDNN

- It can be used in academic tasks, startup prototypes, and even industrial applications in speech, vision, and multimedia.

3. Pytorch

Pytorch is an open-source ML library that is based on the Torch library [102]. The AI research lab at Facebook developed Pytorch which has a Python and C++ interfaces.

Pytorch has the following features:

- Production ready.

- Distributed Training.

- Robust ecosystem.

- Cloud support.

F.   Deep Learning in Medical Imaging

Deep learning can be used to build computer-aided diagnosis (CAD) systems for disease diagnosis and classification. Deep learning can be used in such systems because it has shown great successes in many medical imaging applications such as image classification, image segmentation, image registration, image restoration, and image generation and transformation. In the following sections we will discuss medical image classification and segmentation which are the focus of this dissertation.

1.   Medical Image classification

Image classification is a task where we need to assign one label from a fixed group of categories to the input image. In supervised deep learning, we train the model by a set of training images and their associated ground truth labels. Then, we test the model by testing images and evaluate its performance by comparing the actual model's output with the ground truth labels of the testing images. One important task for physician is proper differential diagnosis for the medical images of patients. This task is a classification task which involves a wide range of applications from detecting the presence or absence of a disease to determining the type of malignancy. CNN has showed great successes in image-based classification applications such as kidney diagnosis [103], diagnosis of diabetic retinopathy [104], and diagnosis of skin cancers [105]. Figure 22 shows that CNNs can be trained

39

by various medical imagery such as radiology, dermatology, pathology, and ophthalmology. The values of the softmax layer (the output vector) can refer to the probabilities of the presence of disease or the type of abnormality.
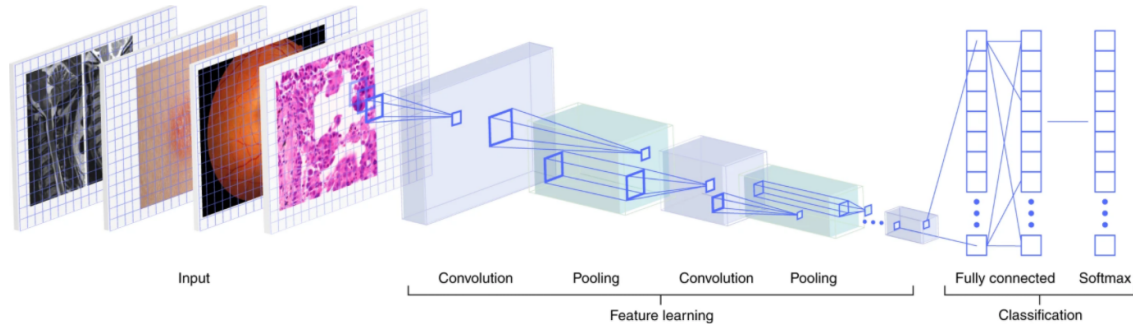
FIGURE 22: Convolutional neural networks can be trained by various medical imagery such as radiology, dermatology, pathology, and ophthalmology. [11]

2.   Medical Image Segmentation

Image segmentation involves assigning a label to each pixel in the input image. Segmentation task can be considered classification task in the spatial domain of the input image. Therefore, images segmentation is a dense classification because we classify each pixel in the input image. Encoder-decoder architectures segment medical images with high accuracy such as the segmentation of the heart [106], and the kidney [107]. Figure 23 shows an application of encoder-decoder architecture for segmenting the heart from cardiac magnetic resonance images.
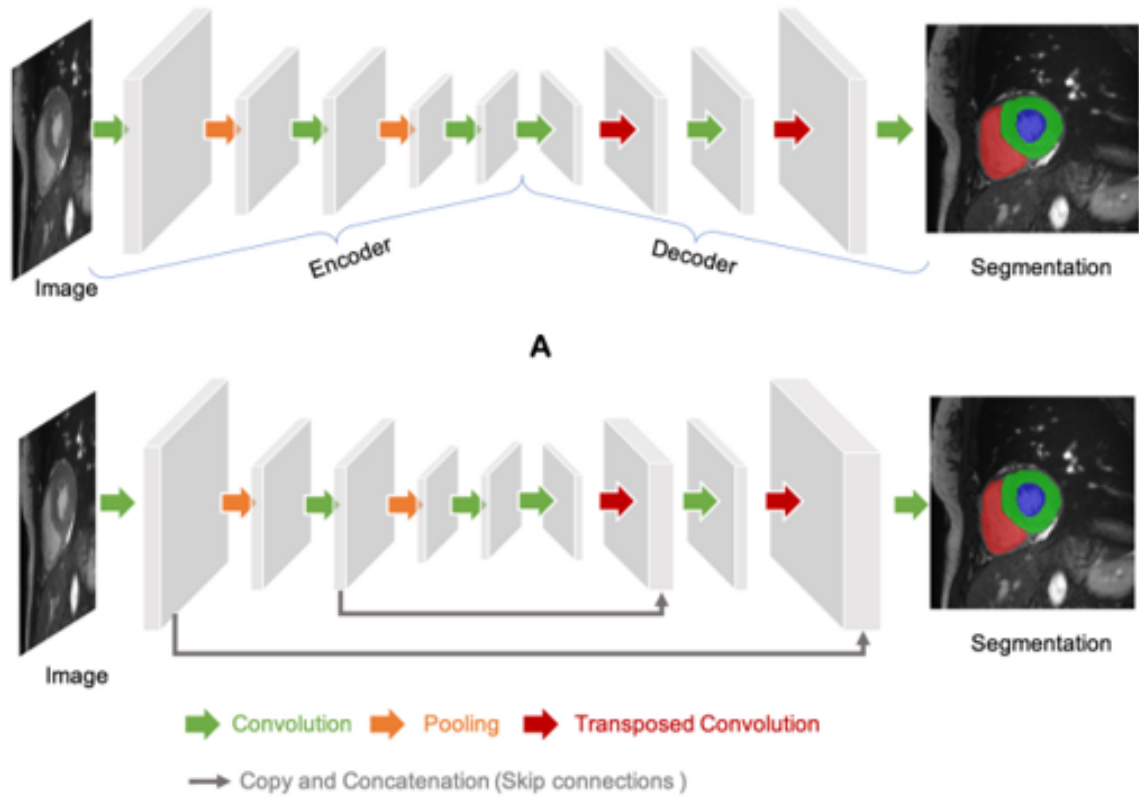
FIGURE 23: Encoder-decoder architectures can be used for medical image segmentation. The figure shows two architectures one without skip connections and one with skip connections. The blue, green, red colors refer to the left ventricle blood pool, left ventricle, and right ventricle, respectively [12]

CHAPTER III
EARLY DETECTION OF ACUTE RENAL TRANSPLANT REJECTION USING DIFFUSION WEIGHTED MRI AND CLINICAL BIOMARKERS

In this chapter, a novel machine learning framework for the classification of non-rejection (stable) versus acute renal rejection status of kidney transplants is proposed. The proposed framework uses the diffusion weighted (DW) MRI and clinical biomarkers to perform its function. From the DW-MRI, we estimate the image markers which is the apparent diffusion coefficients (ADC) which represent the perfusion of the blood and the diffusion of the water in renal graft. The clinical biomarkers are creatinine clearance (CrCl) and serum plasma creatinine (SPCr) which are considered kidney functionally indices. Therefore, our framework is a non-invasive diagnostic tool for the early detection of acute renal transplant rejection. The framework consists of three main steps. In the first step, we apply preprocessing for the DW-MRI data where we perform histogram equalization and region of interest (ROI) extraction. In the second step, we perform feature extraction where we estimate the 3D ADC maps and fuse the clinical biomarkers with the ADC to produce new maps with more discriminatory power. In the third step, we employ a CNN based classification system to differentiate between non-rejection and acute renal rejection. The framework is trained and evaluated using DW-MRI scans obtained from 56 cases who belong to geographically diverse populations. Furthermore, different imaging systems and protocols were used. The proposed framework achieved an overall accuracy of 92.9%, sensitivity of 93.3%, and specificity of 92.3% in distinguishing non-rejection from acute renal rejection kidneys. Our results established the potential of the proposed framework for the reliable and non-invasive diagnosis of kidney transplant status for any DW-MRI scans, regardless of imaging protocol and/or geographical differences. Furthermore, our framework

does not require kidney registration and segmentation steps which add complexity in the system as previously presented in the literature.

A. Kidney Function, Anatomy, and Transplanted Kidney Diseases

The kidneys are important organs that perform various functions in the human body. First, kidneys excrete metabolic waste products, drugs, foreign chemicals, and hormone metabolites. The waste products that the kidneys eliminate are no longer needed by the body. Examples of these products are urea, creatinine, uric acid, metabolites of various hormones, and end products of hemoglobin breakdown. These products should be removed from the body as fast as they are produced. The kidneys also remove foreign substances that are created by the human body or ingested, such as drugs, pesticides, and food additives. Second, kidneys regulate water and balance electrolytes. They maintain homeostasis by excreting water and electrolytes with an amount equal the intake. The person's eating and drinking habits govern the intake of water and electrolytes and the kidneys adjust the removal rate to equal the intake of different substances. Third, kidneys regulate arterial pressure. Kidneys excrete variable amount of sodium and water. Therefore, they regulate the arterial pressure in a long-term. The kidneys also regulate the arterial pressure in a short-term by secreting hormones and vasoactive substances. Fourth, kidneys regulate acid-base balance. Fifth, kidneys regulate erythrocyte production. Finally, the kidneys play an important role in glucose synthesis from amino acids during fasting [13].

The two kidneys lie outside the peritoneal cavity on the posterior wall of the abdomen (Figure 24). The kidney of the adult person has a size of a clenched fist and weighs 150 grams approximately. The hilum is an indented region that lies at the medial side of each kidney. The renal artery and vein, nerve supply, lymphatics, and ureter pass through the hilum. The urine is carried by the ureter from the kidney to the bladder. The urine remains in the bladder until it is emptied through the urethra. The inner structures of the kidney are protected by a tough capsule that surrounds the kidney. If we bisected the kid-

ney from top to bottom, we will find two major regions which are the inner medulla and the outer cortex regions. The medulla contains cone-shaped structures of tissue named renal pyramids. Approximately 22 percent of the cardiac output flows to the two kidneys through the renal artery which enters the organ through the hilum.
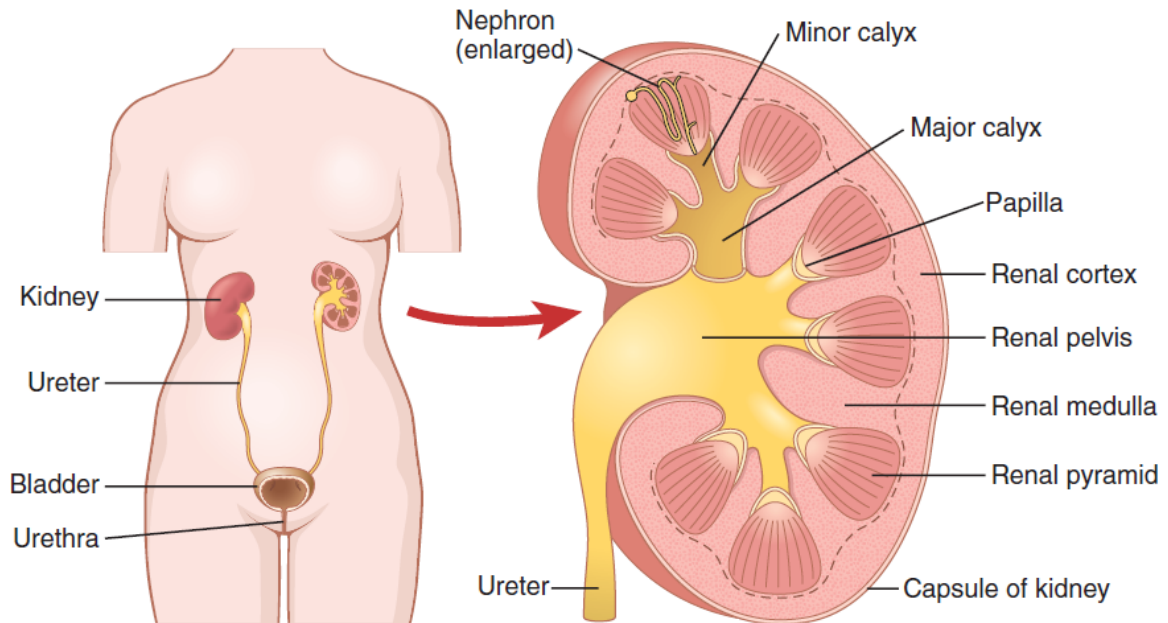
FIGURE 24: Organization of the urinary system and the kidneys [13].

The functional unit of the kidney is the nephron. A person's normal kidney contains approximately 800,000 to 1,000,000 nephrons. The nephron is responsible for urine formation. The number of functioning nephrons gradually decreases with normal aging, renal injury, or disease. Each nephron has two main components: (1) a glomerulus which is a tuft of glomerular capillaries. In the glomerulus large amounts of the fluid are filtered. (2) a long tubule where the urine is formed from the filtered fluid (see figure 25). The glomerulus contains a network of capillaries that have high hydrostatic pressure. The glomerulus is covered by Bowman's capsule that receives the fluid filtered from the capillaries of the glomerulus. Then, the fluid enters the proximal tubule which is located in the kidney's cortex (Figure 26). After leaving the proximal tubule, the fluid enters the loop of henle. The loop consists of a descending limb and an ascending limb. Then, the fluid travels in many

tubular segments until it reaches the collecting duct as shown in Fig. 26. The normal urine of well-functioning kidney must be free from blood cells, protein, and glucose.
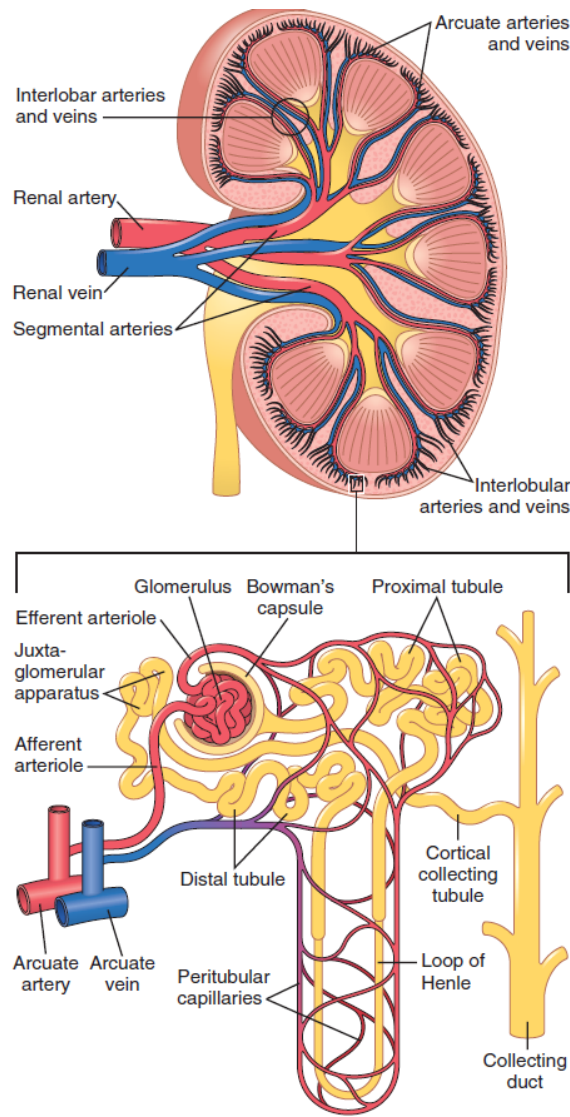
FIGURE 25: Schematic illustration for a section in a human kidney that shows the microcirculation of the nephron and the main vessels that are responsible for blood supply [13].

Large amounts of fluid is filtered from the glomerular capillaries to Bowman's capsule and this represents the first step in urine formation where about 180 liters are filtered each day. However, most of the filtered fluid is reabsorbed and only one 1 liter is excreted each day. The glomerular filtration rate (GFR) is controlled by two factors (1) the balance
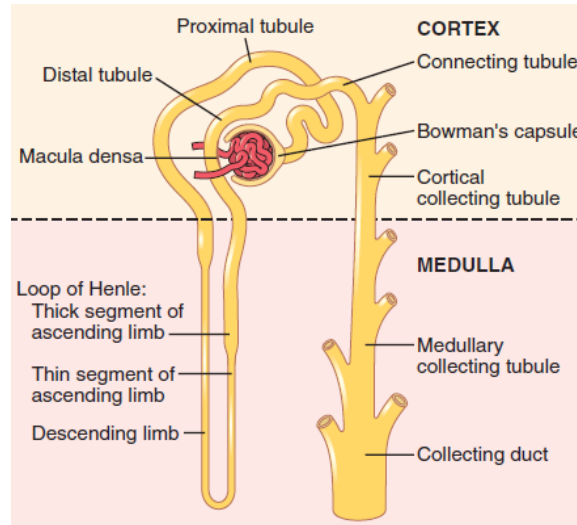
FIGURE 26: The main tubular segments of the nephron [13].

of colloid osmotic and hydrostatic forces and (2) the capillary filtration coefficient. The GFR of an average adult human is about 125 ml/min, or 180 liters/day. Approximately 20% of the plasma entering the kidney is filtered by the glomerulus capillaries.

The transplanted kidney can experience various renal complications that deteriorate its function, as shown in Figure 27 there are 6 categories: (1) fluid collections, (2) urologic complications, (3) Neoplasmas, (4) vascular complications, (5) recurrent disease, (6) graft dysfunction [108]. Transplant fluid collections include urinomas, lymphoceles, hematomas, and abscesses. These complications are reported in up to fifty percent of renal transplantations and their size, location, and growth possibility significantly affect their clinical relevance [109]. Within two weeks from transplantation, urologic complications can happen and they are manifested as urine leaks with discharged urinomas. Urinary obstruction and calculous disease are additional complications that the renal transplant patients can experience. Cancer development is increased by kidney transplantation, especially with extended immunosuppression period. Examples of neoplasms risks are renal cell carcinoma and lymphomas [110]. Vascular complications encompass transplanted artery infarction, stenosis, renal vein thrombosis, and arteriovenous fistulas and pseudoaneurysms. The prevalence of these complications is only 10% of transplantation subjects.

However, they are associated with grave graft dysfunction with high death rate [111]. The prevalence of the recurrent diseases is very low in the early stage after transplantation. However, recurrent diseases can be found after a long-term in patients who suffer from diabetes, cystinosis, or amyloidosis [112]. Graft dysfunction is another type of renal complications which is considered one of the major problems that causes graft loss [113]. Graft dysfunction can cause acute tubular necrosis, drug nephrotoxicity, and rejection [39]. Acute tubular necrosis is related to the donor kidney and is widely detected in patients whose renal transplant are from living relatives [34]. Drug toxicity is another form of the problems that deteriorate renal graft function. Cyclosporine is an immunosuppressive drug that has a high nephrotoxic potential with an effect on glomerular arterioles [33]. A major cause of renal transplant dysfunction is acute renal rejection which represents the response of the immune system to the foreign organ. Acute renal rejection can lead to a complete loss of the transplanted kidney. In the next section, we will provide more details regarding acute renal rejection, which is the focus of this dissertation. We will also introduce current approaches that aim to the early detection of this kidney disease [114].
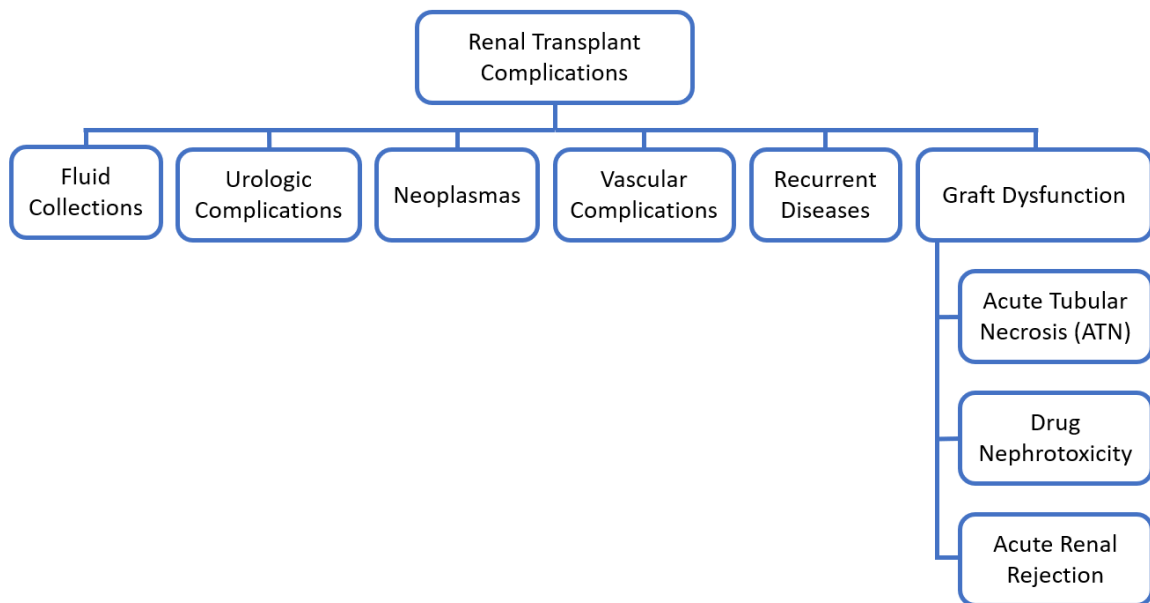
FIGURE 27: The different complications of renal transplant.

B.   Acute renal transplant rejection

Acute renal transplant rejection is a critical problem in the urology field. In the United States, approximately 22,393 kidney transplants were performed in 2018 [32]. Usually, the number of kidney donors is limited. Therefore, it is important to maintain the transplanted kidney in a healthy state. Acute renal rejection can occur in up to 40% in kidney recipients within 3 weeks after transplantation. Typically, high-dose steroids or antibody therapy can treat acute renal rejection. A gradual damage in the transplanted kidney function leads to chronic rejection, which starts about at 3 months after transplantation [33]. Given that providing an effective therapy for chronic rejection is an open area of research, avoiding acute renal rejection is the most promising way to prevent chronic rejection [34]. Therefore, it is of paramount importance to early detect rejection to intervene with the suitable medical and immune therapy in kidney transplant recipients [35].

Currently, initial assessment of kidney transplant function is based on blood and urine tests such as plasma creatinine and creatinine clearance. creatinine clearance is a laboratory test that can be used to evaluate kidney function, such as GFR. Urologists consider creatinine clearance as a practical measurement to evaluate kidney function. However, we can not rely on such index to effectively detect renal rejection due to the following reasons: (1) creatinine clearance gives an overall idea about the function of both kidneys together not about a specific kidney [36], (2) we can observe a significant change in creatinine level only after 60% loss of kidney function [37]. Biopsy is considered the gold standard to evaluate graft function, but only as the last option because it is associated with potential morbidity and high costs. Also, the invasive nature of this procedure imposes a risk of infection and bleeding to the patients. Furthermore, the sample that we get from the needle biopsy is small and might lead to underestimation or overestimation of the degree of inflammation in the entire graft [38]. Therefore, a technology that is based on accurate and noninvasive techniques is so much needed to evaluate transplanted kidney function and to early detect acute renal rejection.

48

Various noninvasive imaging techniques have been used in clinical settings to evaluate the function of transplanted kidney. Graft function can be evaluated quantitatively and qualitatively by an imaging modality called radionuclide imaging which is a traditional renal imaging method used to screen for common complications [39]. However, this technique suffers from limited spatial resolution, therefore, it does not show accurate anatomical information of functional abnormalities inside the kidney [40]. Additionally, radionuclide imaging is associated with exposure to radiation which limits the range of its applications. For example, it can not be used to monitor diseases such as acute tubular necrosis or cyclosporin [115]. Ultrasound imaging can be used to assess the function of the graft early in the postoperative period and it is suitable for long-term evaluation of the graft. Ultrasound is non-nephrotoxic modality and relatively cheap. However, this modality suffers from the following drawbacks: (1) investigator's experience greatly affects the sensitivity and reliability of this technique, (2) it has low signal-to-noise ratio, (3) it suffers from shadowing artifacts and speckles which decrease the diagnostic confidence [41]. Computed tomography (CT) is successful in diagnosing post transplantation complications [42]. CT is better than radionuclide imaging in terms of safety. However, it has low specificity and its process contains nephrotoxic agent. Different magnetic resonance imaging (MRI) techniques have been introduced to perform assessment for kidney transplant function. For instance, dynamic contrast enhanced (DCE) MRI utilizes contrast agents, such as gadolinium to assess tissue perfusion, which gives an idea about kidney function [43]. Despite the fact that good anatomical and functional information can be obtained from DCE-MRI, the use of the contrast agent (gadolinium) can lead to nephrogenic systemic fibrosis [44], especially when $GFR < 30\,ml/min$ per $1.73\,m^2$. Blood oxygen level dependent (BOLD) MRI which assess the renal oxygenation level has been utilized for the evaluation of kidney transplant function and the detection of acute renal rejection [45]. However, BOLD-MRI has several limitations such as susceptibility to breathing motion artifact and bowel gas artifacts [46, 47].

Diffusion weighted (DW) MRI is an imaging modality that is not associated with contrast agents. DW-MRI has been successful in detecting and characterizing tumors, neuroimaging, and kidney transplant function evaluation [48]. DW-MRI measures water's motion inside the tissue, and thus it is considered a way to evaluate the diffusion characteristics for the tissue. We can obtain apparent diffusion coefficient (ADC) from DW-MRI. ADCs are quantitative maps which represent the diffusion and can be measured at different magnetic field strengths and duration noted as b-values [48].

C.   Related Work

The applicability of DW-MRI in kidney transplant function assessment. has been investigated in multiple studies [116–126]. In these studies, they used the ADC as a discriminatory feature to classify between biopsy proven healthy renal allografts and unstable allografts that have problems such as acute tubular necrosis and acute renal rejection. The results of those studies demonstrated that allografts with deteriorated function have lower ADC values compared to stable allografts. Furthermore, a positive correlation was found between the estimated GFR and the ADC values. On the other hand, these studies have drawbacks. First, the ADCs were estimated from the middle or the largest cross-section at a specific b-value only. Second, they did not study the result of integrating clinical and imaging biomarkers to classify between stable and allografts with acute renal rejection. Finally, they did not investigate the potential of deep learning as a valuable tool for disease diagnosis and image classification.

The application of machine learning and deep learning in kidney diagnosis and classification has gained some attention. Pedraza et al. [127] investigated the use of CNN that is based on a pre-trained AlexNet model [128]. Their study aimed at classifying glomerular tissue versus non-glomerular tissue. They trained their network by images obtained from tissue slides extracted from kidney biopsies and they achieved a performance of 0.999 (F-score). Yang et al. [129] proposed a method based on CNNs to classify histological kid-

ney images adapted from tissue microarrays. They prepared the microarrays from kidney biopsies from tumor and notmal subjects. Their classification accuracy was 97%. Kolachalama et al. [130] investigated the use of CNN to predict the stage of chronic kidney disease, nephrotic-range proteinuria, and baseline serum creatinine. Their study was based on the analysis of richrome-stained images adapted from kidney biopsy samples and reported a classification accuracy that is comparable to an expert nephropathologist. March et al. [131] proposed an extended version of a pre-trained model to classify non-sclerosed versus sclerosed glomeruli. From the analysis of frozen sections obtained from kidney biopsies, they aimed to assess the eligibility of a donated kidney before renal transplantation to minimize the complications of post-transplantation. In spite of the success of the mentioned computational methods [127, 129–131], they rely on an invasive procedure (renal biopsy). Shehata et al. [1] proposed a deep learning framework that is based on stacked nonnegative constrained autoencoders. The framework aimed at classifying stable renal allografts versus allografts with acute renal rejection from DW-MRI data. However, their method contains a segmentation step for the kidney where they use deformable model. This step adds a computational burden to the overall system.

D. Methods

In this chapter, we propose a machine learning framework that aim at classifying stable transplanted kidneys (non-rejection transplants) versus acute renal rejection transplants by analyzing DW-MRI data, see Figure 28. Our framework is efficient in terms of it does not require kidney segmentation, and non invasive in terms of it requires only imaging and clinical biomarkers. Furthermore, our framework is intended to early detect acute renal transplant rejection. It is composed of the following three steps: (1) preprocessing where we perform histogram equalization and region of interest (ROI) extraction, (2) feature extraction where we estimate the 3D ADC maps and fuse the clinical biomarkers, (3) CNN based classification to non-rejection and acute renal rejection transplants.
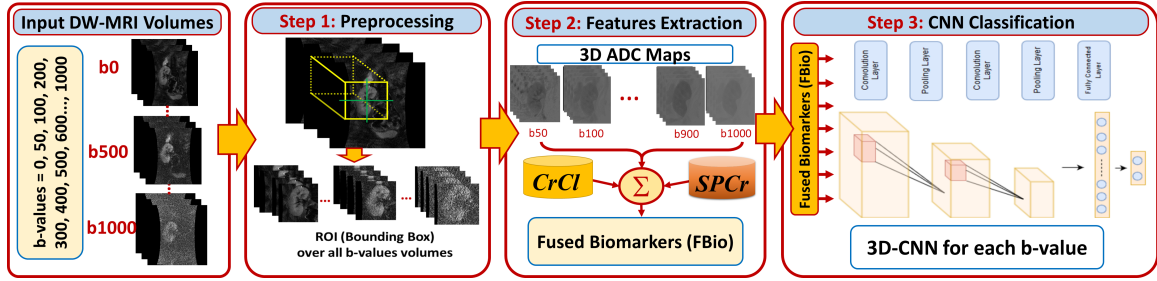
FIGURE 28: The proposed machine learning framework for early detection of acute renal rejection using DW-MRI analysis. The framework is composed of three main steps. In the first step, we equalize the histogram to reduce the inhomogeneity and noise. Then, we construct an ROI that encloses the kidney for each subject. In the second step, we estimate the 3D ADC maps for the chosen ROI and we fuse the clinical biomarkers (creatinine clearance (CrCl) and serum plasma creatinine (SPCr)). Finally, the fused markers (3D volumes of size 150x150x24) are inputted to the CNN classifier to classify non-rejection versus acute rejection allografts.

1. Patient Data and MRI Acquisition Protocol

In this study, we collected data for 56 patients who had renal transplantation. The data is composed of DW-MRI images (image markers) and clinical biomarkers, i.e., the creatinine clearance (CrCl) and serum plasma creatinine (SPCr). We used the collected data to train and evaluate our proposed machine learning framework. The protocols followed in our experiments were approved by the Institutional Review Boards (IRB) of University of Louisville, USA; the University of Michigan, USA; and University of Mansoura, Egypt. We carried out the experiments in accordance with the relevant regulations and guidelines. A written and/or a verbal consent were obtained from the participants after fully informing them regarding the aims of the study. The number of males was 38 participants and the number of females was 18 participants. The average of their age was 34 ± 15.62 with a range of 12 to 65. We divided the subjects into two groups: patients with stable allograft functionality or Non-rejection group (N = 26) and patients with acute renal rejection (N =

30).

The DW-MRI data was collected from different sources/scanners. The first source was 3T scanner: MRI Ingenia, Philips Medical System, Amsterdam, Netherlands, where we collected 17 DW-MRI scans in the United States. The MRI sequence was gradient single-shot spin-echo echoplanar with body coil. Imaging parameters were as follows: TR/TE = 8000/93, slice size = 256x256 pixels, inter-section gap = 0 mm, section thickness = 4 mm, field of view = 360×360×152 mm^3. To cover the whole kidney, 38 coronal cross-sections were acquired in 60 s. The second source was 3T scanner: MRI Ingenia, Philips Medical System, Amsterdam Netherlands, where we collected 5 DW-MRI scans in Egypt. The MRI sequence was gradient single-shot spin-echo echoplanar with body coil. Imaging parameters were as follows: TR/TE = 4400/82, slice size = 176×176 pixels, intersection gap = 0 mm, section thickness = 4 mm, field of view = 220×195×96 mm^3. To cover the whole kidney, 24 coronal cross-sections were acquired in 30-60 s at each b-value. The third source was 1.5T scanner: SIGNA Horizon, General Electric Medical Systems, Milwaukee, WI where we collected 34 DW-MRI scans in Egypt. The MRI sequence was gradient single-shot spin-echo echoplanar with body coil. Imaging parameters were as follows: TR/TE = 8000/61.2, slice size = 256×256 pixels, intersection gap = 0 mm, section thickness = 4 mm, field of view = 360×360×152 mm^3. To cover the entire kidney, 50 coronal cross-sections were acquired in 60-120 s. In addition to the DW-MRI data, we collected clinical biomarkers (CrCl and SPCr) for each patient.

Kidney motion will not affect our analysis because surgeons implanted the allograft in the iliac region, and therefore, the implanted kidney is less sensitive to respiration than a native kidney. Furthermore, operators asked the patients to hold their respiration during image acquisition to minimize any possible respiratory motion artifacts. We believe that single direction will reduce the acquisition time to about 12 min compared to three direction which needs about 31 min. Therefore, we were able to acquire 12 b-value scans. Lower b-values are responsible for capturing the blood perfusion while higher b-values

are responsible for capturing water diffusion effects. The acquisition of 12 b-values will enhance the final diagnostic performance because our decision will be based on integrating individual decisions from the 12 b-values. The acquired b-values are 0, 50, 100, 200, 300, 400, 500, 600, 700, 800, 900, and 1000 s mm^{-2}. Figures 29, 30 show examples of cross-sections for non-rejection and acute rejection kidneys at all b-values, respectively.
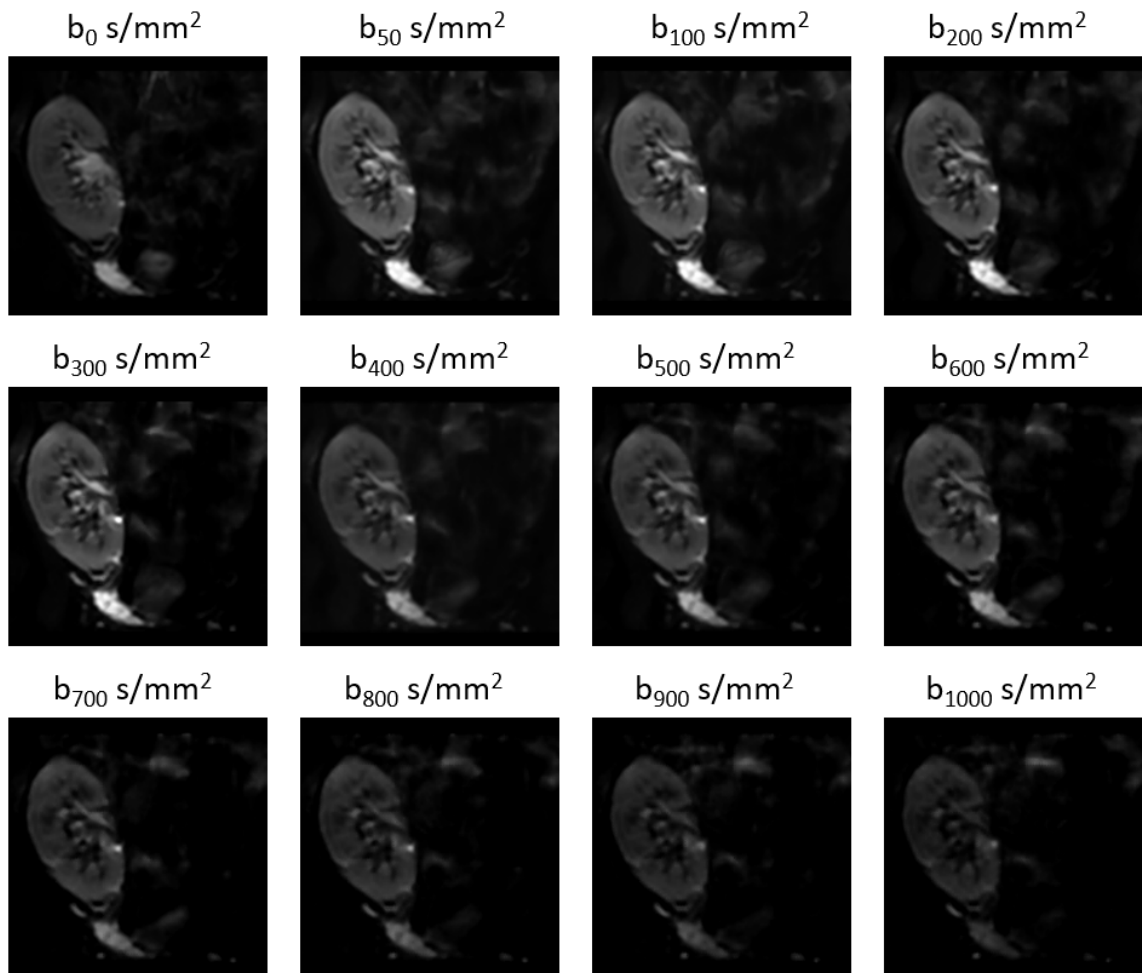
FIGURE 29: A coronal kidney cross-section of a non-rejection kidney (stable) at the mentioned 12 b-values.
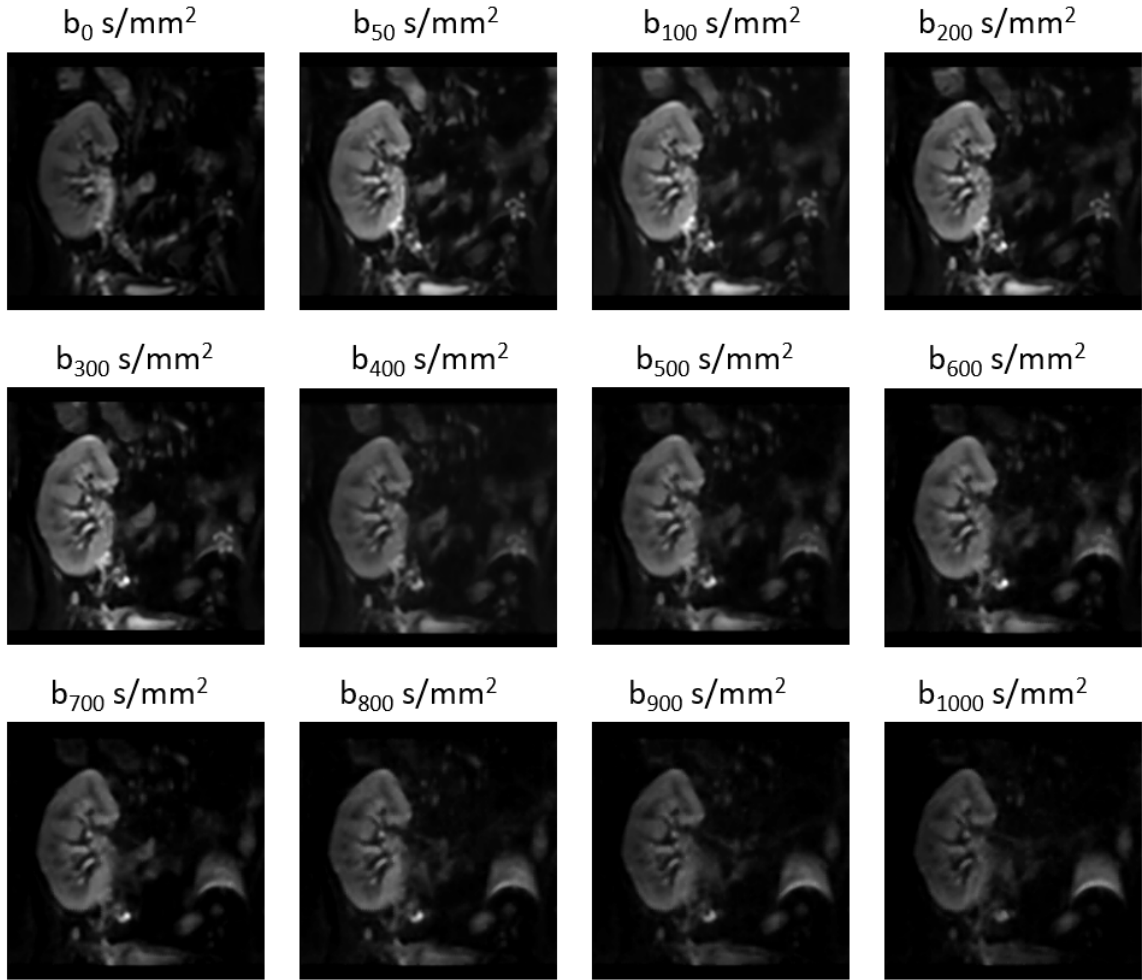
FIGURE 30: A coronal kidney cross-section of an acute rejection kidney at the mentioned 12 b-values.

2. Data Preprocessing and ROI selection

A histogram equalization with non-parametric bias field correction was applied on the DW-MRI as a preprocessing step to reduce inhomogeneity of intensities or the noise and inconsistencies resulted from low-frequency non-uniformity. After preprocessing, we construct an ROI that encloses the kidney for each subject. The largest kidney size was used to determine the size of the ROI and this size was kept fixed for all subjects. To construct an ROI that encompasses the entire kidney of a specific subject, we choose the slice that contains the largest cross section and a software user is asked to determine the approximate

centroid (center of the ROI) of this cross section. Then, for this specific subject, we crop the DW-MRI kidney volumes based on this rectangular ROI at all b-values. Cropping the DW-MRI scans has two advantages. First, complex registration and segmentation of the kidney are no longer required, and thus, the complexity of the system is reduced. Second, the computational cost of network training and testing is reduced because, the size of the input data is reduced. In our study, the size of the original kidney volumes ranges from 176×176×24 to 256×256×38 voxels. After ROI-cropping, the size of the kidney volumes became 150×150×24 voxels.

3. ADC Maps and Clinical Biomarkers

As illustrated in Figure 31, an ADC map which quantifies the diffusion at a certain b-values can be estimated from two DW-MRI images: one at the baseline signal intensity ($b = 0 \, \mathrm{s \, mm^{-2}}$) and the other one at this specific b-value, as follows:

$$\mathrm{ADC} = \frac{-(\ln(S_b) - \ln(S_0))}{b} \tag{7}$$

Where ADC, S_b, S_0 are the apparent diffusion coefficient, signal intensity at b-value = b, base line signal intensity, respectively. The ADC maps were estimated for all of the cropped DW-MRI kidney volumes at all b-values.

In the clinical practice, different biomarkers are used for kidney function evaluation. One of these markers is creatinine which is the metabolic waste result of creatine in muscle. The level of creatine in the blood and urine can be measured by SPCr, and CrCl tests, respectively. Therefore, we can use these tests as indicators for GFR and kidney function. Usually, urologists carry out the SPCr and CrCl tests for patients after kidney transplantation [132].

To increase the discriminatory power of our ADC data, we fused the measured ADC maps with the clinical biomarkers (the CrCl and SPCr values) which are obtained during
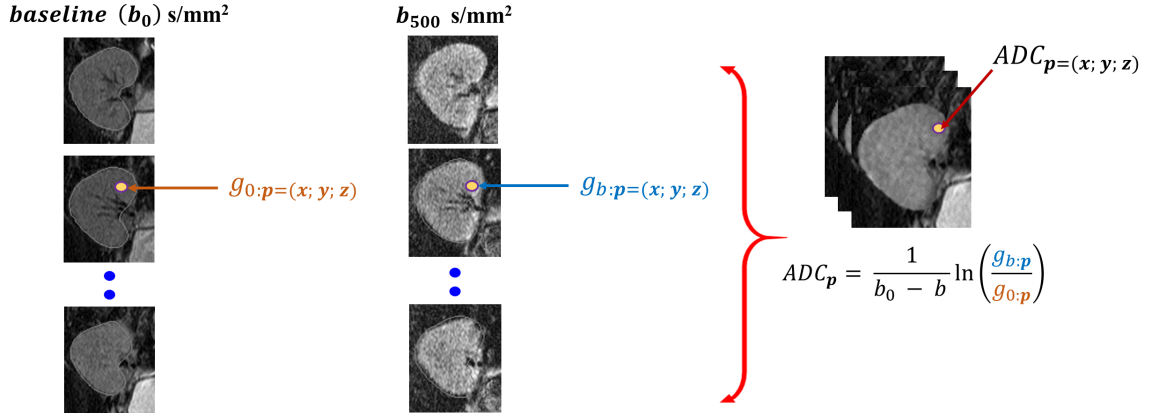
FIGURE 31: Voxel wise ADC estimation at a voxel (x, y, z) for b-vlaue = $500 \, \mathrm{s \, mm^{-2}}$: (left) the cropped ROI kidney volume at the base line b_0, (middle) the cropped ROI kidney volume at the base line b_{500}, and (right) The estimated ADC maps for the constructed ROI.

routine monitoring after transplantation. We trained our machine learning framework with the fused data which we call FBio. As shown in Figure 32, for each subject we added his clinical biomarkers to his ADC maps at all b-values. We refer to the fused data as FBio. To underline the advantages of the integration process of both markers, we adopted two main scenarios. First, we used the estimated ADC maps alone to evaluate the functionality of the allograft. Second, we used the FBio maps resulted from the fusion process of the estimated ADC maps and the clinical biomarkers to evaluate the functionality of the allograft. In the two scenarios, the maps were constructed at the 11 b-values. Then, the two types of maps were used to train and evaluate our CNN-based classifier usine a leave one subject out (LOSO) cross-validation strategy. We discuss the proposed CNN-based classifier in the next section.

4.   The Proposed Machine Learning Framework

Our machine learning framework is based on CNNs and SVM. The input which consists of 11 3D volumes is first fed to 11 CNNs that share the same architecture. Then, the decisions of the 11 CNNs is fused using a SVM classifier. Therefore, the input 3D maps

3D ADC Maps at different b-values (Imaging Markers)

b= 50 s/mm² · b= 500 s/mm² · b= 1000 s/mm²

Clinical Biomarkers: CrCl → Σ ← SPCr

Normal Kidney Function — Poor Kidney Function

Fused Biomarkers

a) Example for a Non-rejection Kidney

b) Example for an Acute Rejection Kidney
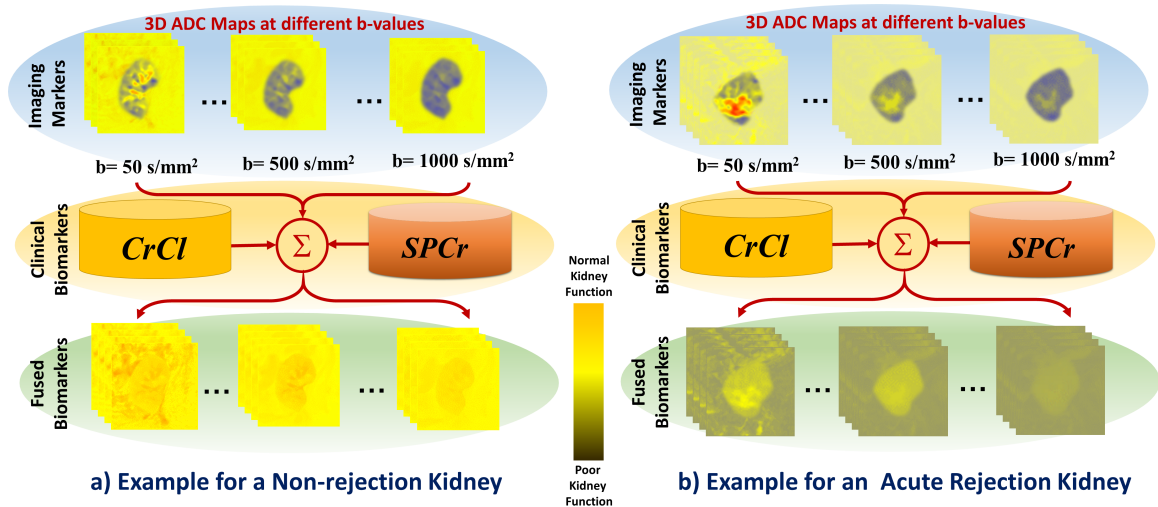
FIGURE 32: Illustration of the efficacy of the integration (fusion) of image markers with the clinical markers. In this study, the measured ADC maps are fused with CrCl and SPCr which both are one-dimensional values obtained from routine monitoring after transplantation. We first normalize the clinical biomarkers for each subject with respect to the maximum values of each marker. Then, we add the normalized values to the ADC maps at all the b-values by voxel-wise operations. In the figure, pixel values were color coded. As demonstrated in the figure, it is challenging to discriminate between the non-rejection kidney (a), and acute rejection kidney (b) using the ADC maps alone due to the overlap between the ADC values between these two cases. After the fusion process of both markers, we can visually differentiate between the two cases. Poor kidney function appears as a dark green color (i.e. low CrCl + high SPCr + low ADCs), while normal kidney function appears as an orange-yellowish color (i.e. high CrCl + low SPCr + high ADCs).

are first processed by convolution layer. Figure 33 illustrates how the 3D input is process by the convolution layer. First, we will discuss how we reached the best CNN architecture. Second, we will discuss how we trained our machine learning framework. Finally, we will discuss how we fused CNN decisions which leads to final diagnosis.

To find the optimal CNN architecture for our problem, we used a total of 20 subjects (10 non-rejection and 10 acute rejection) as an independent set for training the CNN. This

Convolution Layer

Convolution in x and y with Kernel #1

150 · 150 · x · y · 24 · z

3D Input volume

Kernel #1 with size $(k, k, 24)$ and learned weights w^1

w_1^1 k · w_2^1 k · w_{24}^1 k

2D Feature Map #1

Stacking of feature maps

$(150-k+1)/s$ · $(150-k+1)/s$

Output from convolution layer: 3D feature map, where s is the stride of the kernel in x or y directions

Convolution in x and y with Kernel #n

150 · 150 · x · y · 24 · z

3D Input volume

Kernel #n with size $(k, k, 24)$ and learned weights w^n

w_1^n k · w_2^n k · w_{24}^n k

2D Feature Map #n

FIGURE 33: Schematic illustration for how a single volume is process by a convolution layer in the CNN. The figure shows how an input of size $150 \times 150 \times 24$ voxels is convolved with a kernel of size $K \times K \times 24$. Ususlly, we use multiple kernels and the shown operation is performed again for each kernel to result in several feature maps. The result of the convolution layer is a volume of feature maps.

set was not used during the validation of the framework. With the classification accuracy is the metric to be optimized, we adopted a grid search approach to find the optimal values for the following: number of the convolution layers (search range 2 to 5), number of convolutional kernels (search range 3 to 15), kernel stride (search range 1 to 3), and convolutional or pooling kernel size (search range 3 to 7). The searching operation resulted in a global average accuracy of 96.4% in a 10 training iterations in a row. Our network was organized as a processing blocks. The order of the layers in each block was as follows: convolution layer, batch-normalization layer to speed-up the CNN training [133], ReLU activation layer, and possible average pooling layer based on the result of the grid search. Our proposed CNN has 3 processing blocks followed by concatenation layer and a fully connected layer with two neuron for two classes, as shown in Figure 34(a). The input to

the first block of our CNN is the 3D fused maps with size of 150×150×24 voxels. This input is procssed by the CNN until it is converted to a vector of seven neurons after the third block in the concatenation layer. The output of the concatenation layer is fed to a fully connected layer with two neurons and finally to a soft-max layer which performs two class classification. Please refer to Figure 34(a) and Table 1 for more information about the CNN configuration.

TABLE 1: The proposed CNN configuration

Layer	kernel	Stride	Depth	Spatial Size	Parameters
Input	—	—	24	$150 \times 150 \times 24$	0
1. Conv.	5×5	2×2	3	$73 \times 73 \times 3$	$5 \times 5 \times 24 \times 3$
2. Conv.	5×5	2×2	3	$34 \times 34 \times 3$	$5 \times 5 \times 3 \times 3$
3. Avg	5×5	2×2	3	$15 \times 15 \times 3$	0
4. Conv.	5×5	2×2	7	$6 \times 6 \times 7$	$5 \times 5 \times 3 \times 7$
5. Avg	6×6	1×1	7	$1 \times 1 \times 7$	0
6. Concat.	—	—	1	7×1	0
7. Full	—	—	1	2×1	7×2

Total number of parameters = 2564

We optimized our CNNs using stochastic gradient descent (SGD) optimizer. We started by a learning rate of 0.1 which was reduced to 0.0001 over 70 epochs. The complex architecture of CNNs makes them prone to overfitting problem. We said that a network is overfitting the training data when the network learns features special to the training set and on the other hand, fails to generalize well to unseen input. To override overfitting, the dropout technique was proposed [134]. Applying the dropout technique allows modeling various representations for the relationships found in the input data. Typically, we achieve this by deactivating a proportion of neurons in a random manner in each iteration during training. By deactivating these neurons, we avoid complex co-adaptations during training.

a) The architecture of proposed CNN

Kernel Size: 5x5x24, 5x5x3, 5x5x3
No. of Kernels: 3, 3, 7
Stride: 2, 2, 2

Cx Convolution Layer | Nx Normalization Layer | Rx ReLU Layer | Px Average Pool Layer | Dx Dropout Layer | Fx Fully Connected Layer

b) The fusion of the CNN decisions for each individual b-value
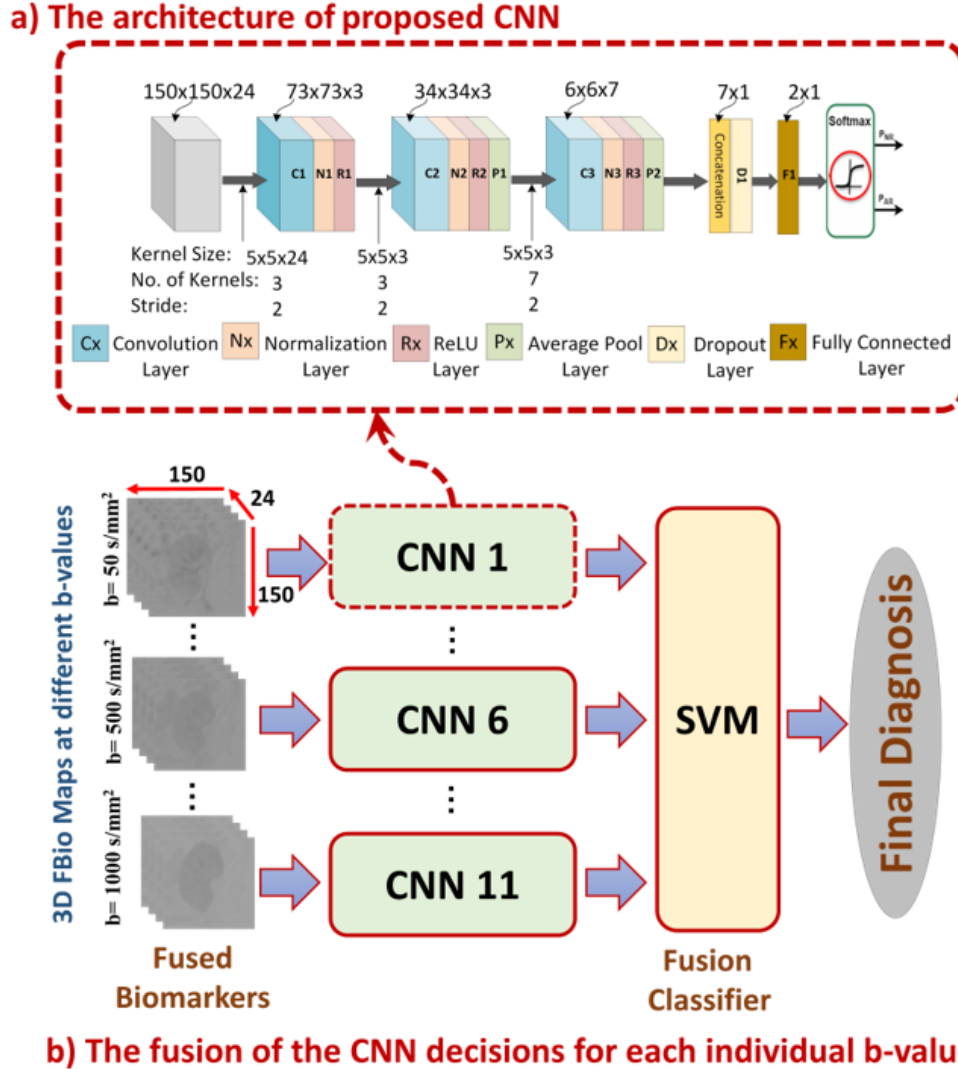
FIGURE 34: The architecture of the proposed machine learning framework, where (a) shows the configuration of the designed CNN and its processing blocks, and (b) illustrates the process of the fusion of CNN decisions using SVM classifier. At each b-value, we trained and validated the CNN apart form the other b-values. At each b-value, we have 56 samples and each sample represents a 3D volume for a specific subject. The volume size is $150 \times 150 \times 24$ voxels. We feed the 3D volume to the CNN as a one sample of the set. By combining the output probabilities of all CNNs (all b-values), we get the final classification result. The fusion is performed using SVM where each sample inputted to the SVM has 22 features (two probabilities for 2 classes resulted from the CNN at a certain b-value times 11 b-values)

61

Although, some neurons are randomly deactivated in each iterations, these neurons may be active in the next iterations. In our designed CNN, a dropout of factor 0.5 was used. PyTorch deep learning framework [135] was used to implement our network. All experiments were done using NVIDIA Quadro P4000 GPU.

FBio maps (the fused image and clinical biomarkers) were used as input to the CNN at 11 different b-values. Then, we collect the predicted probabilities given for each class at the 11 b-values. To enhance the classification accuracy, we fused all of the CNN decisions (11 decisions for 11 b-values) by using SVM classifier. In other words, each subject has 22 probabilities (2 classes times 11 b-values). We feed each subject as a single sample with 22 features to the SVM classifier to obtain the final diagnosis. To sum up, we use data fusion and decision fusion for better diagnosis accuracy.

E. Experimental Results

We used LOSO and 10 fold cross validation to evaluate our proposed pipeline. For the NR group, the average of the CrCl values was 72.62 ± 17.58 ml/min, and the average of the SPCr values was 1.16 ± 0.23 mg/dl. For the ARR group, the average of the CrCl values was 39.63 ± 11.98 ml/min, and the average of the SPCr values was 2.23 ± 0.71 mg/dl. Figure 32 shows the efficacy of the integration between image and clinical markers. This fusion helps in differentiating a NR subject from ARR subject. As shown in Figure 32, it is challenging to discriminate between the normal and abnormal cases using the ADC alone, demonstrated in Table 2. We can justify this difficulty by the large degree of overlap between the ADC values of both subjects, which can be better manifested by the color coded maps. These two subject can be visually discriminated after fusing the clinical biomarkers (FBio maps). Therefore, We use FBio maps as our input discriminatory features for kidney transplant function assessment.

We adopted two training and testing scenarios: one using ADC maps only and one with FBio maps which result from the fusion process. For all available subjects, we

TABLE 2: A comparison between NR (non-rejection) and ARR (acute renal rejection) cases based on the average value of the ADC maps alone at the 11 b-values.

Average ADC Values at Individual b-values: mean(std) $\approx$											
b-value	b_{50}	b_{100}	b_{200}	b_{300}	b_{400}	b_{500}	b_{600}	b_{700}	b_{800}	b_{900}	b_{1000}
NR	3.0(1.17)	2.63(0.72)	2.59(0.50)	2.46(0.41)	2.43(0.37)	2.21(0.33)	2.10(0.08)	2.09(0.26)	2.05(0.25)	1.95(0.23)	1.90(0.21)
ARR	2.85(1.79)	2.97(0.98)	2.81(0.53)	2.33(0.42)	2.28(0.29)	2.12(0.25)	2.09(0.24)	1.93(0.22)	1.90(0.22)	1.87(0.19)	1.83(0.17)

constructed these maps at the 11 individual b-values. Table 3 reports the individual accuracies for each b-value and the accuracy resulted from the fusion process, for both the mentioned scenarios. To make sure that our approach is robust, We performed a 10-fold cross-validation strategy using the FBio maps and our CNN-based classifier. The 10-fold cross-validation resulted in accuracy = 91%, sensitivity = 90%, and specificity = 92%. To find the contribution of clinical biomarkers vs. imaging markers to the overall diagnostic accuracy, we conducted a comparison between the performance of the CNN using FBio, CNN using ADC maps alone, and the SVM classifier using the clinical biomarkers alone (ClinBio). Table 4 summarizes the resultant performance in terms of accuracy, sensitivity, and specificity. The overall diagnostic performance of our proposed machine learning framework is compared with other deep learning approach that aims at evaluating transplanted kidney function. Specifically, our approach was compared with an approach that uses stacked auto-encoders (SAEs) with input data that is composed of the cumulative distribution function of the ADC of the segmented kidney [1]. Table 4 summarizes the results from our approach and the other approach in the literature. As demonstrated in Table 4, our machine learning framework outperforms the SAEs method. The advantages of our method over the SAEs approach is that our framework dos not require kidney registration or segmentation which are difficult processes because of inter-patient anatomical differences the diffused boundaries. Therefore, our proposed framework is efficient in terms of time and computational complexity.

Furthermore, receiver operating characteristic (ROC) was used to assess the robustness of our machine learning framework. The ROC curves of individual b-values and their

TABLE 3: The diagnostic classification accuracy of individual b-values and fused b-values (F_{11}) resulted from three scenarios. The first scenario (S_1) uses the ADC maps alone and our machine learning framework. The second scenario uses the fused data FBio and our machine learning framework. The third scenario uses the cumulative distribution function of the ADCs of the segmented kidney along with stacked auto-encoders (SAEs) [1]

Different b-values Classification Accuracy % $\approx$												
Method	b_{50}	b_{100}	b_{200}	b_{300}	b_{400}	b_{500}	b_{600}	b_{700}	b_{800}	b_{900}	b_{1000}	**F_{11}**
$SAEs$(CDFs) [1]	68	50	**77**	68	82	68	73	64	75	86	68	86
S_1(ADC only)	59	50	64	59	59	59	68	64	77	73	68	82
S_2(FBio)	68	**68**	73	**80**	**83**	**73**	**85**	**86**	75	86	**86**	**93**

TABLE 4: A comparison between the diagnostic performance of four scenarios. The first scenario (S_1) uses the ADC maps alone and our machine learning framework. The second scenario uses the fused data FBio and our machine learning framework. The third scenario uses the cumulative distribution function of the ADCs of the segmented kidney along with stacked auto-encoders (SAEs) [1]. The fourth scenario uses the clinical biomarkers (ClinBio) only with the SVM classifier. The performance is reported in terms of accuracy, sensitivity, specificity, and area under curve (AUC).

Performance of the Final Diagnosis				
Method	**Accuracy% $\approx$**	**Sensitivity% $\approx$**	**Specificity% $\approx$**	**AUC $\approx$**
$SAEs$(CDFs) [1]	86	70	100	0.88
SVM(ClinBio)	77	80	73	0.80
S_1(ADC only)	82	80	83	0.83
S_2(FBio)	93	93	92	0.93
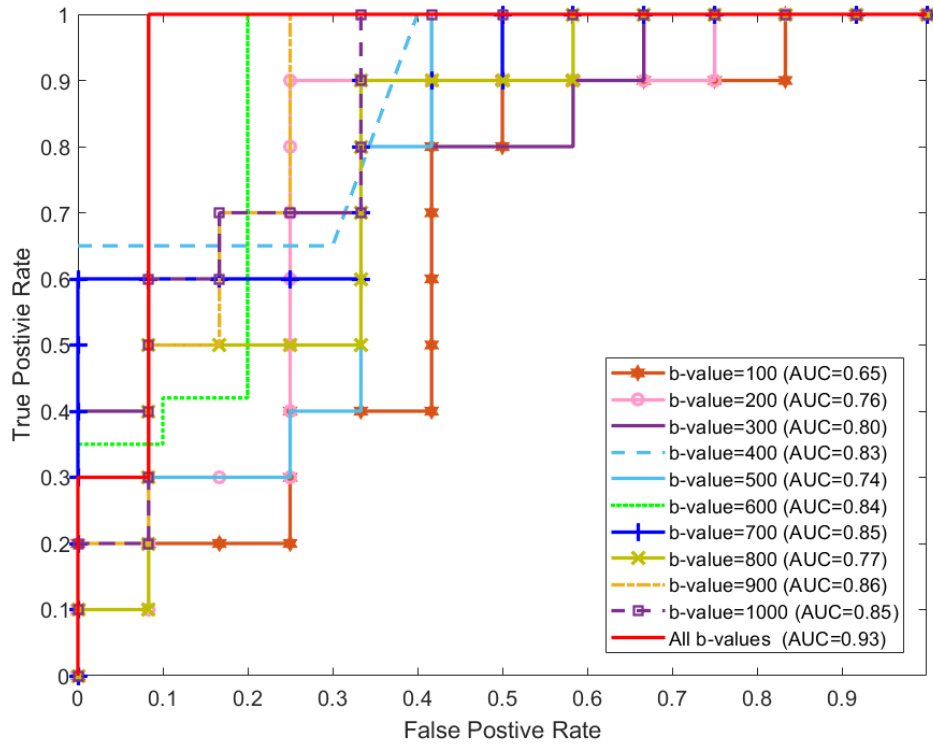
FIGURE 35: Receiver operating characteristics (ROC) curves for the fused image and clinical markers (FBio) for individual b-values and their fusion.

fusion for the proposed framework is presented in Figure 35. As shown in Figure 35, decision fusion of individual b-values increased the overall diagnostic accuracy by reporting an AUC = 0.93. Furthermore, we analyzed the ROC curves of all scenarios mentioned in this chapter and the previous method that was based on SAEs, as shown in Figure 36. Particularly, the SAEs approach achieved an AUC that is less than our approach by 0.05. The above analysis and comparisons have established the potential of our framework for the early detection of ARR

F.   Discussion

Early detection of ARR can prolong the renal graft normal function, because the early detection of such a problem allow the urologists to intervene with an appropriate
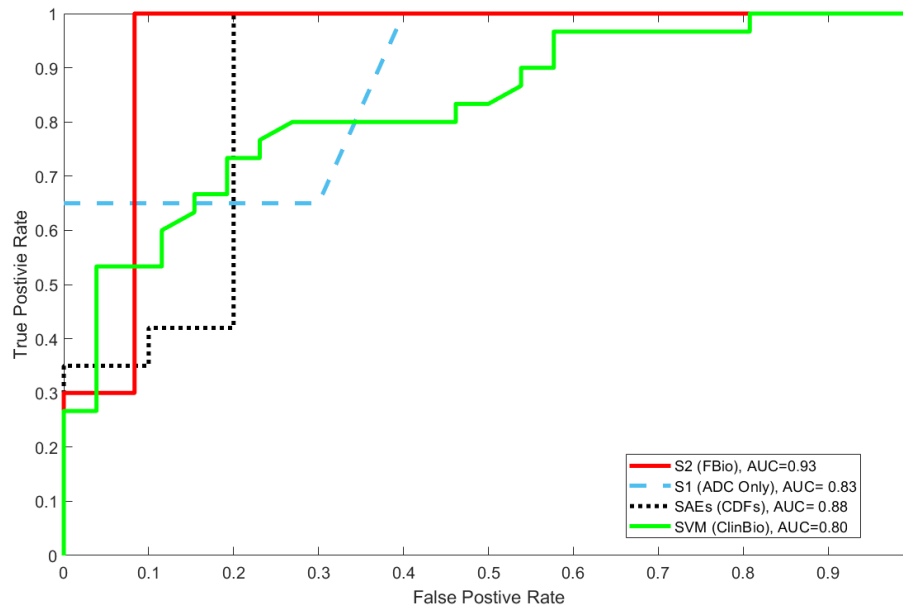
FIGURE 36: Receiver operating characteristics (ROC) curves for our CNN-based framework (for both scenarios S_1 and S_2), the support vector machine (SVM) classifier with clinical biomarkers (ClinBio), and the stacked auto encoders (SAEs) system [1].

treatment. Typically, there various types of ARR and the selection of the suitable treatment depends on the type of rejection. For instance, urologists treat acute cellular rejection with a high dose of corticosteroids, taken intravenously as a first line treatment [136, 137]. The administration of methylprednisolone for 3 consecutive days is the most popular regimen [136]. A second biopsy is needed in the following cases: (1) there is a persistent renal deficiency with the steroid and/or antithymocyte globulin, (2) there is a new defect in kidney function after curing ARR. The second biopsy will help urologists to find additional causes of kidney dysfunction. ARR episodes that do not respond to steroids and aggressive vascular cellular rejection are treated with T-cell depleting antibodies [138]. On the other hand, plasmapheresis, intravenous immunoglobulin, immunoadsorption, or monoclonal antibodies are suggested for antibody mediated rejection [139].

It is worth noting that most of the clinical studies do not estimate the ADC at 11 b-

values and only a few b-values are selected [116–126]. Usually, these studies choose a low b-value and high b-value along with the baseline b_0. Typically, low b-values are responsible for blood perfusion, while high b-values are responsible for water diffusion [140–143]. To efficiently discover the differences in blood perfusion and water diffusion between NR and ARR groups, we used 11 b-values. And to get more accurate diagnosis we fused the decisions from the 11 individual b-values. It is worth mentioning that this fusion helped with tackling any problems that may occur during the acquisition process of one or two b-values because of existing artifacts or chemical shifts.

To summarize, a machine learning framework that is based on CNN and SVM was proposed. This framework was developed as a non-invasive evaluation of kidney transplant status using the integration between DW-MRI derived markers and clinical biomarkers. Our framework is a non-invasive tool to differentiate between NR (stable allograft) and allograft with ARR. This integration process resulted in a well-separated maps that were fed to CNN as kidney status discriminatory features. These maps affected the individual and global diagnostic accuracy of the proposed framework. Our framework is considered a computer aided diagnostic system for kidney transplant function evaluation. Our results established the potential of this CAD system as a non-invasive and reliable kidney transplant diagnostic tool. The CAD system is independent of imaging protocol and/or the scanner type that has been used to acquire the DW-MRI. Furthermore, it does not depend on the geographical area where the DW-MRI data were acquired. Complex kidney registration and segmentation are not required in our CAD system. Therefore, it is more efficient in terms of reduced processing time without compromising the final diagnosis accuracy.

Currently, we are using 11 different b-values that include low and high b-values. This helped us to collect all possible information that leads developing a non-invasive and accurate diagnostic tool. The tool is intended to be an alternative to biopsy for kidney transplant function assessment. On the other hand, we plan to perform a statistical analysis to identify what is the most informative b-values. When we achieve that goal, we can remove

some of the b-values from our analysis and this will leads to a reduction in DW-MRI ac-quistion time. Future work includes the use of a larger sample size collected from different transplantation centers and/or various imaging systems and protocols. Furthermore, we are planning to explore additional biomarkers.

CHAPTER IV
LEFT VENTRICLE SEGMENTATION AND QUANTIFICATION USING DEEP LEARNING

Cardiac MRI is a widely used noninvasive tool that can provide us with an evaluation of cardiac anatomy and function. It can also be used for heart diagnosis. Heart diagnosis through the estimation of physiological heart parameters requires careful segmentation of the left ventricle (LV) from the images of cardiac MRI. Therefore, we aim at building a new deep learning method for the automated delineation and quantification of the LV from cine cardiac MRI. Our goal is to reach lower errors for the calculated heart parameters than the previous works by introducing a new deep learning cardiac segmentation method. Our pipeline starts with an accurate LV localization by finding LV cavity center-point using a fully convolutional neural network (FCN) model called FCN1. Then, from all heart sections, we extract a region of interest (ROI) that encompasses the LV. A segmentation for the LV cavity and myocardium is performed from the extracted ROIs using FCN called FCN2. The FCN2 model is associated with multiple bottleneck layers and uses less memory footprint than traditional models such as U-net. Furthermore, we introduced a novel loss function called radial loss which works on minimizing the distance between the ground truth LV contours and the predicted contours. After myocardial segmentation, we estimate the functional and mass parameters of the LV. We used the Automated Cardiac Diagnosis Challenge (ACDC-2017) dataset to validate our pipeline, which provided better segmentation, accurate calculation of heart parameters, and produced fewer errors compared to other approaches applied on the same dataset. Additionally, our segmentation approach showed that it can generalize well across different datasets by validating its performance on a locally collected cardiac dataset. To sum up, we propose a novel deep

learning framework that we can translate it into a clinical tool for cardiac diagnosis.

A.   Heart: Anatomy, Function, and Diseases

The heart is an important muscular organ that has two main functions: (1) the continuous collection of the blood from different parts of the body and the pumping of the blood to the pulmonary system and (2) the continuous collection of the blood from the pulmonary system and the pumping of the blood to all body's tissues. The heart lies at the center of the circulatory system. The circulatory system is composed of a complicated network of blood vessels, such as arteries, arterioles, capillaries, and veins. These vessels are responsible for carrying the blood to and from the tissues of the body [14].

An electrical system that is composed of electrical signals is responsible for giving the heart its contractile function. These electrical signals allow the heart to contract and pump the blood into the blood vessels. The heart contains valves that force the blood to move in the correct direction.

The heart is crucial for the health of the human and his tissues. Without a working heart, the body loses the circulation of the blood to the tissues. In order to work and grow normally, the organs need oxygen and nutrients carried and delivered by the blood. Carbon dioxide and waste products are also carried by the blood to the lungs where the body gets ride of them into the air. A certain amount of blood at a certain rate is required to be pumped by the heart to different body parts in order to maintain a healthy status. If there is a disease or an injury in the heart, the body tissues will not receive its sufficient amount of blood for normal working [14].

1.   Location, Size and Shape of the Heart

The heart is located in the protective thorax and occupies the area between lungs in a compartment called the mediastinum [144]. The mediastinum is the area inside the mem-

brane enclosing the heart, the pericardium. The heart is located posterior to the sternum and costal cartilages and above the surface of the diaphragm at the fifth or the sixth rib. It has an oblique position in the chest and its two-thirds lies in the left of the midline. Figure 37 shows the location of the heart in the thorax. The heart is roughly the size of the fist of the man and weigh about 250-350 grams [145]. The heart has the shape of an inverted cone. The apex of the heart is the narrow end and is located above the diaphragm. While the base of the heart is the broad end which is located at the level of the second rib.
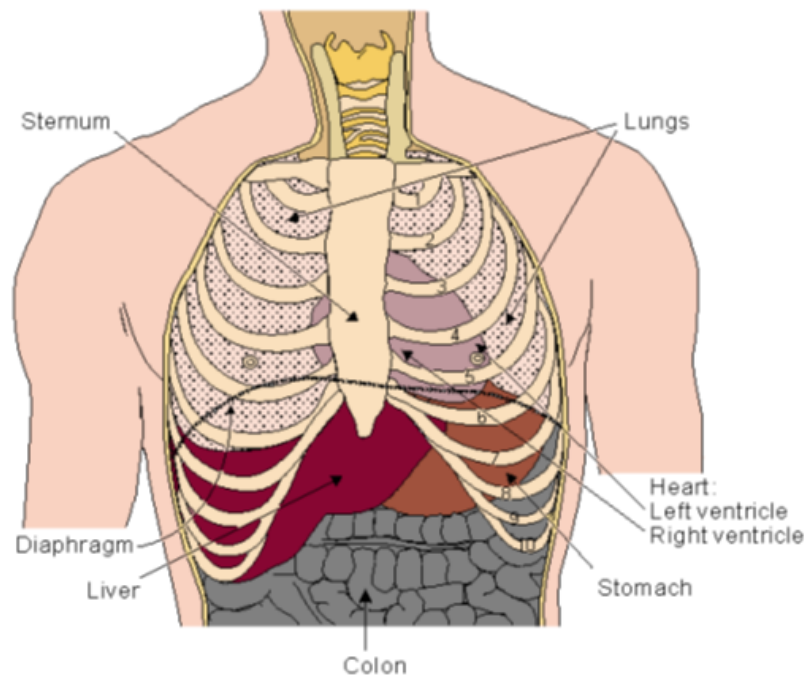
FIGURE 37: Location of the heart in the chest. Its boundaries are the sternum, lungs, diaphragm, and esophagus [14].

The pericardium is a double-layered fibrous sac that encloses the heart and contains the roots of the great vessels (Superior vena cava, Inferior vena cava, Pulmonary arteries, Pulmonary veins, Aorta). It has two layers: the outer layer called fibrous pericardium and the inner layer called serous pericardium. The serous pericardium in turn consists of two layers: the parietal and visceral pericardium. There is a cavity in the pericardium called pericardial cavity which contains pericardial fluid. The fluid is a serous fluid secreted by

the serous pericardium into the cavity. Figure 38 describes the structure of the pericardium. The pericardium has multiple functions. It places the heart into its position in the thorax. The fluid forms as a lubricant to the outer wall of the heart. Therefore, the heart beats without friction. Also, the pericardial sac prevents infections and over-expanding to the heart.

The heart wall consists of three layers: epicardium (visceral pericardium), myocardium, and endocardium, as shown in Figure 38. The epicardium is the outer layer in the heart wall. The epicardium is known also as the visceral pericardium (the inner surface of the pericardium). The epicardium has two tissue layers: the outer surface made of simple squamous tissue which secrets a fluid for lubrication into the pericardial cavity, and the inner surface made of areolar tissue.

The myocardium is the thick layer that lies at the center of the heart wall. It is composed of abundant cardiac muscle fibers that envelope the heart wall. When the myocardium contracts, the aorta and the pulmonary arteries receive the pumped blood from the heart.

The endocardium covers the inner surface of the heart wall and it also lines the heart valves and tendons. It meets the endothelium that covers the blood vessels that linked to the heart. Like the epicardium, simple squamous and areolar tissues constitute the composition of the endocardium

2. Anatomy of The Heart and Circulation System

The heart contains four chambers that consist of cardiac muscle or myocardium. At the upper level there are two chambers: the right atrium (RA) and the left atrium (LA), as shown in Figure 39. The function of the atria is to collect the blood. At the lower level, there are another two chambers: the right ventricle (RV) and the left ventricle (LV). The function of the ventricle is to pump the blood [144]. The function of the right atrium and right ventricle is the collection of the blood from body parts and pumping it to the
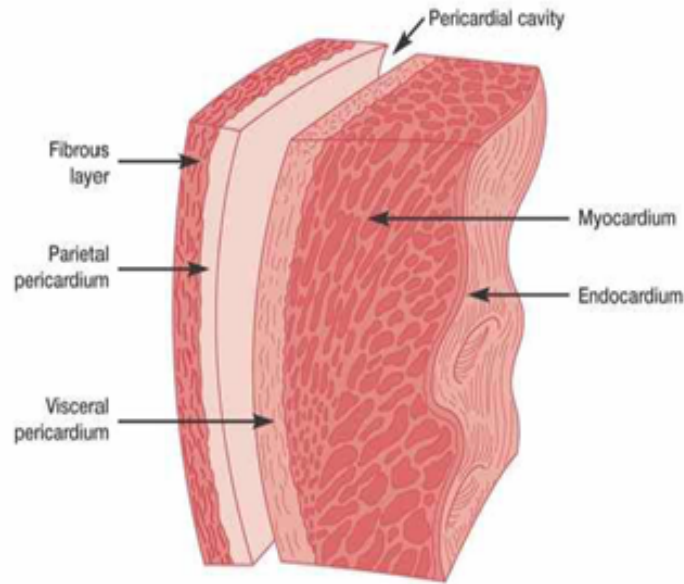
FIGURE 38: Structure of the pericardium and the heart wall. [14].

pulmonary system. The function of the left atrium and left ventricle is the collection of the blood from the pulmonary system and pumping it into the tissues of the body. A set of four valves are used to maintain a one-direction of blood movement throughout the heart. The tricuspid and bicuspid valves (atrioventricular valves) assure that the blood moves only from the atria to the ventricles. The pulmonary and semilunar valves assure that the blood moves from the ventricle to the great arteries.

Despite the fact that the heart contains a high amount of blood, the heart gives a small amount of blood to its tissues. The heart tissues receive the blood by an independent vessel supply. From the aorta arise the left and the right coronary arteries that supply the heart with nourishment and oxygen. Cardiac veins receive the deoxygenated blood from the heart tissues and return it to the right atrium.

The superior vena cava and inferior vena cava transport deoxygenated blood systemic circulation to the right atrium. Deoxygenated blood from the head and upper extremities is transported via superior vena cava while deoxygenated blood from lower extremities, abdomen, and thorax are transported via inferior vena cava. The blood moves

from right atrium to the right ventricle through tricuspid valve. The blood leaves the right ventricle through pulmonary artery which forks into the right and left pulmonary arteries which transport deoxygenated blood to the lungs. Then, gas exchange happens in the lung. The oxygenated blood from the right pulmonary vein and left pulmonary vein moves to the left atrium. The blood travels from the left atrium to the left ventricle through the bicuspid valve. The left ventricle pumps the oxygenated blood to the circulatory system through the aorta.
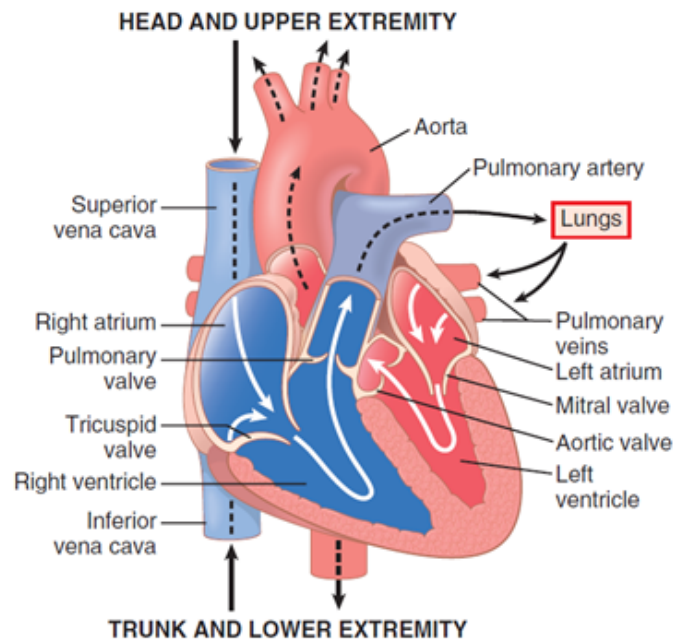
FIGURE 39: Anatomy of the heart and the movement of the blood through the heart chambers, valves, and arteries. [14].

3. Cardiac Cycle

The events of the heart that happens from the start of one heartbeat to the start of the next are referred as the cardiac cycle. The action potential of the sinus node initiates spontaneously the cardiac cycle. The action potential moves through the right and left atrium and then in the A-V bundle to the ventricles. The cardiac cycle is composed of two

periods: diastole and systole. The diastole is a relaxation period in which the blood fills the heart. The systole is a contraction period in which the heart pumps the blood. The heart rate (HR) is the reciprocals of the cardiac cycle (sum of the diastole and systole periods). The normal heart rate is 72 beats/minute, where the cardiac cycle is 0.833 second per beat [1]. The function of the heart depends greatly on the performance of the left ventricle. The performance of the left ventricle is assessed mainly by some volumetric parameters. Let us define some volumetric measures related to the left ventricle.

1. End Diastolic Volume (EDV): The volume of the LV at the end of the diastole. Normal values are 142 mL ($\pm$ 21 mL) [146].

2. End Systolic volume (ESV): The volume of the LV at the end of the systole. Normal values are 47 mL ($\pm$ 10 mL) [146].

3. Stroke Volume (SV): The volume of blood pumped by the LV per minute. $SV = EDV - ESV$. Normal values are 95 mL ($\pm$ 14 mL) [146].

4. Cardiac Output (CO): The volume of the blood pumped by the left ventricle per unit time. It is the product of the HR and SV, i.e. $CO = HR * SV$. Normal values are 4.0–8.0 L/minute [146].

5. Ejection Fraction (EF): The volumetric fraction pumped from the LV with each heartbeat. $EF = \frac{EDV - ESV}{EDV} * 100$. Normal values are 67% ($\pm$ 4.6%) [146].

4. Cardiovascular Diseases

Cardiovascular diseases (CVDs) are a group of diseases associated with heart and blood vessels. Each disease has its own mechanism. Here, we will mention some of the most common CVD.

- Coronary Artery Disease: It is known as ischemic heart disease [147]. This is disease is caused due to the reduction of the blood supply to the cardiac muscle when there is

a build-up of plaque (atherosclerosis) in the heart arteries. The types of this disease are:

- Angina: Pressure or chest pain due to insufficient blood supply to the heart.

- Myocardial Infarction: It is known as heart attack. It is a damage to the heart muscle which happens when the blood supply decreases or stops to an area of the heart. The symptoms of this type include chest pain which might traverse to the arm, neck, or back.

- Cardiac Arrest: Sudden loss of blood supply caused by the inability of the heart to pump normally. Symptoms involve the loss of consciousness and breathing. If the treatment not delivered in minutes, it leads to death.

• Heart Failure: It happens when the heart is unable to pump sufficient blood to meet the body needs. It can happen after myocardial infarction that reduces the heart performance to pump blood. Symptoms involve short breath, intense tiredness, and leg swelling. Heart failure causes a reduction in the EF. The European Society of Cardiology posted guidelines for the diagnosis and treatment of acute and chronic heart failure [148]. These guidelines defined the categories of the heart failure based on the value of the EF as follows.

- Normal ejection fraction: $EF \geq 50\%$

- Moderately reduced ejection fraction: $40\% \geq EF \geq 49\%$

- Reduced ejection fraction: $EF < 40\%$

In the united states a threshold of 30% for EF is used by the authorities to present disability benefits to the patients.

B. Left Ventricle Segmentation and Quantification

76

CVDs are dangerous heath issues as they are responsible for the highest rate of mortality worldwide [49]. About 836,546 patients in the United States die each year because of CVDs [50]. A non-invasive quantitative evaluation for the heart functionality can be available for the cardiologist from Cardiac magnetic resonance (CMR) which is an important imaging technique. The cardiologist can get functional heart indexes from performing segmentation for the LV in CMR short-axis view cine images [51]. These indexes are ESV, EDV, EF, wall mass, and regional indexes (e.g., wall thickening) which are essential for heart diagnosis and treatment. However, to estimate these parameters, accurate delineation of myocardial walls is required. Manual segmentation of LV in CMR short-axis view images requires significant effort and time. Furthermore, manual segmentation is prone to inter- and intra-operator variability [52]. Therefore, an alternative technique to manual segmentation is needed to perform automated segmentation for the LV cavity and myocardium. Furthermore, this technique should be accurate to obtain left ventricular functional metrics with high accuracy.

Currently, deep learning has achieved great success over conventional image processing methods in many medical imaging analysis domains [29, 103]. Deep learning is a subdomain in ML that has the power of automated representation learning from the raw data itself without the need for manually crafting the appropriate features from the data. Convolutional neural network (CNN) is a deep learning algorithm that is widely used for processing image data and has demonstrated great success in computer vision tasks [128, 149]. Researchers initially used CNNs to classify images by assigning the whole image to a specific class. On the other hand, CNNs are now redesigned to perform image segmentation. The new design is the fully convolutional neural network (FCN) which replaces the fully connected layers in the CNN by convolution layers. In a FCN algorithm, we perform dense classification in the image domain where each pixel is assigned to a certain class with the highest predicted probability.

The need for accurate and automated estimation of left ventricular parameters, and

the success of the CNNs motivated us to build a framework for the automated functional assessment of the heart. This framework should segment and quantify the LV with a comparable accuracy to the human operator. In this chapter, we propose a novel fully automated framework that segment and quantify the LV from CMR cine images. We performed segmentation to the LV cavity and myocardium to estimate the heart physiological parameters. Our main achievement is the accurate quantification of the LV indexes and reaching a lower error compared to other previous techniques applied on the same dataset. Our framework that is based on FCN has the following contributions:

- We propose an efficient method that is based of FCN to extract a region-of-interest (ROI) that contains the LV from CMR cine images. We perform this extraction process in the beginning to alleviate the class-imbalance problem and reduce the memory and computational requirements.

- After ROI extraction, we use a novel FCN model for cardiac segmentation. The architecture of this novel model is inspired from the U-net; i. e., the input is passed to a contracting path followed by an expanding path. The addition to the conventional model is the incorporation of multiple bottleneck layers that describe the input by multiple representations. The bottleneck layers are up-sampled and combined to estimate the final segmentation. Our proposed model requires less fewer number of parameters than the state-of-the-art models such as U-net, yet it showed better performance.

- We propose a novel loss function that works on minimizing the difference between the ground truth LV contours and the predicted contours. We refer to the new loss by radial loss and we incorporate it with the cross-entropy loss.

- We evaluated the generalization strength of our proposed segmentation method by estimating the segmentation performance of our method when we use the ACDC 2017 data as a training set and a local dataset as a testing set. Our segmentation

approach demonstrated good segmentation accuracy that is comparable to another model that uses only our local dataset.

C. Related Work on Left Ventricle Segmentation and Quantification

Over the past years, considerable attention has been devoted to the problem of LV segmentation from CMR cine images. Firstly, semi-automatic segmentation methods have been proposed by researchers, see the review by Petitjean et al. [150]. These approaches such as presented in [151–154] used graph cut, active contours, dynamic programming, or atlas-based methods. On the other hand, significant user intervention is required in semi-automatic methods. Therefore, they are not suitable when fast segmentation feature is required. To remedy that limitation, researches proposed fully-automatic approaches for heart segmentation. Queiros et al. [155] and Liu et al. [156] proposed automatic approaches that are based on the level set. Despite the high accuracy they obtained from their approaches, level-set is associated with initialization and it is designed to delineate one anatomical structure only. Wang et al. [157] and Ringenberg et al. [158] proposed methods that are based on traditional image processing such as edge detection, thresholding, and morphology procession. However, when the prior assumptions of these methods are not satisfied, these methods do not work well. Furthermore, shape priors models have been added to the cardiac segmentation methods, such as the work of [159–161]. However, sub-optimal segmentation is achieved when the prior information is imperfect. The shape prior may be restricted by certain assumptions that do not work well with the underlying testing image. On the other hand, we can obtain automatic representation learning from deep learning given sufficient training dataset. Therefore, with the power of deep learning, we do not need shape prior information because automated segmentation for the test image can be directly obtained from the FCN.

Deep learning has been successfully applied on the problem of cardiac segmenta-

tion. Tran et al. [162] proposed the first application of FCN in cardiac image segmentation. They extracted an ROI which was centered at the LV cavity before segmentation. They assumed that the cavity center is at the image center which may lead to inaccurate results. A recurrent FCN was proposed by Poudel et al. [163]. Their network was a modified version of U-net. the spatial dependencies between slices were used during the segmentation of the left ventricular endocarduim. Linear regression using CNN was proposed by Tan et al. [164] to segment the LV. Their method was composed of two main stages: finding the left ventricular center followed by the calculation of the endocardial contour (EnC) radius and epicardial contour (EpC) radius in polar space. A regularized CNN was proposed by Oktay et al. [165]. The regularization is performed by the addition of anatomical prior. Their model is suitable for cardiac image analysis problems such as enhancement and segmentation. An iterative heart segmentation approach was proposed by Zheng et al. [166]. Their approach starts from the base of the LV to the apex. In each iteration, a new model of the U-net performs segmentation to the heart and the output is utilized for the prediction of the next slice and thus they maintain 3D consistency. An FCN that was inspired from VGG-16 [167] was proposed by Bai et al. [168] for the segmentation of the LV and right ventricle (RV) from CMR short-axis view, and right atrium and left atrium from CMR long-axis view. A novel computationally efficient DenseNet was proposed by Khened et al. [2]. Their model is based on FCN for heart segmentation. To build a system for cardiac disease classification, they calculated the clinical indexes from the segmentation maps. The U-net was employed in the work of Tao et al. [169]. Their work aimed at the calculation of LV parameters from CMR images. They evaluated the performance of their approach by a multi-vendor and multi-center cardiac data.

D.   Methods

Our proposed framework for the automated segmentation and quantification of the LV is shown in Figure 40. The quantification is performed by the calculation of left ven-

tricular functional parameters and mass. Our framework consists of four main steps: (1) ROI extraction using a FCN called FCN1, where the center of the ROI is at the center of left ventricular cavity, (2) Image cropping for all CMR images using the extracted ROI, (3) Segmentation of the left ventricular cavity and myocardium using a FCN called FCN2, and (4) Calculation of left ventricular functional parameters and mass. The following section explains our framework in detail.
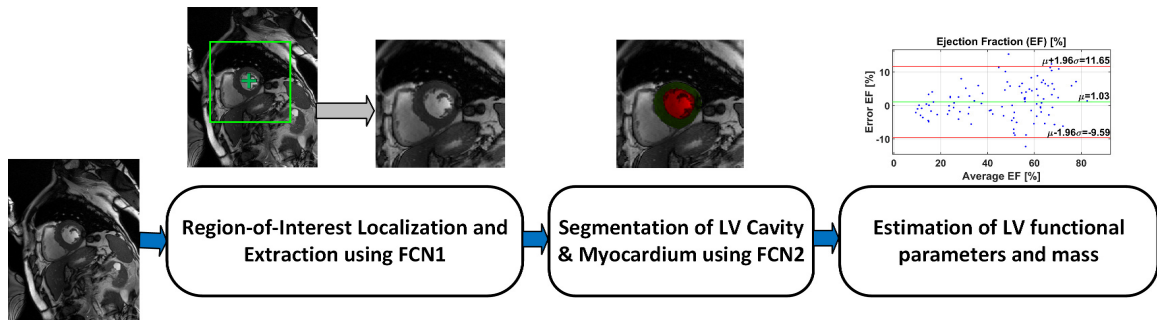
FIGURE 40: Illustration for the proposed framework for automated left ventricle segmentation and quantification (calculation of LV functional indexes and mass).

1. Region-of-Interest (ROI) Extraction

In CMR short-axis view image, the hear occupies a small proportion compared to the surrounding tissues that occupies larger proportion. In this situation, deep learning models that make dense classification in the spatial domain of the image become biased towards the surrounding tissues which is considered the majority class. Therefore, it is a necessary processing step to extract an ROI that encompasses the heart tissues before performing the final segmentation. Also, ROI extraction leads to a reduction in the computational load and boosts speed. In our framework, ROI extraction is performed using a bounding box of size 128×128 pixels. The center of the bounding box was at the center of the left ventricular center point. We calculated the left ventricular center point by a deep learning method that is based on a FCN, called FCN1, which is similar to the U-net. Figure

41 depicts the ROI extraction process.
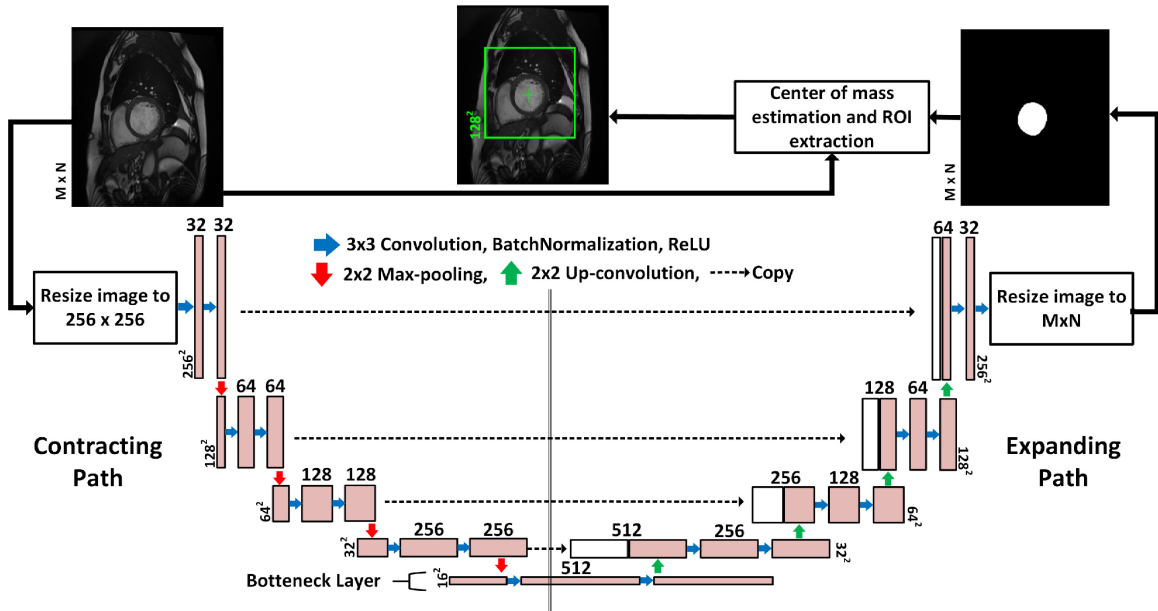
FIGURE 41: The extraction of the LV-ROI is performed using a FCN called FCN1. The original 2D CMR cine image is the input to the network. All images are resized to 256×256 pixels. The output of the network is a segmentation map for the left ventricular cavity. The map is resized to the same spatial dimensions of the input image ($M \times N$). The blue arrow refers to the successive operations of convolution (filter size 3×3), batch normalization, ReLU. For the convolution operation, the number of filters increases from 32 to 512 in the encoder path, and decreases from 512 to 1 in the decoder path. To maintain the same spatial dimension after convolution, we used zero padding. The red arrow indicates max-pooling operation that decreases spatial dimension by a factor of 2. The green arrow indicates up convolution that increases the spatial dimension by a factor of 2. Finally, the contextual information is copied from the contracting path and is concatenated to the expanding path by the dashed arrows that are called skip connections.

We trained the network to segment the left ventricular cavity from the original CMR cine images. Then, the center of mass of the segmented region is estimated. Then, we set the center of the ROI to this center of mass. Finally, the gray image is cropped to a $128x128$

image using the estimated ROI center. The output of the model might suffer from the class imbalance problem that leads to a high number of false negative (FN) pixels. However, this network is not intended to segment the cavity ideally, but it is intended to provide an estimate for the left ventricular center point. The left ventricular cavity in the apical slices occupies a very small part of the entire image. Therefore, due to the high degree of the class imbalance, the model may fail to segment the cavity. To overcome this issue, we use the center of the cavity of the previous slice as a center for the slice that leads to a black map from FCN1. By adopting this strategy, the 3D consistency of the LV is maintained without negatively affecting the overall performance because the extracted ROIs that truly not associated with LV tissues will again lead to black maps from FCN2.

Despite the effort required to train a FCN and to tune its hyper-parameters, the trained network takes little time to extract the LV-ROI. Our method is much faster than other approaches that use Hough Transform [2].

E. Cardiac Segmentation

Figure 42 shows our proposed model for the segmentation of left ventricular cavity and myocardium. The model is inspired from the FCN that was employed for segmentation tasks [98]. By convention, the input of the FCN is passed through an encoder path followed by a decoder path. In the encoder path, we reduce the spatial dimensions progressively until we reach a bottleneck layer where the input is described by an abstract and dense representation. On the other hand, in the decoder path, we restore the original input dimensions from the bottleneck layer by applying transposed convolutions. The bottleneck layer is located between an encoder and decoder path, and it involves a representation for the input in a reduced dimensionality form [10]. Our proposed model has multiple encoder paths that restore the input dimensions from multiple bottlenecks with various representations to the input. Then, the output of each encoder path is concatenated into a single layer that is fed to an inception module that is based on Google research [170]. Figure 43 depicts

83

the used inception module. The learning of multiple-scales features can be obtained from the integration of the inception module that has filters with different sizes. Small cardiac regions are detected by filters with small sizes while lager cardiac areas are detected by the larger filter sizes. Also, filters with large sizes remove the false positive regions that are similar to the targeted cardiac areas. Finally, we process the output of the inception module with a convolution layer to get the segmentation map. We apply a sigmoid layer on the output of the network. Network FCN2 involves various versions of FCN1 with different depths. Typically, FCN2 has 4 versions of FCN1's architecture and they all have the same encoder path. We concatenated the output of each network into one layer to obtain the final segmentation. Figure 44 demonstrates the relationship between FCN1 and FCN2.

1. Loss Function

The class imbalance problem is mitigated by the extraction of a LV-ROI. Furthermore, the extraction of a ROI boosts the performance of the CE loss. We kept using CE loss for its advantages such as smooth training and we proposed a novel loss called the radial loss which gives us a good segmentation for the left ventricular contours. Other loss functions have been proposed in deep learning literature and utilized segmentation metrics such as Dice loss [171]. However, these loss functions result in under-segmented areas with many false positives [171]. Therefore, we excluded these functions from our analysis. Let θ indicates the parameters of the trained model, $X = \{X_1, X_2, .., X_N\}$ indicates to the set of training images with size N, and $Y = \{Y_1, Y_2, .., Y_N\}$ indicates the set of ground truth segmentation labels. Then, the CE is given by the following equation:

$$L_{CE} = -\log p(Y_i|X_i, \theta) = -\sum_{c=1}^{C} \sum_{p_j \in X_i} Y_{i,c,p_j} \log \hat{Y}_{i,c,p_j} \tag{8}$$

Where $p(Y_i|X_i, \theta)$ refers to the probabilistic map predicted by the network after the sigmoid layer. X_i refers to the network's input. c indicates the class index. p_j represents a pixel in
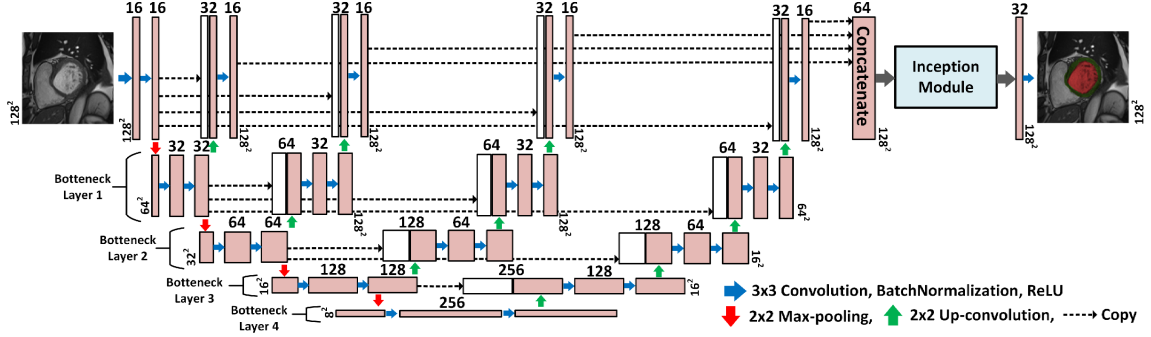
FIGURE 42: The proposed model for cardiac segmentation. The model is fed the extracted LV-ROI of size 128×128 pixels to produce an output that represent the segmentation map for the input ROI. In the map, red refers to LV cavity and green refers to the LV myocardium. The blue arrow refers to the successive operations of convolution (filter size 3×3), batch normalization, ReLU. For the convolution operation, the number of filters increases from 16 to 256 in the encoder path, and decreases from 256 to 1 in the decoder path. To maintain the same spatial dimension after convolution, we used zero padding. The red arrow indicates max-pooling operation that decreases spatial dimension by a factor of 2. The green arrow indicates up convolution that increases the spatial dimension by a factor of 2. Finally, the contextual information is copied from the contracting path and is concatenated to the expanding path by the dashed arrows that are called skip connections.

image X_i. Y_{i,c,p_j} represents the true probability that p_j in the class c, and $\hat{Y}_{i,c,p_j}$ refers to the predicted probability that p_j in the class c.

We exploited the fact that LV is associated with a radial shape to propose a new radial loss function. We define the radial distance (RD) at a specific angel by the distance between the center of mass of a segmented area to its surface at a specific direction. Thus, if we have a ground truth surface G and a segmented area surface S, the error of the local radial distance d at an angle θ is demonstrated in Figure 45 and is defined in Equation 9

$$d = s_\theta - g_\theta \qquad (9)$$

85
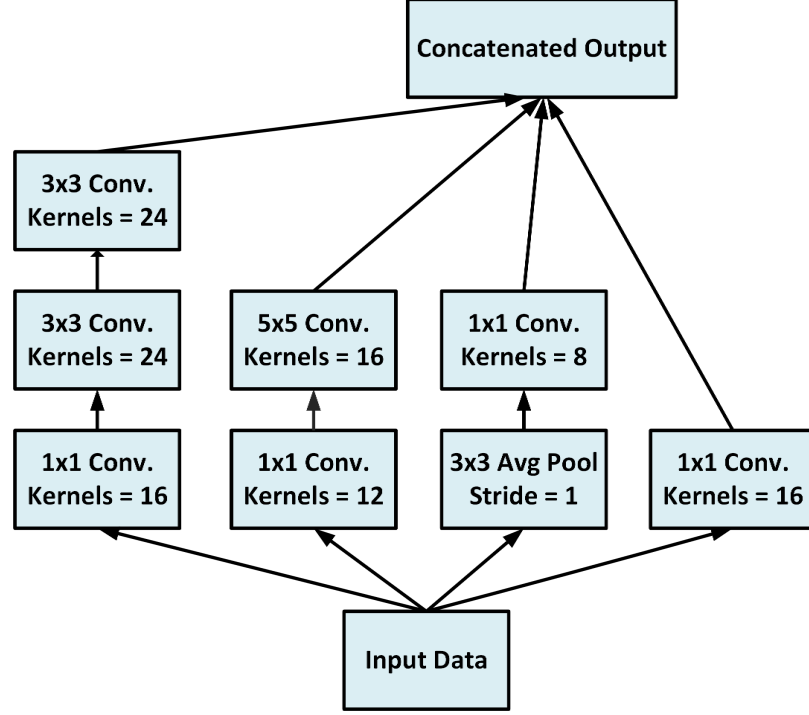
FIGURE 43: The structure of the employed inception module in FCN2 network. The module contains parallel processing paths with kernels of various sizes, i.e., 1×1, 3×3, and 5×5 convolutions, and average pooling operations with kernel of size 3×3. The feature maps obtained from these paths are then concatenated in the final layer.

Where the RDs from the center of mass point to the surfaces G and S are described by g_θ and s_θ, respectively. Now, if we constructed an equi-spaced radial lines, we could calculate the RDs for the surfaces G and S, and save them in the same radial order in vectors g and s, respectively. We can define the RD loss as L2 penalty:

$$L_{RD} = \frac{1}{M} \|s - g\|_2 \tag{10}$$

Wher M refers to the number of radial lines. By applying a Sobel filter on the ground truth Y_i and the predicted probabilistic map $\hat{Y}_i$ we can obtain the surfaces G and S, respectively for an input image X_i. The loss function is an Euclidean norm, thus, it is differentiable by a deep learning library. The loss function L_{RD} can be employed for both EnC and EpC of the LV as follows:
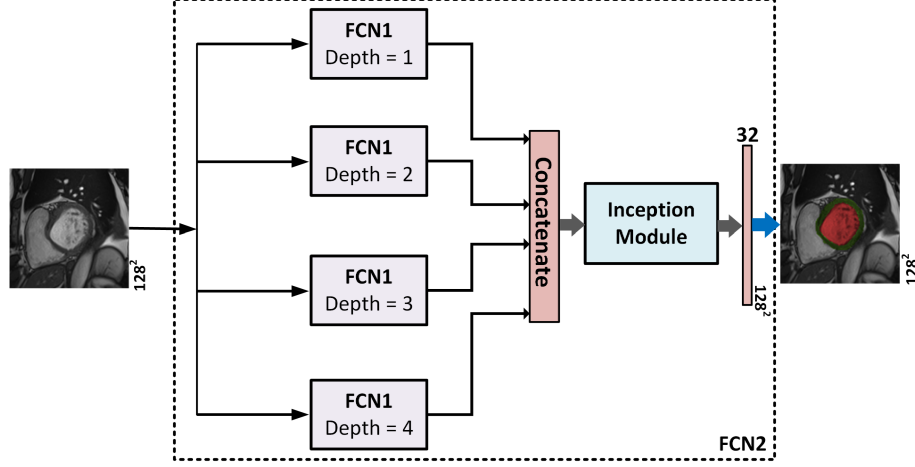
FIGURE 44: The relationship between model FCN1 and FCN2. Model FCN2 contains 4 FCN1s with different depths and the output of each model is concatenated to obtain the final segmentation.

$$L_{RD} = \frac{1}{M} \left\| s_{EnC} - g_{EnC} \right\|_2 + \frac{1}{M} \left\| s_{EpC} - g_{EpC} \right\|_2 \tag{11}$$

We can define the final loss function as follows:

$$L = L_{CE} + L_{RD} \tag{12}$$

2. Network Training Settings

We used Pytorch deep learning framework to build both FCN1 and FCN2. Kaiming initialization [172] was used to initialize the weights of the convolutional layers. The variables that must be set before training the network are called network's hyper-parameters. These variables identify the architecture of the network such as the number of filters. They also determine network's training such as learning rate. The optimal values of the hyper-parameters were calculated using a grid search approach when the segmentation accuracy is our criterion to optimize. The search space for the initial number of filters = {8, 16, 32}. We used Adam optimizer with a learning rate that has the search space = {0.01, 0.001,
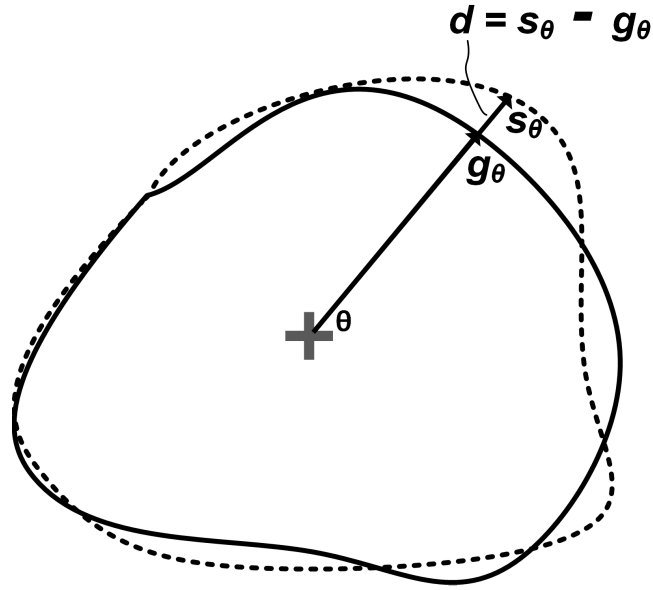
FIGURE 45: The solid line refers to the rue contour while the dashed line refers to the predicted contour in polar space. The local radial distance error at an angle θ is described by d.

0.0001} and the learning momentum = {0.9}. The search space of the batch size = {8, 16, 32}. The search space for the number of the epochs = 100:50:300. The size of the training set can be increased using the data augmentation approach. Data augmentation also helps in overcoming the problem of over-fitting during network's training. Thus, a data augmentation strategy that was associated with random translations, scaling, and rotation was adopted.

F.   Experimental Results

1.   Cardiac Datasets

The left ventricular segmentation and function quantification were performed on two different cardiac datasets. Namely, the ACDC cardiac dataset which is publicly avail-

able from MICCAI 2017 challenge [173], and a locally-acquired cardiac dataset.

ACDC-2017 Dataset: The dataset is composed of 150 exams for various patients. The patients are divided into 5 categories according to physiological heart parameters. The categories are (1) normal subjects, (2) patients with dilated cardiomyopathy, (3) patients with previous myocardial infarction, (4) patients with abnormal right ventricle. (5) patients with hypertrophic cardiomyopathy. The data providers divided the dataset into two sets: (1) A set for training which is composed of 100 cases along with their ground truth manual segmentation at the ED and ES phases in all acquired heart slices; (2) A set for testing which is composed of 50 cases without annotation. The two sets have even arrangement of patients categories. The cardiac cine images were acquired in breath hold with a retrospective or prospective gating and with a SSFP sequence in short-axis orientation. The LV was covered entirely by the short-axis slices. The imaging parameters were: slice thickness equals 5 or 8 mm, inter-slice gap equals 5 or 8 mm, and spatial resolution equals 1.37-1.68 mm^2/pixel.

Locally-acquired dataset: This dataset was utilized to answer the question whether our segmentation method is generalizable by evaluating its performance on this dataset which was not used in training. In this set, cross-sectional cardiac cine images were acquired from 11 patients with known history of myocardial infarction. The institutional review board (IRB) approved our study. The dataset is composed of 26 cardiac scans that cover various heart sections. Twenty-five frames were captured at each section to cover the cardiac cycle. The sum of the 2D images was about 6,000 images.

2. Framework Training and Validation

The ACDC dataset was used to analyze the performance of our novel framework for the automated left ventricular segmentation and quantification through the calculation of physiological heart indexes introduced in IV.A.3. The dataset is composed of 100 patients along with their ground truth manual annotation. A ten-fold cross-validation was adopted to train and validate FCN1 and FCN2. In each fold, we had an equal number of patients

from the five introduced heart diagnoses. This was achieved by stratified sampling. We can reformulate our description of network evaluation by saying that we trained the networks ten times and in each time we used a training set of 90 cases (average of 1800 2D images) and a testing set of 10 cases (average of 200 2D images). Furthermore, the generalization of our segmentation approach was tested using the locally collected dataset.

A probabilistic map is the result of our segmentation network. This map is composed of pixels and each pixel is given a probability of belonging to the object. We applied Otsu thresholding [174] on the probabilistic map to get the final segmented binary mask. Additionally, To eliminate the false positive pixels, we looked for the connected components in the binary mask. Finally, morphological operations, such as gap filling were applied on the resulted binary segmentation. We used segmentation metrics such as Dice score and Hausdorff distance (HD) to assess the segmentation accuracy of our approach.

3.   Evaluation of LV-ROI Extraction

To assess the performance of our novel approach of LV-ROI extraction, we trained and validated FCN1 by the ACDC dataset using a ten-fold cross-validation strategy. In each fold, we estimated the center of mass of the segmented left ventricular cavity P_s for each image. Then, we used two measures/metrics to evaluate the network performance; namely (i) the Euclidean distance between the center point of our predicted segmentation P_s and the center of mass of LV cavity from manually annotated segmentation P_m, and (ii) the percentage of the images with ROI prediction that encompasses all pixels of left ventricular cavity and myocardium. the statistics of the Euclidean distance between P_m and P_s is shown in Table 5. The statistics are for 1902 images from the ACDC dataset. Our method resulted in a good accuracy and surpassed the method in [2]. Our approach takes on average a 700 msec to extract the desired ROIs of one case at the phases of ED and ES. Furthermore, the resulted ROIs encompassed all the left ventricular cavity and myocardium tissues,

TABLE 5: The statistics of difference in pixels between the predicted left ventricular center of mass point (P_s) and the center of mass resulted from manual annotation (P_m). The table compares our approach with the method of [2]. STD refers to standard deviation.

	Mean	STD	Max.
Hough transform [2]	4.00	3.83	36.24
Proposed (FCN1)	1.41	1.65	5.00

4.   Evaluation of the Proposed Loss Function

After the extraction of the ROI, cardiac segmentation was performed using FCN2. We used L_{CE} loss only in one time and the proposed loss $L_{CE} + L_{RD}$ in another time. A comparison between the segmentation metrics resulted from the two loss functions are shown in Table 6. The table shows the segmentation for both LV cavity and myocardium (MYO). We can notice that when we used L_{CE} alone, we obtained a good segmentation performance because the issue of class-imbalance was mitigated by the ROI extraction step. While, our proposed novel loss function resulted in a superior performance in terms of the used segmentation metrics. Furthermore, a better segmentation quality resulted from our proposed loss, as shown in Figure 6. Thanks to the capability of the RD loss in which the distance between the predicted contours and the true contours is minimized.

TABLE 6: The average segmentation accuracy of our segmentation approach when using two different loss functions in a ten-fold cross-validation strategy applied on the ACDC dataset.

Loss Function	Dice Coeff.		HD (mm)	
	LV Cavity	MYO	LV Cavity	MYO
L_{CE}	0.93	0.86	9.52	11.41
$L_{CE} + L_{RD}$	0.94	0.89	6.71	7.13
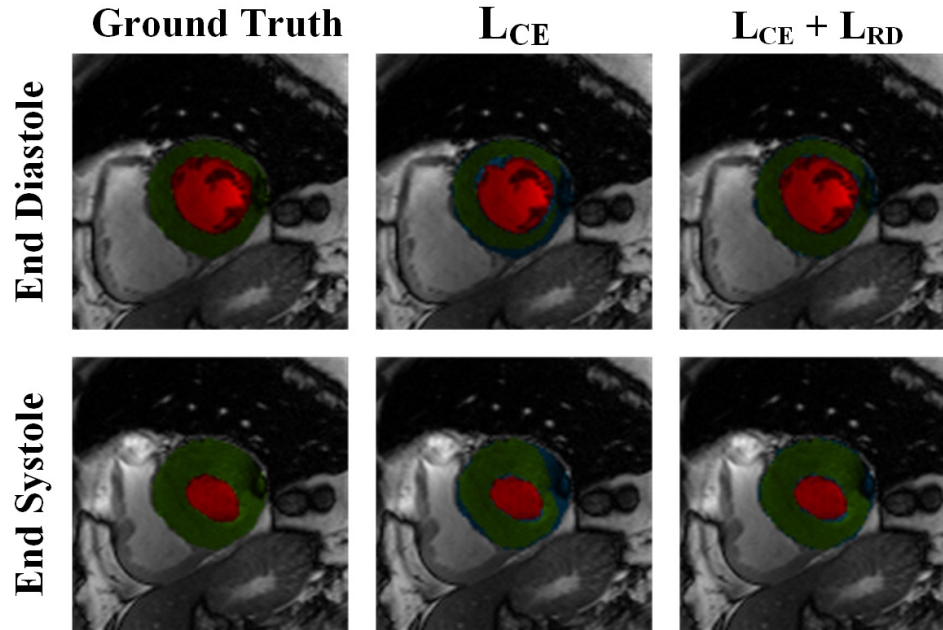
FIGURE 46: Comparison between the ground truth annotation of the LV in phases of ED and ES, the predicted LV delineation from FCN2 with L_{CE} loss, and the predicted delineation from FCN2 with our loss ($L_{CE} + L_{RD}$). Green and red areas indicate the LV myocardium and cavity, respectively. Blue color indicates the segmentation error. We can notice visual qualitative improvement for the delineation of our proposed loss function.

5. Evaluation of the Proposed Network Model FCN2

To evaluate the performance of our proposed network FCN2, a comparison was made with two other methods: (i) the original model of the U-net [10] with 4 layers and initial convolutional layers that had 64 filters, (ii) the ConvDeconv model introduced in [175]. We trained all of the three networks with the same configurations as explained in section IV.F.2 in a ten-fold cross-validation strategy. Table 7 shows a comparison in terms of the resultant segmentation accuracy between the proposed FCN2 model against the other models. The table also shows the number of the learnable parameters required by each model. This reflects the required computational cost by each model. As indicated, FCN2 resulted in the best segmentation performance for all segmented regions, while ConvDeconv model re-

sulted in the lowest performance. The inferiority associated with the ConvDeconv net may
be due to the lack of skip connections that works on adding high resolution features to the
expanding path. Furthermore, FCN2 works better than the original U-net model that starts
with 64 filters. Therefore, U-net with fewer filters were excluded with our comparison.
Another advantage for our proposed FCN2 network is that it requires fewer parameters,
and thus, it requires less training time and GPU memory usage.

TABLE 7: A comparison between three different segmentation methods in terms of the
segmentation metrics (Dice and HD). The values of the metrics are presented as the average
values when we trained the model using a ten-fold cross-validation strategy. The number
of learnable parameters are also shown.

Method	Dice Coefficient				HD (mm)				# of Param.
	LV Cavity		MYO		LV Cavity		MYO		
	ED	ES	ED	ES	ED	ES	ED	ES	
U-net [10]	0.94	0.90	0.83	0.85	8.22	10.53	9.81	11.51	31M
ConvDeconv net [175]	0.92	0.88	0.80	0.83	9.14	11.34	10.81	11.95	252M
FCN2 (proposed)	0.96	0.92	0.88	0.89	6.31	7.42	7.11	7.25	2.5M

Segmentation results of FCN2 are shown in Figure 47. The figure contains three
different short-axis slices of a heart along with their ground truth segmentation at phases
of ED and ES of the cardiac cycle. Overall, our method (FCN2) resulted in accurate seg-
mentation with some errors at apical slices. After all of the described analysis, we decided
to choose the segmentation results of FCN2 for the estimation of heart's physiological pa-
rameters.

6. Generalization Evaluation

After obtaining good segmentation results from our approach applied on ACDC
dataset only, it is time to assess the generalization capability of our method. We have two
models using our method:

1. Our method when we use the local dataset for both training and testing.
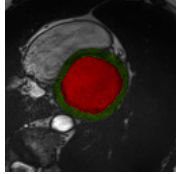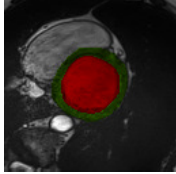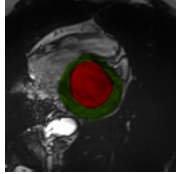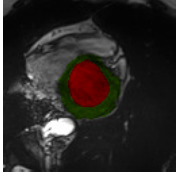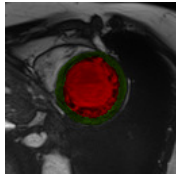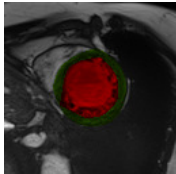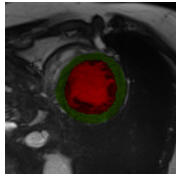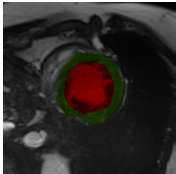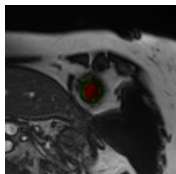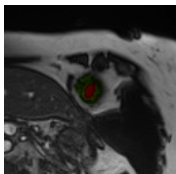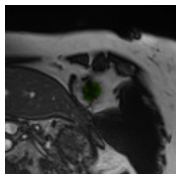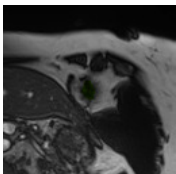
FIGURE 47: Left ventricular segmentation of FCN2 model at ED and ES phases of one case during the ten-fold cross-validation. Green, and red regions refer to the LV myocardium, and cavity, respectively. The basal, mid-cavity, and apical slices are indicated by the letters B, M, and A, respectively.

2. Our method when we use ACDC dataset for training and the local dataset for testing.

The segmentation performance of the two models is shown in Table 8. In the second model, although for training we used ACDC dataset which consists of about 1.4K images acquired at the phases of ED and ES, we obtained good segmentation accuracy for the LV at all cardiac phases of the local dataset. In the first model, the same dataset distribution was used for training and testing. Therefore, the results of the first model are slightly better than the second model that uses different distribution for training and testing. We can conclude that our approach generalizes well to different data. Furthermore, segmentation for the full cardiac cycle was obtained from our approach when trained only using a data that has annotation for the phases of ED and ES.

94

TABLE 8: The segmentation performance of two models. In model A, the local dataset was used for training and testing. In model B, ACDC dataset was used for training and the local dataset was used for testing. The estimates are mean values.

Model	Dice Coef.		HD (mm)	
	LV Cavity	**MYO**	**LV Cavity**	**MYO**
Model A	0.95	0.87	9.31	8.52
Model B	0.94	0.85	11.12	9.74

7. Physiological Parameters Estimation

After performing segmentation for the LV cavity and myocardium from the cardiac images, we estimated five functional parameters; namely the EDV, the ESV, the LVM, the SV, and the EF. These parameters were described in IV.A.3. To calculate the degree of agreement between the estimated values from the output segmentation and the estimated values from ground truth segmentation, we constructed Bland-Altman plots [176], as shown in Figure 48. These figures demonstrate the bias μ (mean difference) and the 95% agreement limits ($\sigma \pm$ 1.96 SD). In Bland-Altman plots, we must check the normality of the differences. Therefore, Shapiro-Wilk test for normality was used with 5% significance level. The obtained p-values were 0.082, 0.052, 0.061, 0.154, 0.787 for the EDV, the ESV, the LVM, the SV, and the EF, respectively. Given that the p-values are higher than 0.05, the test accepted normality. Our calculated parameters have a mean of only 3 outlying points which is considered only 3% of the involved cases.

The error statistics are summarized in Table 9. The table shows the statistics for the EDV, the ESV, and the EF estimations of our approach and other approaches applied on the ACDC dataset. The lowest bias and standard deviation were obtained from the errors of the EDV and the ESV measures. Additionally, the error of the EF estimate has a lower standard deviation than the approach in [177]. Overall, our method resulted in acceptable
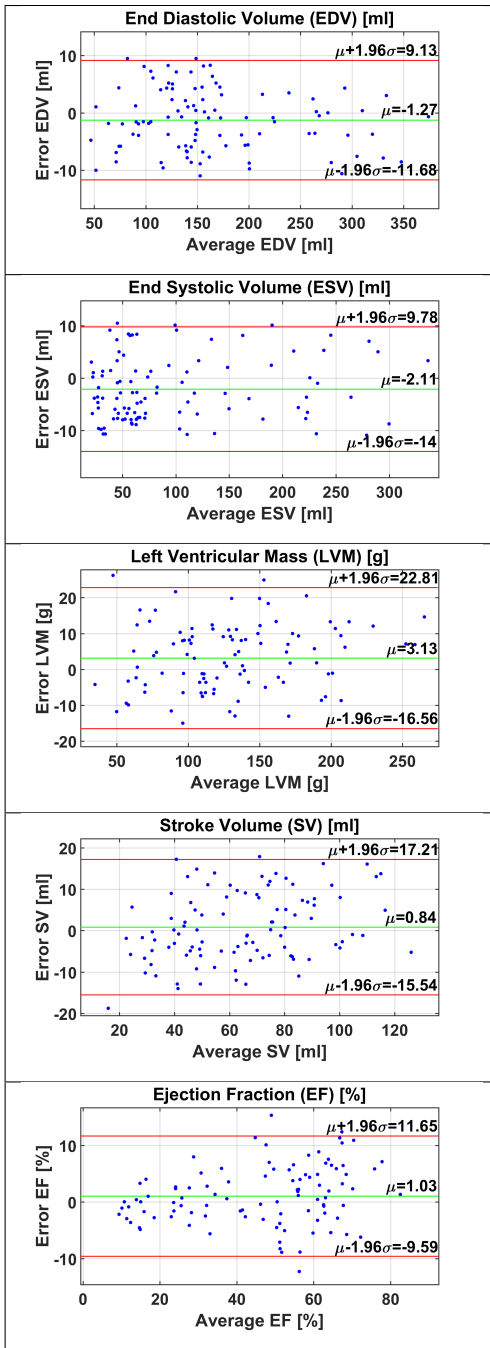
FIGURE 48: Bland-Altman plots for physiological parameters. The figure shows EDV, ESV, LVM, SV, and EF from top to down. The bias of the calculated values from the ground truth is indicated by μ. $\mu \pm 1.96\sigma$ indicates the 95% confidence interval. To obtain a good agreement, the bias should be near the zero value and the error points should be within the confidence interval.

differences that are comparable with intra- and inter-subject variability associated with the manual calculation of functional parameters from cardiac images as reported in [178, 179].

TABLE 9: A comparison between our approach and other methods that aim at the automated estimation of functional LV parameters. The values indicate the average (std.) of the differences between the automated and manual estimation.

Reference	EDV (ml)	ESV (ml)	EF (%)
Wolterink et al. [177]	-1.57 (6.31)	-2.51 (7.66)	1.23 (4.99)
Grinias et al. [180]	1.43 (9.95)	2.61 (17.60)	-0.05 (8.61)
Proposed	-1.27 (5.31)	-2.11 (6.06)	1.03 (5.42)

G. Discussion

In this chapter, we aimed at developing a deep learning based method for efficient and accurate segmentation to the LV myocardium and cavity. Our method is fully automated and provides accurate LV quantification through the estimation of LV functional parameters which are widely used for heart functional evaluation. Our framework features a novel network architecture and major contributions are explained below.

Firstly, we introduced the idea of building an initial network (FCN1) for the task of automated ROI extraction from original CMR images. It turned out that this idea was successful in providing accurate estimates for LV center point. This idea was also fast during detection. A comparison was made between our approach and other approach that has the same aim of LV localization and ROI-extraction; however, it uses Hough transform. Table 5 demonstrates how our approach surpassed an approach which relies on Hough transform for ROI extraction. Additionally, the ROI of one patient at ED and ES were extracted by our method in only 700 msec. Our extracted ROIs contained all of the desired LV cavity and myocardium tissues. Thus, our extraction method is fast, accurate, and

97

reliable.

We selected the suitable ROI size based on the minimum size that resulted in myocardial segmentation without clipping, assuring that there is no clipping for the LV is necessary for further accurate calculation of heart physiological parameters. In our experiments, an ROI of size 128×128 was chosen because this size provided the smallest area that encompasses the LV tissues in our CMR dataset. Extracting an ROI of size 128×128 has two advantages: (1) it reduced the time and computational requirements during the training and inference of the network, (2) it mitigated the class imbalance problem by discarding the unwanted surrounding tissues. Anyway, our implementation can work with arbitrary ROI sizes.

Secondly, our proposed final segmentation is constructed by network FCN2 from various bottleneck layers. These bottleneck layers are different representations to the input image with different dimensionalities. Our network FCN2 resulted in accurate segmentation for the LV cavity and myocardium as demonstrated in Figure 47. It also surpassed state-of-the-art models such as U-net and ConvDeconv networks in terms of the segmentation accuracy as demonstrated in 7. Another advantage of our network is the efficient use of time and memory by requiring a fewer number of learnable parameters.

Moreover, we carefully chose the components of our deep learning model because this is essential for the overall success of the model. For instance, a novel loss function was implemented and we called it radial loss which is appropriate for LV segmentation task because of the underlying radial shape of EpC and EnC. Our final loss function is the sum of both the baseline cross-entropy and the radial loss. Cross-entropy is popular in the domain of deep learning segmentation because it is associated with the ability to provide smooth training and it has nice differentiable properties. Researchers consider cross-entropy as the standard loss in various applications, especially image segmentation. As demonstrated in Table 6 and Figure 45, our loss provided a better segmentation performance when compared to cross-entropy alone. Our loss function showed superior performance due to the

fact that radial loss minimizes the distances between the actual contours and the predicted contours of the LV. The problem of over-fitting can be avoided by a smart way called data augmentation. Data augmentation also works on increasing the training samples which is useful in case of scarce annotated data. By adopting data augmentation, we reached good mean Dice and HD values the delineation of LV cavity and myocardium.

Regarding Figure 48, in the EDV and ESV measurements, there are small negative biases which mean that we slightly underestimated this measurement. We can also notice that EDV had a lower bias than the ESV and, consequently, the calculated EF had a positive bias. If there are points outside the confidence interval, we might have a wrong diagnosis for patients with CVD due to these errors. Therefore, an accurate LV segmentation approach should not result in outlying points by reducing the segmentation errors that are transferred to the next step of functional parameters estimation.

Finally, a comparison has been made between our approach and other approaches that aimed at quantification of LV using the ACDC dataset to establish the merit of our method based on the reported errors of each approach. In general, our framework resulted in acceptable errors for the calculated parameters and surpassed the previous frameworks, as indicated in Table 9. Additionally, we would like to note that our framework is not limited to a specific dataset; we proved that our method can provide good results by using a different dataset, even with a network trained by a sparse data (ACDC-2017) and validated on a dense data that covers the whole cardiac cycle.

CHAPTER V

CARDIAC STRAIN ANALYSIS FROM CINE MRI IMAGING IN MICE

Cardiac magnetic resonance (MR) imaging is a useful form of imaging that can be used to assess cardiac function. Strain analysis enables us to comprehensively assess diastolic myocardial function that is not indicated by estimating systolic functional indices utilizing a normal cine imaging module. Because of the small size of mice heart, it is not feasible to perform appropriate tagged MRI to assess strain. In this chapter, we implemented and developed a novel artificial intelligence framework to automatically quantify myocardial strain from cine cardiac MRI. Our framework starts by efficient segmentation of the LV from the cine MR images by fully convolutional neural network (FCN) architecture. To estimate myocardial strain, a Laplace-based approach was developed to track the LV wall points. The tracking is performed by obtaining a solution for the Laplace equation between the LV boundaries of each two consecutive image frames over the cardiac cycle. After tracking, the strain calculation is performed by adopting a Lagrangian-based method. Our novel framework for myocardial strain analysis was validated by performing a comparison between the outcome of these analyses with the outcome of tagged MRI obtained from the same mice. Notably, our developed algorithm, which uses cine MRI, produced strain values that are quite similar to the values obtained from tagged MRI. Furthermore, our novel algorithm established the potential of determining the strain difference between diseased and normal hearts.

A. Introduction

Mice are the most preferable species for cardiovascular modeling for several rea-

sons including the low procedure costs, ease of handling, and the capability to manipulate the mouse genome [53]. By utilizing animal models, we increase our knowledge for improving the diagnosis and treatment of cardiovascular diseases. For cardiac function quantification, MRI is considered the most accurate imaging technique (gold standard) [54–58]. The major advantages of MRI are ability to rebuild 3D structure and function, clear tissue structure, and high reproducibility [59]. currently, cardiac MRI is clinically utilized to evaluate myocardial function, heart structure, perfusion, and viability in humans [60]. Multiple mouse works by our group and others [54,61–63] got benefit from the use of standard cardiac cine MRI to evaluate heart structure and functional indexes, e.g., EDV, ESV, and EF. However, tagged cardiac MRI is required for strain analysis of myocardial contractility. Tagged cardiac MRI takes considerable time and difficult to interpret, especially in mice. It is difficult to interpret because of the small size of the mice heart, which prevent the creation of multiple $1mm^2$ diamonds within the wall of the ventricle for strain analysis. Therefore, researchers abstain from performing strain analysis from tagged MRI. On the other hand, the picture of the heart function is fully completed by strain analysis which is an essential component because it is the only parameter that distinguish between active and passive movement of myocardial segments, assess longitudinal myocardial shortening, and quantify intraventricular dyssynchrony. Furthermore, evaluating myocardial strain gives us a comprehensive evaluation of diastolic myocardial function, that is not inferred from EF [64].

Deep learning approaches, particularly, CNN [181] showed successful results in medical imaging tasks that range from traditional image processing problems such as registration and segmentation to the development of advanced computer aided diagnostic systems for disease diagnosis [182, 183]. From only convolutional layers, we can build FCN which is a deep learning model that can be trained to provide object segmentation or semantic segmentation for the input image [98]. Several implementations of FCNs have been proposed in the literature for automatic segmentation and quantification of the human LV

from cine cardiac MRI images [168, 184–186]. Although, these techniques resulted in successful development of tools for segmentation and quantification of the LV, there has been a very limited research in the domain of deep learning applied on cardiac data of mice [187].

The primary aim of this chapter is the development of an automated framework for the accurate estimation of strain from standard cardiac cine MRI of mice. The used mice are sham and myocardial infarction (MI) cases. Our work is an alternative to tagged imaging of mice. Myocardial segmentation is an important initial step for the estimation of LV functional indexes. Therefore, a deep-learning based technique was proposed for the automated segmentation of LV borders. Subsequently, we developed a Laplace based technique to track myocardial points along the cardiac cycle to accurately evaluate the strain from the standard cardiac cine MRI. We validated the Laplace-based technique by performing a comparison between the Laplace-based technique and the analyzed strain outcome resulted from the tagged imaging which is considered the current gold standard for strain analysis. Finally, we also utilized our framework to calculate other global and local cardiac indexes.

B. Methods

We propose an automated framework that quantify LV functional and structural parameters. This framework consists of two primary stages. The first stage is the segmentation of the LV cavity and myocardium using a deep learning approach that is based on a FCN. The second stage is the calculation of the functional and structural heart parameters. The parameters include global volumetric indexes (e.g., EDV, ESV, EF) and local indexes involving wall thickening and myocardium strain. Fig. 49 depicts the overall framework. The development and validation of the framework will be described in detail in the upcoming subsections.
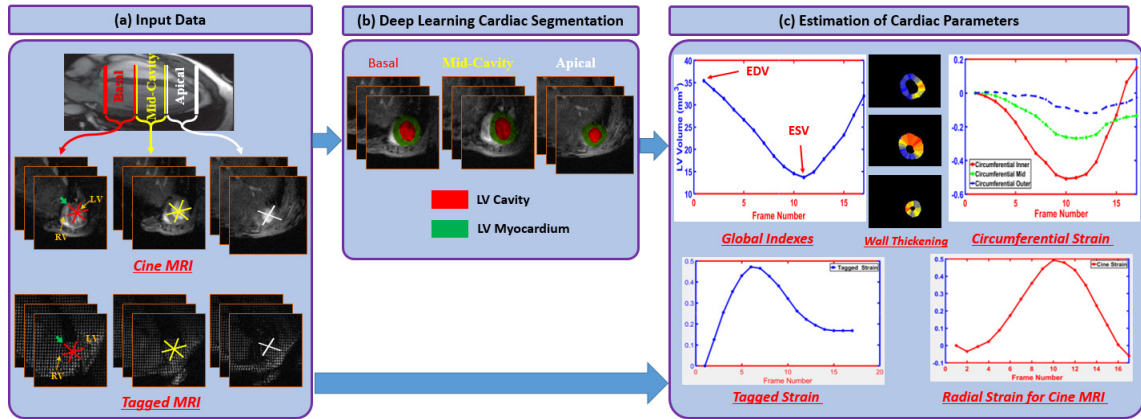
FIGURE 49: The proposed framework for the quantification of left ventricular functional and structural parameters.

1. Mice Data

The procedures of all animals were performed in accordance with the guide of the National Institutes of Health for the Care and Use of Laboratory Animals. Furthermore, our procedures were approved by the University of Louisville Institutional Animal Care and Use Committee. We performed the surgeries as described in [62, 188–190]. The dataset is composed of cross-sections cine MRI and tagged MRI of 6 cases. Approximately, seventeen temporal frames for each cross-section through the cardiac cycle were acquired. Eight cross-sections were obtained to completely cover the LV for each subject. The dataset contains 3 sham and 3 MI subjects.

All MRI mice data were obtained utilizing Agilent $9.4T$ horizontal bore system. Agilent $205/120$ HD gradient coil (Agilent Technologies, Santa Clara, CA, USA) operates with the MRI system. For signal transmission, we used a RAPID 72-mm volume coil. For signal detection, we used a 4-channel mouse heart surface coil. We positioned the surface coil below the mouse body. During in-vivo scans, We anesthetized the mice with 1.5% isoflurane in $100\%O_2$. Breath rate, Electrocardiography (ECG) signals, and body temperature were continuously monitored utilizing a small animal monitoring and gating system (Model 1030, SA Instruments, Inc., Stony Brook, NY, USA). We gated the cine and

tagged MRI scans with ECG and breathing. High resolution short-axis view cine images were acquired utilizing the following imaging parameters: TR/TE equals $5.0/1.6$ msec; matrix size equals $30 \times 30 \ mm^2$; flip angle equals 15°; field of view (FOV) equals 256×256 mm^2; eight slices with a thickness equals $1.0mm$; and number of averages equals two. For black blood tagged cine, we used the following scanning parameters: TR/TE equals 5.0/1.9 msec; matrix size equals $30 \times 30 \ mm^2$; flip angle equals 20°; FOV equals $128 \times 128 \ mm^2$; eight slices with a thickness of $1.0mm$; number of averages equals four; tagging resolution equals $0.3mm$ with $0.6mm$ separation; and tag time equals 16.24 msec.

2. Development of an Automated Approach for LV Segmentation

Recently, traditional image segmentation techniques have been outperformed by deep learning based segmentation techniques. In this work, we utilized an FCN that is is similar to the U-net which was implemented to achieve medical image segmentation [10]. Figure 000 shows the architecture of the used FCN. The input to the network is a 2D cardiac image. The image passes through a contracting route followed by an expanding route. In the contacting route, a processing by successive blocks occurs to the image, where each block consists of convolutional layer, RelU activation, and max-pooling layer. The result of this processing is the the reduction in the spatial dimensions and the production of an abstract representation for the input image. This representation is the end of the the contracting route in a layer called bottleneck. In the expanding route, we restore the spatial dimensions by applying up-convolution. The resolution of the input is increased by up-convolution operation which uses a deconvolution filter which can be learned during the training of the network [98]. Furthermore, skip connections have been added to the network. These connections copy and concatenate the maps that contain high resolution feature from the contacting route to the expanding route. Skip connections work on providing fine segmentation. We train the network by pairs of cardiac cine MRI images, X, and their corresponding manual segmentation for LV (binary masks), Y. The outcome of

104

the segmentation network to an input image x is a one layer segmentation map $\hat{y}$, where we classify each pixel in the map by assigning a probability to it. This value represents the probability of being an object. The segmented object in our work is the LV myocardium. A sigmoid layer is employed as the final layer in the FCN to obain the map $\hat{y}$. Then, we use a loss function to make a comparison between the predicted map and the true binary image y.

A dense classification is performed in the spatial domain of the image when segmenting the LV myocardium using FCN. A big obstacle is that the network is prone to class-imbalance problem when the majority of the pixels in the image belong to the surrounding tissues. This class-imbalance problem is found in our segmentation task because the myocardium comprises small proportion in the image compared to the other tissues. When using the BCE loss alone, the network is susceptible to class-imbalance issue [191]. Therefore, we propose a new loss function that is composed of two terms to present a solution to the class-imbalance problem. First term is the conventional BCE loss function defined as follows:

$$L_{BCE} = \sum_{i=1}^{N} -(y_{oi} \log(\hat{y}_{oi}) + y_{bi} \log(\hat{y}_{bi}))) \tag{13}$$

Where N refers to the number of pixels. y_{bi}, and y_{oi} are the true annotations for pixel i resulted from manual segmentation y (y_{bi} is 1 when i lies in the background and 0 for the object, and vice versa with y_{oi}). $\hat{y}_{bi}$ and $\hat{y}_{oi}$ are the predicted probabilities that a pixel i in the background and the object, respectively. The second term is a loss function that combines the sensitivity and specificity and is described as follows:

$$
\begin{aligned}
L_{ss} &= 2 - (Sensitivity + Specificity) \\
&= 2 - \left(\frac{\sum_{i=1}^{N} \hat{y}_{oi} y_{oi}}{\sum_{i=1}^{N} \hat{y}_{oi} y_{oi} + \sum_{i=1}^{N} \hat{y}_{bi} y_{oi}} + \frac{\sum_{i=1}^{N} \hat{y}_{bi} y_{bi}}{\sum_{i=1}^{N} \hat{y}_{bi} y_{bi} + \sum_{i=1}^{N} \hat{y}_{oi} y_{bi}} \right)
\end{aligned}
\tag{14}
$$

By definition, high specificity and sensitivity require minimization of false positives (FPs) and false negatives (FNs). Our final loss function L is the summation of the two terms as

follows:

$$L = \alpha L_{BCE} + \beta L_{ss} \qquad (15)$$

Where α and β are hyper-parameters that govern the assigned weights for the two terms in the equation. $\alpha + \beta = 1$ in our formulation. We assigned higher value to β compared to α, this will help us to reduce the class-imbalance effect resulted from BCE and in the same time assign a higher weight to the minimization of the summation of FNs and FPs. We adopted a grid search strategy to calculate the optimal values of α and β when the optimization criterion is the value of DSC. The search space of α and β ranges from 0 to 1 with 0.1 step and $\alpha + \beta = 1$ constraint. A novel loss called Tversky loss has been introduced by Salehi et al. [192] to account for the class-imbalance issue in deep learning segmentation. To assess the potential of our new loss, we compared its performance with other existing methods.

When the input of the network is a test image x^*, the output of the network will be the corresponding predicted segmentation map $\hat{y}^*$. In this map, each pixel is assigned a probability of belonging to the object by the sigmoid layer. A binary segmentation map is then obtained by applying Otsu thresholding [174] on $\hat{y}^*$. We applied extra post-processing on the binary map: (1) Connected region with the maximum number of pixels was kept and the remaining FP regions were discarded, (2) gap filling to fill the gaps. Following post-processing, the segmentation quality of our method was inferred for both LV cavity and myocardium.

Cardiac cine MRI images of 6 mice were used to train and test our proposed FCN in a leave-one-out cross-validation strategy. In this strategy, six iterations of successive training and testing were made. In each iteration, the 2D images of 5 cases (about 750 images) were used as a training set for the network and the remaining images of the last case (about 150 images) were used as a testing set. For segmentation performance evaluation, we estimate the averages of the segmentation metrics along the tested images. This
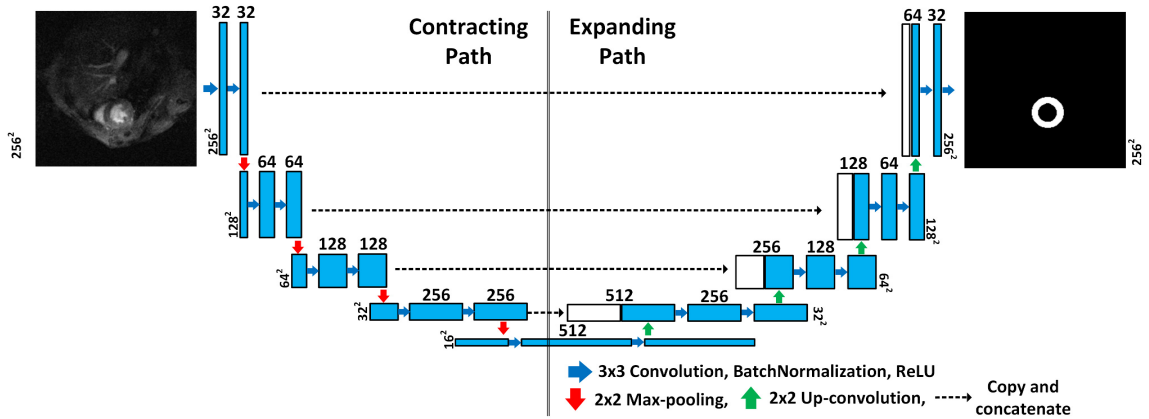
FIGURE 50: The architecture of the used FCN for LV segmentation.

methodology was repeated 6 times and the final averages of segmentation metrics were calculated. We reported these values to assess the performance of our segmentation method. We re-scaled the images to the spatial dimension of 256×256. A data augmentation scheme that involves random translation, rotation, and scaling was adopted. The initialization of the convolutional layers was performed using The Kaiming initialization [92]. We used the Adam optimizer [172] with a learning momentum of 0.9 and a learning rate of 0.001. A total of 100 epochs was used with a stopping criterion of five waiting iterations. To make an early stoppage, we utilized the loss value of a validation set that is composed of twenty percent of the training set. Therefore, our network makes an early stoppage before the end of the 100 epochs, if the loss value of the validation set did not improve for five successive epochs. Finally, the lowest loss value on the validation gave us the best model that will be used for testing. We used Pytorch software to develop, train, and test our deep learning model.

3. Algorithm for Calculating Cardiac Function and Structure Parameters

Following the delineation of the LV myocardial boundaries, both local and global physiological heart parameters can be estimated for heart functional evaluation. Accurate

segmentation of the inner and outer borders is required for the calculation of the local and global ventriculometrics. The estimation of Strain which is a local ventriculometric requires tracking of the contour points of the myocardium. The estimation of wall thickness requires the accurate co-allocation of the corresponding points on the heart wall. Therefore, the crucial step is the careful localization of the myocardial points. To anticipate for the problems associated with the lack of strong edges and intensity variations between subjects' scans, we adopted a method that benefits from geometric features instead of image intensities. This process tracks myocardial points or co-allocates the corresponding border pairs through the cardiac cycle. We applied a geometric technique to determine geometric features and by solving the Laplace equation between two consecutive contours, we match myocardial points through various time frames [193–195]:

$$\nabla^2 \gamma = \frac{\partial^2 \gamma}{\partial x^2} + \frac{\partial^2 \gamma}{\partial y^2} = 0 \qquad (16)$$

The estimated electric field between the inner and outer wall boundaries is $\gamma(x, y)$. In general, the solution of Eq. 16 leaded to intermediate equipotential surfaces and streamlines. These streamlines are described as being everywhere orthogonal to all equipotential surfaces and they also establish natural pixel correspondences between the boundaries. To calculate $\gamma(x, y)$, we used a 2nd order central difference technique and the iterative Jacobi method as given by:

$$\gamma^{i+1}(x, y) = \frac{1}{4} \{ \gamma^i(x + \Delta x, y) + \gamma^i(x - \Delta x, y) + \gamma^i(x, y + \Delta y) + \gamma^i(x, y - \Delta y) \} \qquad (17)$$

During the ith iteration, the estimated electric field at (x, y) is defined as $\gamma^i(x, y)$. Δx and Δy are defined as the step length or resolution in directions of x and y, respectively. Algorithm 1 demonstrates basic necessary steps for the co-allocation of pixel-wise correspondences utilizing the Laplace equation.

Algorithm 1 The solution of Laplace equation between wall boundaries (tracking over the time series)
---
1. Delineate the LV myocardium boundaries from the input cardiac data.

2. Initial condition: set the minimum and maximum potential γ at the reference boundary and the corresponding target boundary, respectively.

3. Calculate γ between both boundaries by Eq. 17 (Jacobi method) and the initial condition at Step 2.

4. Iterate Step 3 till convergence is reached (i.e., there is no change in the calculated γ values between iterations).
---

a.   Algorithm for estimating myocardial strain from cine MR data   Various heart diseases including coronary atherosclerosis can be detected by accurately estimating myocardial strain. By convention, tagged MRI data is utilized to estimate myocardial strain. Acquisition of tagged MRI is usually performed during the acquisition of cardiac cine MRI for the same subject. However, because of changes in the protocols of acquisition, estimation of strain using cardiac cine MRI from mice is proposed. Specifically, a single modality will be used for complete analysis of a subject. By adopting this strategy, we can avoid problems such as different slice locations, an unequal number of slices, and different number of frames per slice. We estimated the circumferential and radial strain.

Strain calculation from cine cardiac images depends primarily on tracking the LV wall geometry. In this chapter, we implemented a Laplace-based method to track the LV wall points. The tracking is performed by solving the Laplace equation between LV contours of each two consecutive frames over the heart cycle, as described in Algorithm 1. After tracking, the strain calculation was done utilizing a Lagrangian-based method. We used the Lagrangian strain estimation for finite small displacement to estimate the strain [196]:

$$\epsilon_L = \begin{pmatrix} \epsilon_{x_1} & \epsilon_{x_1 x_2} \\ \epsilon_{x_2 x_1} & \epsilon_{x_2} \end{pmatrix} \tag{18}$$

where ϵ_{x_1} and ϵ_{x_2} refer to the normal strain components. $\epsilon_{x_1 x_2}$ and $\epsilon_{x_2 x_1}$ refer to the shear strain components. According to the cardiac data in Fig. 51, we consider our strain cycle complete within the seventeen frames of the cine and tagged images. Therefore, our analysis is focused on the first frame to the seventeenth frame of the cardiac cycle. The basic necessary steps for estimating myocardial strain are summarized in Algorithm 2 for radial and circumferential directions.
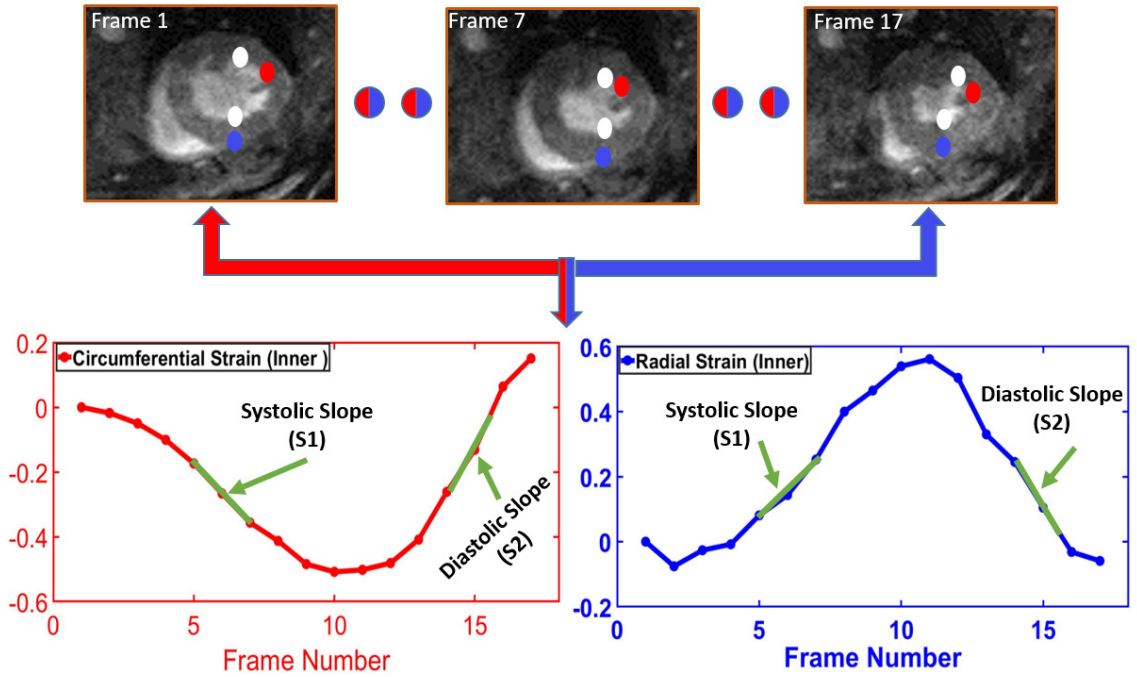
FIGURE 51: Explanation of the process of tracking throughout the heart cycle to calculate the radial and circumferential strains. The figure also shows systolic slope (S1) and diastolic slope (S2) which are two functional metrics on the circumferential and radial strain.

    b.   Algorithm for estimating myocardial strain from tagged MR data   The standard way for strain analysis is performed using tagged MRI. Therefore, strain was estimated from tagged MRI. Myocardial tagging is considered an MRI method specialized in

---
Algorithm 2 Necessary steps for estimating the circumferential and radial strain
---

1. Track the boundary at inner, mid, and outer wall along the time series images by Algorithm 1.

2. Calculate the circumferential strains by performing tracking for the fractional change in Euclidean distance between two adjacent reference correspondence points on the same boundary, i.e. inner, mid, or outer wall (refer to Fig. 51).

3. Calculate the average radial strain (between inner and mid walls, called inner strain, or between mid and outer walls, called outer strain) by performing tracking for the fractional change in Euclidean distance between the two adjacent tracked reference points along the radial direction (refer to Fig. 51).

---

the evaluation of cardiac contractile function. In this technique, the cardiac motion is captures by making a spatial arrangement of saturated magnetization in the wall of the heart at a specific time. Then, the deformation of that arrangement is captured during the cardiac cycle. When performing cardiac strain analysis, we prefer tagged MRI over conventional MRI, because the latter has no well identifiable landmarks within the heart wall. On the other hand, the tagged MRI data of mice has a very low signal-to-noise ratio and poor resolution compared to human tagged MRI. These disadvantages constitute an obstacle towards using commercially available tools to get reliable strain measurements. Therefore, we implemented a tracking technique to calculate the myocardial motion during the cardiac cycle. Valuable information can be obtained from myocardial motion and this information is used for cardiac strain assessment. We identified the tracking points of tagged MRI by determining the maximum correlation between a 3×3 window and a 5×5 window for pixel (i, j) in frame n and frame $n + 1$, respectively, as explained in Fig. 52. Furthermore, the basic necessary steps for calculating myocardial strain from tagged MRI are summarized in Algorithm 3.
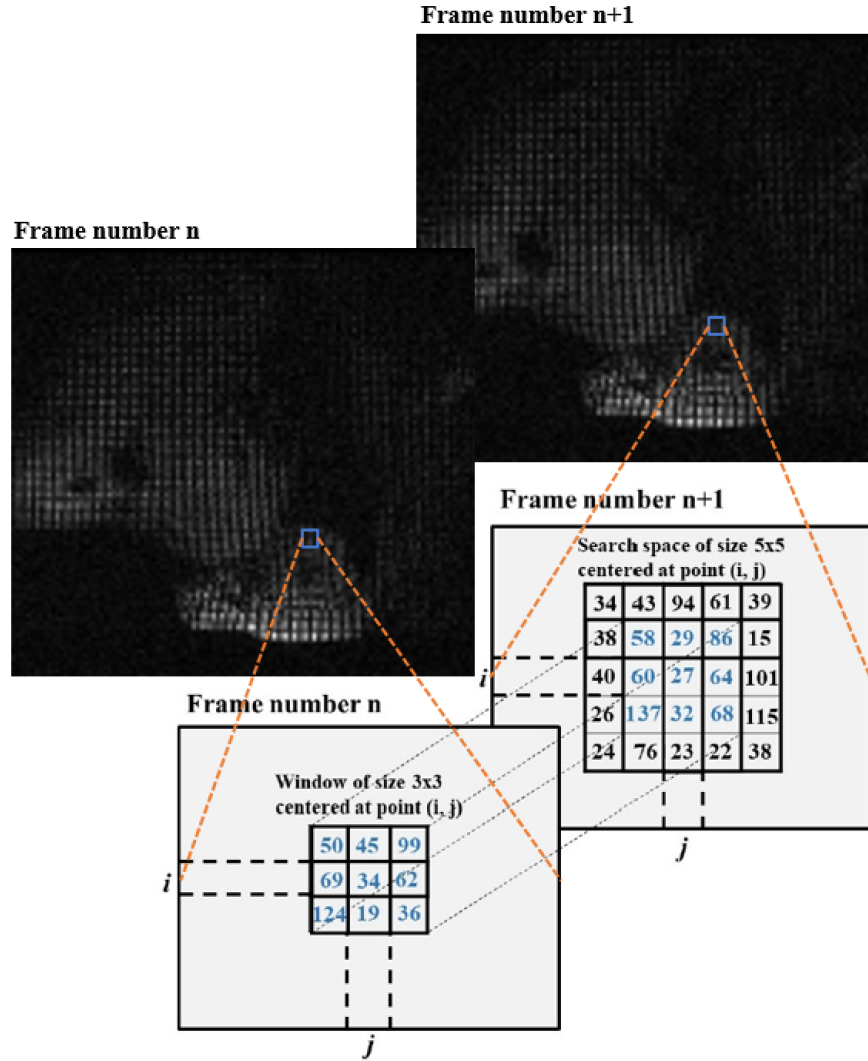
FIGURE 52: Tracking of tagged MRI pixel (i, j) in frame n by identifying the maximum correlation between a 3×3 window and a 5×5 window in frame $n + 1$.

  c. *Wall thickness and thickening*   In addition to the estimation of local strain, we estimated the wall thickness during systole and diastole. Systole leads to an increase in the thickness of the LV wall. In order to determine the wall thickness and thickening, we used the distances between the corresponding points on the inner and the outer wall boundaries. By solving Laplace equation between the segmented heart boundaries (refer to Algorithm 1), we obtain a field vector from which we estimate the co-localization of the corresponding points on the heart wall.

Algorithm 3 Strain estimation using tagged MRI
---

1. Divide the LV myocardium into 4 even sectors in the 1st frame.

2. Choose a candidate point from each sector and label it as a starting point for Performing myocardial tracking through the frames.

3. Find the maximum correlation between a 3×3 window centered at the starting point in the 1st frame, and a window in the 2nd frame that shifts in 5×5 searching space (explained in Fig. 52). By this process we perform myocardial tracking between two consecutive frames.

4. Estimate the myocardial strain from the tracked points across the cardiac cycle.

Following thickening estimation, we used a color-coded parametric map for visualization. A schematic demonstration for the estimation and visualization of the thickening is shown in Fig. 53. In addition to parametric maps, we conducted an analysis for the wall thickness utilizing a seventeen-segment model. This model dictates that the heart should be divided into 6 segments at the basal and mid-cavity levels, and should be divided into 4 segments at the apical level. At the tip of the heart, there is the last segment where no analysis is performed (refer to Fig. 54). The goal of this analysis is to find changes in myocardial function.
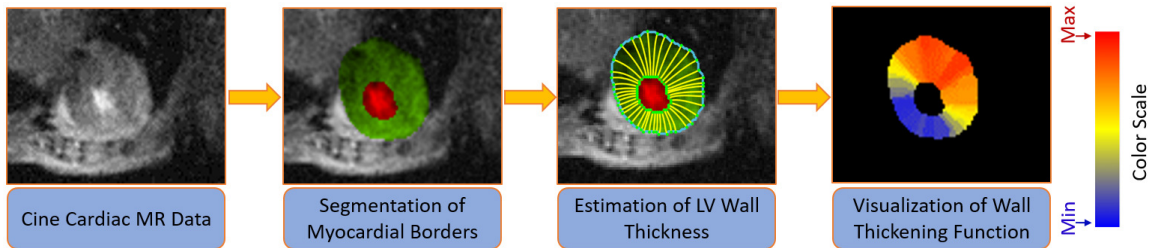
FIGURE 53: The consecutive steps for the wall thickening analysis and visualization utilizing parametric maps.
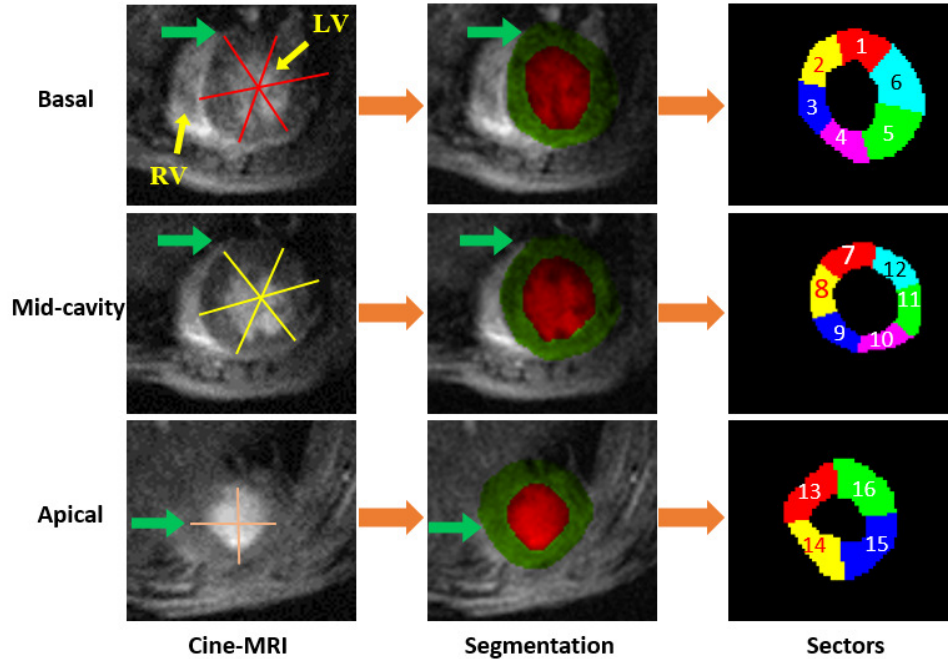
FIGURE 54: Demonstration of the seventeen segments model [3] for one case. Section division starts counterclockwise from the green arrows.

d. Global Ventriculometrics In addition to the local ventriculometrics, global volumetric parameters were estimated to reliably evaluate the heart function. Global parameters are EDV, ESV, and EF. The goal of our methodology is to provide an accurate and automated way to assess global heart function. Therefore, the parameters related to the LV function and mass [51] were estimated. Please refer to chapter 000 section 000 for the definition of these parameters.

C. Experimental Results

The main goal of this chapter is to implement a tool that is able to estimate myocardial strain from cine MRI of mice without the need for tagged MRI. Additionally, the proposed technique will allow an accurate estimation of the correlation coefficients between the strain index and other global indices obtained from cine data, such as EF, and

114

local indices (e.g., wall thickening).

1. Myocardial segmentation

Myocardial segmentation is the first step and important part in the analysis of cardiac cine data. LV myocardial boundary segmentation was performed using our proposed deep learning framework demonstrated in section V.B.2. An expert performed the manual segmentation of the LV wall borders. Therefore, we now have a ground truth for the cine data. Figure 000 shows sample results obtained from our proposed deep learning segmentation method.
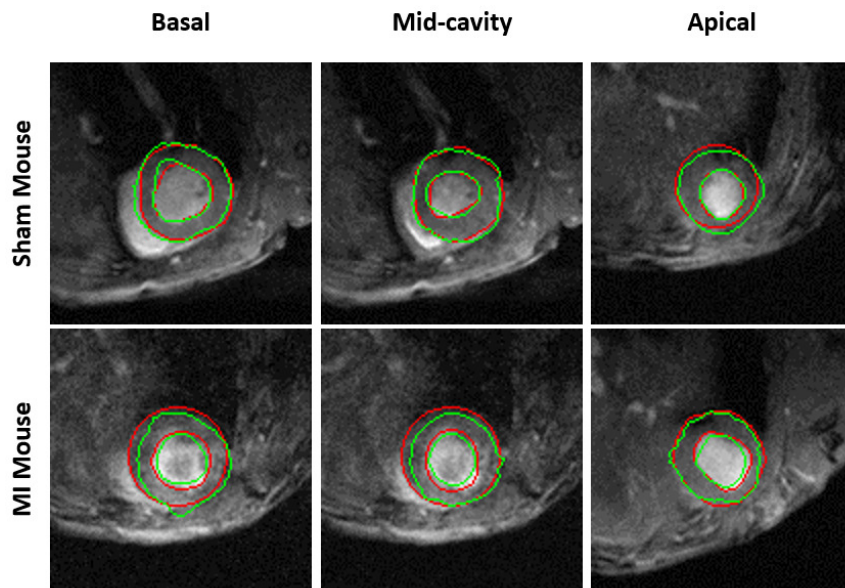
FIGURE 55: Sample LV segmentation results for subjects with sham and myocardial infarction (MI) statuses at various cross-sections of the heart. The delineation shows the LV cavity and myocardium utilizing our framework (green) and manual segmentation (red).

To assess the performance of our deep learning segmentation framework, the DSC and HD metrics were used. The DSC characterize the spatial overlap between the ground truth and the segmented region for each LV slice. The HD works on characterizing the closeness of the segmented borders to their ground truth counterparts, for each LV slice.

Table 10 summarizes the overall segmentation accuracy for the delineation of LV cavity and myocardium with respect to the manual segmentation. The table also compares between our proposed loss function and other loss functions. The table shows the mean delineation accuracies per LV slice. The optimal performance for our proposed segmentation technique was obtained when $\alpha = 0.2$ and $\beta = 0.8$.

TABLE 10: The segmentation performance of our proposed framework with the novel loss in comparison with other other techniques. The performance is characterized by the DSC and HD metrics and they are shown in the form of average$\pm$SD for the LV cavity and myocardium (Myo) of the heart.

	BCE Loss		Proposed Loss		Tversky Loss	
	Cavity	Myo	Cavity	Myo	Cavity	Myo
DSC (%)	90.02$\pm$2.11	87.33$\pm$2.22	96.45$\pm$ 1.56	93.26$\pm$ 2.21	94.55$\pm$1.41	92.32$\pm$1.74
HD (mm)	0.91$\pm$0.16	1.23$\pm$0.14	0.59 $\pm$ 0.05	0.87 $\pm$ 0.15	0.63$\pm$0.07	0.91$\pm$0.13

2. Global ventriculometrics

The quantification of cardiac function and the detection of local and global heart diseases can be accurately obtained from functional indices. A commonly used global evaluation of the heart are EDV, ESV, EF, LVM, and SV indices. To obtain these parameters, we construct LV volume curves at any time-point during the cardiac cycle. Then, we can derive relevant parameters from those curves. Figure 56 illustrates how the EDV and ESV are estimated from the LV curves. The LVM, SV, and EF can be computed using Eqs. 00 00 00, respectively. Figure 57 (a, b) shows the LV curves for all sham and MI subjects, respectively. Furthermore, the results for all functional parameters (EDV, ESV, EF, LVM, and SV) are computed and summarized as a bar plot of the mean of all cases for these parameters and their standard error for mean (SEM) bar (refer to Fig. 57 (c))

We performed the two-sample Student's t-test to compare between each functional
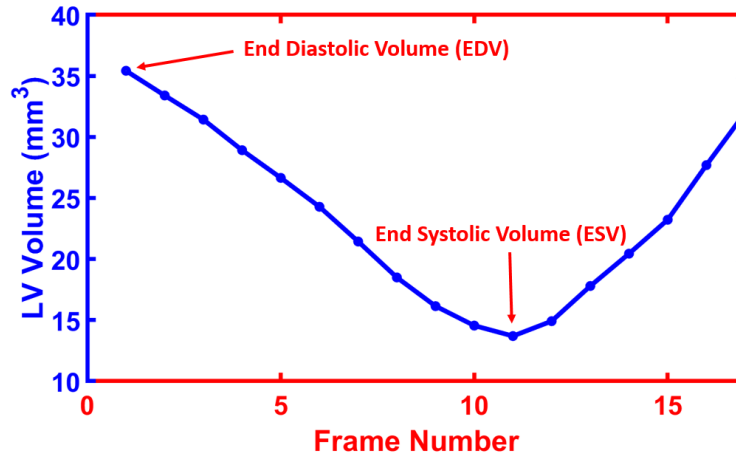
FIGURE 56: The construction of the LV curve by estimating the cavity volume across the cardiac cycle.

parameter for sham and MI groups (see Fig. 57 (c)). In the animal experiments, 3 mice per group were used. We applied log transformation, because there were high standard deviations between the two groups. From the test results, there were statistically significant differences (P-value ¡ 0.05) in EDV, ESV, and EF between the two groups. A higher EF was noticed in the sham group compared to the MI group. The values of EDV and ESV of the MI group were much higher than those of the sham group. On the other hand, there was not any statistically significant difference (P-value ¿ 0.05) when comparing the LVM and SV of the two groups. The average value of the SV of the sham group is approximately equal the average value of the MI group.

3. Wall thickening

Figure 53 demonstrates the calculation of the wall thickness of the LV from the point-to-point correspondences on the segmented cardiac images. A pixel-wise (color-coded) parametric map was used for visualization evaluation of the wall thickening. Figure 58 shows a representation example for wall thickening color-map. The maximum a posterior (MAP) estimates of a three-dimensional generalized Gauss-Markov random field

117
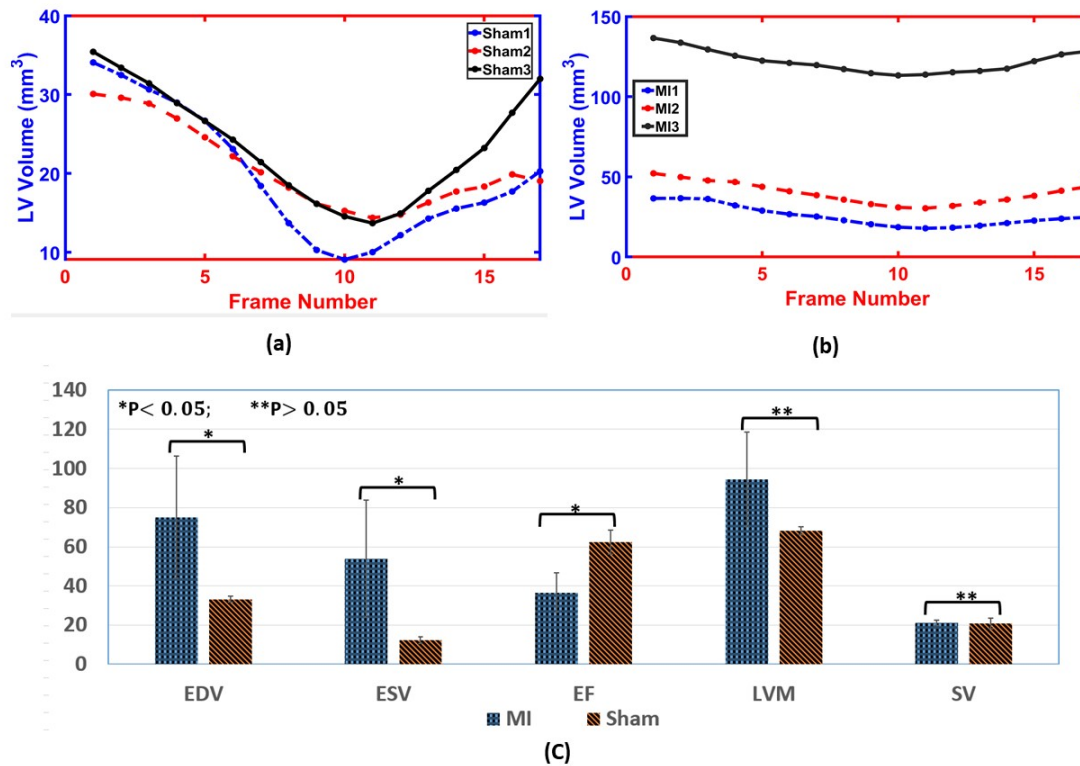
FIGURE 57: Global functional parameters for the sham and myocardial infarction (MI) cases: (a) left ventricular curves for the sham cases, (b) left ventricular curves for MI cases, and (c) a bar plot of the mean of all cases for functional parameters (EDV, ESV, EF, LVM, and SV) and their standard error for mean (SEM) bar. A t-test was performed for comparison.

(GGMRF) model [197] was employed as a continuity analysis to reduce the effect of the noisy calculations of the wall thickening measures and to preserve consistency. We did not use only the parametric-maps visualization, but also local wall thickening which can be calculated in each of the seventeen segments of the heart [3]. Table 11 presents the estimations for the seventeen segments of the LV myocardium for sham and MI groups. For further explanation for the seventeen segments, please refer to Fig. 54.
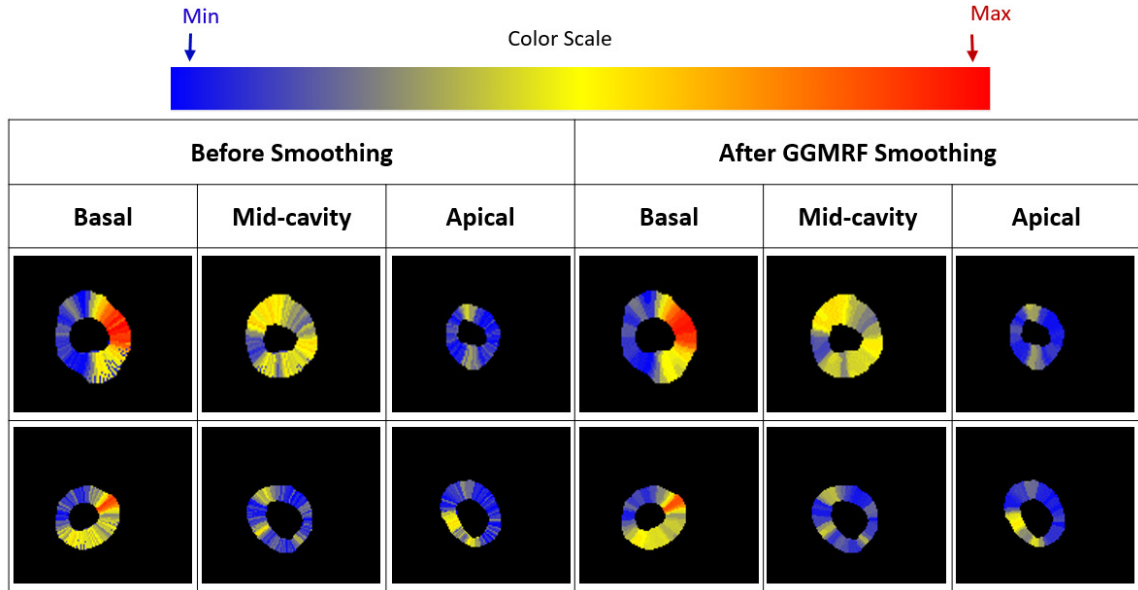
FIGURE 58: Wall thickening represented as a color-map. The first row shows the representation for a sham subject, while the second row shows the representation for a myocardial infarction (MI) subject. The maximum and minimum thickening relate to the red and the blue hues of color scale, respectively. For continuity, we modeled the estimated maps utilizing a Gauss-Markov random field (GGMRF) model.

4. A Comparison between myocardial strain resulted from tagged MRI images vs. cine

Important insights regarding cardiac function can be obtained from strain analysis. A geometrical-based method was used for tracking the myocardial boundary to calculate functional strain from cine MRI. We used Algorithm 2 to calculate the strain (both circumferential and radial strains) and generate strain curves from cine images. The circumferential and radial strains can be estimated by a tracking process through the cardiac cycle, shown in Fig. 51. To estimate the strain from tagged MRI images, Algorithm 3 was used. For cine and tagged MRI images, we present in Fig. 59 the estimated average radial strain curves for each cardiac cross-section (i.e., basal, mid-cavity, apical) for the two animal categories (i.e., sham and MI). From Fig. 59, a similar trend and a good agreement can be noticed from the cine and tagged MRI strain curves. It is worth noting that the end of

119

TABLE 11: Wall thickening estimations for all cases utilizing the standardized myocardial seventeen-segments model [3].

	Section Number	Sham 1	Sham 2	Shame 3	MI 1	MI 2	MI 3
Normalized wall thickening	1	0.63	0.35	0.44	0.58	0.27	0.63
	2	0.43	0.35	0.44	0.58	0.27	0.43
	3	0.21	0.35	0.44	0.58	0.27	0.21
	4	0.26	0.35	0.44	0.58	0.27	0.26
	5	0.41	0.35	0.44	0.58	0.27	0.41
	6	0.64	0.35	0.44	0.58	0.27	0.64
	7	0.15	0.82	0.65	0.44	0.46	0.15
	8	0.14	0.63	0.43	0.25	0.39	0.14
	9	0.22	0.21	0.33	0.36	0.24	0.22
	10	0.26	0.18	0.49	0.43	0.17	0.26
	11	0.20	0.31	0.54	0.38	0.08	0.20
	12	0.25	0.62	0.44	0.46	0.23	0.25
	13	0.16	0.60	0.15	0.27	0.14	0.16
	14	0.16	0.60	0.15	0.27	0.14	0.16
	15	0.16	0.60	0.15	0.27	0.14	0.16
	16	0.16	0.60	0.15	0.27	0.14	0.16

the cardiac cycle is not shown in the figure and we showed up to frame seventeen out of twenty-two frames. This is because of the existence of artifacts in the tagged MRI images at the end of the heart cycle. This problems occur because of the small size of the heart of the mice and the incompatibility associated with the diamond structure of tagged MRI images.

Besides the visual comparison, a statistical metric and Bland-Altman plots were used to evaluate curves agreement. First, we estimate the mean squared error (MSE) and the correlation coefficient (CorCoef) between the two curves. The average CoreCoef and MSE values are shown in Table 12. These values compare the strain curves estimated from Cine MRI and tagged MRI for both the two animal groups at different cross-sections. As shown in the Table, a high CorCoef was obtained with a range between 0.737 and 0.966 for all the animal groups. Additionally, a low MSE was obtained with a range between 0.024 and 0.08. As shown in Fig. 59, the same trend exists between strain curves from cine MRI and
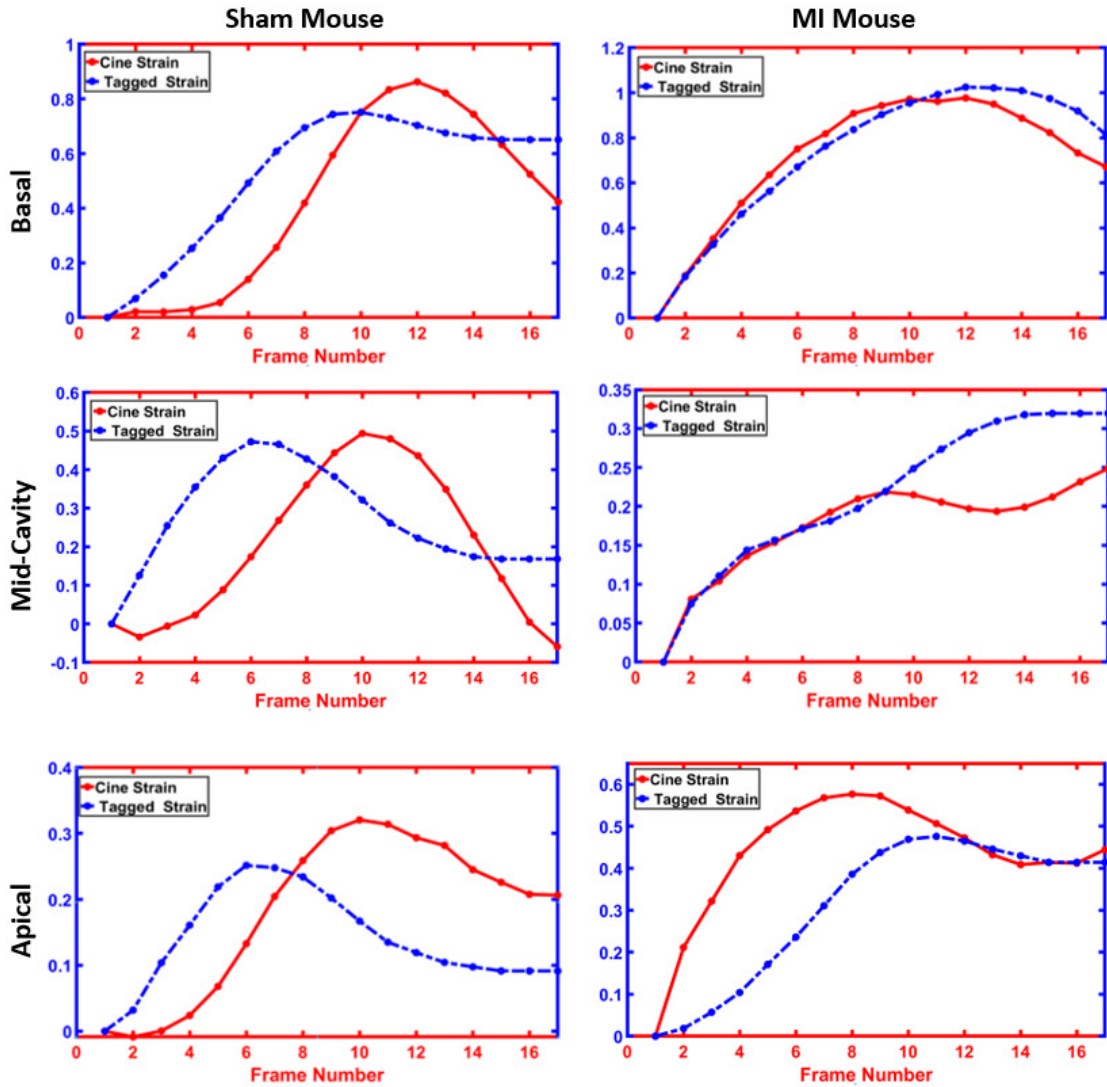
FIGURE 59: Visualization of the constructed average strain curves from cine images (radial strain) and the tagged MRI images. The rows show the strain at different cross-sections. The columns represent different animal groups (i.e., sham and MI).

tagged MRI, which is aligned with the quantitative results in Table 12. The Bland-Altman analysis is a good tool to assess measurements agreement. Therefore, we constructed the Bland-Altman plots to validate the agreement of the curves. We performed this analysis for the slope of diastolic phase (S1), the slope of diastolic phase (S2), and the peak strain value for the calculated strain curves. As demonstrated in Fig. 60, there are excellent

agreements for the estimated metrics for both the animal groups. In the Bland-Altman analysis, we must verify the normality of the differences. Therefore, the Shapiro-Francia test was applied to check for normality of the differences with 5% significance level. The P-values were (0.0914, 0.9767), (0.9434, 0.7446), and (0.0609, 0.1241) for S1 (sham, MI), S2 (sham, MI), and peak (sham, MI), respectively. We obtained P-values which are larger than 0.05, therefore, the test accepted normality. Furthermore, data normality were accepted using DAgostino Pearson, Jarque-Berat, and Kolmogorov-Smirnov [198] tests.

TABLE 12: A quantitative evaluation for the degree of agreement between the strain calculated from cine MRI (radial) and tagged MRI for both animal groups. CorCoef and MSE refer to the correlation coefficient and mean squared error, respectively.

	Basal		Mid-Cavity		Apical	
	CorCoef	MSE	CorCoef	MSE	CorCoef	MSE
Sham	0.909	0.028	0.966	0.023	0.812	0.039
MI	0.920	0.025	0.737	0.033	0.783	0.079

a.   Circumferential strain from cine MRI   Algorithm 2 was used for the estimation of the circumferential strain and the construction of strain curves. Figure 61 shows an example for the calculation of the circumferential strain for a sham and an MI case. The strain was calculated for all cases for the inner, mid and outer walls at different cardiac cross-section. As we can notice from the figure, the strain curves of the sham group differ from the MI group. A reduced strain value is manifested in the MI group particularly at the outer wall. Furthermore, different locations of the heart show different circumferential strain values. For instance, a higher range was obtained for the inner wall in comparison with the mid and outer walls. We performed quantitative analysis for the resulted strain curves by the calculating S1 and S2 for the curves as shown in Fig. 51. Figure 62 shows a bar plot for the mean and SEM of S1 and S2 for the basal, mid-cavity, apical cross-sections for the sham and MI groups. Additionally, the two sample student's t-test was used to
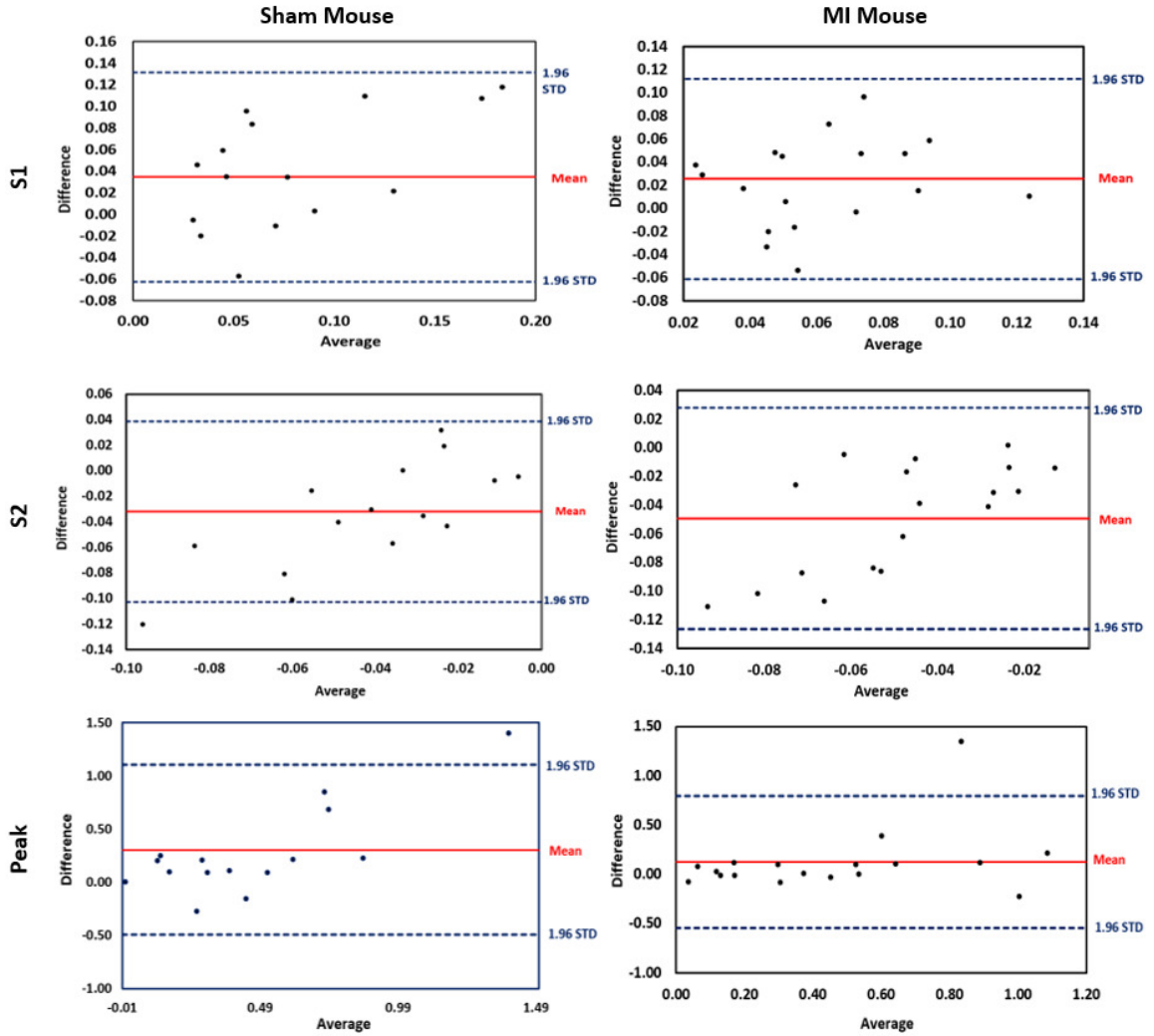
FIGURE 60: The Bland-Altman analysis for the systolic phase (S1), the diastolic phase (S2), and the peak value of the strain curves from the cine and tagged MRI images. The analysis was performed for the two animal groups. Note that STD refer to standard deviation.

perform statistical analysis for each cross-section, see Fig. 62. Three slices were included for the basal and mid-cavity sections, and only two were included in the apical section. Then, we obtain the average for each section for each case. These data were utilized to investigate the group difference between the two animal groups. To make measurement variable follow a normal distribution with homogeneous data, we used log transformation.

As shown in Fig. 62, all sections showed a statistically significant difference (P-value $<$ 0.05) between the sham and the MI groups.
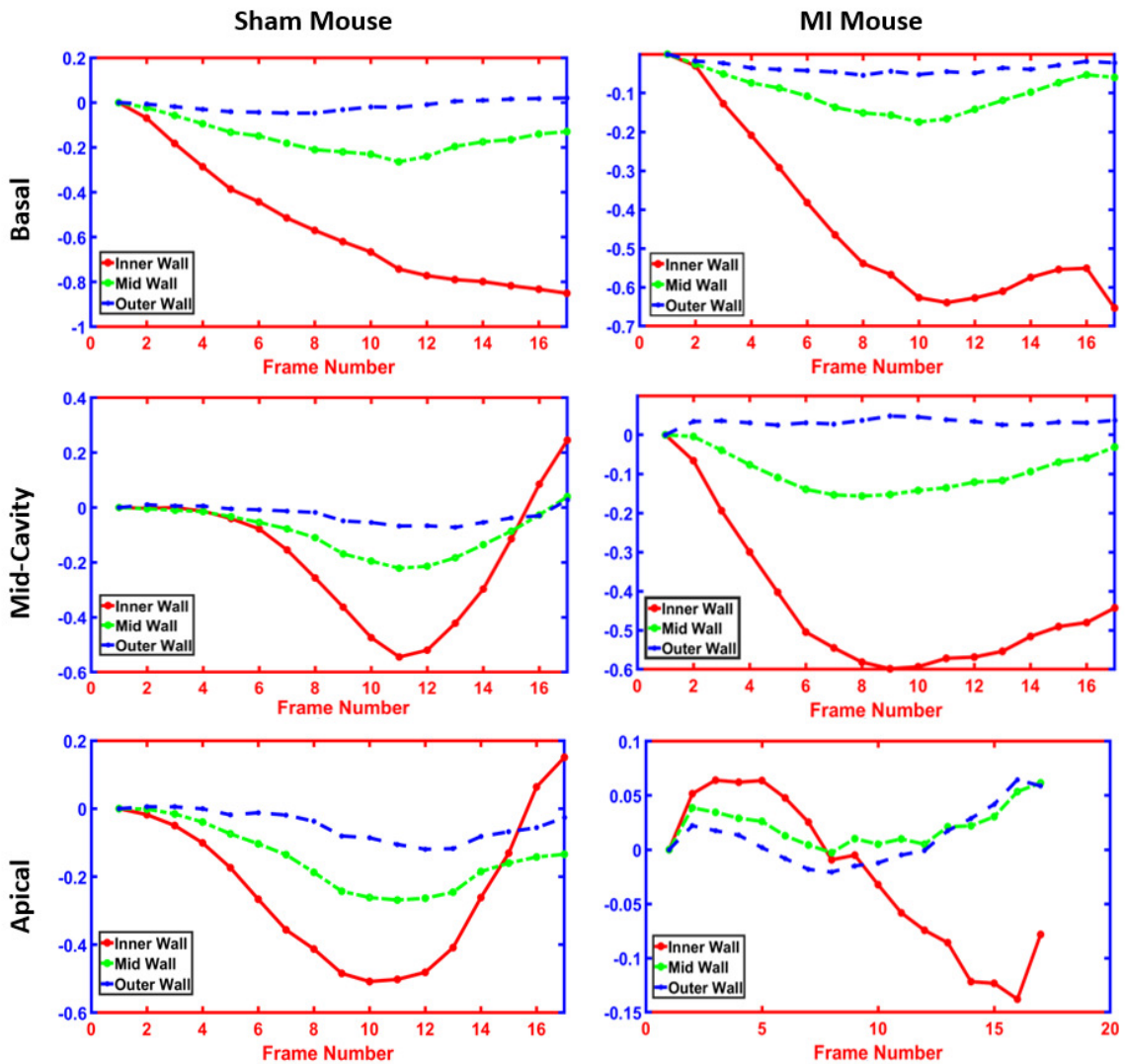
FIGURE 61: The calculated circumferential strain of the LV for the inner, mid and outer walls (at different slices) for a sham case (left) and a myocardial infarction (MI) case (right).

b. Radial strain from cine MRI We estimated the radial strain for the inner and the outer walls, besides the circumferential strain. The outer radial strain is estimated between the mid and outer walls , while the inner radial strain is estimated between the inner and mid walls. Fig. 63 shows the radial strain results for two cases at different cross-sections. A high level of variation exists between the inner and the outer radial strains.
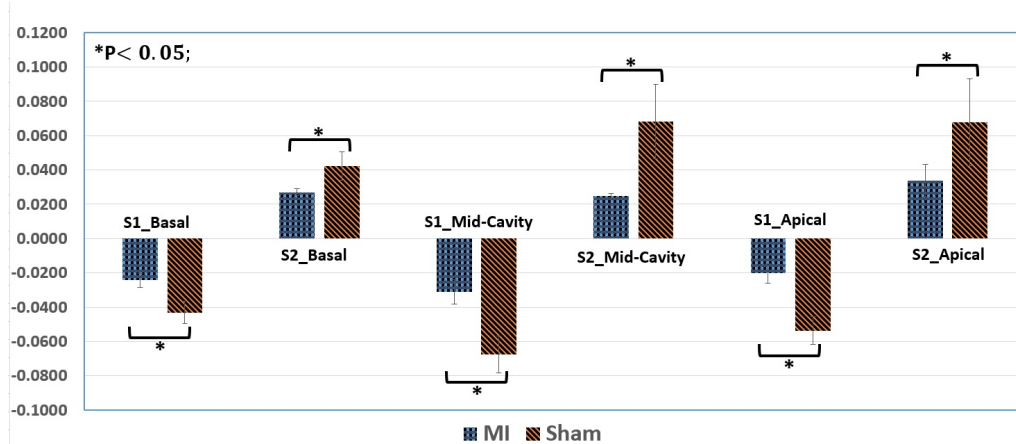
FIGURE 62: A bar plot representation for the mean and standard error of the mean (SEM) of systolic slope (S1) and diastolic slope (S2) for the circumferential strain calculated for the three cross-sections (basal, mid-cavity, apical) to both animal groups (sham and my-ocardial infarction (MI)). The student's t-test was used to obtain the P-values.

Approximately, the value of the outer strain is zero at all the frames. on the other hand the inner radial strain exhibits convex curves. Additionally, a bar plot for the mean and SEM of S1 and S2 was plotted for the radial strain for the basal, mid-cavity, and apical cross-sections for both animal groups, see Fig. 64. We performed our the statistical analysis of this part in a similar way such as the analysis of the circumferential strain in Section V.C.4.a. There are a statistically significant differences (P-value < 0.05) between both animal groups for all section except the apical section for slope S2.

    c.   *Sector strain*   Besides the circumferential and the radial stain curves, we can estimate local strain in each of the seventeen segments represented in Fig. 54. Fig. 65 shows the strain for sixteen-sectors at the inner wall for an MI and a sham cases. The mid-cavity and basal levels contain 6 curves, with a curve for each sector, and apical level contains 4 curves. Fig. 66 shows the strain for sixteen-sectors at the outer wall for an MI and a sham cases.

    d.   *Validation for Strain estimation*   Two validation experiments were performed to validate our proposed methods for strain estimation using real and synthetic data. Syn-
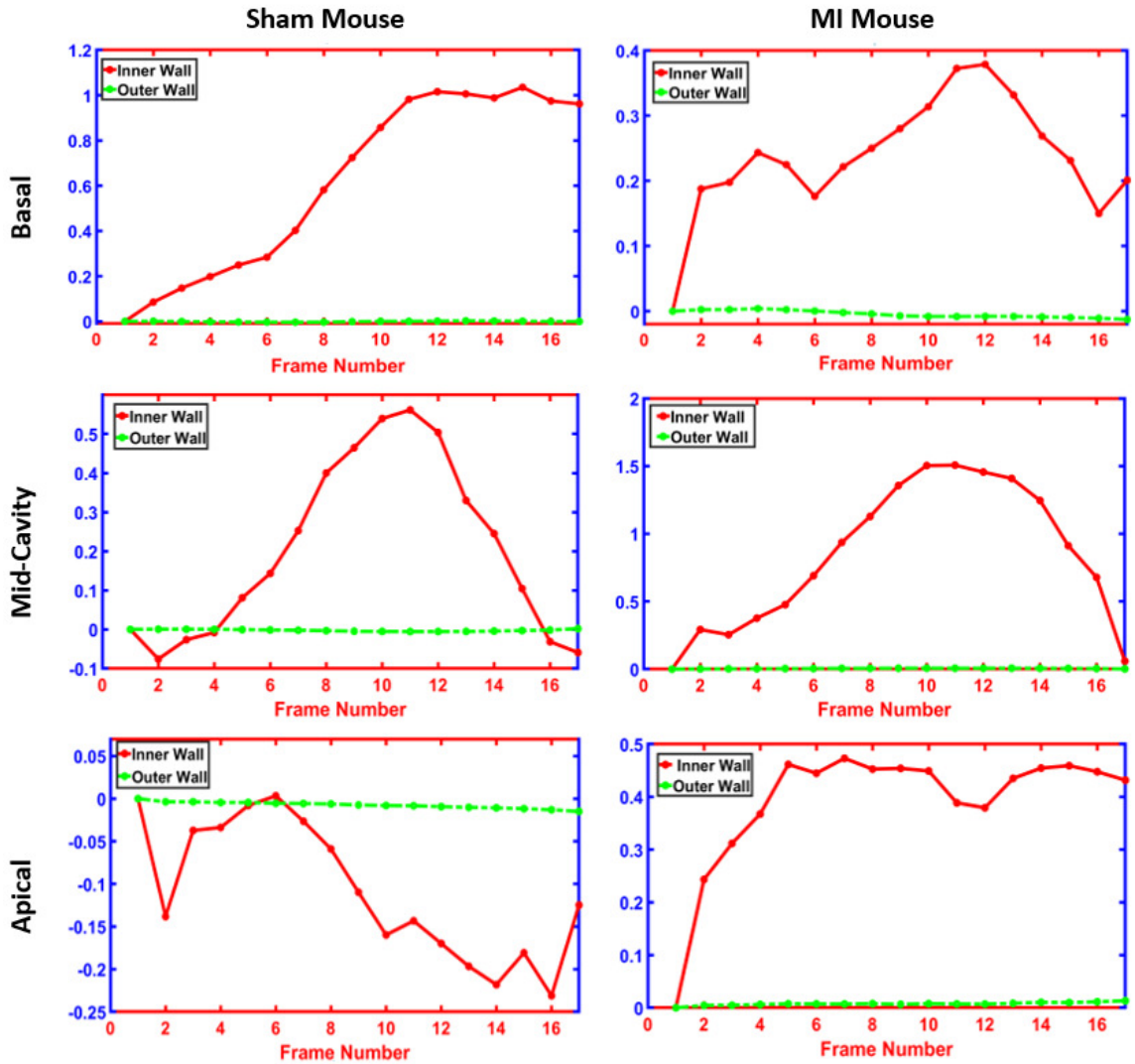
FIGURE 63: The calculated inner radial strain and outer radial strain of the LV at different cross-sections for a myocardial infarction (MI) case (right) and a sham case (left).

thetic phantom images were used to validate the strain calculation from cine MRI. The LV response and the physiological features during the cardiac cycle affect the phantom model. To represent the LV motion, a geometric transformations that include translation, shearing, torsion, rotation, and compression were utilized. Using these transformations, we map each location in the LV model to corresponding spatial point at a specific time instant [199,200]. An inverse motion map was generated utilizing these transformations and the map is esti-
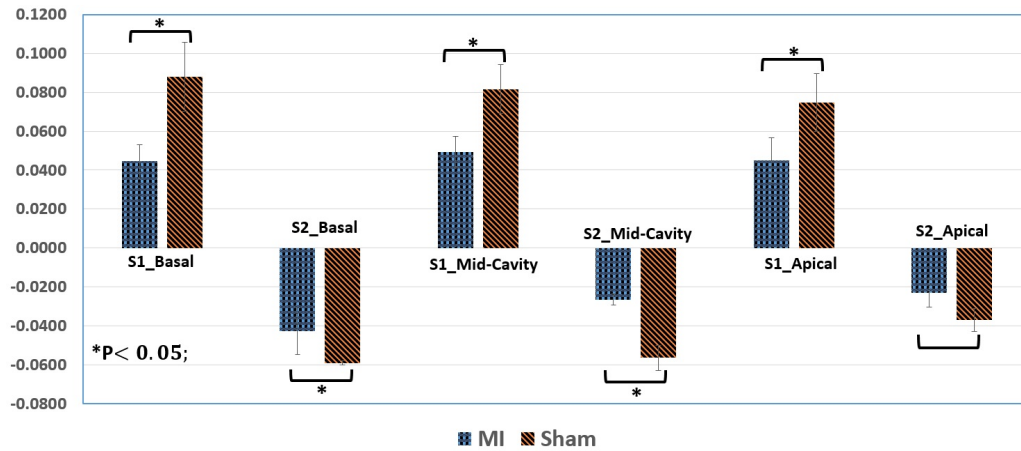
FIGURE 64: A bar plot representation for the mean and standard error of the mean (SEM) of systolic slope (S1) and diastolic slope (S2) for the radial strain calculated for the three cross-sections (basal, mid-cavity, apical) to both animal groups (sham and myocardial infarction (MI)). The student's t-test was used to obtain the P-values.

mated analytically. Therefore, we can establish correspondences between 2 points at any 2 time instants.

Figure 67 compares the calculated strain of the mid-wall from our developed method with the ground truth strain. Both the original and the normalized strain values are illustrated in the figure. The strain curves from the two methods show a good agreement. Furthermore, the agreement was confirmed by statistical metrics such as CorCoef and the MSE. There was a high correlation (CorCoef = 0.89) between the estimated strain and the ground truth, and there was a low MSE (2.7%). This validates the usefulness of our proposed strain estimation method.

Additionally, we also validated the strain estimation method from tagged MRI. Three mice subjects were used to validate our strain estimation method from tagged MRI. Myocardial points in the tagged MRI were chosen and an expert manually tracked these points over the cardiac cycle for all image frames of the three subjects. Then, we constructed the strain curve from the tracked points. Then, our strain estimation method was applied on the tagged MRI data using the chosen points of the first frame. Our method
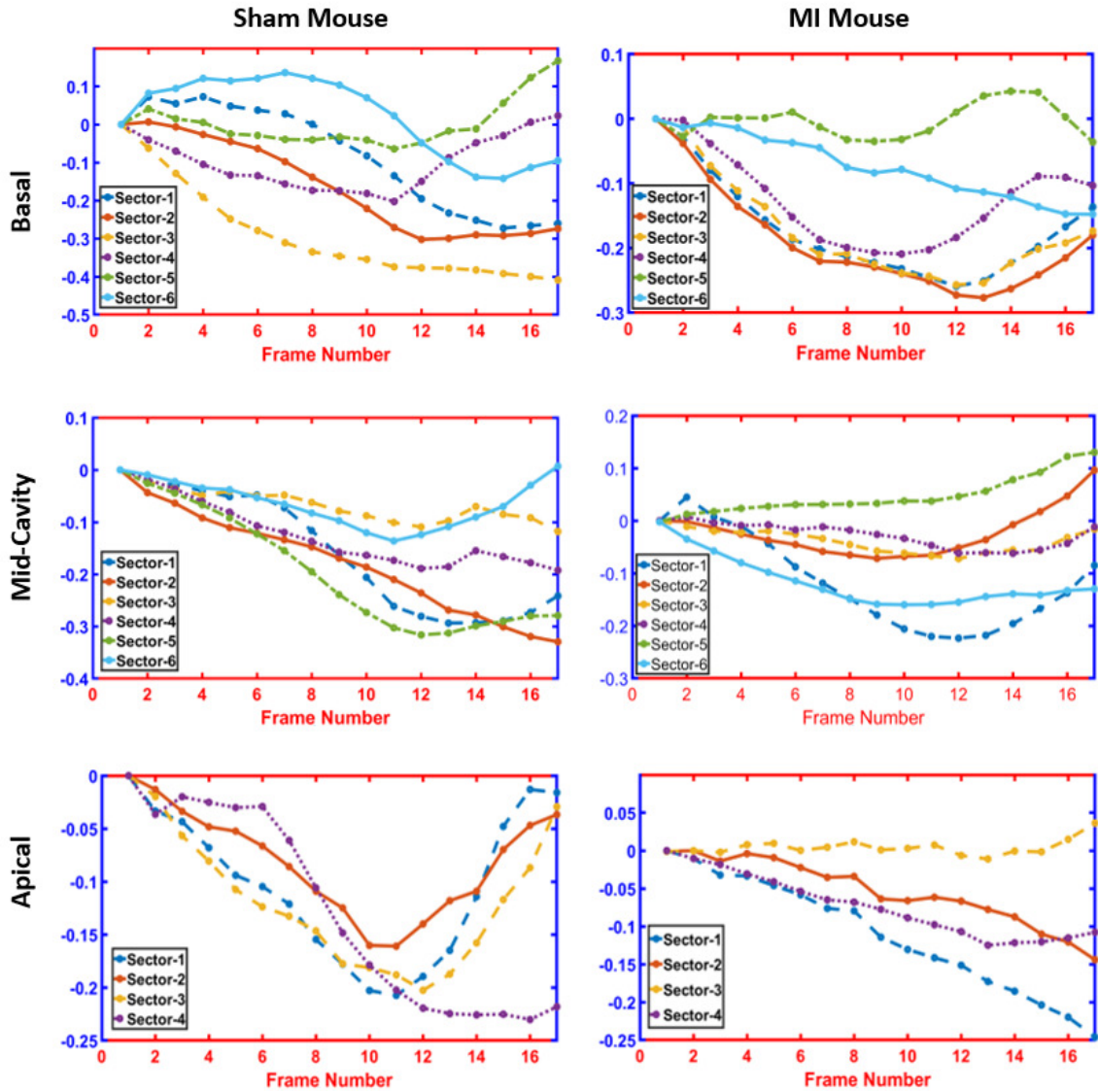
FIGURE 65: Examples for sector-strains calculated at the inner wall for myocardial infarc-
tion (MI) and sham cases.

resulted in a strain curve. We performed quantitative comparison between the results of
the two methods. The CorCoef was 0.941 and the MSE was 15e-4. Furthermore, Fig. 68
demonstrates the visual comparison between the curves resulted from the two methods for
a mouse subject.

D. Discussion
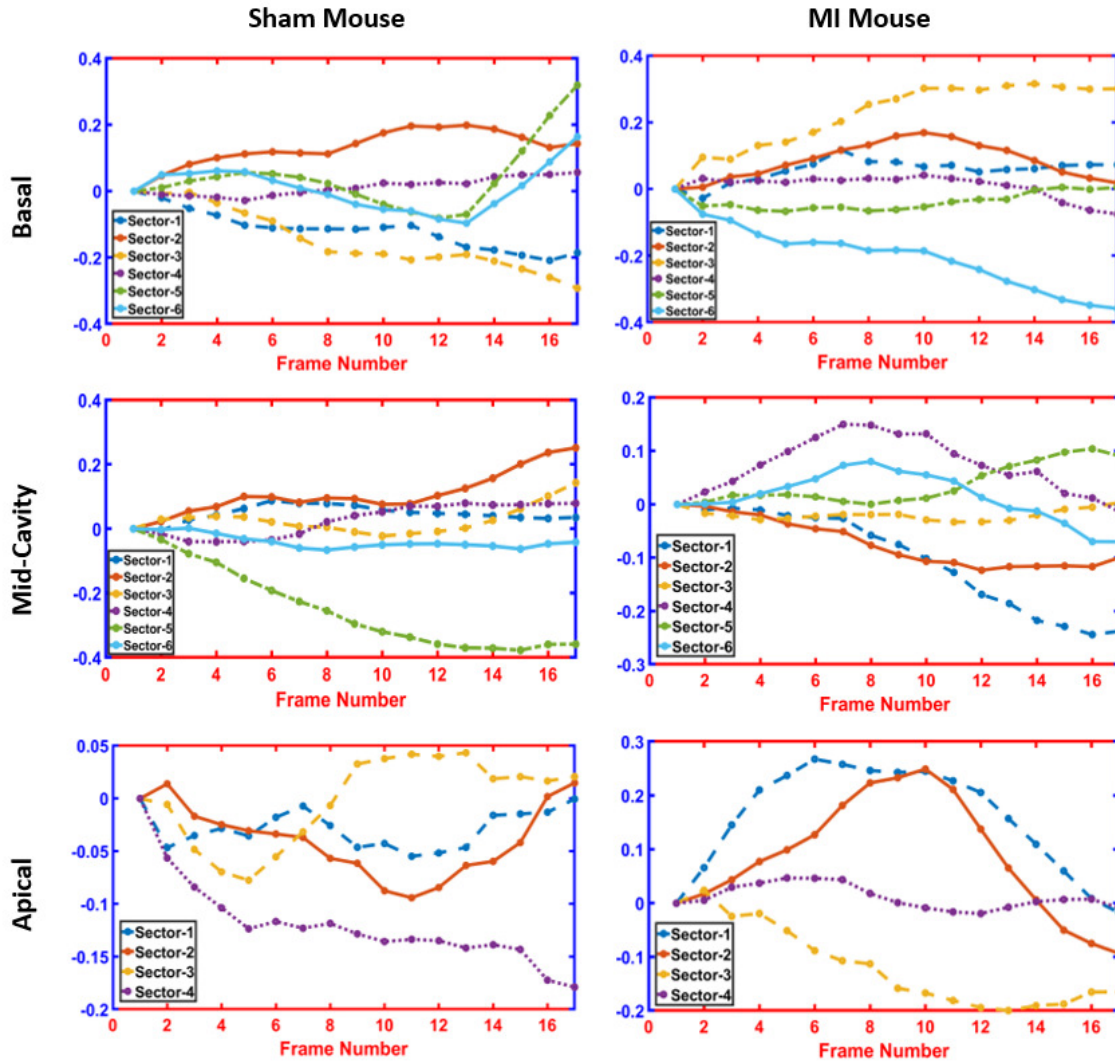
FIGURE 66: Examples for sector-strains calculated at the outer wall for myocardial infarction (MI) and sham cases.

An automated and accurate algorithm for the evaluation of global LV function, LV structure, and LV localized strain was developed utilizing cine MRI in mice. Our experiments showed the potential of our technique to efficiently work on mice hearts, regardless of their tiny sizes. Recent studies utilized cine MRI for estimating strain for human hearts. Tommaso MAns et al. [201] used the diffeomorphic demons which is non-linear registration algorithm to estimate 3D myocardium strain from cine images by inserting 2 physical constraints. By constraining the deformations to be divergence-free, Myocardium near in-
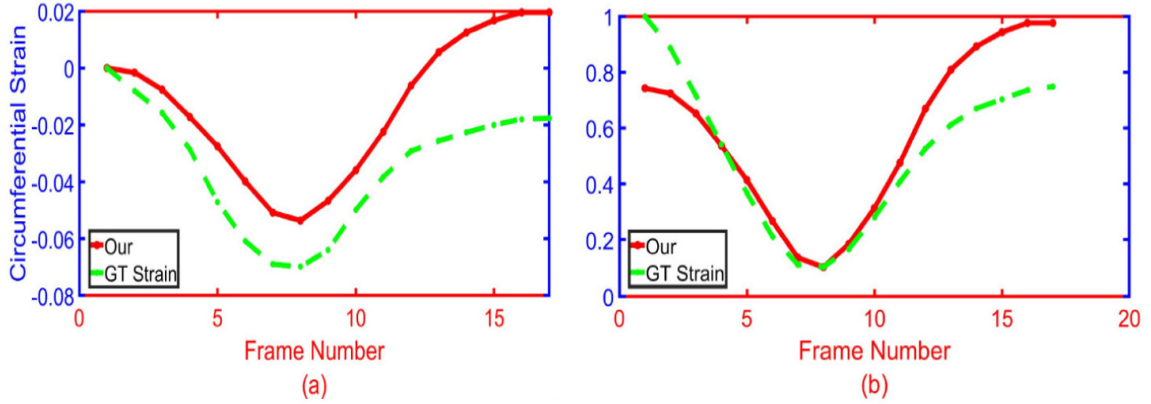
FIGURE 67: A comparison between our estimated strain and the ground truth strain (resulted from the phantom's deformation). (a) demonstrates the comparison with the original values of the strain (raw data). (b) demonstrates the comparison with the normalized strain values (between zero and one).
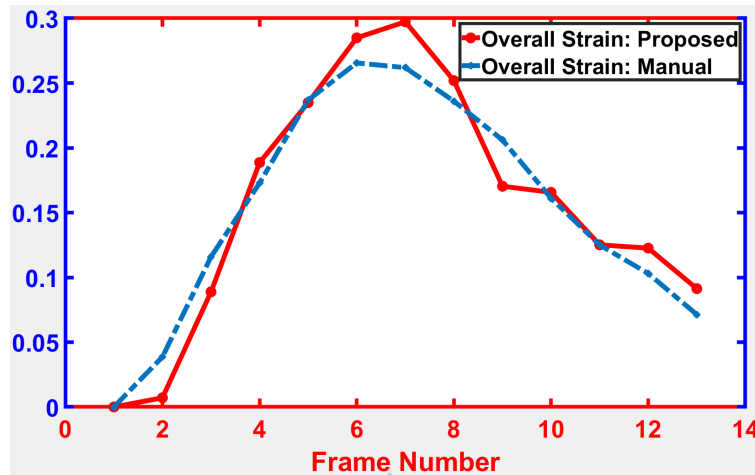
FIGURE 68: A visual comparison between the strain estimated by automated tracking of tagged MRI (our method) and the strain estimated from manual tracking by an expert.

compressibility was ensured. They also use smooth vector filters to model myocardium elasticity. Hao Gao et al. [202] used a b-spline deformable image registration technique to estimate myocardial strain from cine images. They determined both LV strain, structure, and both regional and global function. The advantages of the proposed method are the reduction of the cost and MRI scanning time. It also estimates the myocardial indices

reliably. The novel algorithms proposed in this chapter allow the opportunity to reliably analyze myocardial strain in the mice heart. We can extend our application to estimate myocardial strain in humans.

Different anatomical structures can be segmented efficiently using FCN architecture. In this chapter, the application of FCN in segmenting the LV from cine MRI of mice (both sham and MI) was proposed. As demonstrated in Table 10, our method resulted in accurate LV segmentation in terms of DSC and HD, which are popular segmentation performance metrics. Furthermore, our results in Fig. 55 indicate that our method achieved good segmentation results at different LV slices, including the slices where the LV occupies a small area in the image. These slices are at the apical part of LV which are difficult to segment. The LV tissue occupies a smaller area compared with other surrounding tissues in cine MRI. This problem makes the deep learning model biased towards the majority category (background tissues). To solve this problem, we developed a novel loss function that accounts for this category-imbalance and penalizes the action of the cross-entropy loss which is susceptible to category-imbalance problem. The potential of our novel loss function is established by providing good segmentation. As shown in Table 10, in comparison with the BCE loss (where the issue of category-imbalance is more severe), better results were obtained from our loss function. Furthermore, the results obtained from our loss function is better than the results of Tversky loss that was presented to overcome category-imbalance. Besides our novel loss, data augmentation is another important factor that enhanced the performance of our approach. We obtained a higher sample size by adopting data augmentation, and thus the over-fitting is reduced. Human cine MRI is the main focus for the majority of the recent heart image analysis works [153]. The human cine MRI has better contrast and higher signal-to-noise (SNR) ratio than mice cardiac MRI. Therefore, the task of segmenting the LV is easier in human cardiac MRI. Due to the fact that mice MRI has low SNR, we can conclude that the prime merit of our work is the application of convolutional neural networks on mice data to obtain preclinical and accurate tool for heart

segmentation.

To fully characterize cardiac function and structure, local (regional) and global indices are important to estimate. EF, SV, and LVM are global indices, while myocardial strain, wall thickness, and thickening are local indices. To estimate cardiac structure and function indices, different algorithms that are based on the Laplace equation were used. We solved the Laplace equation between wall boundaries and this enabled us to co-allocate corresponding pairs on the boundaries to calculate the wall thickness and thickening. Previous studies focused on radial-based or center-line-based methods [203]. These approaches are not suitable for noisy images and result in high errors because of the lack of strong edges. Furthermore, these approaches assume heart circular symmetry, and this assumption can increase the error. To obtain a high accuracy for myocardial points co-localization, Algorithm 1 was adopted as it depends on geometric features, refer to V.C.3.

Myocardial strain is another important local index, that is usually obtained from tagged MRI to evaluate the contractile function of the heart. However, in the cases of rapid motion between consecutive image frames or large displacement, the methods that are based on tagged MRI fail [204]. Due to the fact that strain is not the only parameter that quantifies the functionality and workings of the heart, other performance parameters are required (e.g., wall thickness, EF) and are obtained from cine cardiac MRI. The aim of our study is to implement and develop a framework that is able to calculate the myocardial strain and other cardiac parameters from a single modality (i.e., cine cardiac MRI). This approach has advantages of avoiding the inter-slice variability and producing better correlated parameters for heart function. Specifically, we used the solution of Laplace equation to track myocardial points over the heart cycle between consecutive frames. The proposed approach demonstrated the capability of tracking the movement and rotation of the heart. Therefore, our approach puts a limit for the effect of the noise that comes from cardiac motion and enables accurate strain calculation utilizing the Lagrangian-based method (refer to Algorithm 2). Our results and statistical analyses in Figs. 59, 60 and Table 12, demon-

132

strated that we can obviate the need of tagged imaging for strain calculation by using cine MRI. In this chapter, we addressed the ability of our approach to assess strain from cine cardiac MRI to avoid performing tagged MRI for the small mice hearts. There are high consistency between myocardial strain obtained from tagged MRI and cine MRI.

One of the limitations of our work is the small size of our cohort (n=6). Despite the small size of subjects, our developed methodology demonstrated the ability of accurately estimating strain and other heart functional parameters from a single cardiac MRI modality (cine MRI). It is important to note that, for the MI group, we intentionally induced different degrees of MI, to assess the capability of our algorithm to detect the changes in the myocardial strain in different degrees of heart failure. In our MI cohort, MI-1 has more serious heart failure and dilation in comparison with the remaining two animals (refer to Fig. 57). In future, we plan to adopt a larger cohort size to test and validate our algorithm potential. In spite of these limitations, our methodology demonstrated the feasibility of accurately calculating the strain indices, which might obviate the need for tagged MRI clinically.

CHAPTER VI
A PYRAMIDAL DEEP LEARNING PIPELINE FOR KIDNEY WHOLE-SLIDE HISTOLOGY IMAGES CLASSIFICATION

Histopathological images can provide histological features that are important for disease diagnosis and these images are considered the gold standard for cancer diagnosis and recognition. The manual analysis of histopathological images takes time and effort from the pathologist. On the other hand, automated analysis helps the pathologist diagnose the disease and alleviates his work load. Renal cell carcinoma is the most prevalent kidney cancer. Renal Cell carcinoma has various subtypes with different clinical behavior. In this work, we demonstrate how a deep learning approach can be utilized for the automated classification of kidney histopathological images. Our approach can classify four classes: the fat, the parenchyma, the clear cell renal cell carcinoma, and the unusual cancer which has been discovered recently, called clear cell papillary renal cell carcinoma. Our framework consists of three convolutional neural networks and the whole-slide kidney images were divided into patches with three different sizes to be inputted to the networks. Our approach can provide patch-wise and pixel-wise classification. The kidney histology images consist of 44 image slides that belong to the four mentioned tissue types. Our framework results in an image map that classify the slide image on the pixel-level. Furthermore, we applied Gauss-Markov random field smoothing to maintain consistency in the map. Our approach classified the four classes accurately and surpassed other state-of-the-art methods such as ResNet (pixel accuracy: 0.89 Resnet18, 0.93 proposed). We conclude that deep learning has the potential of alleviating the work-load on the pathologist by providing accurate automated classification for histopathological images.

A. Introduction

Kidney cancer is an abnormal cell growth in the kidney. Kidney cancer is ranked fourteenth among common cancers and contributes about 2.4% of the total number of diagnosed cancers [65]. A recent estimate by the American Cancer Society revealed that in the United States about 76,080 new patients will be diagnosed with kidney cancer and about 13,780 patients will die from the disease [66]. Cancers of the kidney are of several types, renal cell carcinoma (RCC) being the most prevalent, accounting for nearly 85% of kidney cancers. RCC is itself a heterogeneous set of cancers with distinct molecular characteristics, responses to treatment, and clinical outcomes [67]. The most prevalent subtypes of RCC include clear cell (70%—80%), papillary (14%-–17%), chromophobe (4%—8%), and clear cell papillary RCC (4%) [68]. We can classify RCC subtypes by utilizing the morphological features observed on histopathological slides stained with hematoxylin and eosin. However, there is a significant morphological overlap between clear cell RCC and clear cell papillary RCC subtypes due to the presence of clear cells. It is critical to be able to classify between clear cell RCC and clear cell papillary RCC to determine the correct patient management. Clear cell RCC is associated with poor prognosis because it is likely to metastasize, while clear cell papillary RCC is considered an indolent neoplasm, which is not associated with metastatic spread [69]. When there is a significant overlap in morphological features, immunohistochemistry helps in separating the two subtypes of the tumor. However, we cannot find immunohistochemistry in all areas of the world. We can use deep learning algorithms to differentiate between the types of tumor by analyzing histopathological images.

Recently, deep learning algorithms have achieved outstanding success in various fields. Convolutional neural networks (CNNs) are deep learning models that are excellent at processing 2D and 3D images with great success in object detection [205], image recognition [128,167,206], and image segmentation [10,207]. Particularly, there have been many applications of CNNs in the domain of histopathological image analysis. Cruz-Roa

et al. [208] presented a new method that is based on CNN for the detection of invasive breast cancer in whole-slide images. Their analysis involved various sites and scanners. They assessed their approach for classifying breast cancer without direct comparison with pathologists. Han et al. [209] proposed a technique for the multi-way classification of breast cancer using histopathological images. Their technique was based on structured learning and achieved a high accuracy of 93.2%. They used pre-trained models that use imagenet database. Coudray et al. [210] trained a CNN called inception v3 to classify between normal lung tissue, adenocarcinoma, and squamous cell carcinoma using whole-slide images. They achieved an accuracy that is comparable with the pathologist. However, they did not study less common lung cancers such as large-cell carcinoma and small-cell lung cancer. Wang et al. [211] implemented a CNN to recognize tumor regions in lung pathology images. They extracted features from tumor regions and associated them with the survival outcome of the patients. However, their CNN model is sensitive to out-of-focus tissue such as macrophages, red blood cells, and stroma cells. Khosravi et al. [212] demonstrated the use of deep learning techniques to identify cancer subtypes such as lung cancer. However, there was discordance between pathology results and their findings due to the small number of the used tumor images. Jimenez–del–Toro et al. [213] used CNN to classify prostate histology from whole-slide images. Their method was evaluated on a limited settings as a binary classification problem on a limited set of Gleason scores. Karimi et al. [214] proposed a CNN based method for the automated grading of prostate cancer using a limited number of pathology images. They implemented a new data augmentation strategy to improve the classification accuracy. However, their method has a low classification accuracy when differentiating between grade 4 and grade 5 tumors. Tabibu et al. [215] implemented a deep learning approach based on Resnet to classify between three subtypes of RCC. However, they did not study clear cell papillary RCC. As we can see, there has been a limited work in the kidney histopathological image analysis.

We propose a computer aided diagnostic system that uses CNNs for the automated

classification of kidney tissues and RCC. Our framework can partition histology images into four classes including fat, parenchyma, clear cell RCC, and clear cell papillary RCC. Our framework has the following contributions:

- The first study to discriminate between clear cell RCC and clear cell papillary RCC. This classification has high clinical relevance.

- We propose a pyramidal deep learning model that utilizes a hierarchy of three CNNs that process different image sizes. The deep learning improves the precision of the diagnosis and decrease human error. Furthermore, it produces reproducible results and objective assessment.

- Our approach can provide both patch-wise classification and pixel-wise classification.

- We incorporate a statistical approach based on Markov-Gibbs random field (MGRF) to remove inconsistencies in the final pixelwise classification.

FIGURE 69: Schematic illustration of the proposed computer aided diagnostic system for automated classification of kidney tissues.

B.   Materials and Methods

We propose a computer aided diagnostic system (shown in Fig. 69) that is based on deep learning for the automated classification of four kidney tissues: fat (class 1), parenchyma (class 2), clear cell RCC (class 3), and clear cell papillary RCC (class 4). We first start with dividing the whole slide images into small patches in order to be fed to the CNNs. Then, we preprocess these patches to enhance their visual appearance and features. Then, the patches are used to train and test our deep learning framework.

The institutional review board (IRB) approved our study. Before digitization, we de-identified all the slides which are stained with hematoxylin and eosin. We removed all identifiers that link to the patients from all the used image files. From the institution files, we randomly chose 30 cases which are diagnosed with clear cell RCC and 22 cases which are diagnosed with clear cell papillary RCC. A pathologist with expertise in genitourinary pathology reviewed these cases. From each case, we selected one representative slide from each case. Then, we selected 7 slides for the parenchyma tissue type and 5 slides for the fat tissue type. Then, we scanned the slides with a Philips UFS. A pathologist manually segmented the images into 4 tissue classes: fat, renal parenchyma, clear cell RCC, and clear cell papillary RCC.

1.   Patch Generation

Our dataset consists of 64 image slides (30 slides of clear cell RCC, 22 slides of clear cell papillary RCC, 7 slides of parenchyma, and 5 slides of fat). To follow the best practice in validating our deep learning framework, the data were divided into two sets that contain different image slides. We kept the first set, which is composed of forty-four image slides, for training and testing. We kept the second set, which is composed of 20 image slides, for final validation. Table 13 describes how we divided our data. The image slides is very large in size to be processed by a CNN. Therefore, generating image

patches form the image slides produces a suitable image size for the CNN and creates high sample size which is necessary for obtaining good accuracy from the deep learning framework. Overlapping patches were generated from the slides of the first set where we create three different patch sizes: small size = 250×250, medium size = 350×350, and large size = 450×450. Therefore, we create three folders for each image slide where each folder contains a specific patch size. A fifty percent overlap was kept between each patch and the next one. The overlap between patches results in learning multiple viewpoints within the tissue by the deep learning framework. We removed the patches that are dominated by background pixels. We kept about 70% of the patches for training and the reminder for the testing keeping in mind that the slides used for training are different from the slides used for testing. Figure 70 shows samples for the patches of the first set at different sizes. Overlapping patches were generated for the second set (i.e. validation set). Similarly, for each image slide we created three different sizes in three different folders but the degree of overlap was higher than of the first set: there was a 5-pixels shift in both dimensions between each patch and the next one. During inference by our deep learning framework, we assign a label for each input patch and we can obtain multiple labels for the same pixel because a pixel can belong to multiple patches.

TABLE 13: The number of slides in each set.

	Number of Slides		
	Training	Testing	Final Validation
Fat	2	1	2
Parenchyma	4	1	2
Clear Cell Papillary RCC	14	2	6
Clear Cell RCC	18	2	10
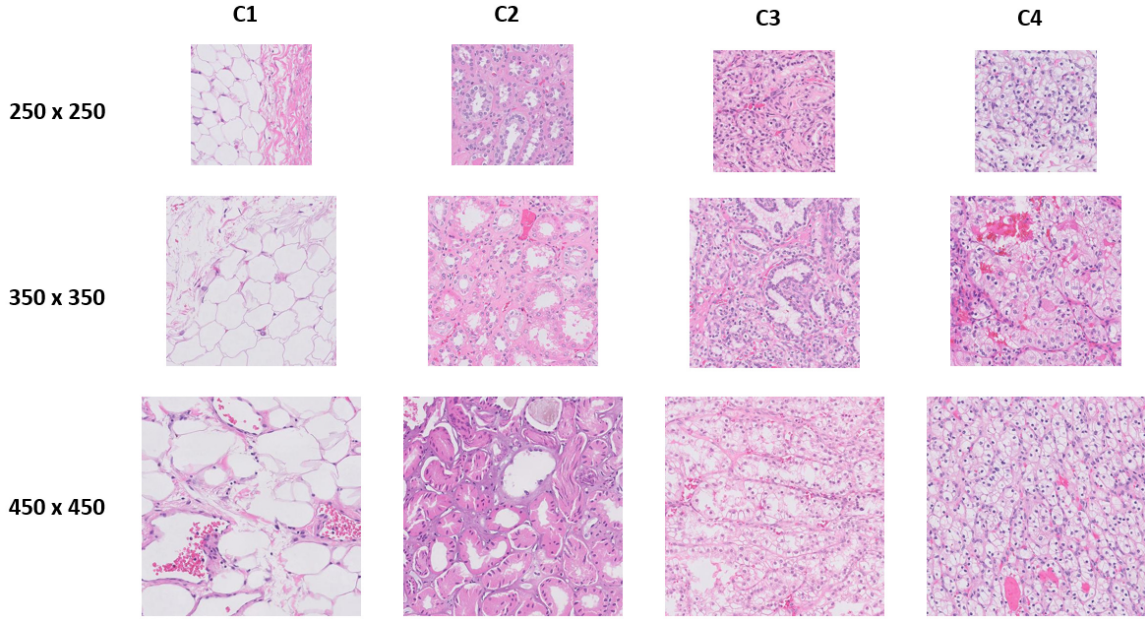
FIGURE 70: Samples of the generated patches from different image slides at different patch sizes. Note that C1, C2, C3, and C4 refer to class 1, class2, class 3, and class 4, respectively.

2. Preprocessing

To enhance the visual appearance of the patches, we used two preprocessing techniques for the first and the second datasets. Fist, we applied adaptive histogram equalization. Second, we applied edge enhancement.

a. Adaptive Histogram Equalization We used adaptive histogram equalization as an image processing technique to improve the contrast of the image patches [216]. In this technique, we address different parts in the image where we compute a histogram for each part. However, in ordinary histogram equalization, only one histogram is computed for the whole image. In adaptive histogram equalization, we use the multiple histograms to redistribute the pixel values of the image. By doing this redistribution, the local contrast and the edges are enhanced in each part of the image. However, adaptive histogram equalization is prone to noise amplification within homogeneous regions of the image where the

histogram is highly concentrated. To prevent noise amplification a variant is utilized which puts a limit on the amplification. This variant is called contrast limited adaptive histogram equalization. After the equalization step in contrast limited adaptive histogram equalization, the artificially generated boundaries are removed by employing bilinear interpolation between neighbouring regions. Figure 71 shows examples for histogram equalization of the image patches.
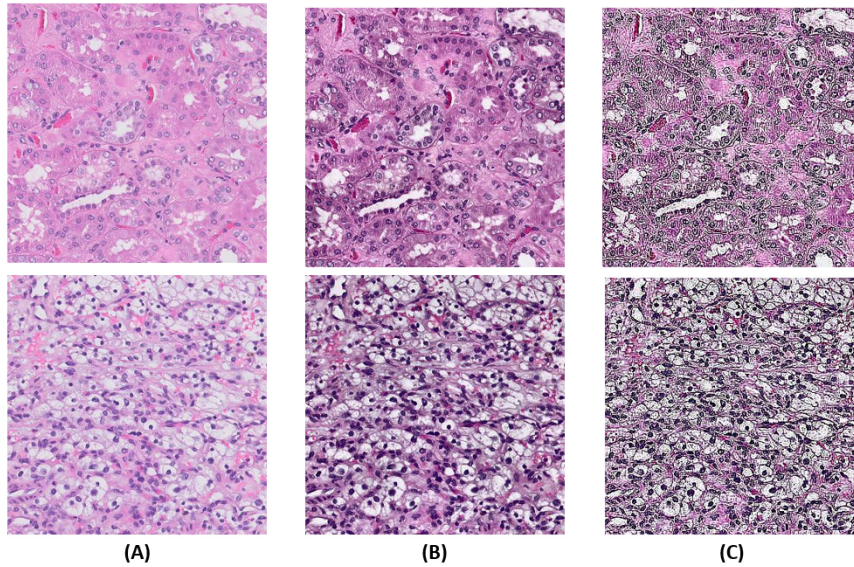
FIGURE 71: Examples of preprocessing step where original patches are shown in (A) and their histogram equalized versions are in (B) followed by edge enhancement (C).

    b.   Image Edge Enhancement   The edges of the first and the second datasets are enhanced. In this operation we enhance the edges of the important objects in image patches. The result of image edge enhancement is the improvement of the visual perception of the image patches. The improvement results from the suppression of low frequency components in the image. This high-pass filtering can be done in the spatial or the frequency domain; we filter the image patches by performing a convolution with a sharpening filter (shown in Fig. 72) in the spatial domain. Suppose that the filter is $h(m, n)$ and the input image patch is $x(m, n)$, the filtered image patch is given by Eq. (19).

$$y(m, n) = h(m, n) * x(m, n) \qquad (19)$$

where $*$ is the convolution operator. Figure 71 shows examples for edge enhancement of the image patches.

-1	-1	-1
-1	10	-1
-1	-1	-1

FIGURE 72: Weights of employed sharpening filter.

3. The Proposed Deep Learning Framework

The proposed deep learning framework is shown in Fig 73. The framework is composed of three CNNs. Each CNN is designed to process a specific patch size. We refer to the three CNNs by CNN_S, CNN_M, and CNN_L which process patches with small (250 x 250), medium (350 x 350), and large (450 x 450) sizes, respectively. We designed the three CNNs to have the same architecture except the input size. As shown in Fig. 73, the architecture of our CNNs is composed of a series of convolutional blocks where each block contains two convolutional layers followed by max-pooling layer. After convolutional blocks, there are two fully connected layers. Finally, the output of the fully connected layers was fed to a soft-max layer. The purpose of the convolutional layer is to extract feature maps by convolving the input image with a group of trainable kernels/filters. The feature maps contain features that describe the input objects in the input image. Each convolutional layer produces a volume of feature maps because we use a group of filters. In our design, we use filters of size 3x3 and stride of 1. The spatial dimensions of the feature maps were reduced by a factor of two in max-pooling layers. The purpose of the max-pooling layers is to keep the most prominent features and discard those less important. Furthermore, max-pooling layers reduces the training time and the computational cost. We used a stride of 2 in max-pooling layers. To summarize, each CNN contains 8 convolutional layers and 4 max-pooling layers. The first fully connected layer was composed of twelve neurons and

the second fully connected layer was composed of four neurons for four class classification. The purpose of the soft-max layer is to take the output of the last fully connected layer and convert it to class probabilities in the range of zero to one. The neuron which has the highest probability is the classification result for the input image patch. Table 15 shows our CNN configuration for an input patch of size 250×250. Similarly, the same concepts can be applied on the other patch sizes. Each CNN was trained by finding the minimum of the cross-entropy loss. In this loss the cross-entropy between the predicted class probabilities and the ground truth labels is minimized. The cross-entropy loss is defined as follows:

$$L_{BCE} = -\sum_{i=1}^{M} y_{o,i} \log(P_{o,i}) \tag{20}$$

Where M is the number of classes. $y_{o,i}$ is a binary indicator (0 or 1) which indicates the correct classification that observation o belongs to class i. $P_{o,i}$ is the predicted probability that observation o belongs to class i. To overcome network overfitting, we used a drop out with a rate 0.2 in the convolutional and fully connected layers.

a.   Output Of The Deep Learning Framework   The input image patches are classified to one of the four classes. Then, we assembled the classification of the patches to obtain a classification map for the whole-slide image. Each CNN in our deep learning framework can provide a patch-wise and pixel-wise accuracy. Our CNNs provide patch-wise accuracy because it can classify the input patch to one of the four classes. For one slide image, the patch-wise accuracy is defined as follows:

$$\text{Patchwise Accuracy} = \frac{\text{number of correctly classified patches}}{\text{total number of patches}} \tag{21}$$

We estimated the patch-wise accuracy for all slide images. Then, we estimated the average accuracy along slide images.

Patches of the second dataset had a five pixels shift between each other. Therefore, each pixel in the image can belong to several patches. Normally, when we classify a single patch, we assign a label for this patch then we give the same label to the pixels of this patch. When we assign the patch label to all pixels in the patch, we check if there are background

143
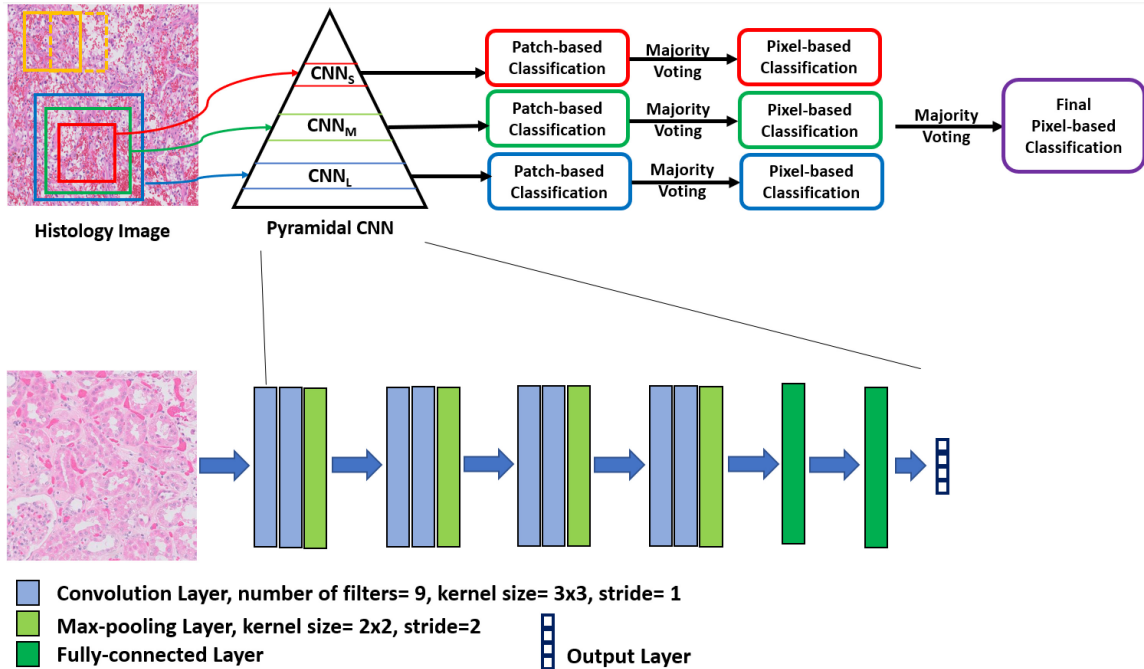
FIGURE 73: The pyramidal deep learning framework which consists of three convolutional networks for the automated classification of kidney histopatological images.

pixels (pixels with green color) in the patch. Then, the background pixels are assigned a background label which is kept until the final labeling. Given that pixels belong to many patches, each pixel can have several labels. We used majority voting to convert the several labels of a pixel into one label. Then, a labelled slide image (on the pixel level) was created for each slide image. For a single slide image, the pixel-wise accuracy is defined as follows:

$$\text{Pixelwise Accuracy} = \frac{\text{number of correctly classified pixels}}{\text{total number of pixels}} \quad (22)$$

The average of accuracies for all slide images was then estimated. To get an improved pixel-wise accuracy, we combined the result of the three CNNs. Given that our framework contains three CNNs, we obtained three labels for each pixel in an inputted slide image. Again, we adopted a majority voting strategy to get one label for each pixel. Finally, the final pixel-wise accuracy and the average accuracy over all slide images were estimated after combining CNNs results.

TABLE 14: The proposed CNN configuration for an input patch of size 250×250.

Layer	Depth	kernel	Stride	Spatial Size	Parameters
Input	3	—	—	$250 \times 250 \times 3$	0
1. Conv.	9	3×3	1×1	$248 \times 248 \times 9$	$3 \times 3 \times 3 \times 9$
2. Conv.	9	3×3	1×1	$246 \times 246 \times 9$	$3 \times 3 \times 9 \times 9$
3. Max-pool	9	2×2	2×2	$123 \times 123 \times 9$	0
4. Conv.	9	3×3	1×1	$121 \times 121 \times 9$	$3 \times 3 \times 9 \times 9$
5. Conv.	9	3×3	1×1	$119 \times 119 \times 9$	$3 \times 3 \times 9 \times 9$
6. Max-pool	9	2×2	2×2	$59 \times 59 \times 9$	0
7. Conv.	9	3×3	1×1	$57 \times 57 \times 9$	$3 \times 3 \times 9 \times 9$
8. Conv.	9	3×3	1×1	$55 \times 55 \times 9$	$3 \times 3 \times 9 \times 9$
9. Max-pool	9	2×2	2×2	$27 \times 27 \times 9$	0
10. Conv.	9	3×3	1×1	$25 \times 25 \times 9$	$3 \times 3 \times 9 \times 9$
11. Conv.	9	3×3	1×1	$23 \times 23 \times 9$	$3 \times 3 \times 9 \times 9$
12. Max-pool	9	2×2	2×2	$11 \times 11 \times 9$	0
13. Concat.	1	—	—	1089×1	0
14. Full	1	—	—	12×1	1089×12
15. Full	1	—	—	4×1	12×4
16. Softmax	1	—	—	4×1	0

Total number of parameters = 18462

TABLE 15: The proposed CNN configuration for an input patch of size 250×250.

b. Gauss-Markov Random Field Smoothing   We assigned a label for each pixel in the slide image. Therefore, the result is a labelled image for the whole slide image. To preserve continuity and remove inconsistencies (smooth) the labelled image, we considered the estimated labels (denoted δ) as samples generated from a generalized 2-D Gauss-

Markov random field (GGMRF) model [197]. There is eight-neighborhood for each pixel as shown in Fig. 74. The voxel-wise relaxation [217] and maximum a posteriori estimates amplify the continuity of the δ values:

$$\hat{\delta}_s =_{\hat{\delta}_s} \left\{ |\delta_s - \hat{\delta}_s|^\alpha + \rho^\alpha \lambda^\beta \sum_{r \in \nu_s} \eta_{s,r} |\hat{\delta}_s - \delta_r|^\beta \right\} \tag{23}$$

Where the original label and its expected estimates are denoted by δ_s and $\hat{\delta}_s$, respectively, at $s = (x, y)$ which is the observed 2D location. The GGMRF potential is denoted by $\eta_{s,r}$. The eight-neighborhood voxel set is denoted by ν_s (Fig. 74). ρ and λ are scaling factors. $\alpha \in 1, 2$ is a parameter that determines the Laplace ($\alpha = 1$) or the gaussian ($\alpha = 2$) estimator's prior distribution. β is a parameter that controls the degree of smoothness. We set $\alpha = 2$, $\beta = 1.01$, $\rho = 1$, $\lambda = 5$, and $\eta_{s,r} = \sqrt{2}$. Our hypothesis is that GGMRF smoothing can increase the pixel-wise accuracy of the labelled slide image.
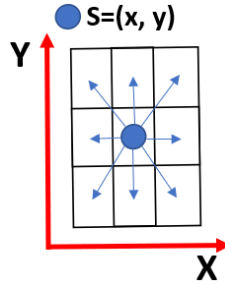
FIGURE 74: For eight neighbors, the figure shows the Pairwise voxel interaction in a 2-D GGMRF image model.

Algorithm summarizes how our deep learning framework produced a pixel-wise classification map for the tested slide image and how we estimated classification accuracy. We reported in the Results section the average accuracy over all slide images.

Algorithm 4 Pixel-wise Classification

A pixel-wise classification map Divide the input slide images into overlapped patches each CNN Assign a label for each patch Estimate the patch-wise accuracy by Eq. 21 Assign labels for each pixel Apply majority voting for the labels of each pixel Apply GGMRF smoothing Estimate the pixel-wise accuracy by Eq.22 Apply majority voting for the results of the CNNs (fusion) Apply GGMRF smoothing Estimate the pixel-wise accuracy by Eq.22

C.   Results

The deep learning library TensorFlow [218] was used to develop our deep learning framework. The parameters of our deep learning framework, such as network architecture, should be optimized to obtain the optimal accuracy. We performed this optimization process using grid search strategy by searching for the framework parameters that give the best system performance. The searched parameters are: 1) the number of convolutional layers, 2) kernel size, 3) initialization of the convolutional kernels, 4) number of filters, 5) stride, 6) patch size, 7) number of epochs, 8) learning rate, 9) type of optimizer. The best parameters from the grid search is presented in Table 16. During training of our deep learning framework, we kept 20% of the training data for validation. After each epoch we estimate the validation accuracy and validation loss. The best final model is the one that gave the highest validation accuracy. Data augmentation is a way to increase training data to avoid overfitting. We adopted a data augmentation strategy that consists of random rotation, scaling, and flipping.

We trained our framework by 70% of the patches of the first dataset and the remaining 30% was used for testing. During training and testing, each CNN of the three CNNs was fed by the appropriate patch size. After testing, we calculated the patch-wise accuracy for the four tissue types. Table 17 shows the patch-wise accuracy for each tissue type. Fur-

TABLE 16: Optimal parameters of our deep learning framework.

Parameter	Value
Number of convolutional layers	8
Kernel size	3×3
Kernel initialization	He initialization [172]
Number of filters	9
Stride	1 (convolution), 2 (max-pooling)
Patch size	32
Number of epochs	60
Learning rate	0.001
Optimizer	Adam

thermore, Fig. 75 shows samples of correctly and wrongly classified patches for the four kidney tissue types.

TABLE 17: The estimated patch-wise accuracy for the testing set. The table shows the accuracy for the four tissue types at different patch sizes.

Tissue Type	Patch Size		
	250×250	350×350	450×450
Fat	0.89 ± 0.01	0.91 ± 0.02	0.90 ± 0.11
Parenchyma	0.88 ± 0.11	0.91 ± 0.12	0.90 ± 0.24
Clear cell papillary RCC	0.86 ± 0.23	0.90 ± 0.21	0.89 ± 0.21
Clear cell RCC	0.85 ± 0.02	0.89 ± 0.31	0.89 ± 0.03

After testing the framework using the testing set, we performed final validation using the second dataset. In the second dataset we made 5 pixels shift between successive patches because we want to perform pixel-wise classification besides the patch-wise classification. We feed the second dataset to our framework one slide at time. In other words,
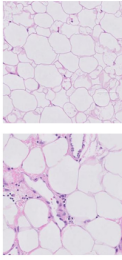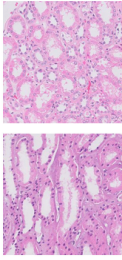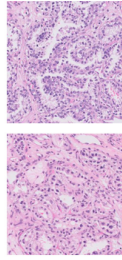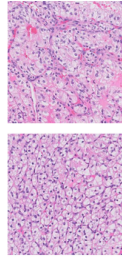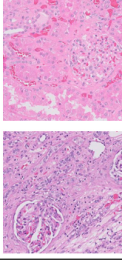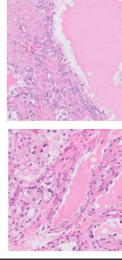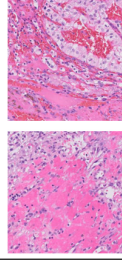
	Fat	Parenchyma	Clear Cell papillary RCC	Clear Cell RCC
Correctly Classified				
Misclassified				

FIGURE 75: Samples of correctly and wrongly classified patches for the four tissue types.

we divide the slide image into the three mentioned patch sizes, and we perform patch and pixel classification for that slide image. Then, we report the average accuracy over slide images for each tissue type in the data. Tables 18 shows the average of patch-wise accuracy and the average of pixel-wise accuracy of fat, parenchyma, clear cell papillary RCC, and clear cell RCC cases, respectively. We combine the classification of the three CNNs. Therefore, each pixel in the tested slide image can have three labels and by applying majority voting we obtain one label for each pixel. Then, we estimate the pixel-wise accuracy for the labelled slide image. To remove inconsistencies, we apply GGMRF smoothing on the labelled slide image. To assess the performance of our method after obtaining the final labeling, we constructed a confusion matrix from which we can estimate the accuracy, sensitivity and specificity for detecting each tissue type. Table 19 shows the confusion matrix of our proposed method and Table 22 shows the performance metrics for detecting each tissue type.

To compare the performance of our proposed approach with other deep learning models, we used two pre-trained models: ResNet18 and ResNet34. In these two models,

TABLE 18: For the fat, parenchyma, clear cell papillary RCC, and clear cell RCC cases of the second dataset, we estimated the average patch-wise accuracy and the pixel-wise accuracy. The table shows the accuracy at different patch sizes.

Tissue		Patch Size		
		250×250	350×350	450×450
Fat	patch-wise	0.87 ± 0.04	0.90 ± 0.11	0.89 ± 0.05
	pixel-wise	0.89 ± 0.01	0.91 ± 0.08	0.91 ± 0.11
Parenchyma	patch-wise	0.85 ± 0.11	0.88 ± 0.21	0.87 ± 0.03
	pixel-wise	0.88 ± 0.05	0.90 ± 0.08	0.89 ± 0.16
Clear cell papillary RCC	patch-wise	0.84 ± 0.14	0.87 ± 0.06	0.86 ± 0.07
	pixel-wise	0.86 ± 0.11	0.90 ± 0.05	0.90 ± 0.21
Clear cell RCC	patch-wise	0.83 ± 0.20	0.86 ± 0.01	0.86 ± 0.02
	pixel-wise	0.85 ± 0.05	0.89 ± 0.04	0.88 ± 0.11

we replaced the last layers with two output layers. Then, fine-tuning was performed by using the kidney data. Tables 20, 21 show the confusion matrices for the labeling resulted from the use of ResNet18, and ResNet 34, respectively. Table 23 demonstrates a quantitative comparison in terms of accuracy, sensitivity, and specificity between our framework and ResNets. To establish a statistical significance for our method, we performed a t-test to show that there is a significant difference between our approach and ResNets. We obtained a p-value of 0.01 for ResNet18 and a p-value of 0.02 for ResNet34. The p-values are less than 0.05. Therefore, the null hypothesis can be rejected in the two tests. We can conclude that the differences between our method and other models are statistically significant. Figures 76, 77, 78, 79 show the labeling on the pixel level for the four tissue types.

D.   Discussion

TABLE 19: Confusion matrix based on the final labeling obtained from our proposed approach, where class 1, class 2, class 3, and class 4 refer to fat, parenchyma, clear cell papillary RCC, and clear cell RCC. Values are shown as percentages.

		Actual Label			
		Class 1	Class 2	Class 3	Class 4
	Class 1	92.98%	2.68%	1.75%	1.69%
	Class 2	2.22%	92.41%	3.75%	2.66%
Predicted Label	Class 3	1.92%	2.51%	91.35%	4.38%
	Class 4	2.87%	2.41%	3.15%	91.30%

TABLE 20: Confusion matrix based on the labeling obtained from ResNet18, where class 1, class 2, class 3, and class 4 refer to fat, parenchyma, clear cell papillary RCC, and clear cell RCC. Values are shown as percentages.

		Actual Label			
		Class 1	Class 2	Class 3	Class 4
	Class 1	90.13%	3.78%	3.31%	3.08%
	Class 2	3.52%	90.00%	4.70%	4.04%
Predicted Label	Class 3	2.59%	3.13%	88.50%	4.87%
	Class 4	3.75%	3.09%	3.49%	88.00%

A computer aided diagnostic system that is able to automatically classify kidney histopathological images was proposed in this chapter. In particular, we target the challenging task of classifying between clear cell papillary RCC and clear cell RCC. The morphological features of these tumor subtypes overlap; however, they have different prognosis. Therefore, our task is necessary to best determine the appropriate clinical management. We used deep learning algorithms to build our system of classification. Our design of pyrimidal CNN successfully determined the normal tissues and abnormal tissues and managed to differentiate between clear cell papillary RCC and clear cell RCC. Our approach also

TABLE 21: Confusion matrix based on the labeling obtained from ResNet34, where class 1, class 2, class 3, and class 4 refer to fat, parenchyma, clear cell papillary RCC, and clear cell RCC. Values are shown as percentages.

		Actual Label			
		Class 1	Class 2	Class 3	Class 4
Predicted Label	Fat	89.10%	4.31%	2.91%	2.78%
	Parenchyma	3.83%	89.00%	5.94%	2.87%
	Clear Cell Papillary RCC	2.99%	3.79%	87.50%	7.35%
	Clear Cell RCC	4.08%	2.89%	3.66%	87.00%

TABLE 22: Performance metrics for the labeling obtained from our framework for each tissue type. The values in the table are based on the confusion matrix in Table 19.

Tissue	Accuracy	Sensitivity	Specificity
Fat	0.98	0.93	0.98
Parenchyma	0.97	0.92	0.97
Clear cell papillary RCC	0.95	0.91	0.96
Clear cell RCC	0.94	0.91	0.97

generates classification maps that classify the kidney histopathological images on pixel-level. Finally, our method remove inconsistencies from the generated maps using GGMRF smoothing.

An enormous amount of information can be found in the whole-slide histopathological images. Therefore, the pathologist spends significant time and effort to manually examine the histology image. On the other hand, fast and accurate analysis of histology images is needed because there is a growth in the number of diagnosed cancer cases. Our study addresses this shortcoming by proposing a computer aided diagnostic system that is able to automatically classify kidney tissues and detect kidney cancers. Our approach can be adopted to other the diagnosis of other cancers.

TABLE 23: Quantitative comparison between our approach and other deep learning models. The values are the estimated averages across tissue types.

	Average Accuracy	Average Sensitivity	Average Specificity
ResNet18	0.942	0.892	0.961
ResNet34	0.937	0.882	0.957
Proposed without GGMRF smoothing	0.946	0.911	0.965
Proposed with GGMRF smoothing	**0.957**	**0.920**	**0.971**
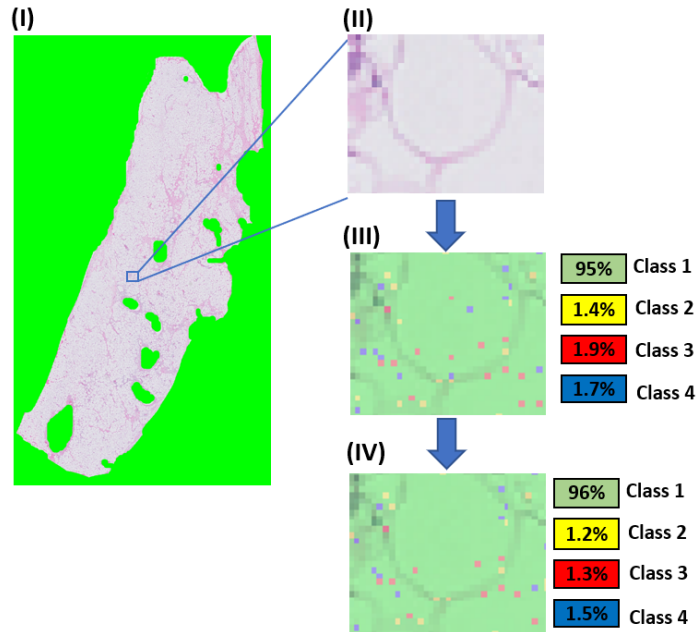
FIGURE 76: Example for the output from our framework. (I) whole-slide fat image. (II) is zoomed region, (III) is the labelled region before GGMRF smoothing, and (IV) is the labelled region after GGMRF smoothing. Gree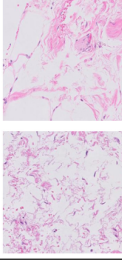n (Class 1), yellow (Class 2), red (Class 3), and blue (Class 4) refer to fat, parenchyma, clear cell papillary RCC, and clear cell RCC, respectively.

As shown in Table 18, our approach gave a high patch-wise accuracy. In other words, our framework can determine tissue types from image patches with different sizes. Patch size of 350×350 resulted in the best patch-wise accuracy. For the cases of fat, parenchyma, and clear cell papillary RCC, patch size of 350×350 resulted in better patch-wise accuracy. For the case of clear cell RCC, patch sizes of 350×350 and 450×450 re-
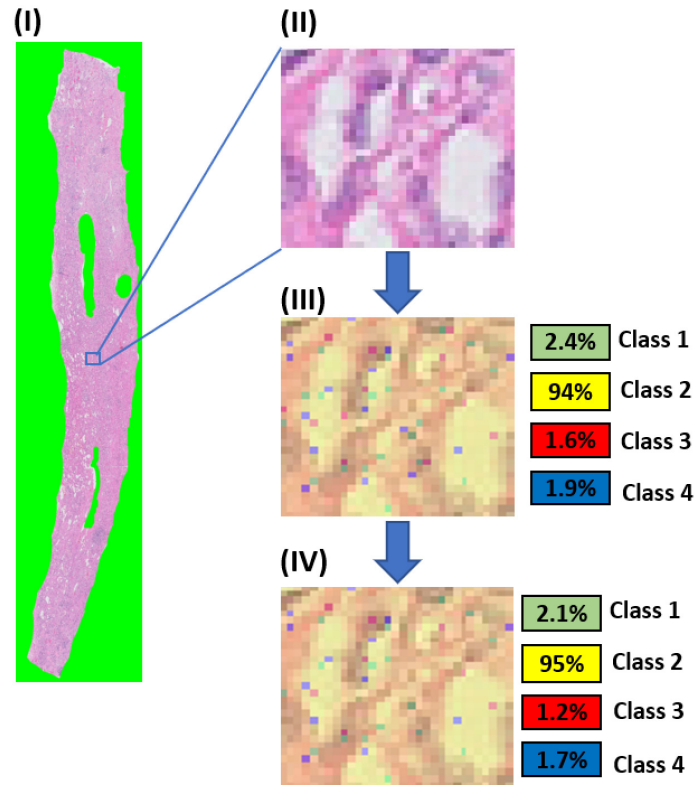
FIGURE 77: Example for the output from our framework. (I) whole-slide parenchyma image. (II) is zoomed region, (III) is the labelled region before GGMRF smoothing, and (IV) is the labelled region after GGMRF smoothing. Green (Class 1), yellow (Class 2), red (Class 3), and blue (Class 4) refer to fat, parenchyma, clear cell papillary RCC, and clear cell RCC, respectively.

sulted in similar patch-wise accuracy; however, patch size of 350×350 gave better accuracy than patch size of 250×250.

Our pyramidal framework allows us to analyze the histopathological images at various spatial scales. Then, we combine the classifications of the three CNNs to obtain a better pixel-wise classification. Combining the results allows us to get better accuracy than the one obtained from a single CNN. Furthermore, we give the pathologist the opportunity to analyze the pathology regions at small scale because our framework results in pixel-level classification. Patch size of 350×350 gave the best pixel-wise accuracy. For the cases of
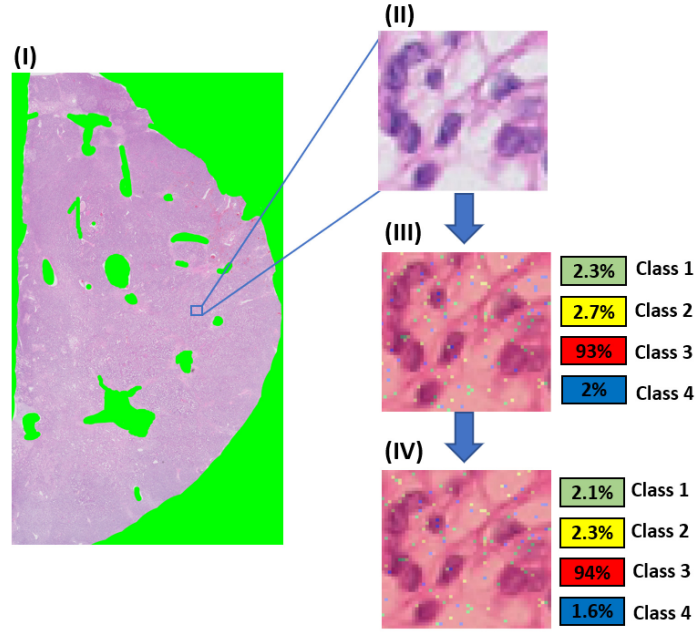
FIGURE 78: Example for the output from our framework. (I) whole-slide clear cell papillary RCC image. (II) is zoomed region, (III) is the labelled region before GGMRF smoothing, and (IV) is the labelled region after GGMRF smoothing. Green (Class 1), yellow (Class 2), red (Class 3), and blue (Class 4) refer to fat, parenchyma, clear cell papillary RCC, and clear cell RCC, respectively.

parenchyma and clear cell RCC, better pixel-wise accuracy can be obtained from patch size of $350{\times}350$. Similar pixel-wise accuracy resulted from patches with sizes $350{\times}350$ and $450{\times}450$ for the cases of fat and clear cell papillary RCC; however, patch size of $350{\times}350$ had better accuracy than patch size of $250{\times}250$.

Table 22 shows that the fusion of CNN classifications produced a good classification for the pixels in terms of accuracy, sensitivity, and specificity. After fusion, we obtained the best performance metrics from fat cases while clear cell papillary RCC and clear cell RCC gave the lowest performance metrics. We can conclude that CNNs extract new features related to the underlying tissue texture from each patch size and fusing between CNN classifications resulted in enhancing the final classification.

Finally, using GGMRF smoothing enhanced the labeling of the input slide image as
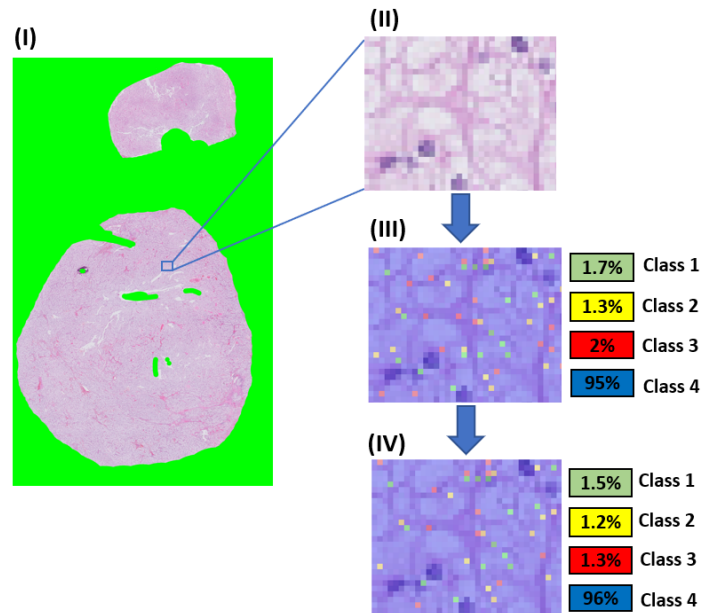
FIGURE 79: Example for the output from our framework. (I) whole-slide clear cell RCC image. (II) is zoomed region, (III) is the labelled region before GGMRF smoothing, and (IV) is the labelled region after GGMRF smoothing. Green (Class 1), yellow (Class 2), red (Class 3), and blue (Class 4) refer to fat, parenchyma, clear cell papillary RCC, and clear cell RCC, respectively.

it works on removing inconsistencies, Table 23. Furthermore, our deep learning framework surpassed other state-of-the-art models such as ResNet18 and ResNet34, Which demonstrates that the notion of fusing CNN decisions is fruitful for obtaining improved classification.

CHAPTER VII
CONCLUSION AND FUTURE WORK

The work presented in this dissertation documents the ability of the deep learning algorithms to provide an automated and efficient tools for the diagnosis of human diseases using medical images. The effectiveness of the deep learning algorithms has been successfully tested for the automated disease diagnosis in four different case studies. The first case study aimed at developing an automated tool for the early detection of acute renal transplant rejection using DWI. The second second case study aimed at developing an automated tool for the segmentation and quantification of the left ventricle from cardiac cine MRI. The third case study aimed at developing an automated tool for the assessment of LV strain using mice cine MRI and tagged MRI. The fourth case study aimed at developing an automated tool for the detection of renal cell carcinoma from histopathological kidney images. Summary of the main contributions of this dissertation are as follows:

- In the first case study, we propose an automated framework that combines the advantages of both DW-MRIs and deep learning to classify renal allografts into non-rejection and acute rejection status by using data generated from the fusion of voxel-wise ADCs and the clinical biomarkers. To the best of our knowledge, this is the first automated non-invasive CAD system of its kind to assess renal transplant status using the integration of the DW-MR image markers and clinical biomarkers along with convolutional neural networks.

- In the second case study, the contributions are (i) The extraction of a region-of-interest (ROI) that encompasses the LV from CMR images using an efficient method that is based on FCN. ROI extraction before final cardiac segmentation alleviates the

class-imbalance problem and reduces the computational and memory requirement. (ii) A novel FCN architecture for cardiac segmentation following ROI extraction. The network follows the same idea of the U-net of passing the input to a contracting path followed by an expanding path. However, it has several bottleneck layers that refer to different representation to the input. The up-sampling of these layers are combined to obtain the final segmentation. The proposed architecture has less number of parameters than the established models such as U-net, yet it demonstrated a better performance. (iii) A novel loss function called radial loss that minimizes the difference between the predicted LV contours and the ground truth contours was incorporated with the cross-entropy loss. (iv) The generalization strength of our proposed segmentation approach was evaluated by measuring the segmentation performance of our approach when trained on the whole ACDC training dataset and tested on another dataset (local dataset). We achieved good segmentation accuracy which was comparable to another model that used only our local dataset.

- In the third case study, we developed an automated pipeline for accurate estimation of strain from standard cine MRI in mice for both sham and myocardial infarction (MI) subjects, to obviate the need for tagged imaging in mice. Since myocardial segmentation is a crucial first step in processing and analyzing LV functional indices, a deep learning-based approach for automatic segmentation of myocardium borders was developed. Subsequently, a Laplace-based approach to track myocardial points through the cardiac cycle was developed to accurately assess the strain from the standard cine cardiac MRI. The Laplace-based method was validated by comparing it to the strain analysis outcome obtained by the tagged imaging which is currently the gold standard for strain analysis. Finally, we also used our pipeline to estimate other global and local cardiac indices.

- In the fourth case study, the contributions are (i) the first study to discriminate be-

158

tween clear cell RCC and clear cell papillary RCC. This classification has high clinical relevance. (ii) We propose a pyramidal deep learning model that utilizes a hierarchy of three CNNs that process different image sizes. The deep learning improves the precision of the diagnosis and decrease human error. Furthermore, it produces reproducible results and objective assessment. (iii) Our approach can provide both patch-wise classification and pixel-wise classification. (iv) We incorporate a statistical approach based on Markov-Gibbs random field (MGRF) to remove inconsistencies in the final pixelwise classification.

Several possibilities for the future work of this dissertation include, but are not limited to, the following:

- In the first case study, we are planning to extend our study by performing a statistical analysis to determine the most informative b-values. Once we reach this point, we could sacrifice some of the b-values that are unsatisfactory informative, which in turn will reduce the DW-MRI acquisition time. Future progress includes the usage of a larger sample size collected at different transplant centers and/or different imaging systems and collection protocols, and the exploration of additional biomarkers, such as genomic information that will augment personalized data in the cohort.

- In the second case study, we are planning to extend our deep learning algorithms to process 3D cardiac data. This will help us to easily estimate left ventricular functional indexes. Furthermore, our future work incorporates a complete computer aided diagnostic system for cardiac disease classification and detection.

- In the third case study, we are planning to test and validate the algorithm potential using a larger cohort. Despite these limitations, this study demonstrates the feasibility of accurately estimating myocardial strain indices, which may obviate the need for tagged MR clinically.

- In the fourth case study, we are planning to include all renal carcinoma sub-types in our classification system. We can also extend our system to assess the grading of kidney cancer.

REFERENCES

[1] M. Shehata, F. Khalifa, A. Soliman, M. Ghazal, F. Taher, M. Abou El-Ghar, A. C. Dwyer, G. Gimel'farb, R. S. Keynton, and A. El-Baz. Computer-aided diagnostic system for early detection of acute renal transplant rejection using diffusion-weighted mri. *IEEE Transactions on Biomedical Engineering*, 66(2):539–552, 2018.

[2] M. Khened, V. A. Kollerathu, and G. Krishnamurthi. Fully convolutional multi-scale residual densenets for cardiac segmentation and automated cardiac diagnosis using ensemble of classifiers. *Medical Image Analysis*, 51:21–45, 2019.

[3] M. D. Cerqueira, N. J. Weissman, V. Dilsizian, A. K. Jacobs, S. Kaul, W. K. Laskey, D. J. Pennell, J. A. Rumberger, T. Ryan, et al. Standardized myocardial segmentation and nomenclature for tomographic imaging of the heart: a statement for healthcare professionals from the cardiac imaging committee of the council on clinical cardiology of the american heart association. *Circulation*, 105(4):539–542, 2002.

[4] I. Goodfellow, Y. Bengio, A. Courville, and Y. Bengio. *Deep learning*, volume 1. MIT press Cambridge, 2016.

[5] F. Emmert-Streib, Z. Yang, H. Feng, S. Tripathi, and M. Dehmer. An introductory review of deep learning for prediction models with big data. *Frontiers in Artificial Intelligence*, 3:4, 2020.

[6] convolutional neural networks for visual recognition.

[7] J. Brownlee. How to Choose an Activation Function for Deep Learning. https://machinelearningmastery.com/choose-an-activation-function-for-deep-learning/, 2021. [Online; accessed 1-July-2022].

[8] S. Saha. A Comprehensive Guide to Convolutional Neural Networks. https://towardsdatascience.com/a-comprehensive-guide-to-convolutional-neural-networks-the-eli5-way-3bd2b1164a53, 2018. [Online; accessed 5-July-2022].

[9] V. E. Balas, B. K. Mishra, and R. Kumar. *Handbook of deep learning in biomedical engineering: techniques and applications*. Academic Press, 2021.

[10] O. Ronneberger, P. Fischer, and T. Brox. U-net: Convolutional networks for biomedical image segmentation. In *International Conference on Medical image computing and computer-assisted intervention*, pages 234–241. Springer, 2015.

[11] A. Esteva, A. Robicquet, B. Ramsundar, V. Kuleshov, M. DePristo, K. Chou, C. Cui, G. Corrado, S. Thrun, and J. Dean. A guide to deep learning in healthcare. *Nature medicine*, 25(1):24–29, 2019.

[12] C. Chen, C. Qin, H. Qiu, G. Tarroni, J. Duan, W. Bai, and D. Rueckert. Deep learning for cardiac image segmentation: a review. *Frontiers in Cardiovascular Medicine*, 7:25, 2020.

[13] J. E. Hall and M. E. Hall. *Guyton and Hall textbook of medical physiology*. Elsevier, 2021.

[14] J. E. Hall and M. E. Hall. *textbook of medical physiology e-Book*. Elsevier Health Sciences, 2020.

[15] R. Smith-Bindman, D. L. Miglioretti, E. Johnson, C. Lee, H. S. Feigelson, et al. Use of diagnostic imaging studies and associated radiation exposure for patients enrolled in large integrated health care systems, 1996-2010. *Jama*, 307(22):2400–2409, 2012.

[16] P. Mansfield. Imaging by nuclear magnetic resonance. *Journal of Physics E: Scientific Instruments*, 21(1):18, 1988.

[17] P. Mansfield. Snapshot magnetic resonance imaging (nobel lecture). *Angewandte Chemie International Edition*, 43(41):5456–5464, 2004.

[18] R. W. Brown, Y.-C. N. Cheng, E. M. Haacke, M. R. Thompson, and R. Venkatesan. *Magnetic resonance imaging: physical principles and sequence design*. John Wiley & Sons, 2014.

[19] R. Bammer. Basic principles of diffusion-weighted imaging. *European journal of radiology*, 45(3):169–184, 2003.

[20] C. Sato, S. Naganawa, T. Nakamura, H. Kumada, S. Miura, O. Takizawa, and T. Ishigaki. Differentiation of noncancerous tissue and cancer lesions by apparent diffusion coefficient values in transition and peripheral zones of the prostate. *Journal of Magnetic Resonance Imaging: An Official Journal of the International Society for Magnetic Resonance in Medicine*, 21(3):258–262, 2005.

[21] P. Sundgren, Q. Dong, D. Gomez-Hassan, S. Mukherji, P. Maly, and R. Welsh. Diffusion tensor imaging of the brain: review of clinical applications. *Neuroradiology*, 46(5):339–350, 2004.

[22] X. Yang and K. Murase. Tagged cardiac mr image segmentation by contrast enhancement and texture analysis. In *2009 9th International Conference on Electronic Measurement & Instruments*, pages 4–210. IEEE, 2009.

[23] E.-S. H. Ibrahim, J. Stojanovska, A. Hassanein, C. Duvernoy, P. Croisille, R. Pop-Busui, and S. D. Swanson. Regional cardiac function analysis from tagged mri images. comparison of techniques: Harmonic-phase (harp) versus sinusoidal-modeling (sinmod) analysis. *Magnetic resonance imaging*, 54:271–282, 2018.

[24] X. Liu and J. L. Prince. Shortest path refinement for motion estimation from tagged mr images. *IEEE Transactions on Medical Imaging*, 29(8):1560–1572, 2010.

[25] L. Axel and L. Dougherty. Heart wall motion: improved method of spatial modulation of magnetization for mr imaging. *Radiology*, 172(2):349–350, 1989.

[26] M. N. Gurcan, L. E. Boucheron, A. Can, A. Madabhushi, N. M. Rajpoot, and B. Yener. Histopathological image analysis: A review. *IEEE reviews in biomedical engineering*, 2:147–171, 2009.

[27] Y. LeCun, Y. Bengio, and G. Hinton. Deep learning. *nature*, 521(7553):436–444, 2015.

[28] Z. Obermeyer and E. J. Emanuel. Predicting the future—big data, machine learning, and clinical medicine. *The New England journal of medicine*, 375(13):1216, 2016.

[29] 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*, 42:60–88, 2017.

[30] D. A. Bluemke, L. Moy, M. A. Bredella, B. B. Ertl-Wagner, K. J. Fowler, V. J. Goh, E. F. Halpern, C. P. Hess, M. L. Schiebler, and C. R. Weiss. Assessing radiology research on artificial intelligence: a brief guide for authors, reviewers, and readers—from the radiology editorial board, 2020.

[31] C. J. Kelly, A. Karthikesalingam, M. Suleyman, G. Corrado, and D. King. Key challenges for delivering clinical impact with artificial intelligence. *BMC medicine*, 17(1):1–9, 2019.

[32] Annual data report.

[33] H. M. Isoniemi, L. Krogerus, E. von Willebrand, E. Taskinen, J. Ahonen, and P. Häyry. Histopathological findings in well-functioning, long-term renal allografts. *Kidney international*, 41(1):155–160, 1992.

[34] G. Baxter. Ultrasound of renal transplantation. *Clinical radiology*, 56(10):802–818, 2001.

[35] B. J. Nankivell and S. I. Alexander. Rejection of the kidney allograft. *New England Journal of Medicine*, 363(15):1451–1462, 2010.

[36] H. Bennett and D. Li. Mr imaging of renal function. *Magnetic resonance imaging clinics of North America*, 5(1):107–126, 1997.

[37] G. L. Myers, W. G. Miller, J. Coresh, J. Fleming, N. Greenberg, et al. Recommendations for improving serum creatinine measurement: a report from the laboratory working group of the national kidney disease education program. *Clinical chemistry*, 52(1):5–18, 2006.

[38] D. Yang, Q. Ye, M. Williams, Y. Sun, T. C.-C. Hu, D. S. Williams, J. M. Moura, and C. Ho. Uspio-enhanced dynamic mri: Evaluation of normal and transplanted rat kidneys. *Magnetic Resonance in Medicine: An Official Journal of the International Society for Magnetic Resonance in Medicine*, 46(6):1152–1163, 2001.

[39] E. D. Brown, M. Y. Chen, N. T. Wolfman, D. J. Ott, and N. E. Watson Jr. Complications of renal transplantation: evaluation with us and radionuclide imaging. *Radiographics*, 20(3):607–622, 2000.

[40] E. L. W. Giele. *Computer methods for semi-automatic MR renogram determination*. Citeseer, 2002.

[41] C. Kolofousi, K. Stefanidis, D. D. Cokkinos, D. Karakitsos, E. Antypa, and P. Piperopoulos. Ultrasonographic features of kidney transplants and their complications: an imaging review. *International Scholarly Research Notices*, 2013, 2013.

[42] C. Sebastià, S. Quiroga, R. Boyé, C. Cantarell, M. Fernandez-Planas, and A. Alvarez. Helical ct in renal transplantation: normal findings and early and late complications. *Radiographics*, 21(5):1103–1117, 2001.

[43] F. Khalifa, G. M. Beache, M. Abou El-Ghar, T. El-Diasty, G. Gimel'farb, M. Kong, and A. El-Baz. Dynamic contrast-enhanced mri-based early detection of acute renal transplant rejection. *IEEE transactions on medical imaging*, 32(10):1910–1927, 2013.

[44] A. Sharfuddin. Renal relevant radiology: imaging in kidney transplantation. *Clinical Journal of the American Society of Nephrology*, 9(2):416–429, 2014.

[45] H. Sayar, A. A. Sharfuddin, T. E. Taber, and R. Mehta. Chronic myeloid leukemia within a year of kidney transplant with elevated alkaline phosphatase correlated with imatinib therapy. *Experimental and clinical transplantation: official journal of the Middle East Society for Organ Transplantation*, 9(5):336–339, 2011.

[46] P. V. Prasad, A. Priatna, K. Spokes, and F. H. Epstein. Changes in intrarenal oxygenation as evaluated by bold mri in a rat kidney model for radiocontrast nephropathy. *Journal of Magnetic Resonance Imaging: An Official Journal of the International Society for Magnetic Resonance in Medicine*, 13(5):744–747, 2001.

[47] N. Grenier, F. Basseau, M. Ries, B. Tyndal, R. Jones, and C. Moonen. Functional mri of the kidney. *Abdominal imaging*, 28(2):0164–0175, 2003.

[48] G. S. Chilla, C. H. Tan, C. Xu, and C. L. Poh. Diffusion weighted magnetic resonance imaging and its recent trend—a survey. *Quantitative imaging in medicine and surgery*, 5(3):407, 2015.

[49] WHO. Cardiovascular diseases (cvds), 2017.

[50] A. H. Association. Heart disease and stroke statistics 2018 at-a-glance. *https://www.heart.org/-/media/data-import/downloadables/heart-disease-and-stroke-statistics-2018—at-a-glance-ucm$_4$98848.pdf*, 2018.

[51] A. F. Frangi, W. J. Niessen, and M. A. Viergever. Three-dimensional modeling for functional analysis of cardiac images, a review. *IEEE Transactions on Medical Imaging*, 20(1):2–5, Jan 2001.

[52] M. Souto, L. R. Masip, M. Couto, J. J. Suárez-Cuenca, A. Martínez, P. G. Tahoces, J. M. Carreira, and P. Croisille. Quantification of right and left ventricular function in cardiac mr imaging: Comparison of semiautomatic and manual segmentation algorithms. *Diagnostics (Basel, Switzerland)*, 3(2):271–282, Apr 2013.

[53] P. D'Orléans-Juste, J.-C. Honoré, E. Carrier, and J. Labonté. Cardiovascular diseases: new insights from knockout mice. *Current opinion in pharmacology*, 3(2):181–185, 2003.

[54] F. Franco, G. D. Thomas, B. Giroir, D. Bryant, M. C. Bullock, M. C. Chwialkowski, R. G. Victor, and R. M. Peshock. Magnetic resonance imaging and invasive evaluation of development of heart failure in transgenic mice with myocardial expression of tumor necrosis factor-α. *Circulation*, 99(3):448–454, 1999.

[55] R. M. Peshock, D. Willett, D. Sayad, W. Hundley, M. Chwialkowski, G. D. Clarke, and R. Parkey. Quantitative mr imaging of the heart. *Magnetic resonance imaging clinics of North America*, 4(2):287–305, 1996.

[56] C. K. Phoon and D. H. Turnbull. Cardiovascular imaging in mice. *Current protocols in mouse biology*, 6(1):15–38, 2016.

[57] M. Joubert, P. Tager, D. Legallois, E. Defourneaux, B. Le Guellec, B. Gerber, R. Morello, and A. Manrique. Test-retest reproducibility of cardiac magnetic resonance imaging in healthy mice at 7-tesla: effect of anesthetic procedures. *Scientific reports*, 7(1):6698, 2017.

[58] D. J. Stuckey, C. A. Carr, P. Camelliti, D. J. Tyler, K. E. Davies, and K. Clarke. In vivo mri characterization of progressive cardiac dysfunction in the mdx mouse model of muscular dystrophy. *PloS one*, 7(1):e28569, 2012.

[59] W. D. Gilson and D. L. Kraitchman. Cardiac magnetic resonance imaging in small rodents using clinical 1.5 t and 3.0 t scanners. *Methods*, 43(1):35–45, 2007.

[60] T. Niendorf, J. Schulz-Menger, K. Paul, T. Huelnhagen, V. A. Ferrari, and R. Hodge. High field cardiac magnetic resonance imaging: a case for ultrahigh field cardiac magnetic resonance. *Circulation: Cardiovascular Imaging*, 10(6):e005460, 2017.

[61] T. M. Mohamed, N. R. Stone, E. C. Berry, E. Radzinsky, Y. Huang, et al. Chemical enhancement of in vitro and in vivo direct cardiac reprogramming. *Circulation*, 135(10):978–995, 2017.

[62] T. M. Mohamed, Y.-S. Ang, E. Radzinsky, P. Zhou, Y. Huang, A. Elfenbein, A. Foley, S. Magnitsky, and D. Srivastava. Regulation of cell cycle to stimulate adult cardiomyocyte proliferation and cardiac regeneration. *Cell*, 173(1):104–116, 2018.

[63] A. V. Naumova, V. P. Chacko, R. Ouwerkerk, L. Stull, E. Marbán, and R. G. Weiss. Xanthine oxidase inhibitors improve energetics and function after infarction in failing mouse hearts. *American Journal of Physiology-Heart and Circulatory Physiology*, 290(2):H837–H843, 2006.

[64] M. Dandel, H. Lehmkuhl, C. Knosalla, N. Suramelashvili, and R. Hetzer. Strain and strain rate imaging by echocardiography-basic concepts and clinical applicability. *Current cardiology reviews*, 5(2):133–148, 2009.

[65] W. C. R. F. International. Worldwide cancer data, global cancer statistics for the most common cancers. https://www.wcrf.org/dietandcancer/cancer-trends/worldwide-cancer-data, 2021-01-27.

[66] A. C. Society. Key statistics about kidney cancer. https://www.cancer.org/cancer/kidney-cancer/about/key-statistics.html, 2021-01-27.

[67] A. Znaor, J. Lortet-Tieulent, M. Laversanne, A. Jemal, and F. Bray. International variations and trends in renal cell carcinoma incidence and mortality. *European urology*, 67(3):519–530, 2015.

[68] G. Kovacs, M. Akhtar, B. J. Beckwith, P. Bugert, C. S. Cooper, et al. The heidelberg classification of renal cell tumours. *The Journal of Pathology: A Journal of the Pathological Society of Great Britain and Ireland*, 183(2):131–133, 1997.

[69] H. Zhou, S. Zheng, L. D. Truong, J. Y. Ro, A. G. Ayala, and S. S. Shen. Clear cell papillary renal cell carcinoma is the fourth most common histologic type of renal cell carcinoma in 290 consecutive nephrectomies for renal cell carcinoma. *Human pathology*, 45(1):59–64, 2014.

[70] L. Deng and D. Yu. Deep learning: methods and applications. *Foundations and trends in signal processing*, 7(3–4):197–387, 2014.

[71] J. M. Baker, L. Deng, J. Glass, S. Khudanpur, C.-H. Lee, N. Morgan, and D. O'Shaughnessy. Developments and directions in speech recognition and understanding, part 1 [dsp education]. *IEEE Signal processing magazine*, 26(3):75–80, 2009.

[72] J. V. Bouvrie. *Hierarchical learning: Theory with applications in speech and vision*. PhD thesis, Massachusetts Institute of Technology, 2009.

[73] Y. Bengio. *Learning deep architectures for AI*. Now Publishers Inc, 2009.

[74] L. Bottou and Y. LeCun. Large scale online learning. *Advances in neural information processing systems*, 16:217–224, 2004.

[75] J. Dean, G. S. Corrado, R. Monga, K. Chen, M. Devin, et al. Large scale distributed deep networks. 2012.

[76] X. Chen, A. Eversole, G. Li, D. Yu, and F. Seide. Pipelined back-propagation for context-dependent deep neural networks. In *Thirteenth Annual Conference of the International Speech Communication Association*, 2012.

[77] O. Vinyals and D. Povey. Krylov subspace descent for deep learning. In *Artificial Intelligence and Statistics*, pages 1261–1268. PMLR, 2012.

[78] B. Kingsbury, T. N. Sainath, and H. Soltau. Scalable minimum bayes risk training of deep neural network acoustic models using distributed hessian-free optimization. In *Thirteenth Annual Conference of the International Speech Communication Association*, 2012.

[79] Y. Bengio, P. Lamblin, D. Popovici, H. Larochelle, et al. Greedy layer-wise training of deep networks. *Advances in neural information processing systems*, 19:153, 2007.

[80] G. E. Hinton, S. Osindero, and Y.-W. Teh. A fast learning algorithm for deep belief nets. *Neural computation*, 18(7):1527–1554, 2006.

[81] B. Yegnanarayana. Artificial neural networks for pattern recognition. *Sadhana*, 19(2):189–238, 1994.

[82] M. van Gerven. Editorial: Artificial neural networks as models of neural information processing. front. comput. neurosci. 11: 114. doi: 10.3389/fncom. 2017.00114 editorial: Artificial neural networks as models of neural information processing marcel van gerven 1* and sander bohte 2 1 department of artificial intelligence. *Donders Institute for Brain, Cognition and Behaviour, Radboud University Nijmegen, Nijmegen, Netherlands, 2 Department of Machine Learning, Centrum Wiskunde and Informatica, Amsterdam, Netherlands*, 2017.

[83] W. S. McCulloch and W. Pitts. A logical calculus of the ideas immanent in nervous activity. *The bulletin of mathematical biophysics*, 5(4):115–133, 1943.

[84] A. R. Webb. *Statistical pattern recognition*. John Wiley & Sons, 2003.

[85] V. Nair and G. E. Hinton. Rectified linear units improve restricted boltzmann machines. In *Icml*, 2010.

[86] S. Lawrence, C. L. Giles, A. C. Tsoi, and A. D. Back. Face recognition: A convolutional neural-network approach. *IEEE transactions on neural networks*, 8(1):98–113, 1997.

[87] F. Rosenblatt. *The perceptron, a perceiving and recognizing automaton Project Para*. Cornell Aeronautical Laboratory, 1957.

[88] S. Sharma. Activation functions in neural networks. *towards data science*, 6, 2017.

[89] Y. Nesterov. A method for unconstrained convex minimization problem with the rate of convergence o (1/k^ 2). In *Doklady an ussr*, volume 269, pages 543–547, 1983.

[90] J. Duchi, E. Hazan, and Y. Singer. Adaptive subgradient methods for online learning and stochastic optimization. *Journal of machine learning research*, 12(7), 2011.

[91] M. D. Zeiler. Adadelta: an adaptive learning rate method. *arXiv preprint arXiv:1212.5701*, 2012.

[92] D. P. Kingma and J. Ba. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014.

[93] J. Kim, O. Sangjun, Y. Kim, and M. Lee. Convolutional neural network with biologically inspired retinal structure. *Procedia Computer Science*, 88:145–154, 2016.

[94] E. Gawehn, J. A. Hiss, and G. Schneider. Deep learning in drug discovery. *Molecular informatics*, 35(1):3–14, 2016.

[95] W. Yin, K. Kann, M. Yu, and H. Schütze. Comparative study of cnn and rnn for natural language processing. *arXiv preprint arXiv:1702.01923*, 2017.

[96] Y. LeCun, Y. Bengio, et al. Convolutional networks for images, speech, and time series. *The handbook of brain theory and neural networks*, 3361(10):1995, 1995.

[97] G. W. Taylor, R. Fergus, Y. LeCun, and C. Bregler. Convolutional learning of spatio-temporal features. In *European conference on computer vision*, pages 140–153. Springer, 2010.

[98] J. Long, E. Shelhamer, and T. Darrell. Fully convolutional networks for semantic segmentation. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 3431–3440, 2015.

[99] L.-C. Chen, G. Papandreou, F. Schroff, and H. Adam. Rethinking atrous convolution for semantic image segmentation. *arXiv preprint arXiv:1706.05587*, 2017.

[100] M. Abadi, A. Agarwal, P. Barham, E. Brevdo, Z. Chen, et al. TensorFlow: Large-scale machine learning on heterogeneous systems, 2015. Software available from tensorflow.org.

[101] Y. Jia, E. Shelhamer, J. Donahue, S. Karayev, J. Long, R. Girshick, S. Guadarrama, and T. Darrell. Caffe: Convolutional architecture for fast feature embedding. *arXiv preprint arXiv:1408.5093*, 2014.

[102] A. Paszke, S. Gross, F. Massa, A. Lerer, J. Bradbury, et al. Pytorch: An imperative style, high-performance deep learning library. In H. Wallach, H. Larochelle, A. Beygelzimer, F. d'Alché-Buc, E. Fox, and R. Garnett, editors, *Advances in Neural Information Processing Systems 32*, pages 8024–8035. Curran Associates, Inc., 2019.

[103] H. Abdeltawab, M. Shehata, A. Shalaby, F. Khalifa, A. Mahmoud, et al. A novel cnn-based cad system for early assessment of transplanted kidney dysfunction. *Scientific reports*, 9(1):1–11, 2019.

[104] V. Gulshan, L. Peng, M. Coram, M. C. Stumpe, D. Wu, et al. Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs. *Jama*, 316(22):2402–2410, 2016.

[105] A. Esteva, B. Kuprel, R. A. Novoa, J. Ko, S. M. Swetter, H. M. Blau, and S. Thrun. Dermatologist-level classification of skin cancer with deep neural networks. *nature*, 542(7639):115–118, 2017.

[106] H. Abdeltawab, F. Khalifa, F. Taher, N. S. Alghamdi, M. Ghazal, G. Beache, T. Mohamed, R. Keynton, and A. El-Baz. A deep learning-based approach for automatic segmentation and quantification of the left ventricle from cardiac cine mr images. *Computerized Medical Imaging and Graphics*, 81:101717, 2020.

[107] 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*, pages 424–432. Springer, 2016.

[108] S. B. Park, J. K. Kim, and K.-S. Cho. Complications of renal transplantation: ultrasonographic evaluation. *Journal of ultrasound in medicine: official journal of the American Institute of Ultrasound in Medicine*, 26(5):615–633, 2007.

[109] S. A. Akbar, S. Z. H. Jafri, M. A. Amendola, B. L. Madrazo, R. Salem, and K. G. Bis. Complications of renal transplantation. *Radiographics*, 25(5):1335–1356, 2005.

[110] M. Scott and R. Sells. Primary adenocarcinoma in a transplanted cadaveric kidney. *Transplantation*, 46(1):157–158, 1988.

[111] C. Hohnke. Vascular complications in 1,200 kidney transplantations. *Transplant Proc*, 19:3691–3692, 1987.

[112] T. H. Mathew. Recurrence of disease following renal transplantation. *American Journal of Kidney Diseases*, 12(2):85–96, 1988.

[113] J. R. Chapman, P. J. O'Connell, and B. J. Nankivell. Chronic renal allograft dysfunction. *Journal of the American society of nephrology*, 16(10):3015–3026, 2005.

[114] J. D. Pirsch, R. J. Ploeg, S. Gange, A. M. D'Alessandro, S. J. Knechtle, H. W. Sollinger, M. Kalayoglu, and F. O. Belzer. Determinants of graft survival after renal transplantation1. *Transplantation*, 61(11):1581–1586, 1996.

[115] J. Heaf and J. Iversen. Uses and limitations of renal scintigraphy in renal transplantation monitoring. *European journal of nuclear medicine*, 27(7):871–879, 2000.

[116] J. Xu, W. Xiao, L. Zhang, and M. Zhang. Value of diffusion-weighted MR imaging in diagnosis of acute rejection after renal transplantation. *Zhejiang da xue xue bao. Yi xue ban = Journal of Zhejiang University. Medical sciences*, 39(2):163–167, 2010.

[117] M. Abou-El-Ghar, T. El-Diasty, A. El-Assmy, H. Refaie, A. Refaie, and M. Ghoneim. Role of diffusion-weighted MRI in diagnosis of acute renal allograft dysfunction: A prospective preliminary study. *The British Journal of Radiology*, 85(1014):e206–e211, 2012.

[118] A. Kaul, R. K. Sharma, R. K. Gupta, H. Lal, A. Yadav, D. Bhadhuria, N. Prasad, A. Gupta, et al. Assessment of allograft function using diffusion-weighted magnetic resonance imaging in kidney transplant patients. *Saudi Journal of Kidney Diseases and Transplantation*, 25(6):1143, 2014.

[119] K. Wypych-Klunder, A. Adamowicz, A. Lemanowicz, W. Szczesny, Z. Włodarczyk, and Z. Serafin. Diffusion-weighted MR imaging of transplanted kidneys: Preliminary report. In *Polish Journal of Radiology*, 2014.

[120] S. Y. Park, C. K. Kim, B. K. Park, S. J. Kim, S. Lee, and W. Huh. Assessment of early renal allograft dysfunction with blood oxygenation level-dependent MRI and diffusion-weighted imaging. *European Journal of Radiology*, 83(12):2114–2121, 2014.

[121] G. Liu, F. Han, W. Xiao, Q. Wang, Y. Xu, and J. Chen. Detection of renal allograft rejection using blood oxygen level-dependent and diffusion weighted magnetic resonance imaging: A retrospective study. *BMC Nephrology*, 15(1):158, 2014.

[122] K. Hueper, A. A. Khalifa, J. H. Bräsen, V. Chieu, V. Dai, et al. Diffusion-weighted imaging and diffusion tensor imaging detect delayed graft function and correlate with allograft fibrosis in patients early after kidney transplantation. *Journal of Magnetic Resonance Imaging*, 44(1):112–121, 2016.

[123] K. Hueper, B. Hensen, M. Gutberlet, R. Chen, D. Hartung, et al. Kidney transplantation: Multiparametric functional magnetic resonance imaging for assessment of renal allograft pathophysiology in mice. *Investigative Radiology*, 51(1):58–65, 2016.

[124] W.-j. Fan, T. Ren, Q. Li, P.-l. Zuo, M.-m. Long, C.-b. Mo, L.-h. Chen, L.-x. Huang, and W. Shen. Assessment of renal allograft function early after transplantation with isotropic resolution diffusion tensor imaging. *European Radiology*, 26(2):567–575, 2016.

[125] P. Steiger, S. Barbieri, A. Kruse, M. Ith, and H. C. Thoeny. Selection for biopsy of kidney transplant patients by diffusion-weighted MRI. *European Radiology*, 27(10):4336–4344, 2017.

[126] Y. Xie, Y. Li, J. Wen, X. Li, Z. Zhang, et al. Functional evaluation of transplanted kidneys with reduced field-of-view diffusion-weighted imaging at 3T. *Korean Journal of Radiology*, 19(2):201–208, 2018.

[127] A. Pedraza, J. Gallego, S. Lopez, L. Gonzalez, A. Laurinavicius, and G. Bueno. Glomerulus classification with convolutional neural networks. In *Annual conference on medical image understanding and analysis*, pages 839–849. Springer, 2017.

[128] A. Krizhevsky, I. Sutskever, and G. E. Hinton. Imagenet classification with deep convolutional neural networks. *Advances in neural information processing systems*, 25:1097–1105, 2012.

[129] X. Yang, S. Y. Yeo, J. M. Hong, S. T. Wong, W. T. Tang, et al. A deep learning approach for tumor tissue image classification. *IASTED Biomedical Engineering*, 2016.

[130] V. B. Kolachalama, P. Singh, C. Q. Lin, D. Mun, M. E. Belghasem, J. M. Henderson, J. M. Francis, D. J. Salant, and V. C. Chitalia. Association of pathological fibrosis with renal survival using deep neural networks. *Kidney international reports*, 3(2):464–475, 2018.

[131] J. N. Marsh, M. K. Matlock, S. Kudose, T.-C. Liu, T. S. Stappenbeck, J. P. Gaut, and S. J. Swamidass. Deep learning global glomerulosclerosis in transplant kidney frozen sections. *IEEE transactions on medical imaging*, 37(12):2718–2728, 2018.

[132] E. Hollis, M. Shehata, F. Khalifa, M. Abou El-Ghar, T. El-Diasty, and A. El-Baz. Towards non-invasive diagnostic techniques for early detection of acute renal transplant rejection: A review. *The Egyptian Journal of Radiology and Nuclear Medicine*, 48(1):257–269, 2017.

[133] S. Ioffe and C. Szegedy. Batch normalization: Accelerating deep network training by reducing internal covariate shift. In *International conference on machine learning*, pages 448–456. PMLR, 2015.

[134] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov. Dropout: a simple way to prevent neural networks from overfitting. *The journal of machine learning research*, 15(1):1929–1958, 2014.

[135] A. Paszke, S. Gross, S. Chintala, and G. Chanan. Pytorch: Tensors and dynamic neural networks in python with strong gpu acceleration. *PyTorch: Tensors and dynamic neural networks in Python with strong GPU acceleration*, 6:3, 2017.

[136] D. Gray, A. Daar, H. Shepherd, D. Oliver, and P. Morris. Oral versus intravenous high-dose steroid treatment of renal allograft rejection: The big shot or not? *The Lancet*, 311(8056):117–118, 1978.

[137] C. Shinn, D. Malhotra, L. Chan, R. L. Cosby, and J. I. Shapiro. Time course of response to pulse methylprednisolone therapy in renal transplant recipients with acute allograft rejection. *American journal of kidney diseases*, 34(2):304–307, 1999.

[138] A. C. Webster, T. Pankhurst, F. Rinaldi, J. R. Chapman, and J. C. Craig. Monoclonal and polyclonal antibody therapy for treating acute rejection in kidney transplant recipients: a systematic review of randomized trial data. *Transplantation*, 81(7):953–965, 2006.

[139] F. Qureshi, H. Rabb, and B. L. Kasiske. Silent acute rejection during prolonged delayed graft function reduces kidney allograft survival. *Transplantation*, 74(10):1400–1404, 2002.

[140] H. C. Thoeny and F. De Keyzer. Diffusion-weighted MR imaging of native and transplanted kidneys. *Radiology*, 259(1):25–38, 2011.

[141] J. L. Zhang, E. E. Sigmund, H. Chandarana, H. Rusinek, Q. Chen, P.-H. Vivier, B. Taouli, and V. S. Lee. Variability of renal apparent diffusion coefficients: limitations of the monoexponential model for diffusion quantification. *Radiology*, 254(3):783–792, 2010.

[142] H.-J. Wittsack, R. S. Lanzman, C. Mathys, H. Janssen, U. Mödder, and D. Blondin. Statistical evaluation of diffusion-weighted imaging of the human kidney. *Magnetic resonance in medicine*, 64(2):616–622, 2010.

[143] L. Lu, J. R. Sedor, V. Gulani, J. R. Schelling, A. O'Brien, C. A. Flask, and K. M. Dell. Use of diffusion tensor MRI to identify early changes in diabetic nephropathy. *American journal of nephrology*, 34(5):476–482, 2011.

[144] P. A. Iaizzo. *Handbook of cardiac anatomy, physiology, and devices*. Springer Science & Business Media, 2009.

[145] S. Shah, G. Gnanasegaran, J. Sundberg-Cohon, and J. R. Buscombe. The heart: Anatomy, physiology and exercise physiology. In *Integrating cardiology for nuclear medicine physicians*, pages 3–22. Springer, 2009.

[146] A. Maceira, S. Prasad, M. Khan, and D. Pennell. Normalized left ventricular systolic and diastolic function by steady state free precession cardiovascular magnetic resonance. *Journal of Cardiovascular Magnetic Resonance*, 8(3):417–426, 2006.

[147] S. K. Bhatia. *Biomaterials for clinical applications*. Springer Science & Business Media, 2010.

[148] P. Ponikowski, A. A. Voors, S. D. Anker, H. Bueno, J. G. Cleland, et al. 2016 esc guidelines for the diagnosis and treatment of acute and chronic heart failure. *European heart journal*, 37(27):2129–2200, 2016.

[149] S. Herath, M. Harandi, and F. Porikli. Going deeper into action recognition: A survey. *Image and Vision Computing*, 60:4–21, 2017.

[150] C. Petitjean and J.-N. Dacher. A review of segmentation methods in short axis cardiac mr images. *Medical Image Analysis*, 15(2):169–184, 2011.

[151] D. A. Auger, X. Zhong, F. H. Epstein, E. M. Meintjes, and B. S. Spottiswoode. Semi-automated left ventricular segmentation based on a guide point model approach for 3d cine dense cardiovascular magnetic resonance. *Journal of Cardiovascular Magnetic Resonance*, 16(1):8, 2014.

[152] D. Grosgeorge, C. Petitjean, J.-N. Dacher, and S. Ruan. Graph cut segmentation with a statistical shape model in cardiac mri. *Computer Vision and Image Understanding*, 117(9):1027–1035, 2013.

[153] P. Peng, K. Lekadir, A. Gooya, L. Shao, S. E. Petersen, and A. F. Frangi. A review of heart chamber segmentation for structural and functional analysis using cardiac magnetic resonance imaging. *Magnetic Resonance Materials in Physics, Biology and Medicine*, 29(2):155–195, 2016.

[154] I. B. Ayed, H.-m. Chen, K. Punithakumar, I. Ross, and S. Li. Max-flow segmentation of the left ventricle by recovering subject-specific distributions via a bound of the bhattacharyya measure. *Medical Image Analysis*, 16(1):87–100, 2012.

[155] S. Queirós, D. Barbosa, B. Heyde, P. Morais, J. L. Vilaça, D. Friboulet, O. Bernard, and J. D'hooge. Fast automatic myocardial segmentation in 4d cine cmr datasets. *Medical Image Analysis*, 18(7):1115–1131, 2014.

[156] Y. Liu, G. Captur, J. C. Moon, S. Guo, X. Yang, S. Zhang, and C. Li. Distance regularized two level sets for segmentation of left and right ventricles from cine-mri. *Magnetic Resonance Imaging*, 34(5):699–706, 2016.

[157] L. Wang, M. Pei, N. C. Codella, M. Kochar, J. W. Weinsaft, J. Li, M. R. Prince, and Y. Wang. Left ventricle: fully automated segmentation based on spatiotemporal continuity and myocardium information in cine cardiac magnetic resonance imaging (lv-fast). *BioMed research international*, 2015, 2015.

[158] J. Ringenberg, M. Deo, V. Devabhaktuni, O. Berenfeld, P. Boyers, and J. Gold. Fast, accurate, and fully automatic segmentation of the right ventricle in short-axis cardiac mri. *Computerized Medical Imaging and Graphics*, 38(3):190–201, 2014.

[159] J. Woo, P. J. Slomka, C.-C. J. Kuo, and B.-W. Hong. Multiphase segmentation using an implicit dual shape prior: Application to detection of left ventricle in cardiac mri. *Computer Vision and Image Understanding*, 117(9):1084–1094, 2013.

[160] Y. Wu, Y. Wang, and Y. Jia. Segmentation of the left ventricle in cardiac cine mri using a shape-constrained snake model. *Computer Vision and Image Understanding*, 117(9):990–1003, 2013.

[161] W. Bai, W. Shi, C. Ledig, and D. Rueckert. Multi-atlas segmentation with augmented features for cardiac mr images. *Medical Image Analysis*, 19(1):98–109, 2015.

[162] P. V. Tran. A fully convolutional neural network for cardiac segmentation in short-axis mri. *arXiv preprint arXiv:1604.00494*, 2016.

[163] R. P. Poudel, P. Lamata, and G. Montana. Recurrent fully convolutional neural networks for multi-slice mri cardiac segmentation. In *Reconstruction, Segmentation, and Analysis of Medical Images*, pages 83–94. Springer, 2016.

[164] L. K. Tan, Y. M. Liew, E. Lim, and R. A. McLaughlin. Convolutional neural network regression for short-axis left ventricle segmentation in cardiac cine mr sequences. *Medical Image Analysis*, 39:78–86, 2017.

[165] O. Oktay, E. Ferrante, K. Kamnitsas, M. Heinrich, W. Bai, et al. Anatomically constrained neural networks (acnns): application to cardiac image enhancement and segmentation. *IEEE Transactions on Medical Imaging*, 37(2):384–395, 2018.

[166] Q. Zheng, H. Delingette, N. Duchateau, and N. Ayache. 3-d consistent and robust segmentation of cardiac images by deep learning with spatial propagation. *IEEE Transactions on Medical Imaging*, 37(9):2137–2148, 2018.

[167] K. Simonyan and A. Zisserman. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*, 2014.

[168] W. Bai, M. Sinclair, G. Tarroni, O. Oktay, M. Rajchl, et al. Automated cardiovascular magnetic resonance image analysis with fully convolutional networks. *Journal of Cardiovascular Magnetic Resonance*, 20(1):65, Sep 2018.

[169] Q. Tao, W. Yan, Y. Wang, E. H. Paiman, D. P. Shamonin, et al. Deep learning–based method for fully automatic quantification of left ventricle function from cine mr images: a multivendor, multicenter study. *Radiology*, 290(1):81–88, 2019.

[170] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna. Rethinking the inception architecture for computer vision. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 2818–2826, 2016.

[171] K. Kamnitsas, W. Bai, E. Ferrante, S. McDonagh, M. Sinclair, et al. Ensembles of multiple models and architectures for robust brain tumour segmentation. In A. Crimi, S. Bakas, H. Kuijf, B. Menze, and M. Reyes, editors, *Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries*, page 450–462. Springer International Publishing, 2018.

[172] K. He, X. Zhang, S. Ren, and J. Sun. Delving deep into rectifiers: Surpassing human-level performance on imagenet classification. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 1026–1034, 2015.

[173] O. Bernard, A. Lalande, C. Zotti, F. Cervenansky, X. Yang, et al. Deep learning techniques for automatic mri cardiac multi-structures segmentation and diagnosis: Is the problem solved? *IEEE Transactions on Medical Imaging*, 37(11):2514–2525, Nov 2018.

[174] N. Otsu. A threshold selection method from gray-level histograms. *IEEE Transactions on Systems, Man, and Cybernetics*, 9(1):62–66, Jan 1979.

[175] H. Noh, S. Hong, and B. Han. Learning deconvolution network for semantic segmentation. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 1520–1528, 2015.

[176] J. M. Bland and D. Altman. Statistical methods for assessing agreement between two methods of clinical measurement. *The Lancet*, 327(8476):307–310, 1986.

[177] J. M. Wolterink, T. Leiner, M. A. Viergever, and I. Išgum. Automatic segmentation and disease classification using cardiac cine mr images. In *International Workshop on Statistical Atlases and Computational Models of the Heart*, pages 101–110. Springer, 2017.

[178] F. Sardanelli, M. Quarenghi, G. Di Leo, L. Boccaccini, and A. Schiavi. Segmentation of cardiac cine mr images of left and right ventricles: interactive semiautomated methods and manual contouring by two readers with different education and experience. *Journal of Magnetic Resonance Imaging: An Official Journal of the International Society for Magnetic Resonance in Medicine*, 27(4):785–792, 2008.

[179] A. Suinesiaputra, M. M. Sanghvi, N. Aung, J. M. Paiva, F. Zemrak, et al. Fully-automated left ventricular mass and volume mri analysis in the uk biobank population cohort: evaluation of initial results. *The International Journal of Cardiovascular Imaging*, 34(2):281–291, Feb 2018.

[180] E. Grinias and G. Tziritas. Fast fully-automatic cardiac segmentation in mri using mrf model optimization, substructures tracking and b-spline smoothing. In *International Workshop on Statistical Atlases and Computational Models of the Heart*, pages 91–100. Springer, 2017.

[181] Y. LeCun, Y. Bengio, and G. Hinton. Deep learning. *Nature*, 521(7553):436, 2015.

[182] J. Ker, L. Wang, J. Rao, and T. Lim. Deep learning applications in medical image analysis. *IEEE Access*, 6:9375–9389, 2018.

[183] H. Abdeltawab, M. Shehatal, A. Shalaby, S. Mesbah, M. El-Baz, et al. A new 3d cnn-based cad system for early detection of acute renal transplant rejection. In *2018*

24th International Conference on Pattern Recognition (ICPR), pages 3898–3903, Aug 2018.

[184] M. Khened, V. A. Kollerathu, and G. Krishnamurthi. Fully convolutional multi-scale residual densenets for cardiac segmentation and automated cardiac diagnosis using ensemble of classifiers. *Medical Image Analysis*, 51:21 – 45, 2019.

[185] C. Zotti, Z. Luo, A. Lalande, and P. Jodoin. Convolutional neural network with shape prior applied to cardiac mri segmentation. *IEEE Journal of Biomedical and Health Informatics*, 23(3):1119–1128, May 2019.

[186] T. Leiner, D. Rueckert, A. Suinesiaputra, B. Baeßler, R. Nezafat, I. Išgum, and A. A. Young. Machine learning in cardiovascular magnetic resonance: basic concepts and applications. *Journal of Cardiovascular Magnetic Resonance*, 21(1):61, 2019.

[187] A. A. Young, D. J. Medway, C. A. Lygate, S. Neubauer, and J. E. Schneider. Accelerating global left-ventricular function assessment in mice using reduced slice acquisition and three-dimensional guide-point modelling. *Journal of cardiovascular magnetic resonance: official journal of the Society for Cardiovascular Magnetic Resonance*, 13(1):49–49, Sep 2011.

[188] L. J. Watson, H. T. Facundo, G. A. Ngoh, M. Ameen, R. E. Brainard, K. M. Lemma, B. W. Long, S. D. Prabhu, Y.-T. Xuan, and S. P. Jones. O-linked β-n-acetylglucosamine transferase is indispensable in the failing heart. *Proceedings of the National Academy of Sciences*, 107(41):17797–17802, 2010.

[189] M. Wysoczynski, S. Dassanayaka, A. Zafir, S. Ghafghazi, B. W. Long, C. Noble, A. M. DeMartino, K. R. Brittian, R. Bolli, and S. P. Jones. A new method to stabilize c-kit expression in reparative cardiac mesenchymal cells. *Frontiers in cell and developmental biology*, 4:78, 2016.

[190] S. Dassanayaka, K. R. Brittian, A. Jurkovic, L. A. Higgins, T. N. Audam, et al. E2f1 deletion attenuates infarct-induced ventricular remodeling without affecting o-glcnacylation. *Basic research in cardiology*, 114(4):28, 2019.

[191] Q. Dong, S. Gong, and X. Zhu. Imbalanced deep learning by minority class incremental rectification. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2018.

[192] S. S. M. Salehi, D. Erdogmus, and A. Gholipour. Tversky loss function for image segmentation using 3d fully convolutional deep networks. In Q. Wang, Y. Shi, H.-I. Suk, and K. Suzuki, editors, *Machine Learning in Medical Imaging*, pages 379–387, Cham, 2017. Springer International Publishing.

[193] S. E. Jones, B. R. Buchbinder, and I. Aharon. Three-dimensional mapping of cortical thickness using laplace's equation. *Human brain mapping*, 11(1):12–32, 2000.

[194] A. Elnakib, G. M. Beache, G. Gimel'farb, and A. El-Baz. New automated markov–gibbs random field based framework for myocardial wall viability quantification on agent enhanced cardiac magnetic resonance images. *The international journal of cardiovascular imaging*, 28(7):1683–1698, 2012.

[195] F. Khalifa, M. Abou El-Ghar, B. Abdollahi, H. B. Frieboes, T. El-Diasty, and A. El-Baz. A comprehensive non-invasive framework for automated evaluation of acute

renal transplant rejection using dce-mri. *NMR in Biomedicine*, 26(11):1460–1470, 2013.

[196] J. Bonet and R. D. Wood. *Nonlinear continuum mechanics for finite element analysis*. Cambridge university press, 1997.

[197] C. Bouman and K. Sauer. A generalized gaussian image model for edge-preserving map estimation. *IEEE Transactions on image processing*, 2(3):296–310, 1993.

[198] M. Öner and İ. Deveci Kocakoç. Jmasm 49: A compilation of some popular goodness of fit tests for normal distribution: Their algorithms and matlab codes (matlab). *Journal of Modern Applied Statistical Methods*, 16(2):30, 2017.

[199] E. Waks, J. L. Prince, and A. S. Douglas. Cardiac motion simulator for tagged mri. In *Proceedings of the workshop on mathematical methods in biomedical image analysis*, pages 182–191. IEEE, 1996.

[200] T. Arts, W. Hunter, A. Douglas, A. Muijtjens, and R. Reneman. Description of the deformation of the left ventricle by a kinematic model. *Journal of biomechanics*, 25(10):1119–1127, 1992.

[201] T. Mansi, J.-M. Peyrat, M. Sermesant, H. Delingette, J. Blanc, Y. Boudjemline, and N. Ayache. Physically-constrained diffeomorphic demons for the estimation of 3d myocardium strain from cine-mri. In *International Conference on Functional Imaging and Modeling of the Heart*, pages 201–210. Springer, 2009.

[202] H. Gao, A. Allan, C. McComb, X. Luo, and C. Berry. Left ventricular strain and its pattern estimated from cine cmr and validation with dense. *Physics in Medicine & Biology*, 59(13):3637, 2014.

[203] F. H. Sheehan, E. L. Bolson, H. T. Dodge, D. G. Mathey, J. Schofer, and H. Woo. Advantages and applications of the centerline method for characterizing regional ventricular function. *Circulation*, 74(2):293–305, 1986.

[204] M. Nitzken, G. Beache, A. Elnakib, F. Khalifa, G. Gimel'farb, and A. El-Baz. Accurate modeling of tagged cmr 3d image appearance characteristics to improve cardiac cycle strain estimation. In *2012 19th IEEE International Conference on Image Processing*, pages 521–524. IEEE, 2012.

[205] 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*, 39(6):1137–1149, 2016.

[206] 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*, pages 770–778, 2016.

[207] R. Girshick, J. Donahue, T. Darrell, and J. Malik. Rich feature hierarchies for accurate object detection and semantic segmentation. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 580–587, 2014.

[208] A. Cruz-Roa, H. Gilmore, A. Basavanhally, M. Feldman, S. Ganesan, N. N. Shih, J. Tomaszewski, F. A. González, and A. Madabhushi. Accurate and reproducible invasive breast cancer detection in whole-slide images: A deep learning approach for quantifying tumor extent. *Scientific reports*, 7(1):1–14, 2017.

[209] Z. Han, B. Wei, Y. Zheng, Y. Yin, K. Li, and S. Li. Breast cancer multi-classification from histopathological images with structured deep learning model. *Scientific reports*, 7(1):1–10, 2017.

[210] N. Coudray, P. S. Ocampo, T. Sakellaropoulos, N. Narula, M. Snuderl, D. Fenyö, A. L. Moreira, N. Razavian, and A. Tsirigos. Classification and mutation prediction from non–small cell lung cancer histopathology images using deep learning. *Nature medicine*, 24(10):1559–1567, 2018.

[211] S. Wang, A. Chen, L. Yang, L. Cai, Y. Xie, J. Fujimoto, A. Gazdar, and G. Xiao. Comprehensive analysis of lung cancer pathology images to discover tumor shape and boundary features that predict survival outcome. *Scientific reports*, 8(1):1–9, 2018.

[212] P. Khosravi, E. Kazemi, M. Imielinski, O. Elemento, and I. Hajirasouliha. Deep convolutional neural networks enable discrimination of heterogeneous digital pathology images. *EBioMedicine*, 27:317–328, 2018.

[213] O. J. del Toro, M. Atzori, S. Otálora, M. Andersson, K. Eurén, M. Hedlund, P. Rönnquist, and H. Müller. Convolutional neural networks for an automatic classification of prostate tissue slides with high-grade gleason score. In *Medical Imaging 2017: Digital Pathology*, volume 10140, page 101400O. International Society for Optics and Photonics, 2017.

[214] D. Karimi, G. Nir, L. Fazli, P. C. Black, L. Goldenberg, and S. E. Salcudean. Deep learning-based gleason grading of prostate cancer from histopathology images—role of multiscale decision aggregation and data augmentation. *IEEE journal of biomedical and health informatics*, 24(5):1413–1426, 2019.

[215] S. Tabibu, P. Vinod, and C. Jawahar. Pan-renal cell carcinoma classification and survival prediction from histopathology images using deep learning. *Scientific reports*, 9(1):1–9, 2019.

[216] K. Zuiderveld. Contrast limited adaptive histogram equalization. In *Graphics gems IV*, pages 474–485. Academic Press Professional, Inc., 1994.

[217] J. Besag. On the statistical analysis of dirty pictures. *Journal of the Royal Statistical Society: Series B (Methodological)*, 48(3):259–279, 1986.

[218] M. Abadi, P. Barham, J. Chen, Z. Chen, A. Davis, et al. Tensorflow: A system for large-scale machine learning. In *12th {USENIX} symposium on operating systems design and implementation ({OSDI} 16)*, pages 265–283, 2016.

CURRICULUM VITAE

Hisham Abouelseoud Elsayem Abdeltawab

Paul C. Lutz Hall, Room # 429

BioImaging Laboratory

Department of Bioengineering

University of Louisville, Louisville, KY, USA

E-mail: *hisham.abdeltawab@louisville.edu*,

habdeltawab1@gmail.com

Tel: (502) 295-7681

Education

2017-Present   Ph.D. Student, Department of Bioengineering, University of Louisville, Louisville, KY 40292, USA

PhD Candidate since August 2020.

2015   M.Sc., Department of Biomedical and Systems Engineering, Cairo University, Cairo, Egypt.

M.Sc. Thesis: *Image restoration techniques in super-resolution reconstruction of MRI images*- Dr. Yasser Kadah, Thesis Advisor.

2012   B.Sc., Department of Biomedical and Systems Engineering, Cairo University, Cairo, Egypt.

Sr. Project: *SurgiLink: Surgical Guidance through Mobile Devices*.

Experience

2017–2022   Graduate Research Assistant, BioImaging laboratory, Department of Bioengineering, University of Louisville, Louisville, KY 40292, USA.

2012–2017   Graduate Research Assistant, Department of Biomedical and Systems Engineering, Cairo University, Cairo, Egypt.

Professional Affiliations and Training

- Student Member, Institute of Electrical and Electronics Engineers (IEEE).
- Member, Biomedical Engineering Society (BMES).

Awards and Recognition

- Best Doctoral Journal Article Award from the School of Interdisciplinary and Graduate Studies, University of Louisville, 2019.
- Graduate Travel Award to attend 2018 Biomedical Engineering Society Annual Event, University of Louisville, 2018.
- University Fellowship Award, School of Interdisciplinary and Graduate Studies, University of Louisville, Aug. 2017 - Aug. 2019.
- Ranked the 1st in the bachelor's graduation of Biomedical Engineering and Systems department, Cairo University, 2012.

Publications

- **Journal Articles**

    1. **H. Abdeltawab**, F. Khalifa, K. Hammouda, J. M. Miller, M. Meki, Q. Ou, A. El-Baz, and Tamer Mohamed. "Artificial Intelligence Based Framework to Quantify the Cardiomyocyte Structural Integrity in Heart Slices." *Cardiovas-*

cular Engineering and Technology (2021): 1-11.

2. S. Haggag, F. Khalifa, **H. Abdeltawab**, A. Elnakib, M. Ghazal, M. Mohamed, H. Sandhu, N. Alghamdi, and A. El-Baz. "An Automated CAD System for Accurate Grading of Uveitis Using Optical Coherence Tomography Images." *Sensors* 21, no. 16 (2021): 5457.

3. K. Hammouda, F. Khalifa, **H. Abdeltawab**, A. Elnakib, G. A. Giridharan, et al. "A New Framework for Performing Cardiac Strain Analysis from Cine MRI Imaging in Mice," *Sci Rep*, vol. 10, no. 1, p. 7725, Dec. 2020.

4. **H. Abdeltawab**, F. Khalifa, F. Taher, N.S. Alghamdi, M. Ghazal, G. Beache, et al. A deep learning-based approach for automatic segmentation and quantification of the left ventricle from cardiac cine MR images. *Comput. Med. Imag. Graph.*, 81 (2020), p. 101717

5. **Hisham Abdeltawab**, Mohamed Shehata, Ahmed Shalaby, Fahmi Khalifa, and et al. A novel cnn-based cad system for early assessment of transplanted kidney dysfunction. *Scientific Reports*,9(1):5948, Apr 2019

- **Peer-Reviewed Conference Proceedings**

1. **H. Abdeltawab**, F. Khalifa, M. Ghazal, A. Haddad, and et al. Predictive biomarkers for neoad-juvant chemotherapy response in muscle-invasive bladder cancer: A survey. In *2018 IEEE Inter-national Symposium on Signal Processing and Information Technology (ISSPIT)*, pages 270–275,Dec 2018

2. **H. Abdeltawab**, M. Shehatal, A. Shalaby, S. Mesbah, and et al. A new 3d cnn-based cad system for early detection of acute renal transplant rejection. In *2018 24th International Conference on Pattern Recognition (ICPR)*, pages 3898–3903, Aug 2018

3. **H. A. Alsayem** and Y. M. Kadah. Image restoration techniques in super-resolution reconstructionof mri images. In *2016 33rd National Radio Science*