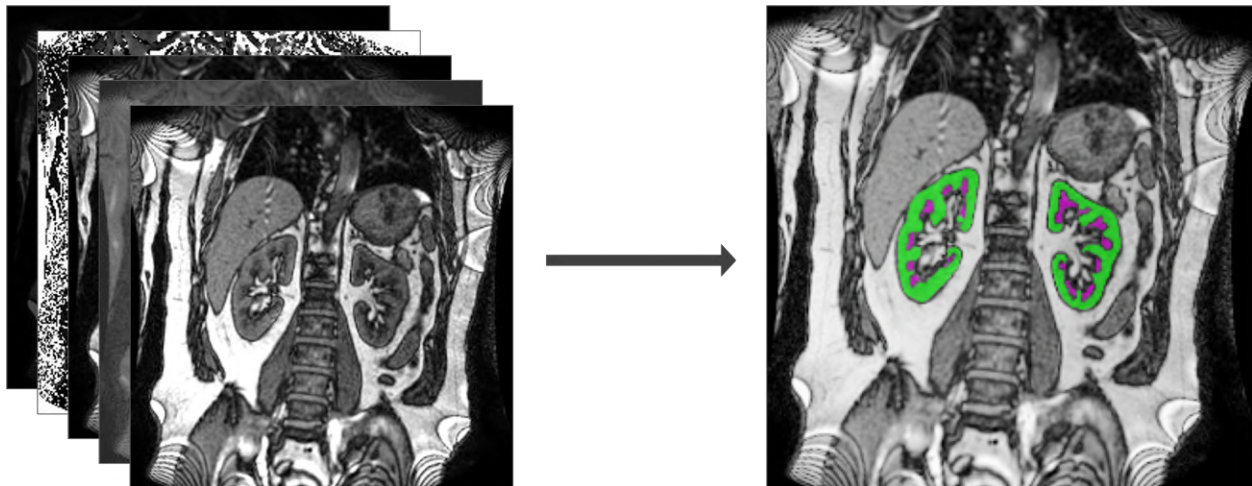




CHALMERS
UNIVERSITY OF TECHNOLOGY



Deep Learning-based Segmentation of Kidneys from MR Images

A Multi-Channel Approach for Automated Segmentation and Quantification in Multi-parametric MRI

Master's Thesis in Biomedical Engineering

CECILIA NORDBERG
VIKTOR LINDFORS

DEPARTMENT OF ELECTRICAL ENGINEERING (E2)

CHALMERS UNIVERSITY OF TECHNOLOGY
Gothenburg, Sweden 2025
www.chalmers.se

MASTER'S THESIS 2025

Deep Learning-based Segmentation of Kidneys from MR Images

A Multi-Channel Approach for Automated Segmentation and
Quantification in Multi-parametric MRI

CECILIA NORDBERG
VIKTOR LINDFORS



CHALMERS
UNIVERSITY OF TECHNOLOGY

Department of Electrical Engineering (E2)
CHALMERS UNIVERSITY OF TECHNOLOGY
Gothenburg, Sweden 2025

Deep Learning-based Segmentation of Kidneys from MR Images
A Multi-Channel Approach for Automated Segmentation and Quantification in
Multi-parametric MRI
CECILIA NORDBERG
VIKTOR LINDFORS

© CECILIA NORDBERG, 2025.

© VIKTOR LINDFORS, 2025.

Supervisor: Bettina Selig, Antaros Medical AB
Supervisor: Kanishka Sharma, Antaros Medical AB
Examiner: Ida Häggström, Electrical Engineering (E2)

Master's Thesis 2025
Department of Electrical Engineering (E2)
Chalmers University of Technology
SE-412 96 Gothenburg
Telephone +46 31 772 1000

Cover: Segmentation of the renal cortex (green) and medulla (magenta) using a 2D ResUNet. The model was trained on multi-channel input from T1-MOLLI images acquired at varying inversion times.

Typeset in L^AT_EX
Printed by Chalmers Reproservice
Gothenburg, Sweden 2025

Deep Learning-based Segmentation of Kidneys from MR Images
A Multi-Channel Approach for Automated Segmentation and Quantification in
Multi-parametric MRI
CECILIA NORDBERG
VIKTOR LINDFORS
Department of Electrical Engineering (E2)
Chalmers University of Technology

Abstract

Chronic kidney disease (CKD) is a progressive condition affecting millions worldwide, and accurate assessment of kidney structure is essential for early diagnosis and monitoring disease progression. Magnetic Resonance Imaging (MRI) has emerged as a powerful, non-invasive technique for visualizing subtle structural and functional changes of the kidneys, providing insights into disease progression and severity. However, manual segmentation of MRI data is both time-consuming and prone to inter- and intra-observer variability, highlighting the need for automated methods.

This thesis presents a deep learning-based approach for automated segmentation of the renal parenchyma, cortex, and medulla using multi-channel and multi-modal MRI data. A 2D ResUNet architecture was implemented with the Medical Open Network for AI (MONAI) framework and trained on a dataset of 37 MRI scans from CKD patients. Two approaches were evaluated: a multi-channel model utilizing T1-weighted Modified Look-Locker Inversion Recovery (T1-MOLLI) images at multiple inversion times, and a multi-modal model incorporating diffusion-weighted imaging (DWI) and T2*-weighted image data. While the multi-channel T1-MOLLI model demonstrated strong agreement with manual annotations, achieving Dice scores of 0.9089 for parenchyma and 0.8552 for cortex, the multi-modal approach underperformed due to spatial misalignment between input images and reference labels.

The proposed segmentation pipeline also enabled reliable quantification of renal parenchyma and cortex volumes, and showed potential for quantifying tissue-specific parametric values relevant to CKD monitoring. However, the reliability of these measurements were highly dependent of the models segmentation performance. Overall, the findings highlight the potential of using deep learning models' with multi-channel MRI input for improving kidney segmentation, serving as a tool to support clinical image analysis workflows and reduce manual effort.

Keywords: deep learning, image segmentation, MRI, kidneys, convolutional neural networks, ResUNet, multi-channel images, chronic kidney disease.

Acknowledgements

We would like to thank Antaros Medical AB for giving us the opportunity to carry out our thesis project within their organization. A special thanks to our supervisors, Bettina Selig and Kanishka Sharma, for your ongoing support throughout the project. You have continuously encouraged us to explore different perspectives and shared your knowledge with us during this spring, which we truly appreciated. Additionally, we would like to extend our thanks to Carl Sjöberg for making this thesis possible, and for the warm welcome and support from the start. To all the colleagues at Antaros, thank you for welcoming us and making us feel like part of the team.

We would also like to thank our examiner, Ida Häggström, for guiding us during our thesis work. Finally, as this thesis marks the conclusion of our five years at Chalmers University of Technology, we would like to give a big thank you to our friends and family, and to everyone who has contributed to our experience during this time.

Cecilia Nordberg & Viktor Lindfors, Gothenburg, June 2025

List of Acronyms

Below is the list of acronyms that have been used throughout this thesis listed in alphabetical order:

ADC	Apparent Diffusion Coefficient
CKD	Chronic Kidney Disease
CMD	Corticomedullary Differentiation
CNN	Convolutional Neural Network
CT	Computer Tomography
DNN	Deep Neural Network
DWI	Diffusion-Weighted Imaging
DX	Dextral
ESRD	End-stage Renal Disease
FN	False Negative
FOV	Field of View
FP	False Positive
GPU	Graphical Processing Unit
IoU	Intersection over Union
MOLLI	Modified Look-Locker Inversion Recovery
MONAI	Medical Open Network for Artificial Intelligence
MR	Magnetic Resonance
MRI	Magnetic Resonance Imaging
ReLU	Rectified Linear Unit
RF	Radio Frequency
ROI	Region of Interest
SD	Standard Deviation
SIN	Sinistral
T1w	T1-weighted
T2*w	T2*-weighted
TE	Echo Time
TI	Inversion Time
TP	True Positive
TR	Repetition Time

Contents

List of Acronyms	ix
List of Figures	xiii
List of Tables	xvii
1 Introduction	1
1.1 Aim	2
1.2 Scope and Limitations	2
1.3 Related Work	2
2 Theory	5
2.1 Kidney Anatomy and Physiology	5
2.1.1 Chronic Kidney Disease	6
2.2 Magnetic Resonance Imaging	7
2.2.1 Physical Principles of MRI	7
2.2.2 MRI Sequences	7
2.2.3 MRI Artifacts	9
2.3 Deep Learning for Image Segmentation	10
2.3.1 Basic Concepts of Deep Learning	10
2.3.2 Convolutional Neural Networks	11
2.3.3 U-Net	14
2.3.4 ResUNet	15
2.3.5 Loss Functions for Medical Image Segmentation	15
2.3.6 Evaluation Metrics	16
2.4 Multi-channel Image Segmentation	17
3 Methods	19
3.1 Data Description	20
3.2 Data Exclusion	21
3.3 Dataset Split	21
3.4 Data Pre-processing	21
3.4.1 Creation of Ground Truth Masks	22
3.4.2 Brightness and Intensity Scaling	22
3.4.3 Splitting into 2D Slices	23
3.4.4 Augmentations	23
3.5 Network Configuration	24

3.6	Model Implementation and Experimental Setup	25
3.6.1	Single-Channel T1-MOLLI-based Segmentation	25
3.6.2	Multi-Channel T1-MOLLI-based Segmentation	25
3.6.3	Multi-Modal Segmentation Integrating DWI and T2*	26
3.7	Training and Evaluation	26
3.8	Post-processing	26
3.9	Volume Quantification	27
3.10	Quantification of Parametric Values	28
4	Results	29
4.1	Data Exploration	29
4.2	Single-channel T1 MOLLI-based Segmentation	30
4.3	Multi-channel T1 MOLLI-based Segmentation	32
4.3.1	Performance of Fine-tuned T1-MOLLI Model	34
4.3.2	Quantification of Kidney Volume	38
4.3.3	Quantification of Median T1 Value	42
4.4	Multi-Modal Kidney Segmentation	43
4.4.1	Quantification of ADC and T2* value	46
5	Discussion	49
5.1	Comparison of 2D and 3D Segmentation	49
5.2	The Effect of Multi-Channel Inputs	50
5.3	The Effect of Multi-modal Integration	51
5.4	Model Strengths and Limitations	51
5.5	Quantifying Volume and Tissue Parameters	53
5.6	Comparison to Existing Works	53
5.7	Future Work	54
6	Conclusion	57
	Bibliography	59
A	Appendix A	I
B	Appendix B	III

List of Figures

2.1	Anatomical illustration of the normal renal structure from a coronal orientation of the left kidney. The renal parenchyma tissue includes cortex and pyramids of medullas. Created with BioRender.com. . . .	6
2.2	Examples of common MRI artifacts that include motion artifact, zebra stripe artifact, and partial volume effect. Images (a) and (b) are adapted from Stadler et al. [28], while (c) is courtesy of Antaros Medical AB.	9
2.3	An example of convolutional layer with kernel size 3×3 , a stride of 1, and no padding.	12
2.4	Max pooling layer with kernel size 2×2 and a stride of 2.	12
2.5	Comparison of activation functions: (a) ReLU, and (b) PReLU. . . .	14
2.6	Illustration of activation functions commonly used for classification: (a) Sigmoid function, and (b) Softmax function.	14
2.7	Illustration of the relationship between ground truth and predicted classifications. The overlapping region represents true positives (TP), where the prediction correctly matches the ground truth. The left (blue) region denotes false negatives (FN), where actual positives were missed. The right (red) region indicates false positives (FP), where the background is incorrectly segmented as foreground.	16
2.8	Example of a multi-channel input image with five different image channels from renal MRI data. The channels provide different views of the same anatomy with complementary tissue contrast information.	18
3.1	Overview of proposed pipeline for kidney segmentation.	19
3.2	Example of ground truth segmentation and corresponding binary masks after applied transformations. From left to right: parenchyma ground truth, parenchyma binary mask, cortex ground truth, and cortex binary mask. The manually segmented ground image shows the left kidney (cyan) and right kidney (green), annotated by image analysts.	22
3.3	Overview of proposed 2D ResUNet architecture. The left side shows the downsampling path with residual units, where convolutional channels increase from 16 to 256 to extract features. The right side correspond to the upsampling path, where channels decrease from 256 to the number of output classes, combined with skip connections. . . .	24

3.4	Visualization of the <code>RemoveSmallObjects</code> transform in MONAI, here removing clusters with 100 pixels or less. Figure adapted from MONAI [45].	27
3.5	Effect of erosion using a 3×3 structural element.	28
4.1	Examples of MRI artifacts and image distortions observed in the dataset. Yellow arrows highlight regions impacted by zebra striping and poorly defined kidney boundaries.	29
4.2	Comparison of segmentation results for (a) renal parenchyma and (b) renal cortex using 2D ResUNet and 3D ResUNet models. The figures illustrate the original image, ground truth segmentation label, and overlay of image, label and model predictions, where true positives (TP) are represented in green, false positives (FP) in red, and false negatives (FN) in blue.	31
4.3	Training and validation performance of the proposed multi-label 2D ResUNet, showing Dice loss and evaluation metrics progress during training (Adam optimizer, learning rate of 10^{-4}). Dice score and IoU are reported separately for the parenchyma and cortex, as well as averaged across both regions.	35
4.4	Representative examples of parenchyma segmentation results from two test cases: a) scan acquired at site A, and b) scan from site B. Each figure shows selected image slices with the predicted segmentation mask, as well as an overlay on the TI 1400 ms image. The color coding in the overlays is as follows: green - true positives, blue - false negatives, and red - false positives.	36
4.5	Examples of renal parenchyma segmentation results, illustrating classification outcomes where the model oversegments in the renal pelvis region. The color coding is as follows: green - true positives, blue - false negatives, and red - false positives.	37
4.6	Examples of renal parenchyma segmentation results, illustrating classification outcomes where the model undersegments in the most anterior and posterior parts of the kidney. The color coding is as follows: green - true positives, blue - false negatives, and red - false positives.	37
4.7	Examples of renal parenchyma segmentation results, illustrating classification outcomes where the model falsely predicts other structures in background-only slices. Red indicates false positives.	37
4.8	Representative examples of cortex segmentation results from two test cases: a) a scan acquired at site A, and b) a scan from site B. Each figure shows selected image slices with the predicted segmentation mask, as well as an overlay on the TI 1400 ms image. The color coding in the overlays is as follows: green - true positives, blue - false negatives, and red - false positives.	38
4.9	Correlation of ground truth and predicted volume for parenchyma and cortex. The identity line represents perfect correlation between the CNN-predicted and ground-truth segmentation. Grey area corresponds to a 5% volume difference.	39

4.10	Bland-Altman plots of agreement for volume prediction and ground truth volume for parenchyma and cortex. Mean and standard deviation (SD) are calculated globally across all data sets. The solid line represents the mean, while the dashed lines indicate the limits of agreement, calculated as the mean ± 1.96 times the SD.	40
4.11	Example of post processing / erosion of segmentation mask for (a) cortex and (b) medulla. From left to right: original segmentation mask and resulting segmentation mask after erosion with a square structuring element of increasing size.	42
4.12	Correlation of predicted and delivered median T1 value for cortex and medulla. The identity line represents perfect correlation between the CNN-predicted and delivered values. Grey area corresponds to a 5% difference in predicted value.	43
4.13	Example of renal parenchyma segmentation results from multi-modal input, illustrating test cases with good image-label alignment. The ground truth labels closely match the kidney boundaries visible in the underlying T2*w image, resulting in fewer false detections. The color coding in the overlays is as follows: green - true positives, blue - false negatives, and red - false positives.	45
4.14	Example of renal parenchyma segmentation results from multi-modal input, illustrating test cases with poor image-label alignment. The segmentation overlays show how the ground truth label either extends beyond the visible kidney region or covers only a portion of the kidney as seen in the underlying T2*w image. The color coding in the overlays is as follows: green - true positives, blue - false negatives, and red - false positives.	45
4.15	Examples of cortex segmentation results using the multi-modal segmentation model, demonstrating classification performance on test images. Images show an overlay of ground truth label and prediction output for several slices, superimposed to T2*w images acquired at TE = 3 ms. The color coding is as follows: green - true positives, blue - false negatives, and red - false positives.	46
4.16	Example of manually delineated regions and generated regions: (a) Manually delineated regions for cortex and medulla on T2* map, (b) generated mask using erosion for cortex, (c) generated mask using erosion for medulla.	46
4.17	Correlation of predicted and delivered median ADC value for cortex and medulla. Model is a proposed multi-modal model (2D ResUNet). The identity line represents perfect correlation between the CNN-predicted and delivered values. Grey area corresponds to a 5% difference in predicted value.	47
4.18	Correlation of predicted and delivered median T2star value for cortex and medulla. Model is a proposed multi-modal model (2D ResUNet). The identity line represents perfect correlation between the CNN-predicted and delivered values. Grey area corresponds to a 5% difference in predicted value.	48

A.1	Example images for T1-MOLLI images acquired with increasing inversion times, ranging from 174 ms to 2574 ms.	I
A.2	Example images for T2*w images acquired with increasing echo times, ranging from 3 ms to 62 ms.	I
A.3	Example images for DWI acquired with increasing b-values, ranging from 0 to 500 s/mm ² . Left to right: b=0, b=50, b=200, b=500. . . .	I
A.4	Examples of parametric maps derived from different MRI sequences. From left to right: T1 map from T1-MOLLI, ADC map from DWI images, and T2* map from T2* mapping.	II

List of Tables

3.1	Summary of imaging parameters by modality, reflecting the variability in acquisition protocols depending on the modality and the scanner setup.	20
4.1	6-fold cross validation results for 2D and 3D ResUNet models. Performance metrics are shown as mean \pm standard deviations across all folds.	30
4.2	Parenchyma segmentation performance using single-channel input images of different inversion time (TI).	32
4.3	Cortex segmentation performance using single-channel input images of different inversion time (TI).	32
4.4	Comparison of the top five multi-channel input combinations for segmentation of the renal parenchyma.	33
4.5	Comparison of the top five multi-channel input combinations for segmentation of the renal cortex.	33
4.6	Comparison of the performance of single-label and multi-label segmentation models trained on multi-channel T1-MOLLI input.	34
4.7	Mean \pm SD of volume difference (in %) between predicted and delivered volumes, across kidney regions and dataset splits.	41
4.8	Overview of input channels used in the multi-modal segmentation model. DWI inputs correspond to multiple b-values and T2* mapping inputs correspond to multiple echo times.	44
4.9	Comparison of the performance of multi-label segmentation models using multi-channel T1-MOLLI or DWI and T2* as input. Models are based on a 2D ResUNet structure, segmenting both cortex and parenchyma simultaneously.	44
5.1	Summary of related work on kidney segmentation.	54
B.1	R^2 correlation values between predicted and actual T1 values for SIN and DX cortex, using erosion with a square structural element of various sizes.	III
B.2	R^2 correlation values between predicted and actual T1 values for SIN and DX medulla, using erosion with a square structural element of various sizes.	III

1

Introduction

Accurate and precise segmentation of anatomical structures in medical images is a critical challenge in the field of medical image analysis. With rapid advancements in deep learning, new and powerful tools for processing and interpreting medical images have been introduced, such as for segmenting vital organs like the kidneys. Precise delineation of kidney regions is essential for advancing the understanding of kidney function and pathology, as well as for supporting effective diagnosis and treatment planning. For instance, segmenting internal kidney structures such as the cortex and medulla allows for measurement of anatomical and functional changes associated with pathology. Such information could support early detection of conditions like chronic kidney disease, a progressive and widespread condition that often goes undetected in its early stages [1]. Among medical imaging techniques, Magnetic Resonance Imaging (MRI) has emerged as a powerful, non-invasive technique for visualizing subtle structural and functional changes of the kidneys, providing insights into disease progression and severity [2].

However, manual analysis and segmentation of kidney regions from medical image data is a time-consuming process that requires high expertise. It becomes even more challenging when there is inter- and intra-subject anatomical variability, limited image contrast between internal tissue types due to limited image quality, and inconsistencies across imaging protocols. These factors make manual annotation both time-consuming and prone to variability. Deep learning offers a promising way to automate the renal segmentation process, saving time and manual effort while also reducing operator bias.

While recent advances in deep learning have led to significant improvements in medical image analysis, existing approaches focus on whole-kidney segmentation and do not often address the challenge of segmenting anatomical substructures of the kidneys. Methods that do attempt to resolve internal regions often struggle with low contrast, leading to insufficient segmentation performance and limited clinical relevance. Thus, there is a clear need for deep learning solutions that can handle the complexity of renal segmentation from imaging data.

This thesis focuses on developing a deep learning-based segmentation pipeline to address the challenge of segmenting the parenchyma and internal renal structures in kidney MRI. By providing predicted segmentations with high accuracy and precision, the need for manual corrections, or any at all, can be minimized. Such advancements have the potential to make the process of analyzing kidney images more efficient, thereby reducing the time, effort, and operator dependency.

1.1 Aim

The aim of this thesis is to develop a deep learning-based solution for automated segmentation of kidney parenchyma, cortex, and medulla from MRI data. Specifically, the objective is to implement a multi-channel approach capable of generating reliable segmentation suggestions for these anatomical regions by leveraging information from multi-parametric MRI. Furthermore, this study aims to evaluate the feasibility of using the proposed deep neural network to estimate renal volume and tissue-specific parametric values, thereby enabling automatic assessment of renal tissue health and disease progression.

This project is carried out in collaboration with Antaros Medical AB, a company specializing in advanced medical imaging technologies for clinical trials and drug development. The desired outcome is a reliable segmentation pipeline that assists their image analysts in the manual segmentation workflow.

1.2 Scope and Limitations

The scope of this study is limited to the automated segmentation of the kidney parenchyma, cortex, and medulla using MRI data. Only these specific renal regions are considered, and no additional structures will be included in the segmentation process. Similarly, the image data is restricted to MRI, and other modalities such as computer tomography (CT) are not within the scope.

The MRI dataset used in the study is pre-acquired and provided by Antaros Medical AB. As a result, this work does not involve any additional image data acquisition. Areas such as motion artifact reduction and image registration techniques are further excluded from the scope.

Given that image segmentation and deep learning are wide research areas, this thesis focuses on investigating a limited set of techniques rather than attempting to explore the entire field. The proposed solution will solely be evaluated against data from the same clinical study, which limits the comparison of results to other medical image data. Lastly, while algorithm performance in terms of efficiency and processing time is discussed, direct comparisons with manual segmentation in terms of time savings are not included in the scope of this thesis.

1.3 Related Work

Automated medical image segmentation has become an increasingly active area of research, driven by advancements in deep learning and computer vision. Early efforts in automatic kidney segmentation are primarily focused on the kidney as a whole, with less attention paid to the internal structures, such as the cortex and medulla. However, research highlights the importance of detailed segmentation of these internal structures due to their clinical importance in diagnosing and monitoring renal diseases [3]. An early study from 2014 emphasized the significance of segmenting the kidney cortex and medulla and proposed a method based on traditional thresh-

olding techniques applied to MRI modalities [4]. With the advancements in deep learning, particularly in the field of convolutional neural networks (CNNs), more refined approaches have been developed to improve segmentation accuracy.

In the context of kidney segmentation, a number of efforts have explored using CNN-based networks to automate and improve segmentation performance in MRI and CT imaging. For example, a study from 2022 introduced a deep learning model based on a modified U-Net architecture for the automated segmentation of the kidney cortex and medulla in abdominal CT scans [5]. Although this approach demonstrated improved accuracy over traditional methods, challenges remain due to the minimal contrast differences between the cortex and medulla. Additional research has applied CNNs for classification and segmentation tasks to support chronic kidney disease (CKD) diagnosis and kidney volume estimation [6], [7], yet poor tissue contrast is consistently highlighted as a challenge. Segmentation accuracy is often reduced near anatomical boundaries where adjacent organs, such as the spleen and liver, blur the distinction between tissues. In patients with CKD, additional challenges arise due to irregular kidney shape and altered contrast, which further complicate delineation of tissue boundaries. Despite these challenges, CNN-based methods demonstrate high accuracy in segmenting kidneys, often outperforming manual segmentation and reducing inter- and intraobserver variability [5], [8].

Deep learning-based segmentation models are typically limited by the quality of input data. For renal segmentation, existing methods rely on either contrast-enhanced or high-resolution imaging. Still, challenges remain due to limited contrast within the kidneys or surrounding tissues. Additionally, models trained on a single imaging modality often experience limited generalizability. Such models tend to become biased toward specific imaging data, reducing their effectiveness when applied to other modalities or diverse clinical datasets. This presents challenges for clinical applications, where robust performance across various imaging techniques is essential.

To overcome such challenges, multi-modal and multi-channel approaches have gained attention. Recently, researchers have explored the integration of multiple imaging modalities with deep learning techniques to improve segmentation performance and robustness. For example, a study from 2024 proposed incorporating multiple imaging modalities into the segmentation process by treating them as separate input channels. This method demonstrated improved segmentation accuracy and better generalization across varying datasets [9]. Such input-level fusion is widely adopted in deep learning-based medical segmentation [10], offering a straightforward yet effective way to combine diverse tissue contrasts within a single model. Other methods involve fusion at a later stage, where integration of different imaging modalities occurs within the network’s internal layers or output stage, offering more flexibility but with higher complexity [11].

Various approaches are represented in the literature, with a general agreement that incorporating data from multiple imaging modalities improves segmentation accuracy and model generalization. Moreover, multi-dimensional U-Net variants have demonstrated strong performance across diverse medical datasets [12]–[14]. Despite the clear advantages, the application of multi-channel and multi-modal approaches to renal image segmentation appears to be limited.

This thesis builds on previous research by using a multi-channel approach that combines data from multiple MRI sequences to segment not only the kidney as a whole, but also its internal structures like the cortex and medulla. Unlike many of the earlier methods that focus on CT or single-modality MRI to segment the entire kidney, this approach combines data from different MRI acquisitions to help separate internal kidney structures, which are often hard to distinguish due to low tissue contrast.

2

Theory

This chapter provides the theoretical background relevant to this study on kidney segmentation using deep learning. It begins with an introduction to kidney anatomy, renal function, and clinical relevance of conditions like CKD. Following this, key concepts related to magnetic resonance imaging will be introduced, highlighting its role in renal imaging and functionality assessments, as well as common artifacts.

Furthermore, the fundamentals of deep learning and neural networks, specifically for the task of semantic segmentation, will be explained. This includes an introduction to state-of-the-art approaches, focusing on convolutional neural networks and architectures like the U-Net and ResUNet. Lastly, this chapter addresses training strategies for deep neural networks, including optimization algorithms, loss functions specific to medical segmentation, and commonly used evaluation metrics.

2.1 Kidney Anatomy and Physiology

The kidneys are paired, bean-shaped organs that are located on either side of the spinal column in the retroperitoneal space, posterior to the abdominal cavity [15]. They play a crucial role in maintaining homeostasis by filtering blood, regulating electrolyte balance, producing hormones, and excreting metabolic waste products and excess water from the bloodstream.

Normally, the human body has two kidneys, identical in structure and function. Each kidney consists of distinct anatomical regions, which are essential for its filtration and regulatory functions. As illustrated in Figure 2.1, the outermost layer consists of the renal cortex, which is where blood filtration begins. The cortex region contains the glomeruli, clusters of tiny blood vessels that initiate the filtration process, and the proximal tubules, which reabsorb essential nutrients and regulate fluid balance [15].

Beneath the cortex lies the renal medulla, organized into cone-shaped renal pyramids. These structures contain loops of Henle and collecting ducts, both important for concentrating urine and regulating water reabsorption. At the tip of each pyramid, urine is drained into the minor calyces, which merge into the major calyces and ultimately form the renal pelvis. The renal pelvis acts as a central collection area, directing urine into the ureter for excretion. Together, the cortex and medulla form the renal parenchyma, the kidney's functional tissue.

Although the kidneys are bilaterally symmetrical in structure, they are positioned

asymmetrically. The left kidney, or sinistral (SIN) kidney, is located slightly higher than the right, or dextral (DX) kidney, because the liver on the right side pushes the DX kidney downward. This anatomical asymmetry, along with natural variations in kidney size and shape due to individual differences or pathological conditions, presents challenges for automated renal segmentation [3], [16].

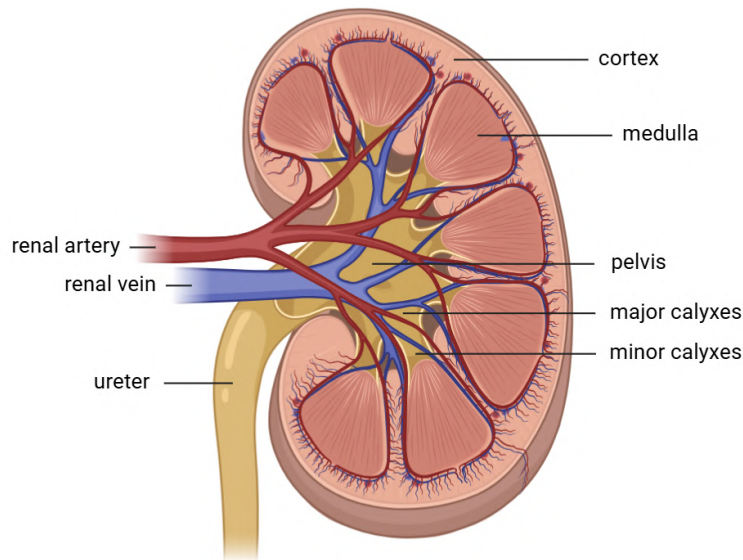


Figure 2.1: Anatomical illustration of the normal renal structure from a coronal orientation of the left kidney. The renal parenchyma tissue includes cortex and pyramids of medullas. Created with BioRender.com.

2.1.1 Chronic Kidney Disease

CKD is a global health challenge, claiming more than 1 million lives annually [2]. This irreversible and progressive condition affects millions worldwide, severely impairing kidney function and, consequently, quality of life. Despite its increasing prevalence, CKD symptoms are subtle, and by the time they become apparent, the disease has often progressed significantly. As a result, early-stage CKD frequently goes undiagnosed, making timely intervention a critical challenge. The increasing incidence and mortality rates underscore the need for better diagnostic tools and more effective management strategies.

As CKD progresses, it can lead to end-stage renal disease (ESRD), which requires expensive kidney replacement treatments, such as dialysis or transplantation. ESRD presents a significant financial burden on healthcare systems worldwide, with dialysis programs growing annually by 6% to 12% over the past two decades [6]. A key barrier to early diagnosis of CKD is the lack of symptoms in the early stages of the disease, highlighting the need for appropriate screening and diagnostic methods. Early detection is crucial for preventing the disease's progression and the development of complications. MRI has emerged as a useful non-invasive tool in assessing CKD progression. For instance, MRI helps detect tissue changes associated with processes such as cyst formation and fibrosis, providing important information on the extent of kidney damage and the effectiveness of treatment [17].

Renal volume is another important biomarker, as its reduction due to glomerular loss correlates with decreased nephron function and impaired kidney filtration, leading to complications like fluid overload and ESRD [18]. Monitoring renal volume provides insights into disease progression, enabling clinicians to adjust treatment plans.

2.2 Magnetic Resonance Imaging

MRI is a widely used medical imaging technique that produces detailed 2D or 3D views of internal organs and structures. It is non-invasive and relies on strong magnetic fields and radio waves to generate detailed images of the body's internal structures, making it particularly useful for imaging soft tissues such as the brain, liver, and kidneys. By providing high-resolution images, MRI plays an important role in diagnosing and monitoring various medical conditions, as well as understanding anatomical and physiological changes. In the context of renal imaging, MRI enables non-invasive tissue characterization and early detection of renal disease progression, helping to predict clinical outcomes and guide treatment decisions [17].

2.2.1 Physical Principles of MRI

The principles behind MRI rely on the interaction between hydrogen nuclei and a magnetic field. During an MRI scan, the strong magnetic field generated by the scanner causes protons in the body to align along the direction of the magnetic field, creating a net magnetization known as equilibrium magnetization. When a radiofrequency (RF) pulse is applied, it temporarily disturbs this equilibrium by tipping the magnetization away from its aligned state, leading to a temporary loss of longitudinal alignment and the development of magnetization in the transverse plane. These changes in magnetization contribute to measurable alterations in the net magnetization. Following the RF pulse, the longitudinal magnetization (M_z) is reduced, while the transverse magnetization (M_{xy}) increases. It is this transverse component that generates a detectable signal in the MRI receiver coils. Once the RF pulse is turned off, the system undergoes a process known as relaxation, during which the magnetization gradually returns to its equilibrium state due to interactions between the hydrogen nuclei and their surrounding environment.

Relaxation occurs on two distinct time scales: M_{xy} decays according to the T2 time constant, while M_z recovers along the T1 time constant [19]. The relaxation times vary across tissue types and are influenced by factors such as magnetic field strength, tissue composition, and water content. Because these relaxation times are tissue-dependent, they form the basis for the contrast seen in the resulting images.

2.2.2 MRI Sequences

MRI uses certain types of sequences to highlight different structural and functional characteristics of tissues. Each sequence is characterized by a specific combination of timing, RF pulses, and gradient fields that manipulate the magnetic properties of protons in the body and enables acquisition of images with precise resolution

and contrast. The choice of imaging sequence determines the type of contrast and, consequently, the kind of physiological or anatomical information captured, such as tissue structure, composition, or water content.

Modified Look-Locker Inversion Recovery

This project utilizes data from several MRI sequences to capture distinct features of kidney anatomy and function. One of them is Modified Look-Locker Inversion Recovery (MOLLI), which is a widely used sequence for T1-weighted (T1w) imaging. The images are acquired with respect to the subjects' heartbeats, which helps to reduce artifacts caused by cardiac pulses or respiratory motion [20]. As the technique is widely available across imaging sites, MOLLI is now a well-established method in renal T1 mapping [17].

In T1 mapping, the MOLLI sequence generates a series of images with different inversion times (TI), which refers to the time between the inversion pulse and signal acquisition. These acquired images enable voxel-wise calculations to derive a quantitative T1 map, where each voxel reflects the T1 relaxation time of the underlying tissue [21]. Shorter T1 values appear brighter, while longer values appear darker, enabling detailed tissue characterization. In the context of renal imaging, T1 mapping is useful for non-invasive assessment of renal microstructure. Changes in T1 relaxation times can indicate pathological changes such as fibrosis or inflammation, making it a promising biomarker for early-stage CKD [2].

T2*-weighted Imaging

Another useful MRI sequence for renal imaging is T2*-weighted (T2*w) imaging, which is a pulse sequence that measures and displays differences in T2* relaxation times across various tissues [22]. The main difference between the T2* relaxation and the conventional T2 relaxation parameter is that T2* relaxation also accounts for magnetic field inhomogeneities from susceptibility differences in various tissues.

By acquiring a series of T2*w images with varying T2* sensitivities and estimating the T2* relaxation times through pixel-wise modeling, a T2* map can be generated [23]. This parametric map visualizes the varying T2*-values for each tissue in the image, with fluids such as water appearing bright. The T2* relaxation time serves as a marker of tissue oxygenation, making T2* mapping valuable for monitoring renal oxygenation. This can in turn be useful for indicating progression of renal diseases or evaluating the effects of drugs or treatments [24].

Diffusion-weighted Imaging

Diffusion-weighted imaging (DWI) is an imaging modality that can provide additional functional data in kidney analyses. This sequence measures diffusion properties and random Brownian motion of water molecules within tissues, offering insights into their cell density and structural integrity [25]. The general principle behind the DWI sequence is that it measures diffusion-related attenuation of the MR signal. It highlights differences in the diffusion of water molecules after application of diffusion-sensitizing gradients, which affect the movement of water molecules in

different directions [26]. The strength and timing of these gradients are quantified by the b -value, a parameter that determines the degree of diffusion weighting. A higher b -value indicates a higher signal attenuation based on diffusion [25]. Darker pixels in the DWI image are thus a result of more signal loss due to more motion of the water molecules.

The resulting data are used to calculate a parametric map that highlights the apparent diffusion coefficient (ADC) within the tissues. The ADC map allows quantification of water movement as contrast in the images reflects differences in diffusion, where a higher ADC value indicates areas with less restricted diffusion and more motion [27]. This can provide functional information that complements anatomical imaging from other imaging sequences. Renal DWI serves as a valuable tool for assessing renal microstructure and circulation, with ADC being a widely used diffusion biomarker and a prognostic indicator for various kidney diseases [2].

2.2.3 MRI Artifacts

In magnetic resonance imaging, artifacts frequently arise from equipment malfunctions, choice of imaging technique, or the inherent physics of the modality [28]. Artifacts are defined as signals that do not correspond to true anatomy or as distortions and deletions of anatomical information.

There are many distinct types of MRI artifacts, each with their own characteristic appearance. One of the most common is the motion artifact, which as its name implies occurs when sudden movement in the ROI produces ghosting or blurring of structures. Another frequent artifact is aliasing, which arises when anatomy extends beyond the field of view (FOV) and its signal is wrapped back into the image, potentially obscuring underlying tissues. When aliasing combines with magnetic field inhomogeneities, it can give rise to the so-called zebra stripe artifact with alternating bright and dark bands across the FOV that may mask anatomical detail. Additionally, the partial volume artifact arises when a single voxel contains multiple tissue types often due to limited resolution or minor motion so that the resulting signal is an average of those tissues, reducing apparent resolution and potentially obscuring small structures. Examples of these artifacts are shown in Figure 2.2.

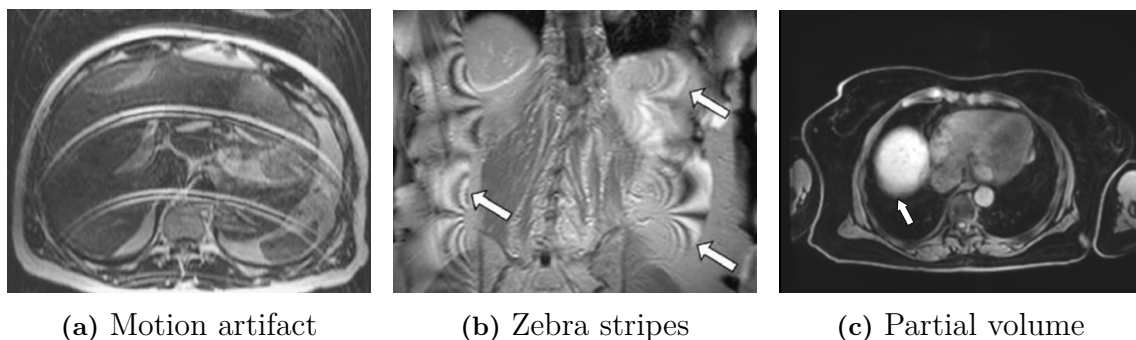


Figure 2.2: Examples of common MRI artifacts that include motion artifact, zebra stripe artifact, and partial volume effect. Images (a) and (b) are adapted from Stadler et al. [28], while (c) is courtesy of Antaros Medical AB.

2.3 Deep Learning for Image Segmentation

Deep learning, a subset of machine learning, allows systems to learn from experience rather than relying on pre-programmed knowledge. Deep neural networks (DNNs) have proven successful in various computer vision tasks, such as image segmentation, where models can automatically delineate regions of interest (ROIs) [29]. This is valuable in medical imaging, such as for organ segmentation. By training DNNs on labeled image data, these models can efficiently segment areas in medical images, providing benefits and guidance for both diagnosis and treatment planning. The process of training an algorithm using labeled datasets is known as supervised learning. One of its advantages is that it generally achieves high accuracy and, with sufficient training data, tends to outperform other approaches such as unsupervised or semi-supervised learning in medical segmentation [30].

Among the various image segmentation techniques, semantic segmentation is one of the most widely used in the medical field. Semantic segmentation refers to algorithms that classify each pixel in an image by assigning it to a specific object class, effectively grouping pixels that belong to the same category [31]. This pixel-level classification enables clear identification and delineation of distinct structures within an image. Semantic segmentation has numerous applications, particularly in medical image analysis. It plays a crucial role in enhancing diagnostic accuracy by identifying and segmenting ROIs, such as tumors, lesions, and anatomical structures [32]. By leveraging deep learning approaches, it enables automation of tasks like annotation and boundary delineation, thereby improving workflow efficiency and reducing the risk of human error.

2.3.1 Basic Concepts of Deep Learning

How the deep learning algorithm works is similar to the human brain. These algorithms use a multi-layered architecture in which artificial neurons are inter-connected between layers to learn hierarchical representations of data [33]. The artificial neuron is an essential component in a DNN that transforms an input vector into a scalar output through a weighted sum followed by a non-linear activation. Mathematically, it can be expressed as

$$z_k = \sum_{i=1}^n w_{ki}x_i + b_k, \quad (2.1)$$

$$a_k = f(z_k), \quad (2.2)$$

where x_i is the i -th element of the input vector, w_{ki} is the weight connecting input i to neuron k , and b_k is the bias term for neuron k . The function $f(z_k)$ denotes the activation function, and a_k is the scalar output of the neuron. In a multi-layer network, passing information between neurons is known as forward propagation, and is repeated across layers l :

$$z^{(l)} = W^{(l)}a^{(l-1)} + b^{(l)}, \quad (2.3)$$

$$a^{(l)} = f(z^{(l)}). \quad (2.4)$$

Once forward propagation is completed across the entire network, the model's output is computed and compared to the actual target value using a function called the loss function, denoted as \mathcal{L} . This error function measures the difference between the predicted and actual values, providing information that guides the model during training to find an optimal set of parameters.

In order to train the neural network and minimize the loss \mathcal{L} , an operation called backpropagation is typically performed, where the gradient of the loss with respect to the weights is computed using the chain rule:

$$\frac{\partial \mathcal{L}}{\partial w} = \frac{\partial \mathcal{L}}{\partial a} \cdot \frac{\partial a}{\partial z} \cdot \frac{\partial z}{\partial w}. \quad (2.5)$$

These gradients, computed through backpropagation, are used to update the weights in the model. By iteratively updating the weights to minimize the loss, the model learns and improves. The optimal solution is reached when the gradients reach zero, indicating convergence to a local or global minimum of the loss function.

For these gradients to be applied during training, they are processed by a component called an optimizer, which adjusts the model's weights to minimize the loss function. Optimizers are typically based on variants of gradient descent, which updates the parameters in the direction of decreasing error. Among the many available optimizers, one of the most commonly used with CNNs and image segmentation tasks is the Adam optimizer. Adam is a type of stochastic gradient descent optimizer that adaptively adjusts the learning rate for each parameter using estimates of the first and second moments of the gradients [34]. The algorithm works by running averages of both the gradients and their squared values, correct the biases introduced during initialization, and use these adjusted values to scale update the parameter values.

2.3.2 Convolutional Neural Networks

Convolutional Neural Networks (CNNs) is a type of artificial neural network that has become important especially for imaging purposes [35]. The CNNs consists of several building blocks that are called layers, each with a specific role in processing data for the network.

Convolutional Layer

The convolutional layer is the primary component of a CNN and the origin of its name [35]. The main purpose of this layer is feature extraction, which is achieved by applying a convolution operation to the input data using a set of learnable filters, also called kernels. As each kernel slides over the input and produces a feature map by capturing local patterns, the network captures spatial hierarchies of features, allowing it to recognize increasingly complex patterns.

The convolution operation, illustrated in Figure 2.3, is a linear operation that involves taking the element-wise product of the kernel and a local region of the input. These values are summed up into a single output value at the corresponding position in the resulting feature map. By applying multiple kernels, the network can extract different types of features from the input image in each layer, such as edges, textures, or shapes. The kernels operate on each input channel separately, producing channel-wise feature maps. A convolutional layer is characterized by its kernel size, stride, and padding, with stride determining how many pixels the kernel moves at each step, and padding adding extra borders around the input to control the spatial dimensions of the output.

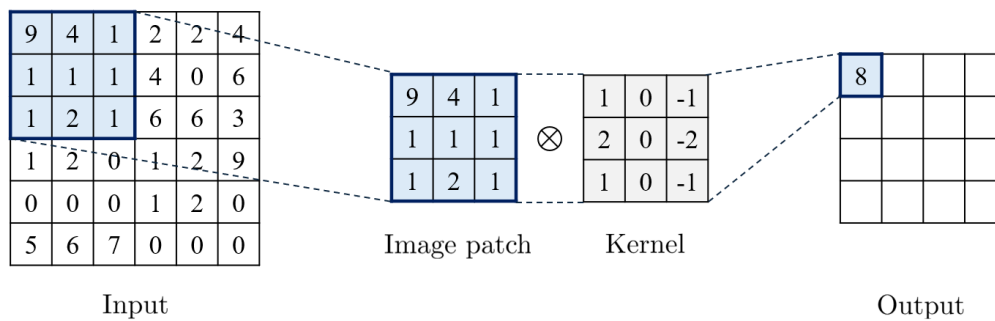


Figure 2.3: An example of convolutional layer with kernel size 3×3 , a stride of 1, and no padding.

Pooling Layer

Pooling layers are another essential component of CNN architectures [35]. Their primary function is to reduce the spatial dimensions of feature maps, while keeping the most important information. Pooling layers help reduce the number of values the network needs to process in the subsequent layers. This helps reduce computational load and makes the network less likely to overfit.

One commonly used pooling layer is max pooling, illustrated in Figure 2.4. It works by dividing the input feature map into smaller patches and keeping only the maximum value from each patch in the output. The stride controls how far the patch moves each step, larger strides result in greater downsampling and smaller output dimensions. This downsampling can also be achieved by using strided convolutions.

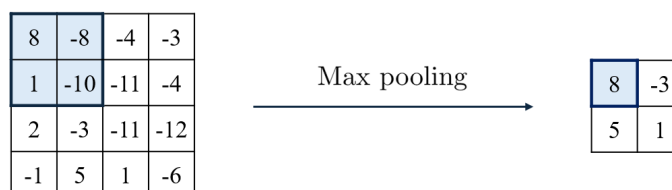


Figure 2.4: Max pooling layer with kernel size 2×2 and a stride of 2.

Drop-out Layer

To combat overfitting in a model, which refers to when the model learns the training data too closely but fails to generalize, the dropout layer is an important component of the architecture. Dropout, introduced as a regularization technique by Hinton et al. [36], works by randomly omitting hidden units in the network during training with a fixed probability. This prevents any single hidden unit from relying too heavily on the presence of others, which helps improve generalization and reduce overfitting.

Activation Functions

Activation functions are applied after convolutional layers to introduce non-linearity into the model, allowing the neural network to capture complex relationships within the data [35]. Among the various activation functions, the Rectified Linear Unit (ReLU) is widely used in CNNs, as it helps reduce the vanishing gradient problem, accelerates training, and enables the model to learn more complex patterns. The ReLU function outputs the input value for positive inputs and zero for negative inputs, mathematically defined as

$$f(x) = \max(0, x). \quad (2.6)$$

To address the issue of zero gradients in the negative input range of ReLU, as shown in Equation 2.6, the Parametric ReLU (PReLU) activation function can be used. The PReLU function is defined as

$$f(x) = \begin{cases} x, & \text{if } x > 0 \\ \alpha x, & \text{if } x \leq 0 \end{cases}, \quad (2.7)$$

where α is a learnable parameter. Unlike the standard ReLU, which outputs zero for all negative inputs, PReLU introduces a small, non-zero slope for negative input values through α , enabling gradients to propagate and improving training stability. This adaptive mechanism has been shown to enhance model performance with minimal additional computational cost [37]. The difference between the ReLU and PReLU activation functions is illustrated in Figure 2.5.

In the final classification layer, the activation function typically differ from those used in earlier layers, as it determines the predicted class of each input [35]. In binary classification tasks, the sigmoid function is typically used, producing a single probability value that indicates the likelihood of the input belonging to the foreground class. For multi-class problems, the softmax function is typically used instead. This function normalizes the raw output of the fully connected layer, converting it into multiple class probabilities, where each value ranges between 0 and 1 and the sum of all values equals 1. These normalized probabilities support effective classification by identifying the class with the highest predicted probability. Visual representations of both activation functions can be seen in Figure 2.6.

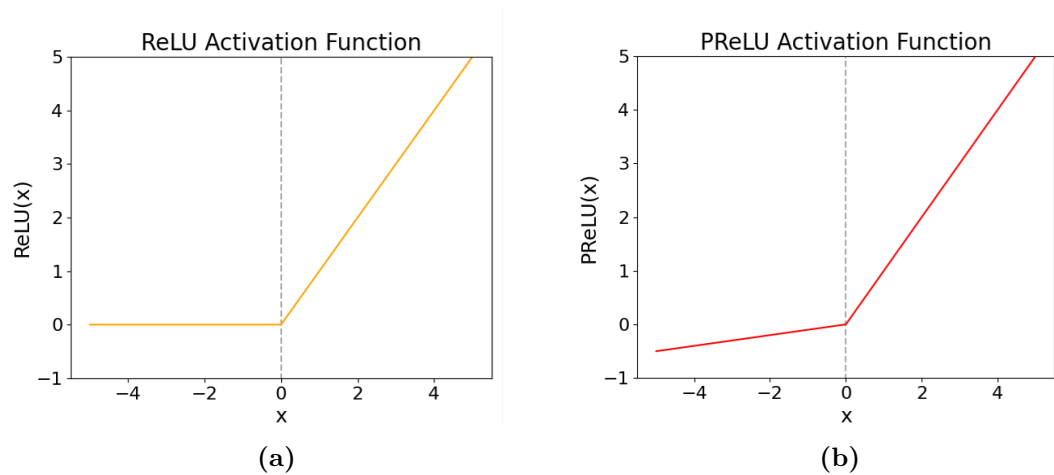


Figure 2.5: Comparison of activation functions: (a) ReLU, and (b) PReLU.

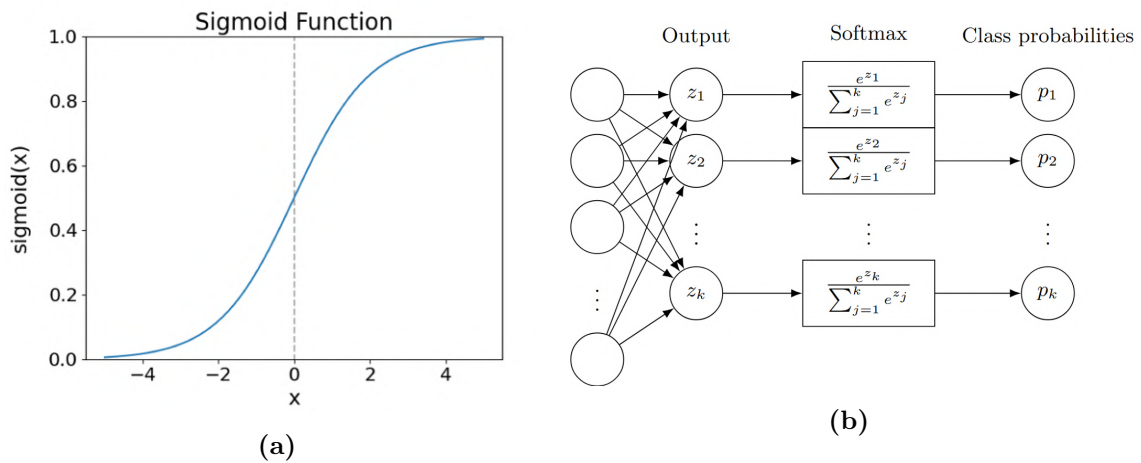


Figure 2.6: Illustration of activation functions commonly used for classification: (a) Sigmoid function, and (b) Softmax function.

2.3.3 U-Net

Among the most commonly used CNN architectures for semantic segmentation is the U-Net, first introduced in 2015 [38]. Originally developed for biomedical image analysis, U-Net was designed to perform well even with limited training data, making it particularly useful in medical imaging applications where annotated datasets often are scarce. U-Net has since become a popular choice for segmenting both 2D and 3D medical images from different modalities, including MRI and CT [39].

The U-Net has a unique, U-shaped architecture consisting of a contracting part (encoder) and an expansive path (decoder). The encoder extract features to learn a more compressed representation of the input. It is built up of stacked encoder blocks, which helps the network learn increasingly abstract representations of the input image at each level. Each encoder block consists of a series of convolutional layers, with ReLU activation function following each convolution. Max-pooling operations are applied at the end of each block, to downsample the feature maps.

The decoder, consisting of mirrored encoder blocks, then reconstructs the segmentation mask by gradually increasing the resolution using transposed convolutions. By using skip connections that link each encoder block to its corresponding decoder block, the network combines both high- and low-level features at each level. This allows for preservation and better learning of spatial information, improving segmentation accuracy. Each decoder block mirrors its encoder counterpart, consisting of a series of convolutional layers with ReLU activations. Finally, an additional 1×1 convolutional layer is used to produce pixel-wise predictions, resulting in the predicted segmentation mask.

2.3.4 ResUNet

ResUNet is a deep learning model that builds on the U-Net architecture by incorporating residual units. This model was first introduced by Diakogiannis et al. in 2020 [40]. ResUNet retains the encoder–decoder structure of U-Net but replaces standard convolutional blocks with residual blocks. These residual units enable better gradient flow during training and make it possible to construct deeper architectures without suffering from vanishing gradient issues. The result is a more stable and efficient model capable of capturing both fine and high-level semantic features.

As the depth of a neural network increases, the training process can become more difficult due to issues such as vanishing gradients and model degradation. Residual connections help address these challenges by allowing the input of a set of layers to bypass those layers and be directly added to the output [41]. The residual connection can be represented by the equation

$$H(x) = F(x) + x, \quad (2.8)$$

where $H(x)$ represents the desired mapping, and $F(x)$ is the residual function learned by the network.

There are many advantages to using residual units. They help reduce training error in deeper architectures, as the identity mapping increases the likelihood of finding suitable initial parameters. An additional benefit of residual connections is that they provide a shortcut for gradients during the backpropagation process, allowing deeper networks to learn more effectively.

2.3.5 Loss Functions for Medical Image Segmentation

The loss function is an essential part of training neural networks, as it quantifies the error between predicted and actual values. Different problems require different loss functions, and in medical image segmentation, several loss functions are commonly used, including cross entropy, Dice loss, Tversky loss, and their variants [42]. Choosing the right loss function is particularly important in medical image segmentation, where datasets often have a class imbalance. This imbalance occurs when there is far fewer foreground pixels compared to background pixels, and is common when the ROI is small in comparison to the full image volume. A suitable loss function

helps the model focus on both the foreground and background, which helps improve the accuracy of the segmentation.

The Dice loss, introduced by Milletari et al. [43], is commonly used in medical image segmentation because it effectively handles class imbalances. The loss function is based on the Dice coefficient (explained further in section 2.3.6), which is a statistical measure of overlap between two sets. The Dice loss, \mathcal{L}_{Dice} , is calculated as follows:

$$\mathcal{L}_{Dice} = 1 - \frac{2 \sum_{i=1}^N y_i \hat{y}_i + \epsilon}{\sum_{i=1}^N y_i^2 + \sum_{i=1}^N \hat{y}_i^2 + \epsilon}, \quad (2.9)$$

where N corresponds to the total number of voxels in the image, \hat{y}_i is the predicted probability of voxel i belonging to the foreground, and y_i is the ground truth value for voxel i . The term ϵ is a small smoothing term added to avoid division by zero in the case of background-only, when both the prediction and ground truth are empty. The Dice loss is typically calculated for each class separately, and the average loss across all classes is then used during training.

2.3.6 Evaluation Metrics

Evaluation metrics are used to quantitatively assess the performance of segmentation algorithms. In the context of medical image segmentation, various metrics exist, each suited to different types of segmentation tasks and applications. For evaluating segmentation overlap, metrics are typically based on pixel-wise comparisons between the predicted segmentation mask and the ground truth, based on certain classification outcomes. True positives (TP) refer to the number of pixels correctly identified as belonging to the foreground object of interest, false positives (FP) are background pixels that are incorrectly labeled as foreground, and false negatives (FN) are foreground pixels missed by the model, misclassified as background. The relationship between these classification outcomes can be visualized in Figure 2.7

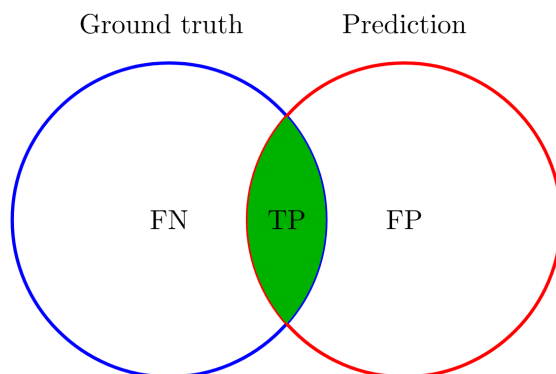


Figure 2.7: Illustration of the relationship between ground truth and predicted classifications. The overlapping region represents true positives (TP), where the prediction correctly matches the ground truth. The left (blue) region denotes false negatives (FN), where actual positives were missed. The right (red) region indicates false positives (FP), where the background is incorrectly segmented as foreground.

Two of the most commonly used metrics that rely on these classification outcomes are the Dice score and the Jaccard index, also known as Intersection over Union (IoU). Both metrics quantify the similarity between predicted and ground truth masks. The Dice score is defined as twice the area of overlap between the predicted and ground truth masks, divided by the total pixel area of both masks:

$$\text{Dice score} = \frac{2 \times |Y \cap \hat{Y}|}{|Y| + |\hat{Y}|} = \frac{2TP}{2TP + FP + FN}, \quad (2.10)$$

The IoU provides a slightly stricter measure of overlap and is defined as the area of overlap divided by the area of the union of the predicted and ground truth masks:

$$\text{IoU} = \frac{|Y \cap \hat{Y}|}{|Y \cup \hat{Y}|} = \frac{TP}{TP + FP + FN}. \quad (2.11)$$

While both metrics assess how closely the prediction matches the ground truth, the IoU penalizes over-segmentation and under-segmentation more strongly than the Dice score, making it particularly useful in applications requiring precise delineation.

2.4 Multi-channel Image Segmentation

In deep learning-based image segmentation, multi-channel and multi-modal input strategies have proven effective in enhancing model performance [10]. A multi-channel image consists of multiple correlated image channels, each capturing different aspects of the same scene or object. For instance, this could be images acquired with different cameras, time points, or acquisition parameters. These channels are stacked as separate input layers, allowing the network to process them simultaneously. By combining multiple input channels, each pixel is represented by a multi-dimensional vector instead of a single intensity-value, which helps CNNs to learn complementary feature representations.

In medical image analysis, this becomes particularly useful. Many anatomical structures are complex, overlapping, or poorly defined in single-channel images. To overcome this, scans are often acquired with varying acquisition parameters. These variations help enhance tissue contrast and allow for the extraction of quantitative measurements like parametric mapping. Multi-modal input further extends the concept of multi-channel input by combining data from different imaging modalities, or from MR sequences with different acquisition parameters. As each image provides unique and complementary information about different tissues, the network learns from images with richer contrast information, which could be useful for differentiating between complex structures. Research indicate that networks trained on multi-channel or multi-modal data often generalize better, particularly across patients and imaging conditions [10], [32]. They reduce the risk of overfitting to channel- or modality-specific features and can even be advantageous for training on smaller datasets, thanks to the diversity of input information to learn from.

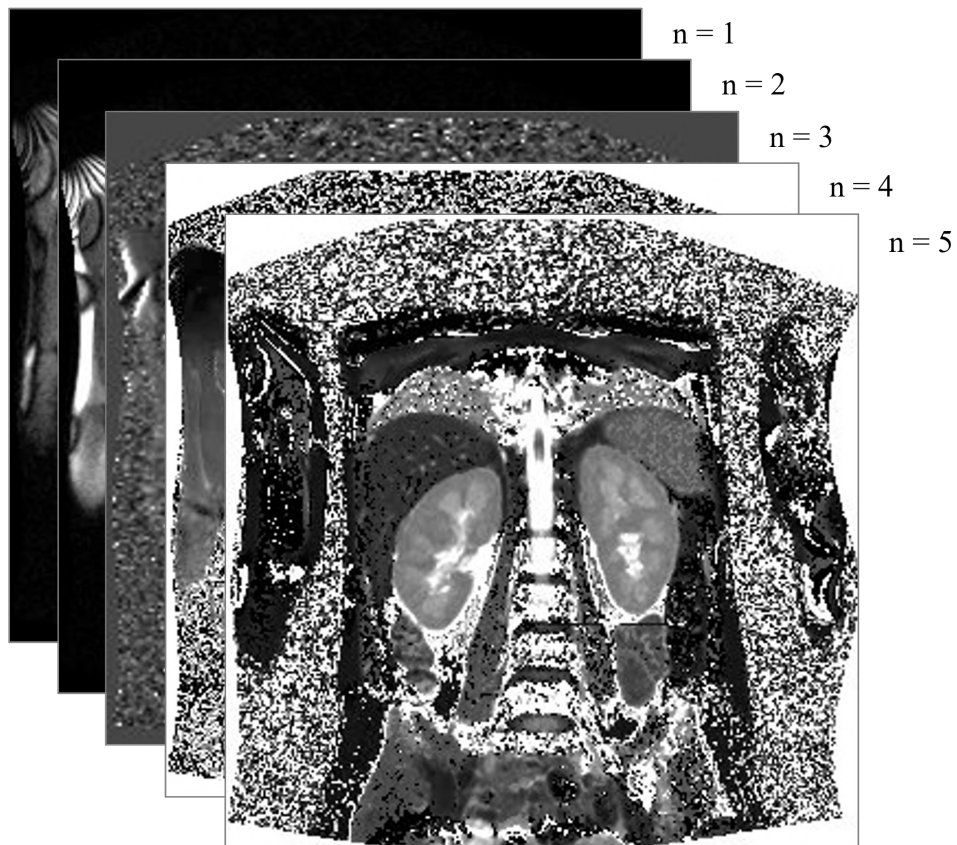


Figure 2.8: Example of a multi-channel input image with five different image channels from renal MRI data. The channels provide different views of the same anatomy with complementary tissue contrast information.

3

Methods

To address the challenges of manual kidney segmentation, this thesis proposes a deep learning-based method for automated segmentation of the kidney parenchyma, cortex, and medulla. Due to the often limited contrast between these internal kidney structures in certain MRI sequences, a multi-channel approach is employed to leverage complementary information from multiple modalities or images acquired with different parameter settings.

Method implementation was done using Medical Open Network for AI (MONAI), an open-source platform optimized for deep learning in medical imaging [44]. The experimental workflow included three main experiments: single-channel segmentation of parenchyma and cortex from T1-MOLLI images, extension to multi-channel using multiple TIs, and finally, multi-modal integration of DWI and T2* mapping. An overview of the proposed segmentation pipeline is visualized in Figure 3.1.

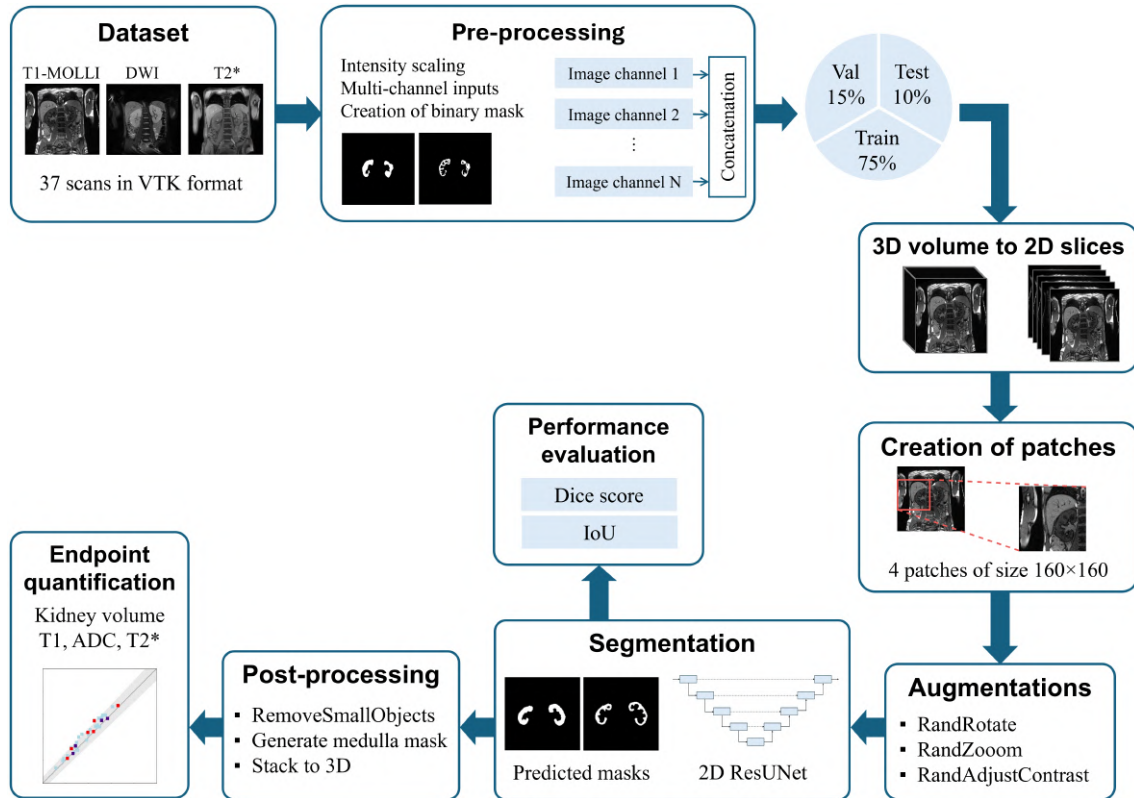


Figure 3.1: Overview of proposed pipeline for kidney segmentation.

3.1 Data Description

The dataset used in this study consists of kidney MR images provided by Antaros Medical AB. The clinical data includes a total of 30 patients, each diagnosed with CKD at varying stages. The patients were divided between two scanner sites, referred to as site A and site B. All patients underwent MRI scans at two separate visits, with each patient being scanned at the same site for both visits. Hence, there were 60 scans available for the study.

MRI acquisitions were performed using 1.5 Tesla scanner (Siemens for site A and GE for site B). Each site acquired scans for three different sequences, T1-MOLLI, DWI, and T2* mapping, with slightly varying acquisition parameters between the two sites. All scans consisted of coronal 2D slices, acquired in sequence, with in-plane pixel resolution ranging from $1.5 \times 1.5 \text{ mm}^2$ to $1.9531 \times 1.9531 \text{ mm}^2$, reflecting differences in scanner settings and FOV requirements. These slice-wise acquisitions were stacked into a 3D image volume, which was then stored and provided in VTK format. Table 3.1 provides an full overview of the imaging parameters for each modality used in this study, including the number of slices, slice spacing, image resolution, and the corresponding parametric maps generated.

Table 3.1: Summary of imaging parameters by modality, reflecting the variability in acquisition protocols depending on the modality and the scanner setup.

Modality	Number of slices	Slice spacing	Image resolution	Parametric map
T1-MOLLI	9-19	5 mm	288×288 (A) 256×256 (B)	T1
DWI	5	10 mm	210×210 (A) 256×256 (B)	ADC
T2*	5	10 mm	288×288 (A) 512×512 (B)	T2*

The T1-MOLLI images were acquired at multiple TIs, ranging from 174 ms to 4452 ms. DWI images were acquired at multiple b-values, ranging from 0 to 500 s/mm^2 . The T2* mapping image data were acquired at multiple TEs, ranging from 3 ms to 62 ms. For each patient scan, the dataset includes MRI data from all sequences, their respective parametric maps, and manually segmented ground truth labels for the cortex and parenchyma. These segmentations were manually delineated by trained image analysts. Additionally, all images and parametric maps were registered prior to this study to ensure spatial alignment across modalities. Specifically, images were registered using the T1-MOLLI image at $\text{TI} = 1300 \text{ ms}$ as fixed reference image. Example images from the T1-MOLLI, DWI, and T2* sequences acquired with varying parameters can be found in Appendix A.

3.2 Data Exclusion

The initial step in the data processing involved handling the input data to support downstream training of the segmentation algorithm. Images from the dataset were manually reviewed to assess image quality, and to do an inventory of which imaging modalities and parametric maps are available for each patient visit. Some visits lacked one or more MRI sequences or were missing corresponding segmentation labels, which are essential for supervised learning. Thus, as a first processing step, data were inspected to exclude potential subjects with incomplete image data. Additionally, images were excluded due to image quality issues, such as poor image, presence of larger imaging artifacts, or abnormalities affecting overall clarity or visualization of the kidney. These exclusion criteria can be summarized as follows:

1. incomplete MRI data (e.g. missing registered T1-MOLLI images or T1 map)
2. missing or incomplete ground truth segmentation maps (e.g. missing cortex segmentation)
3. inconsistent slice count between images and corresponding segmentation map
4. poor image quality, such as pronounced artifacts (e.g. zebra stripes) or low contrast
5. presence of large renal lesions ($\varnothing > 30$ mm)

After applying these exclusion criteria, the final dataset consisted of 37 scans (8 from site A and 29 from site B).

3.3 Dataset Split

Given the limited size of the dataset after exclusion, the split was set to include as many training samples as possible, while still having data for validation and test available. To prevent data leakage, the dataset was split scan-wise, meaning that no data from the same patient scan appears in more than one subset. The MRI dataset was randomly split into three subsets for training (27 scans, 75%), validation, (6 scans, 15%) and testing (4 scans, 10%). Efforts were made to ensure a balanced distribution of renal volumes across the data splits.

3.4 Data Pre-processing

Before being input into the segmentation network, the image volumes go through several pre-processing steps. This pipeline included generating binary ground truth masks from manual segmentations, scaling and normalizing brightness and intensity values, and applying data augmentation techniques to increase variability in the dataset using the MONAI library. To further address the limited size of the dataset, a patch-based approach was used during training. This involved randomly cropping the original images into smaller patches, helping the model learn from a greater variety of spatial contexts.

3.4.1 Creation of Ground Truth Masks

The manually segmented ground truth masks were created by trained image analysts at Antaros Medical, who annotated the original MR images using green and cyan overlays to indicate the parenchyma and cortex in the left and right kidneys. These color-coded annotations served as the reference for generating binary segmentation labels. To convert the color segmentations into useful segmentation labels that serve as ground truth during training of the CNN, a series of transforms were applied. The first transform used was `ForegroundMaskd`, with the input being the color-annotated ground-truth segmentation. Within this transform, a custom HSV filter was applied to isolate the green and cyan annotations from the grayscale background, effectively removing non-relevant pixels. This process generated a binary mask for each endpoint, which was then used as a label for the network.

An example of the of manually segmented ground truth mask and resulting binary mask for both parenchyma and cortex is shown in Figure 3.2.

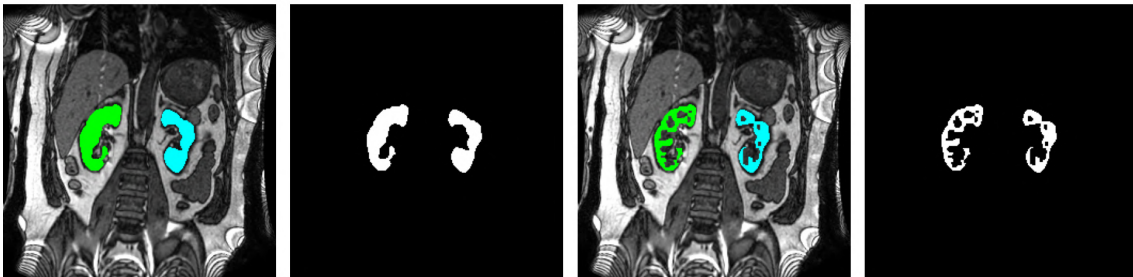


Figure 3.2: Example of ground truth segmentation and corresponding binary masks after applied transformations. From left to right: parenchyma ground truth, parenchyma binary mask, cortex ground truth, and cortex binary mask. The manually segmented ground image shows the left kidney (cyan) and right kidney (green), annotated by image analysts.

3.4.2 Brightness and Intensity Scaling

The MR acquisition and processing technique assigns intensity values to each voxel, reflecting the strength of the signal in the specific region. However, these values can vary significantly between scans due to various external factors, including scanner settings, patient movement, or background noise. Additionally, the tissue pathology also influences image contrast. For example, patients with CKD often exhibit reduced tissue contrast compared to healthy individuals. This inconsistency can affect overall brightness and intensity in acquired images, and generate differences even across scans from the same patient.

When combining data from different MRI sequences or modalities, these variations become even more pronounced. Each sequence can produce images with varying intensity distributions and contrasts, even for the same anatomical structures. To address this, and to be able to combine images with various intensities as multi-channel input in this work, histogram normalization was applied to the MRI data

using `HistogramNormalized`. This technique redistributes intensity values to enhance contrast and rescale them between 0 and 1, which enhanced the visibility of renal structures like cortex and medulla, while making intensity values more consistent across the dataset.

For the parametric maps, which have very different intensity distributions compared to standard MR images, a different strategy was used. The `ScaleIntensityRanged` transform was applied individually to each parametric map. Based on inspection of the corresponding intensity histograms, specific input ranges were chosen and scaled to the $[0, 1]$ interval. This approach enabled customized normalization, ensuring consistent scaling across maps despite large variations in raw intensity values.

3.4.3 Splitting into 2D Slices

The proposed model operates on 2D input slices extracted from volumetric (3D) images. To accommodate this input format, a pre-processing step was required to convert each 3D volume into a series of 2D slices. This was achieved by applying a custom function that systematically iterates through the depth dimension of each volume, extracting one slice at a time. As a result, each 3D image was converted into a series of 2D slices matching the original acquisitions, significantly increasing the total number of training samples available for the model. Additionally, all slices were resized to 256×256 to ensure consistent spatial resolution throughout the dataset.

3.4.4 Augmentations

Data augmentation is a widely used technique in machine learning, particularly useful when working with limited datasets. It helps improve model generalization by introducing variability into the training data. In this work, augmentations were implemented using MONAI’s dictionary-based transforms, ensuring consistent application to both images and their corresponding labels.

A key strategy to address the limited dataset was using a patch-based approach, where the original training images were randomly cropped into smaller patches. This not only increased the number of training samples but also ensured the patches were small enough to fit into GPU memory. Specifically, the `RandCropByPosNegLabeld` transform was used to extract patches based on a defined ratio of foreground and background. For each training image, four patches of size 160×160 were generated. Given the relatively small size of the kidneys in the full input image, patches were sampled with a probability condition ($p = 0.5$) of being centered on foreground regions to ensure sufficient coverage of the ROI.

To further increase the size and variability of the training data, additional augmentations were applied to the extracted patches. These included random scaling, rotation and zooming to simulate anatomical differences. Additionally, random contrast adjustments were introduced to account for variations in acquisition conditions.

3.5 Network Configuration

A 3D ResUNet architecture proposed by Inoue et al. (2023) was identified as a strong baseline for this thesis in kidney segmentation, due to its demonstrated performance on a similar task and its use of residual connections to enhance feature learning [7].

The proposed architecture, seen in Figure 3.3, follows a U-Net structure with four downsampling steps. Each step includes a max pooling layer that reduces the spatial dimensions by a factor of two, resulting in a total downsampling of 2^4 . The number of filters in the convolutional layers increases progressively, using 16, 32, 128, and 256 filters, respectively. A key feature of the network is the inclusion of two residual connections for each convolutional block, which as mentioned in the introduction helps maintain gradient flow and improve training stability. While the original ResUNet was designed for 3D segmentation, this project implements a 2D adaptation of the architecture. Additionally, dropout layers were added to help prevent overfitting during training, which was particularly important given the limited size of the dataset. The proposed network configuration is designed to handle multi-dimensional inputs, making it suitable for processing multi-channel data.

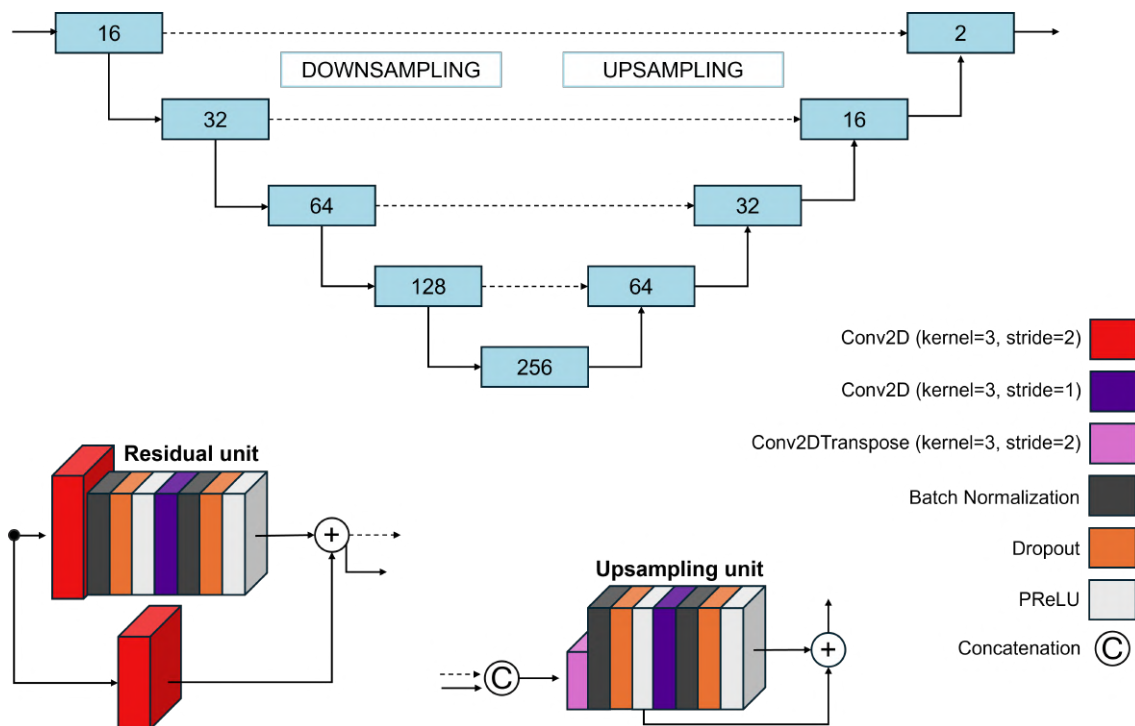


Figure 3.3: Overview of proposed 2D ResUNet architecture. The left side shows the downsampling path with residual units, where convolutional channels increase from 16 to 256 to extract features. The right side correspond to the upsampling path, where channels decrease from 256 to the number of output classes, combined with skip connections.

3.6 Model Implementation and Experimental Setup

The following subsections describe the experimental setup and implementation details of proposed segmentation models based on the above-mentioned network architecture. This includes the design of three main experiments conducted to evaluate different segmentation approaches, and the types of MRI input data used, including single-modality, multi-channel, and multi-modal setups.

3.6.1 Single-Channel T1-MOLLI-based Segmentation

The primary experiment evaluated the performance of the proposed network configuration on T1w kidney images, specifically its ability to accurately segment parenchyma and cortex. The T1-MOLLI sequence was selected as a starting point due to its higher corticomedullary differentiation (CMD), providing visible contrast between the cortex and medulla.

Using the above-mentioned network architecture, two distinct segmentation strategies were implemented and compared. The 2D slice-wise approach was compared to a 3D implementation of the network, similar to the one proposed in the original reference work. This setup allowed for a comparative study between a volumetric segmentation method and a slice-wise approach. Experiments were performed to investigate whether spatial relationships between adjacent slices could be leveraged to improve segmentation performance.

To ensure robustness in performance evaluation, a 6-fold cross-validation was used instead of the conventional train-validation-test split. This offered a more robust assessment of the overall network performance in the early experimental phase by minimizing bias in model evaluation. Specifically, the full dataset was divided into six equally sized folds. In each iteration, one fold served as the validation set while the remaining five were used for training. This iterative process ensures that each subset serves as the validation set exactly once, providing insights to the network performance on data with varying distribution.

3.6.2 Multi-Channel T1-MOLLI-based Segmentation

After evaluating the network’s overall performance on the dataset, an experiment was conducted using combinations of multiple TIs as channel-wise input to the model. This approach aimed to determine whether segmentation performance and generalizability could be improved by leveraging the varying contrasts provided by different TIs. To identify the optimal combination, each TI was individually input into the model and evaluated based on segmentation performance. These were then combined into a multi-channel input, with the goal of utilizing the unique features of each TI. The proposed method utilizes early fusion of the inputs to a multi-channel input image that can be fed into the segmentation network.

The best-performing configuration was then fine-tuned to maximize the model performance, incorporating additional data augmentations and dropout layers to enhance generalization.

3.6.3 Multi-Modal Segmentation Integrating DWI and T2*

As a final experiment, the initial segmentation approach, using only T1-MOLLI image data as input, was extended to incorporate multi-modal MRI data. Instead of relying on a single imaging sequence, the approach was extended to combine both DWI and T2* mapping images as inputs to evaluate whether integration of multiple image modalities could improve segmentation performance and generalizability.

Following a similar strategy to the proposed multi-channel T1-MOLLI approach, multi-channel inputs were constructed by concatenating DWI images acquired at different b-values and T2* mapping images captured at varying echo times. Additionally, parametric maps derived from all three modalities were added as inputs. By allowing a CNN to process these inputs simultaneously, it was investigated whether it can simultaneously process complementary information across modalities to learn a richer and more diverse representation of tissue characteristics.

Since ground truth volume segmentations were not available specifically for the DWI and T2* modalities, segmentation masks were generated by resampling existing MOLLI segmentations. Five MOLLI ground truth slices were selected to match corresponding slices in the DWI and T2* datasets. These served as a new ground truth set for training and evaluating the multi-modal segmentation model.

3.7 Training and Evaluation

All models were trained on a NVIDIA GeForce RTX 2080 Ti GPU until convergence, for up to 4000 epochs. To accelerate the training process, MONAI’s CacheDataset was used, which caches transformed data into memory during training. Adam optimizer was used, with a set learning rate of 10^{-4} . Model performance was continuously monitored using validation loss, where the best-performing model was saved at the lowest validation loss during training.

For evaluation of segmentation performance, the segmentation output was compared to manual ground truth annotations. The output probabilities were passed through a sigmoid activation to obtain binary masks. The segmentation overlap was then evaluated using both Dice score and IoU. These metrics were computed slice-wise during training and validation, and then aggregated to provide an overall mean. Both metrics were computed separately for the renal parenchyma and cortex, offering insight into segmentation accuracy and precision across the two target regions.

During evaluation, the entire image slice was processed step by step using a sliding window with the shape 160×160 and a 50% overlap along each spatial dimension. In overlapping regions, outputs were averaged to produce the final prediction. Zero-padding was applied to the edges of each slice as needed.

3.8 Post-processing

To refine the raw model predictions and prepare them for downstream analysis, a series of post-processing steps were applied. First, small objects were removed from

the output segmentation masks using the `RemoveSmallObjects` transform, which filters out isolated clusters of pixels below a given size threshold. This helped reduce noise and false positives in the resulting segmentation output. An example of this transformation is illustrated in Figure 3.4.

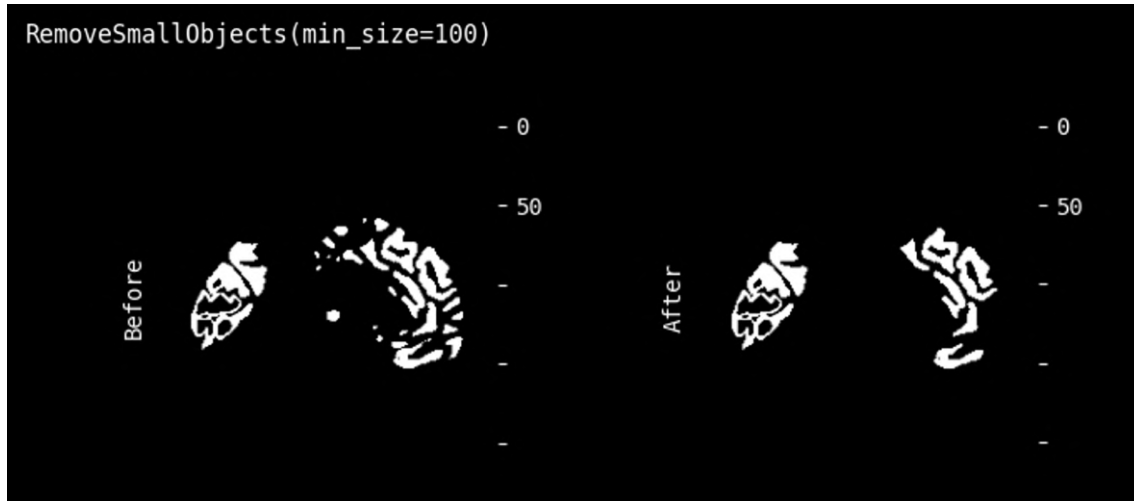


Figure 3.4: Visualization of the `RemoveSmallObjects` transform in MONAI, here removing clusters with 100 pixels or less. Figure adapted from MONAI [45].

Since 2D slices were segmented independently, an essential post-processing step was the aggregation of the predicted slices into a reconstructed 3D volume. This volumetric assembly ensured the predicted segmentation aligned spatially with the original input scan, and was necessary for a more accurate quantification of the endpoints. Since the model segments both kidneys bilaterally, the final prediction volume was divided at the midline to separate the left (SIN) and right (DX) kidneys. This allowed for computation of kidney volume and tissue-specific parameters (T1, ADC, T2*) for each kidney separately.

3.9 Volume Quantification

Once the predicted masks for each scan were generated and post-processed, kidney volume measurements were obtained from the predicted masks by summing the total number of voxels corresponding to parenchyma and cortex separately for each scan. The sum of voxels was then scaled by the voxel dimensions of the respective scan, resulting in the total kidney volume in milliliters for both the SIN and DX kidneys.

The quality of the segmentation and the model’s ability to automatically quantify renal volumes were assessed by evaluating how well the predicted endpoint values matched the reference values, computed from manually created segmentations. This was quantified using the coefficient of determination (R^2), which provides a measurement of the correlation between predicted and reference volumes. The volumetric analysis was performed across all data splits, highlighting differences in how well the model had adapted to the training data, the distribution of training data volumes, and the accuracy of volume predictions on the validation and test data.

3.10 Quantification of Parametric Values

To quantify the parametric values, median values in the renal cortex and medulla respectively were extracted from parametric maps using the predicted segmentation masks. The medulla region was derived by subtracting the predicted cortex mask from the parenchyma mask, both of which were obtained from the model's output. This resulted in a binary medulla mask, excluding the cortical regions.

Before extracting parametric values of the two tissues, both the cortex and medulla masks were refined using binary erosion. Erosion was applied with a square structural element to shrink the ROI, removing edge pixels that may contain noise or partial volume effects. This step ensured that the extracted values were representative of the central, more reliable tissue regions, and not influenced by misclassifications at the outer MR slices. An example of erosion using a 3×3 structural element is illustrated in Figure 3.5

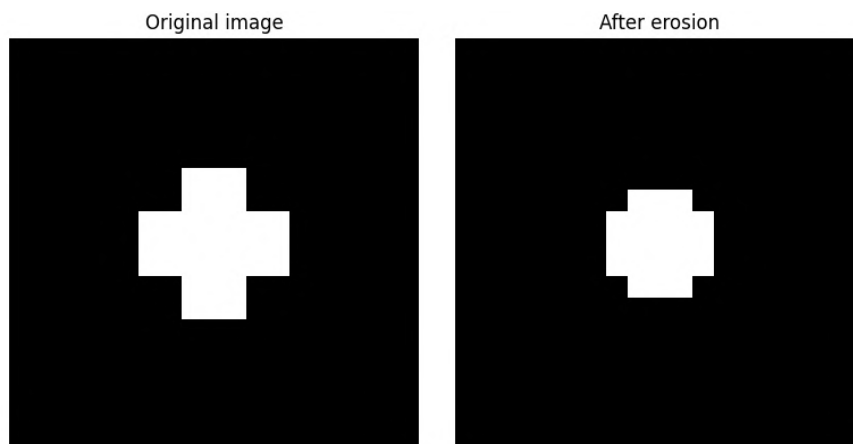


Figure 3.5: Effect of erosion using a 3×3 structural element.

After applying erosion, the refined cortex and medulla masks were applied to the original parametric maps (T1, ADC, and T2*) to calculate the median of the intensity values within the mask. As the parametric maps are sensitive to variations outside the ROI, where intensity values can vary significantly, extracting the median value ensures that the resulting values are more representative of the true tissue characteristics and less sensitive to outliers. For each image volume, pixel-wise parametric values were aggregated across a number of consecutive centered slices to reduce slice-to-slice variability. Lastly, the resulting median values were compared against reference values provided by image analysts using R^2 correlation values.

4

Results

In this chapter, the results from the training and evaluation of the implemented segmentation network are presented. It begins with a brief overview of the dataset and key findings from data exploration. The subsequent sections follow the structure of the experimental setup. First, results from single-channel segmentation of the renal parenchyma and cortex using T1-MOLLI images are presented. This is followed by results from multi-channel segmentation experiments using multiple inversion times from T1-MOLLI, along with the quantification of kidney volume and median T1 values. Finally, the chapter concludes with results from multi-modal segmentation that integrates DWI and T2* input data, including the quantification of ADC and T2* values for both cortex and medulla.

4.1 Data Exploration

Exploration of the dataset revealed the presence of multiple artifacts that affected image quality and the visibility of key anatomical structures of the kidney. Zebra stripe artifacts were among the most prevalent, appearing with differing severity across scans. While images with severe distortions that obscured renal boundaries were excluded due to potential impairment of segmentation performance, small artifacts with minimal impact on the ROI were retained. Examples of such images are presented in Figure 4.1, which illustrates the range of artifacts observed.

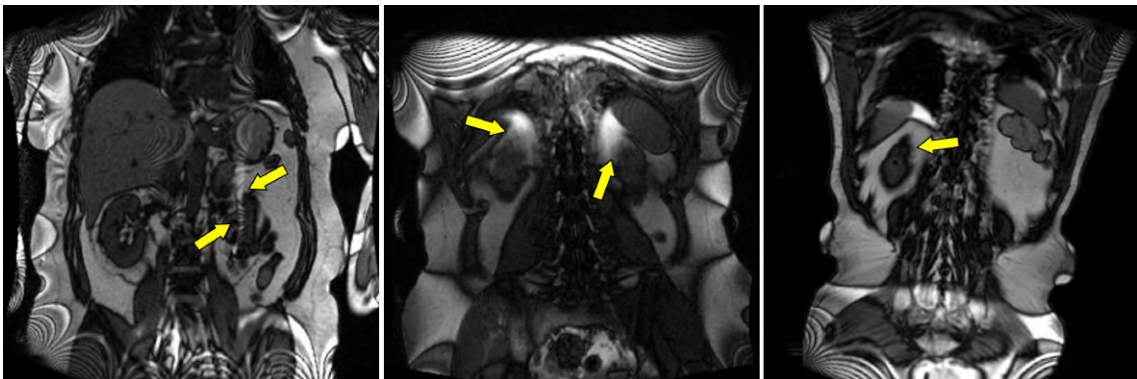


Figure 4.1: Examples of MRI artifacts and image distortions observed in the dataset. Yellow arrows highlight regions impacted by zebra striping and poorly defined kidney boundaries.

The leftmost image shows visible zebra striping. In this case, the artifact was determined to have little impact on the ROI, and the image was retained in the dataset. The second image shows blurring of the upper boundaries of the kidneys, likely caused by banding artifacts in outer slices. The third image shows poorly defined kidney edges, likely from partial volume effects. These distortions were commonly observed throughout the dataset, posing challenges for the segmentation models.

4.2 Single-channel T1 MOLLI-based Segmentation

To evaluate the performance of the 2D and 3D segmentation approaches for renal parenchyma and cortex, four models were trained using 6-fold cross-validation and T11400 as input. Each model was evaluated individually and the results were averaged across folds. The results are presented in Table 4.1, showing the average performance and standard deviation of across the cross validation experiments.

Table 4.1: 6-fold cross validation results for 2D and 3D ResUNet models. Performance metrics are shown as mean \pm standard deviations across all folds.

Network	Endpoint	Dice	IoU
3D ResUNet	Parenchyma	0.8290 ± 0.0254	0.7112 ± 0.0372
	Cortex	0.6925 ± 0.0321	0.5335 ± 0.0372
2D ResUNet	Parenchyma	0.8721 ± 0.0292	0.7930 ± 0.0365
	Cortex	0.7942 ± 0.0248	0.6753 ± 0.0280

As shown in Table 4.1, the 2D ResUNet outperformed the 3D ResUNet in terms of both Dice and IoU. For parenchyma segmentation, the 2D model achieved a mean Dice of 0.8721 ± 0.0278 compared to 0.8290 ± 0.0242 for the 3D model. A similar trend was observed for cortex segmentation, where the 2D model reached a mean Dice of 0.7942 ± 0.0236 , notably higher than the 3D model’s 0.6925 ± 0.0327 . The IoU scores followed the same pattern, further suggesting that the 2D ResUNet was more effective at both segmentation tasks.

Figure 4.2 illustrates a comparison of example segmentation outputs for the center slice of two patient scans. While the parenchyma segmentation are comparable for both 2D and 3D ResUNet, the cortex segmentation output clearly differs between the two architectures. In the parenchyma case, both models struggle to accurately delineate the outer kidney boundary. These challenges become more apparent in the cortex segmentation task, where the renal cortex’s complex structure introduces more intricate boundary regions. The segmentation overlays demonstrate that the 2D ResUNet provides more accurate cortex delineation compared to the 3D ResUNet. In the 3D model overlays, a higher occurrence of false positives is observed, particularly in regions corresponding to the medullary pyramids. This leads to outputs that more closely resemble parenchyma segmentation, indicating a reduced performance in distinguishing cortical boundaries. These observations are consistent with the quantitative performance metrics shown in Table 4.1, which confirm the superior performance of the 2D model in segmenting the parenchyma and cortex.

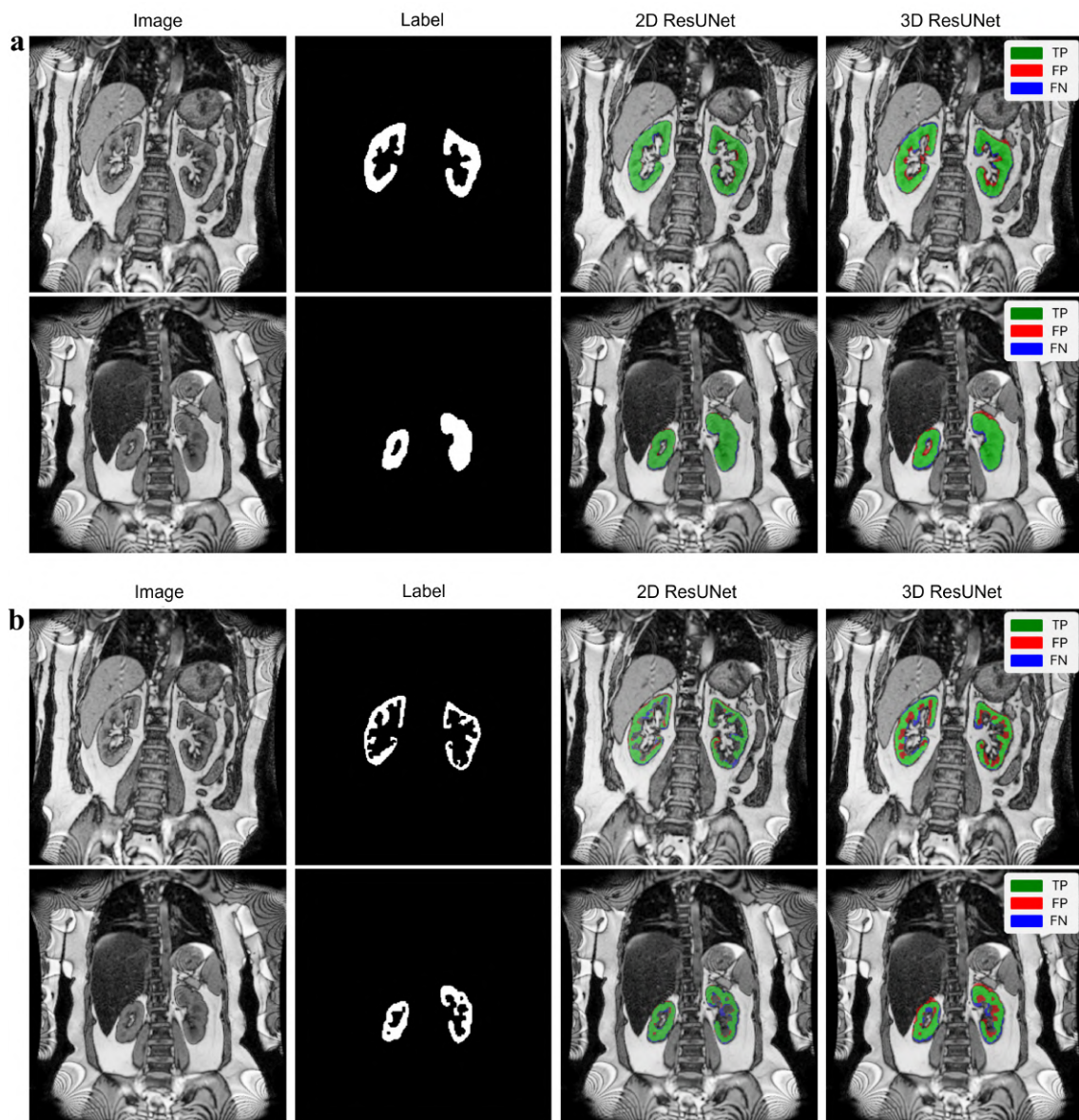


Figure 4.2: Comparison of segmentation results for (a) renal parenchyma and (b) renal cortex using 2D ResUNet and 3D ResUNet models. The figures illustrate the original image, ground truth segmentation label, and overlay of image, label and model predictions, where true positives (TP) are represented in green, false positives (FP) in red, and false negatives (FN) in blue.

4.3 Multi-channel T1 MOLLI-based Segmentation

To explore the impact of multi-dimensional input data, each available T1 image and the T1 map were first evaluated individually as single-channel inputs. This experiment aimed to determine the standalone contribution of each input to segmentation performance. Segmentation performance was assessed for both renal parenchyma and cortex using standard performance metrics on the validation and test sets. The results for each input channel are summarized in Table 4.2 - 4.3.

Table 4.2: Parenchyma segmentation performance using single-channel input images of different inversion time (TI).

Input	Val Dice	Val IoU	Test Dice	Test IoU
TI200	0.8148	0.7109	0.8310	0.7125
TI800	0.8443	0.7676	0.7878	0.7050
TI1400	0.8584	0.7955	0.8488	0.7717
TI2000	0.8517	0.7773	0.7982	0.7037
TI2500	0.8291	0.7513	0.8019	0.7068
T1-map	0.8911	0.8204	0.8520	0.7740

As can be seen in the results, the T1 map consistently achieved the best overall performance among the single-channel inputs. For parenchyma segmentation, the T1 map yielded the highest Dice scores on both the validation and test sets (0.8911 and 0.8520, respectively). The best-performing individual TI was 1400 ms, which also demonstrated strong segmentation accuracy, with validation and test Dice scores of 0.8584 and 0.8488. A similar pattern can be observed for cortex segmentation, where the T1 map again led to the best performance on the validation set, achieving a Dice score of 0.8155 and IoU of 0.7013. However, the best performance on the test set was obtained with TI1400, which reached a Dice score of 0.8196.

Table 4.3: Cortex segmentation performance using single-channel input images of different inversion time (TI).

Input	Val Dice	Val IoU	Test Dice	Test IoU
TI200	0.7024	0.5508	0.6970	0.5498
TI800	0.7491	0.6211	0.7432	0.6163
TI1400	0.7868	0.6749	0.8196	0.7090
TI2000	0.7710	0.6511	0.7588	0.6404
TI2500	0.7543	0.6292	0.7382	0.6134
T1-map	0.8155	0.7013	0.7898	0.6787

Various combinations of inversion times were evaluated as multi-channel inputs to the 2D ResUNet. Starting with TI1400 and T1 map as a base input, additional input channels were systematically added or removed to enhance segmentation performance. The results from the top five performing models for parenchyma and cortex segmentation are presented in Table 4.4 and Table 4.5, respectively.

Table 4.4: Comparison of the top five multi-channel input combinations for segmentation of the renal parenchyma.

Input	Val Dice	Val IoU	Test Dice	Test IoU
Tl1400 + T1map	0.8799	0.8158	0.8691	0.8011
Tl1400, 2000 + T1map	0.9018	0.8399	0.8914	0.8217
Tl200, 800, 1400, 2000 + T1map	0.9055	0.8454	0.8959	0.8275
Tl800, 1400, 2000, 2500 + T1map	0.9139	0.8504	0.8735	0.8027
All input channels	0.9085	0.8470	0.8410	0.7650

Increasing the number of input channels improved model performance, enhancing segmentation accuracy on both validation and test sets. One of the best-performing multi-channel input configurations was the combination of dropping the inversion time of 2500 ms. The model trained on this input combination achieved a validation Dice score of 0.9055 and a test Dice score of 0.8959 for parenchyma segmentation, along with the second-highest IoU scores for both validation and test data.

The model trained with all available input channels showed high validation performance but notably lower test performance. In contrast, models trained with carefully selected input combinations achieved better test results, as indicated by improved evaluation metrics. A similar trend was observed in another top-performing model where the 200 ms inversion time was excluded. This model achieved the highest validation Dice score of 0.9139 and a validation IoU of 0.8507. However, its performance on test data was lower compared to other models.

Table 4.5: Comparison of the top five multi-channel input combinations for segmentation of the renal cortex.

Input	Val Dice	Val IoU	Test Dice	Test IoU
Tl1400 + T1map	0.7992	0.6219	0.8212	0.7110
Tl1400, 2000 + T1map	0.8253	0.7150	0.8182	0.7123
Tl800, 1400, 2000 + T1map	0.8280	0.7219	0.8223	0.7106
Tl200, 800, 1400, 2000 + T1map	0.8242	0.7166	0.8392	0.7396
All input channels	0.8267	0.7196	0.8272	0.7184

For cortex segmentation, the multi-channel experiments showed that combining inversion times that individually yielded the highest segmentation performance improved accuracy. Similar to segmentation of the parenchyma, including all channels led to one of the top-performing models, but with slightly reduced performance on the test data. Notably, the model that included all inversion times except the 2500 ms image demonstrated consistently high performance across both validation and test datasets for cortex segmentation as well. This input configuration achieved a validation Dice score of 0.8267 and IoU of 0.7196, as well as a test Dice score of 0.8272 and IoU of 0.7184. These results indicate a good balance between model accuracy and generalization to unseen data.

4.3.1 Performance of Fine-tuned T1-MOLLI Model

Fine-tuning of the 2D ResUNet architecture was done with the goal of improving accuracy and robustness of the model. A multi-channel input of five channels representing different inversion times and the parametric T1 map, previously identified as optimal for both anatomical regions, were used as input to the network. Given that this input combination proved to be optimal for both parenchyma and cortex, a multi-label approach was also explored, training a single model to segment both ROIs simultaneously. Its performance was compared against single-label models for segmenting parenchyma and cortex separately, as summarized in Table 4.6. To enhance generalization, additional data augmentation were applied, along with dropout ($p = 0.1$) for regularization. Both model configurations were then trained to convergence using the Adam optimizer with a learning rate of 10^{-4} .

Table 4.6: Comparison of the performance of single-label and multi-label segmentation models trained on multi-channel T1-MOLLI input.

Network	Val Dice	Val IoU	Test Dice	Test IoU
Single-label (2D ResUNet)				
Parenchyma	0.9130	0.8534	0.8962	0.8254
Cortex	0.8352	0.7260	0.8466	0.7428
Multi-label (2D ResUNet)				
Parenchyma	0.9298	0.8737	0.9089	0.8453
Cortex	0.8614	0.7598	0.8552	0.7557

As shown in Table 4.6, the multi-label network outperformed the performance of the models segmenting both segmentation endpoints separately. For instance, the multi-label model achieved a Dice score of 0.9298 and 0.8614 for parenchyma and cortex respectively on the validation set, compared to 0.9130 and 0.8352 for the separate segmentation models. Notably, the multi-label model also demonstrated strong generalization to unseen test data, with Dice scores of 0.9089 (parenchyma) and 0.8552 (cortex). The Dice loss and segmentation performance during training the multi-label network is seen in Figure 4.3.

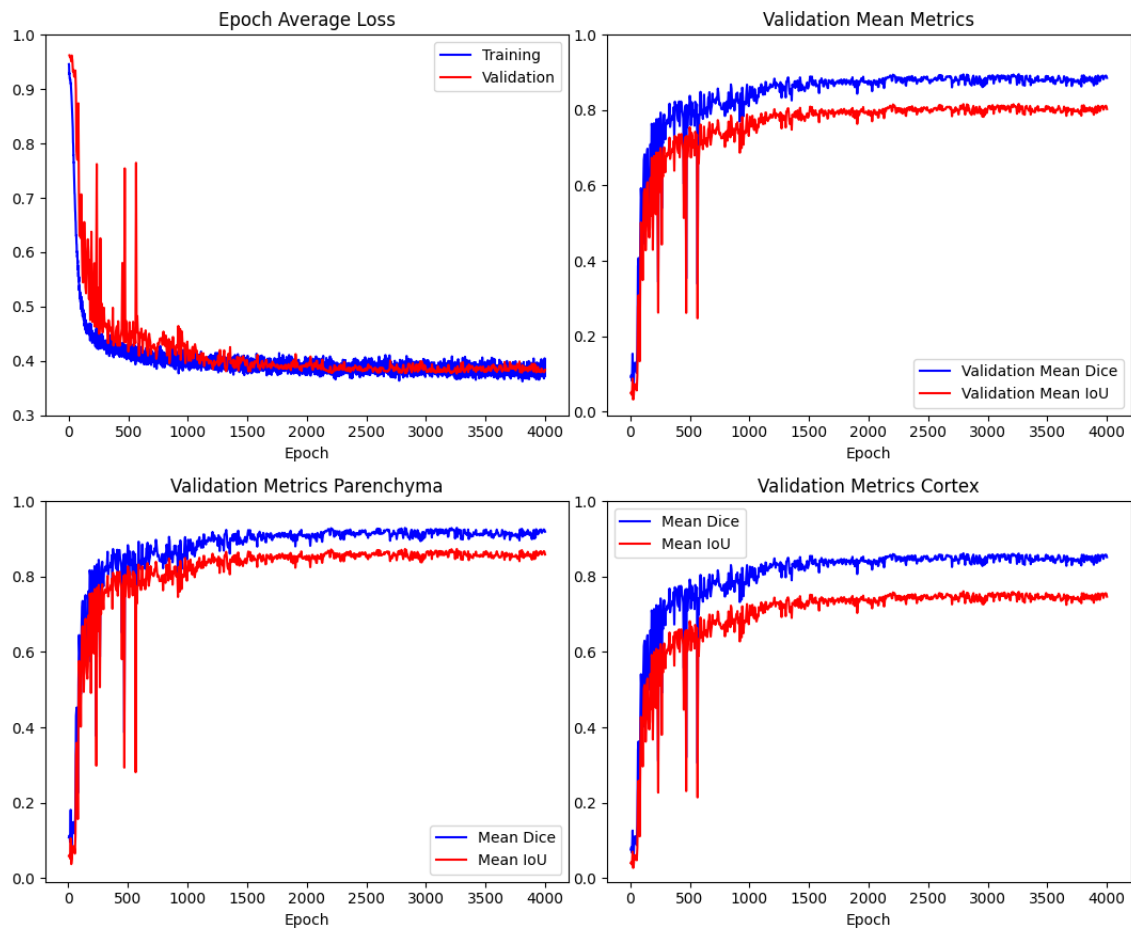


Figure 4.3: Training and validation performance of the proposed multi-label 2D ResUNet, showing Dice loss and evaluation metrics progress during training (Adam optimizer, learning rate of 10^{-4}). Dice score and IoU are reported separately for the parenchyma and cortex, as well as averaged across both regions.

Figure 4.4 shows example segmentations of the renal parenchyma using the proposed multi-channel model. Qualitatively, the model demonstrates a high degree of accuracy in delineating the whole parenchymal region. Despite false detections along the kidney’s outer boundaries, the predicted segmentations are overall comparable to the manual annotations. The segmentation results further indicate that the model is effective in identifying more complex anatomical features, such as regions where the kidney is closing up or contains smaller cysts. Furthermore, the model was able to delineate the kidney in outer slices, where the target region is small. However, performance inconsistencies were observed between scanner types. Specifically, in images acquired with the A-scanner, the model exhibits worse segmentation performance, particularly in the upper portions of the kidney. This effect could not be seen in test images from the B-scanner.

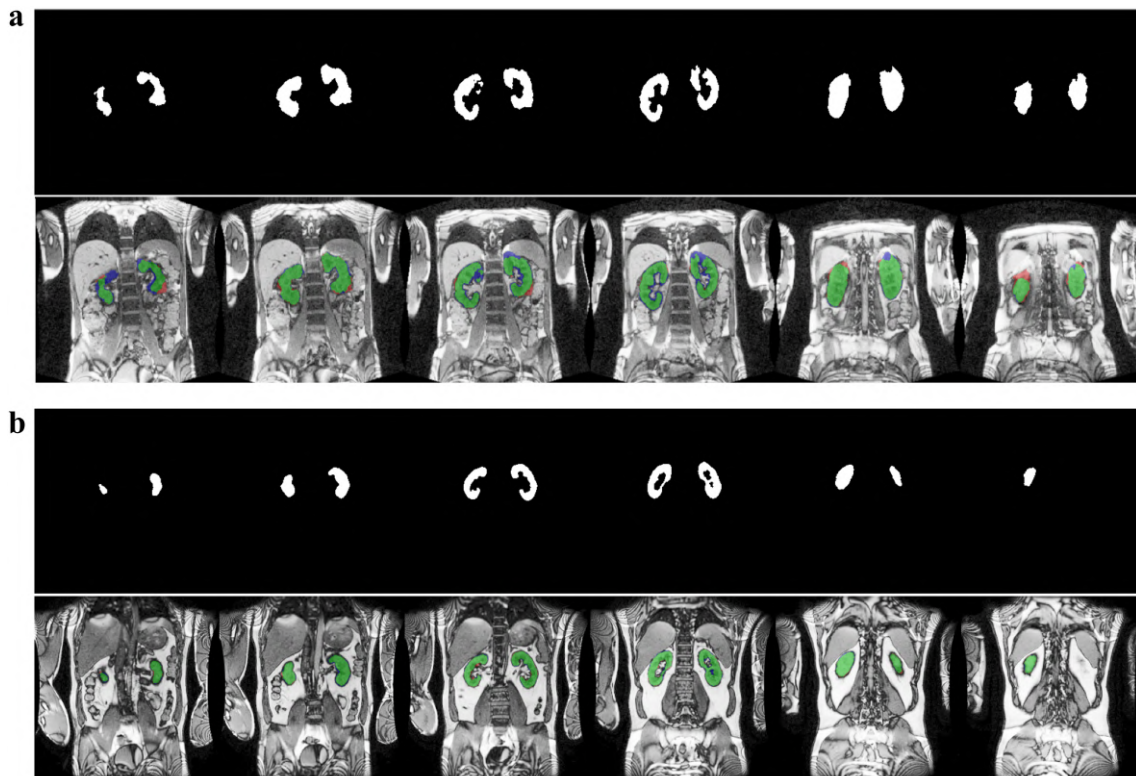


Figure 4.4: Representative examples of parenchyma segmentation results from two test cases: **a)** scan acquired at site A, and **b)** scan from site B. Each figure shows selected image slices with the predicted segmentation mask, as well as an overlay on the TI 1400 ms image. The color coding in the overlays is as follows: green - true positives, blue - false negatives, and red - false positives.

While the multi-channel model provides consistent results on unseen data, certain patterns of failure cases can be observed, as illustrated in Figures 4.5 - 4.7. Particularly, results indicate oversegmentation in the renal pelvis region, which appears as false positives in that area. On the other hand, undersegmentation tends to occur in slices corresponding to the most anterior and posterior parts of the kidneys, where the model is more conservative than the manual annotations. Additionally, the outermost slices containing only background and no kidney tissue, proved challenging for the model, often resulting in false positive predictions where non-kidney structures were incorrectly segmented as kidney tissue.

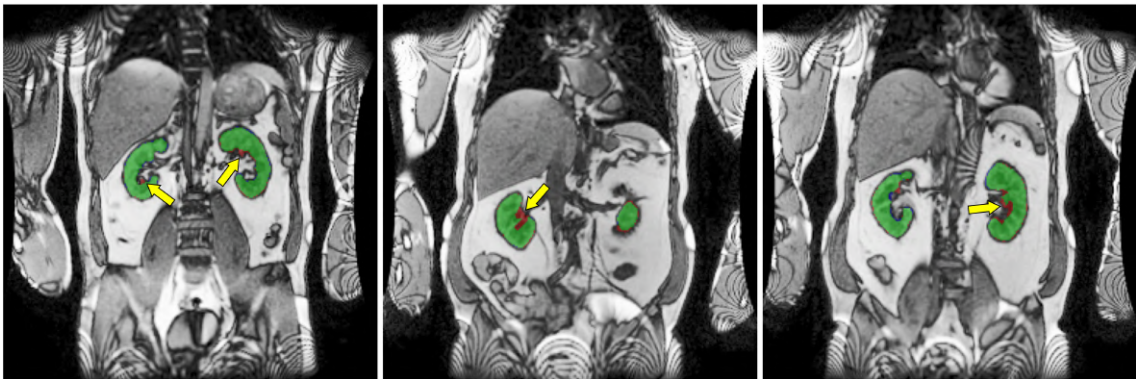


Figure 4.5: Examples of renal parenchyma segmentation results, illustrating classification outcomes where the model oversegments in the renal pelvis region. The color coding is as follows: green - true positives, blue - false negatives, and red - false positives.

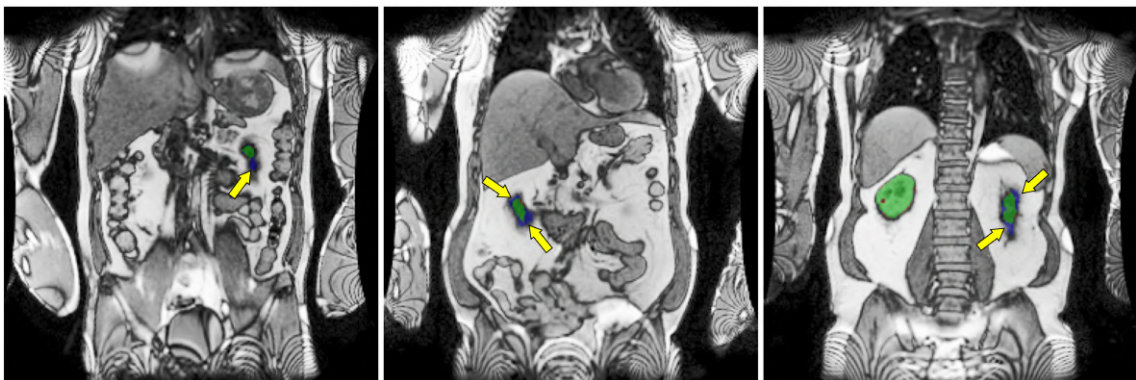


Figure 4.6: Examples of renal parenchyma segmentation results, illustrating classification outcomes where the model undersegments in the most anterior and posterior parts of the kidney. The color coding is as follows: green - true positives, blue - false negatives, and red - false positives.



Figure 4.7: Examples of renal parenchyma segmentation results, illustrating classification outcomes where the model falsely predicts other structures in background-only slices. Red indicates false positives.

Figure 4.8 presents example segmentation results for the renal cortex. By evaluating the model’s ability to predict cortex regions, it can be observed that outer cortical boundaries are generally delineated with high precision. However, it is obvious that the segmentation quality is reduced in the inner cortex regions, particularly near the medulla and renal columns. The results indicate that both FP and FN occur in these regions, indicating a reduced robustness and performance of the model in more complex anatomical regions. This observation aligns with the quantitative results, where both Dice and IoU scores are significantly lower for cortex segmentation.

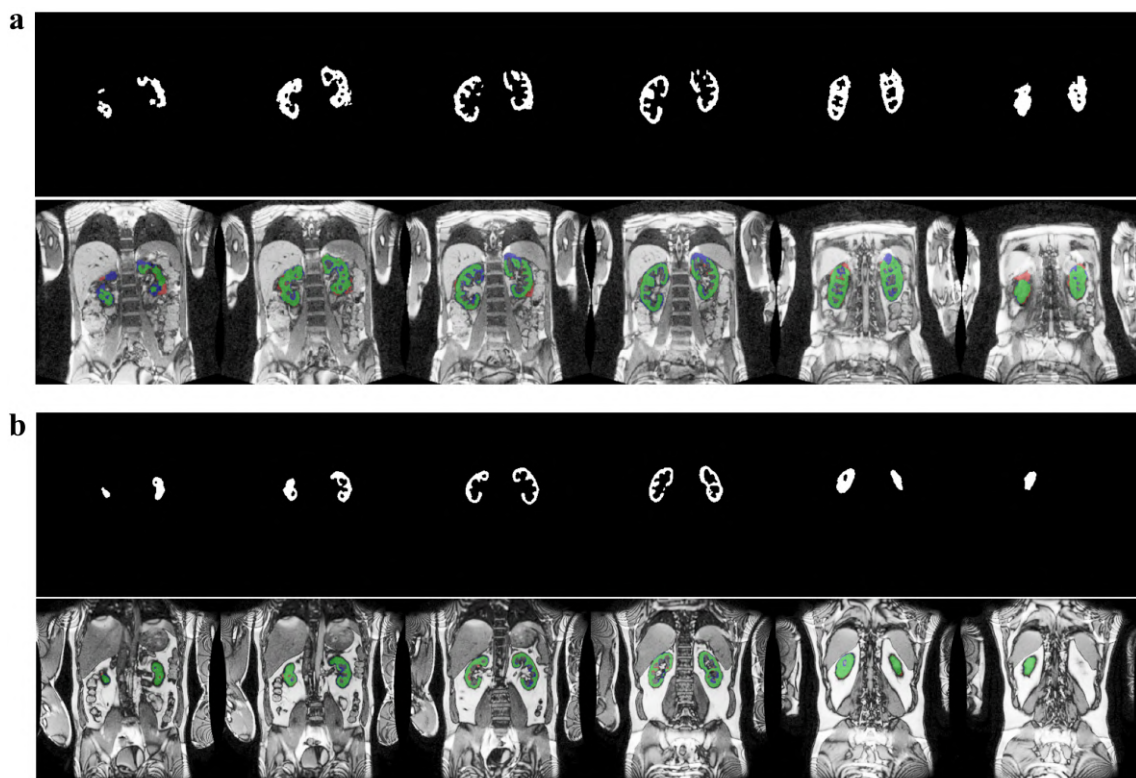


Figure 4.8: Representative examples of cortex segmentation results from two test cases: **a)** a scan acquired at site A, and **b)** a scan from site B. Each figure shows selected image slices with the predicted segmentation mask, as well as an overlay on the TI 1400 ms image. The color coding in the overlays is as follows: green - true positives, blue - false negatives, and red - false positives.

4.3.2 Quantification of Kidney Volume

The fine-tuned, multi-label segmentation model was evaluated on its predictive abilities in terms of quantifying renal parenchymal and cortex volume, as well as the median T1 value of cortex and medulla, respectively. Given the generated output predictions for each ROI, precise volume measurements for SIN and DX kidneys could be calculated. The results of a volume-wise comparison from manual and automatic segmentations are shown in Figure 4.9.

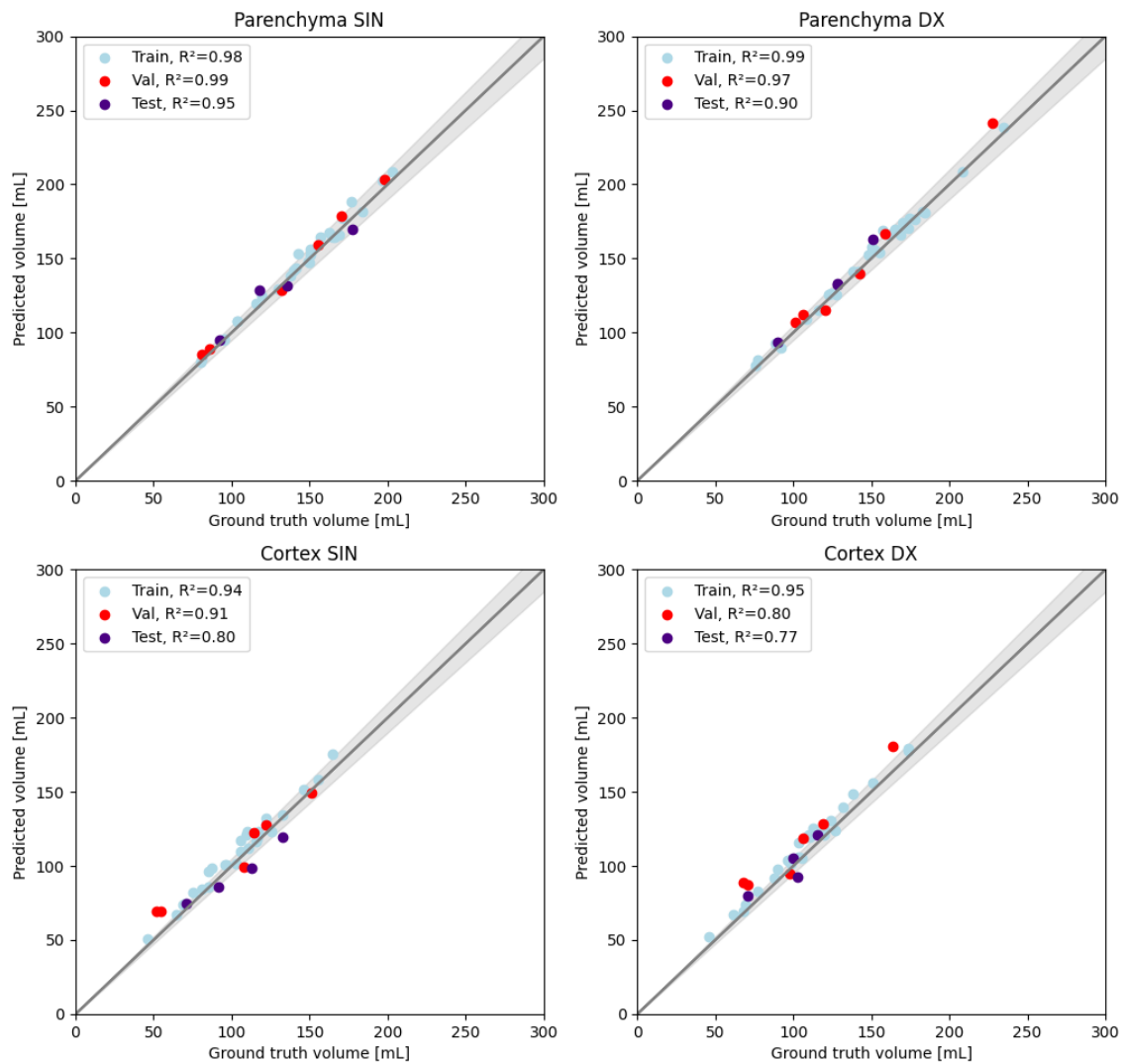


Figure 4.9: Correlation of ground truth and predicted volume for parenchyma and cortex. The identity line represents perfect correlation between the CNN-predicted and ground-truth segmentation. Grey area corresponds to a 5% volume difference.

The parenchymal volume estimates using summation of the voxels in the ground truth segmentations correlate well with the results obtained by summing up voxels annotated automatically by the designed neural network. The coefficient of determination for the validation set was 0.99 (SIN) and 0.97 (DX), while the test set achieved R^2 values of 0.95 (SIN) and 0.90 (DX). Furthermore, the predictions are in good accordance with the line of identity. Overall, the model showed better performance on the left (SIN) kidney, with higher correlation values observed on both the validation and test sets. The majority of data points fall within a 5% difference margin, with some data points towards the outer boundary of this interval.

The correlation is slightly weaker for the cortex volume estimation, particularly in the validation and test set achieving val: R^2 of 0.91 (SIN) and 0.80 (DX) and test: R^2 of 0.80 (SIN) and 0.77 (DX). Although predicted cortex volumes generally follow the identity line, similar to parenchymal estimation, greater volume differences can

4. Results

be observed, resulting in some predictions outside the 5% limit. This suggests reduced model precision in cortex volume prediction compared to parenchyma, as also indicated by the segmentation performance metrics. It can be observed that the variability of the cortex volume prediction is larger than for parenchyma, and that the differences between left and right kidneys are more obvious. For instance, the model tends to underestimate of cortical volume of the SIN kidney in the test set whereas this behavior is less obvious for the DX kidney. Furthermore, there are consistent results between the left and right kidneys, where the SIN kidney gets better correlation values for both segmentation endpoints.

Further analysis on the model's ability to quantify renal volumes are visualized in the Bland-Altman plots, Figure 4.10. These plots illustrate the agreement between the automatically and manually predicted kidney volumes for renal parenchyma and cortex, highlighting percentage differences and any potential systematic bias in the predictions.

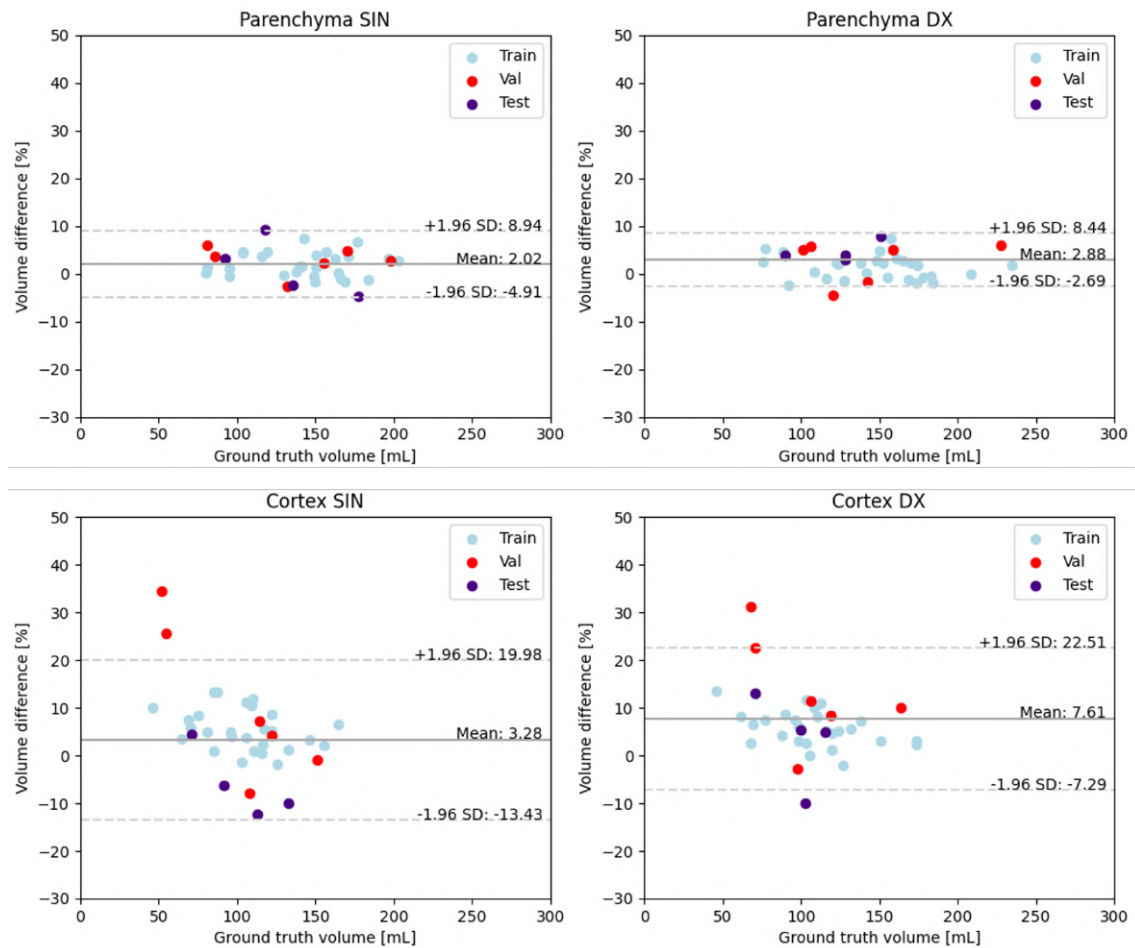


Figure 4.10: Bland-Altman plots of agreement for volume prediction and ground truth volume for parenchyma and cortex. Mean and standard deviation (SD) are calculated globally across all data sets. The solid line represents the mean, while the dashed lines indicate the limits of agreement, calculated as the mean ± 1.96 times the SD.

These plots reveal a consistent overestimation bias in the predicted volumes, observed for both parenchyma and cortex across both kidney lateralities. While this bias is small, this indicates that the model tends to overestimate the volume. Estimation of parenchymal volume, for both left (SIN) and right (DX) kidney, show a mean volume difference (%) across all data splits of +2.02% and +2.88%, respectively, indicating this overestimation. Furthermore, the mean and limits of agreement for both lateralities in parenchyma are similar, indicating that there is no obvious bias depending on the laterality of the kidney.

Similar to parenchymal volume estimation, cortex volume is overestimated as well, looking across all data splits. While cortex estimation demonstrates a small global bias for SIN kidney (+3.28%), the DX kidney have the highest mean volume difference (+7.61%). However, while the training set shows low variability and high consistency in estimating cortical volume, the variability in both validation and test set is substantial, resulting in outliers outside the limits of agreements. For both lateralities, the variability is particularly increased for small-to-mid size kidneys, indicated with larger relative error.

Table 4.7: Mean \pm SD of volume difference (in %) between predicted and delivered volumes, across kidney regions and dataset splits.

Dataset	Parenchyma SIN	Parenchyma DX	Cortex SIN	Cortex DX
Train	1.81 ± 2.61	2.79 ± 5.14	2.11 ± 2.46	3.66 ± 3.42
Validation	2.63 ± 2.86	2.79 ± 3.92	8.42 ± 15.86	9.56 ± 7.16
Test	3.53 ± 4.68	3.38 ± 2.19	-4.58 ± 5.58	0.76 ± 8.89

The Bland-Altman analysis summarized in Table 4.7 evaluates the agreement between predicted and manual kidney volumes across parenchyma and cortex regions, for left (SIN) and right (DX) kidneys, and stratified by dataset split. Overall, parenchymal volume predictions demonstrate consistent performance, with low mean differences and relatively stable variability across training, validation, and test sets. This indicates that the model maintains good generalization when estimating parenchymal volume, regardless of kidney side or dataset split.

In contrast, cortical volume estimates show more variability. While training performance is relatively stable, with mean volume difference is +1.81% and +2.79% for SIN and DX kidney and standard deviation of 2.61% and 5.14% for SIN and DX kidney, validation and test results show higher discrepancies and inconsistent bias. No obvious trends can be seen between left and right kidneys. This suggests that the model struggles to generalize cortical predictions beyond the training data.

4.3.3 Quantification of Median T1 Value

Median T1 values for the cortex and medulla were extracted using refined segmentation masks, where morphological erosion was applied to isolate central regions and reduce edge effects. For an optimal balance between ROI size and anatomical relevance, a 3×3 square structuring element was found suitable for eroding the cortex mask by sufficiently reduce uncertainties from outer boundaries while preserving its anatomical structure. As illustrated in Figure 4.11, the 3×3 erosion successfully reduced the cortex mask without significantly altering its overall shape. In contrast, larger structuring elements removed too much of the ROI, leading to loss of relevant tissue. For the medulla, erosion was found to negatively affect the accuracy of median T1 quantification. See Appendix B for complete results of the effect of various sizes of the structural element.

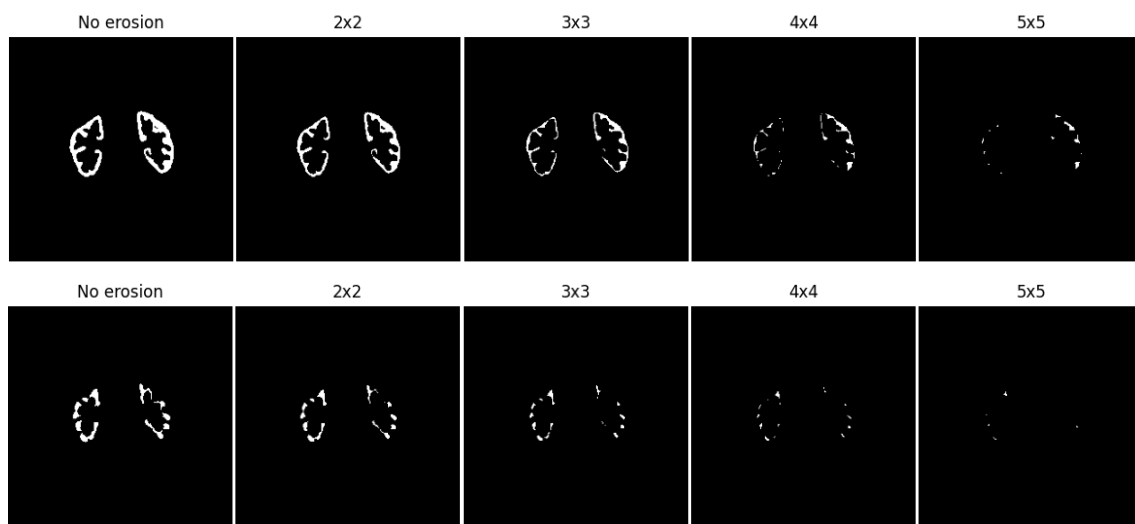


Figure 4.11: Example of post processing / erosion of segmentation mask for (a) cortex and (b) medulla. From left to right: original segmentation mask and resulting segmentation mask after erosion with a square structuring element of increasing size.

Results for quantification of median T1 values are presented in Figure 4.12, showing the correlation between the predicted and delivered reference values for both cortex and medulla. The model demonstrated strong performance in estimating cortical median T1 values for both SIN and DX kidneys. A slight overestimation tendency was observed for the DX kidney, but predictions generally remained within a 5% deviation from the actual values, indicating high predictive accuracy of the segmentation model. For the medulla, correlation was generally lower due to higher variability. Despite this, most predictions still fell within the 5% deviation range. One can observe that, while cortical predictions showed consistency across both kidneys, medullary T1 estimations were less accurate for the DX kidney with notably lower correlation values for all dataset splits. This suggests increased challenges in achieving high-precision quantification in smaller tissue regions like the medulla, which are more sensitive to segmentation errors.

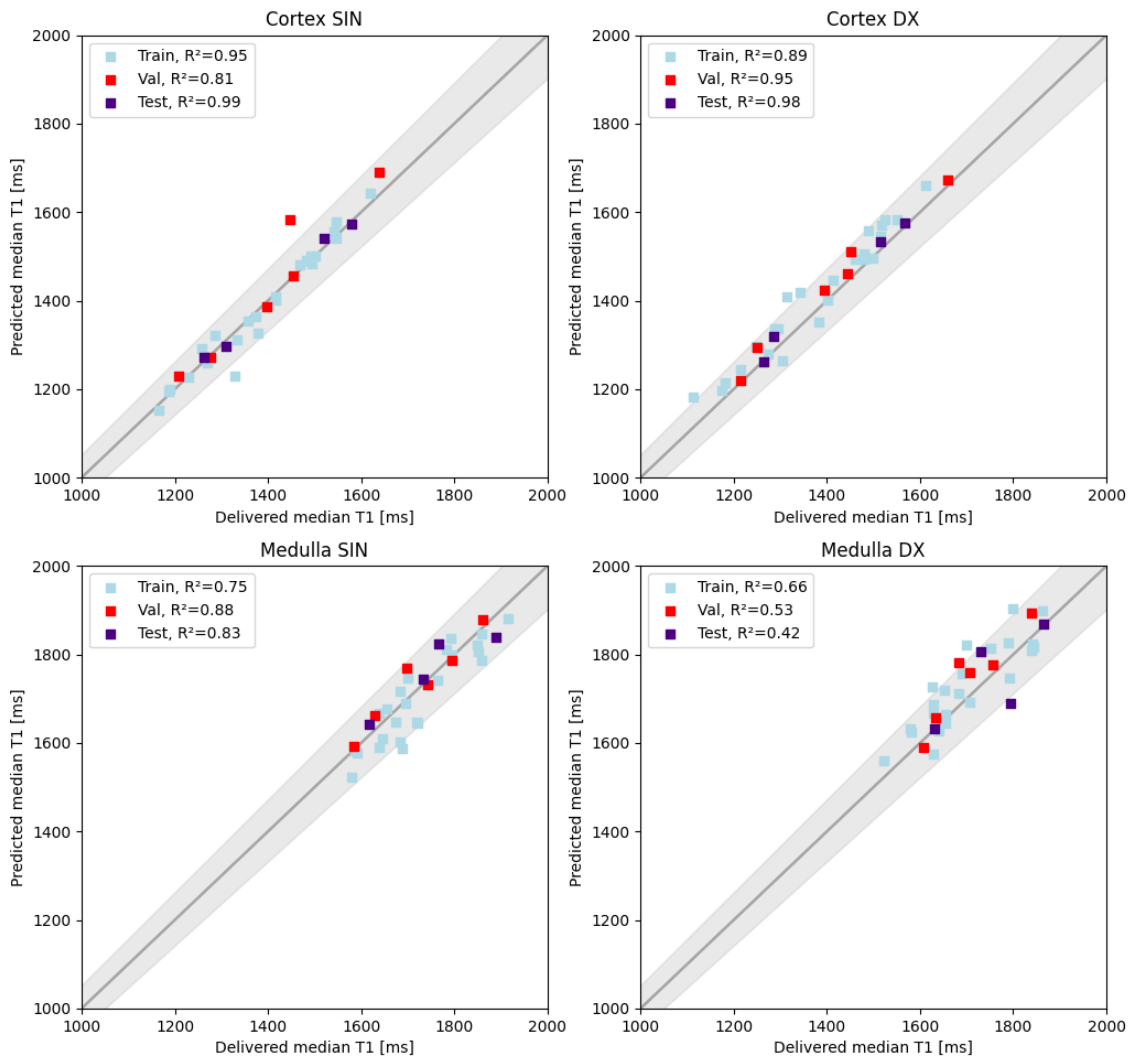


Figure 4.12: Correlation of predicted and delivered median T1 value for cortex and medulla. The identity line represents perfect correlation between the CNN-predicted and delivered values. Grey area corresponds to a 5% difference in predicted value.

4.4 Multi-Modal Kidney Segmentation

A multi-modal model was trained using input from DWI, T2* mapping sequences, and the associated parametric maps ADC, T2*, and T1-MOLLI. Similar to the multi-channel setup, the model received channel-wise input corresponding to multiple b-values from DWI and echo times from T2* mapping. The specific input channels used in the model are listed in Table 4.8.

Table 4.8: Overview of input channels used in the multi-modal segmentation model. DWI inputs correspond to multiple b-values and T2* mapping inputs correspond to multiple echo times.

Input	Image channels
b-values (s/mm ²)	0, 50, 200, 500
Echo times (ms)	3, 8, 14, 20, 25, 30
Parametric maps	ADC, T2*, T1

The performance of the multi-modal model incorporating image data from the DWI and T2* mapping sequences is presented in Table 4.9. Compared to the multi-channel model, the multi-modal model demonstrated lower performance in all cases. The reduced accuracy is due to misalignments between modalities, which negatively impacted the model’s ability to generalize across subjects.

Table 4.9: Comparison of the performance of multi-label segmentation models using multi-channel T1-MOLLI or DWI and T2* as input. Models are based on a 2D ResUNet structure, segmenting both cortex and parenchyma simultaneously.

Model	Val Dice	Val IoU	Test Dice	Test IoU
T1-MOLLI multi-channel				
Parenchyma	0.9298	0.8737	0.9089	0.8453
Cortex	0.8614	0.7598	0.8552	0.7557
Multi-modal (DWI/T2*)				
Parenchyma	0.8639	0.7859	0.8137	0.7353
Cortex	0.7518	0.6220	0.7139	0.5917

After reaching convergence, the multi-modal model achieved a Dice score of 0.8639 and IoU of 0.7859 on the validation set, and 0.8137 Dice and 0.7353 IoU on the test set for parenchyma segmentation. Cortex segmentation yielded 0.7518 Dice and 0.6220 IoU on the validation set, and 0.7139 Dice and 0.5917 IoU on the test set. These results indicate that the model’s ability to generalize to unseen data was moderate, with a consistent decline in performance from validation to test sets for both segmentation targets. Overall, the model underperformed relative to the earlier multi-channel T1-MOLLI-based approach in terms of evaluation metrics.

Figure 4.13-4.14 show example segmentation outcomes for parenchyma using the multi-modal segmentation model. While the model generally tends to capture the full renal parenchyma, obvious failure cases in the segmentation can be observed. The prediction overlays show that the model struggled with boundary accuracy, particularly in cases of poor alignment between the T1-MOLLI-based ground truth and the corresponding DWI or T2* slices. As a result, false detections can clearly be observed across the test cases, explaining the poor segmentation performance metrics. Even in cases where the ground truth labels are accurately aligned with the image, the model struggles more at the kidney boundaries.



Figure 4.13: Example of renal parenchyma segmentation results from multi-modal input, illustrating test cases with good image-label alignment. The ground truth labels closely match the kidney boundaries visible in the underlying T2*w image, resulting in fewer false detections. The color coding in the overlays is as follows: green - true positives, blue - false negatives, and red - false positives.

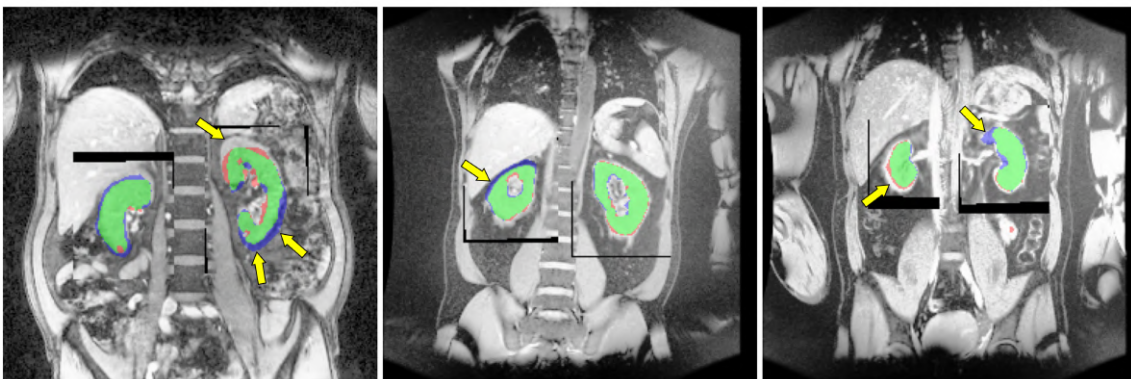


Figure 4.14: Example of renal parenchyma segmentation results from multi-modal input, illustrating test cases with poor image-label alignment. The segmentation overlays show how the ground truth label either extends beyond the visible kidney region or covers only a portion of the kidney as seen in the underlying T2*w image. The color coding in the overlays is as follows: green - true positives, blue - false negatives, and red - false positives.

Figure 4.15 show example cortical segmentation using the multi-modal network. Visual inspection of the overlays showed that, similar to the T1-MOLLI model, the multi-modal model struggled with accurate cortical segmentation, particularly at the cortico-medullary boundary. Numerous false positives were observed along the inner boundaries of the cortex in both central and outer slices, which aligns with the lower Dice and IoU scores for cortex prediction. As with the parenchyma segmentation, the cortex predictions were also affected by poor alignment between the modalities, resulting in spatial mismatches that further contributed to segmentation inaccuracies.

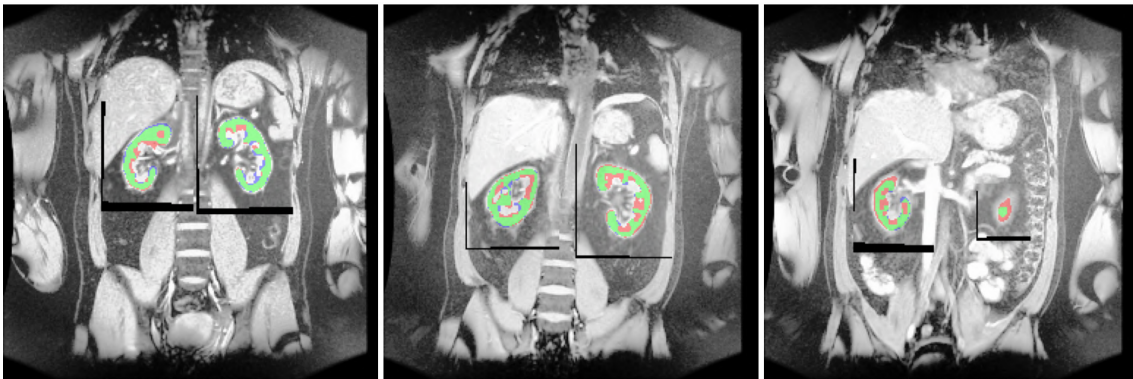


Figure 4.15: Examples of cortex segmentation results using the multi-modal segmentation model, demonstrating classification performance on test images. Images show an overlay of ground truth label and prediction output for several slices, superimposed to T2*w images acquired at TE = 3 ms. The color coding is as follows: green - true positives, blue - false negatives, and red - false positives.

4.4.1 Quantification of ADC and T2* value

Median ADC and T2* values for the cortex and medulla were quantified using segmentation-derived masks. Morphological erosion was applied to refine the ROIs and ensure that measurements were confined to the appropriate tissue regions, following the same approach used for T1 quantification.

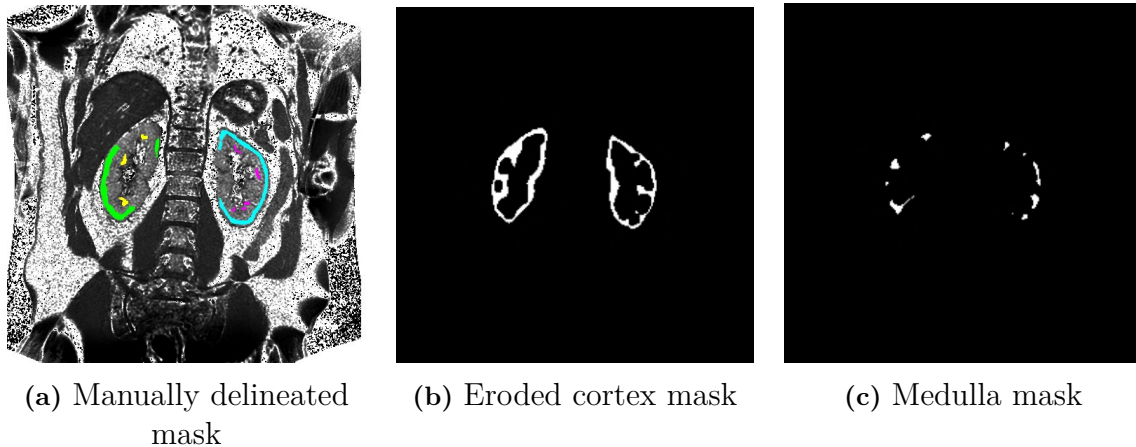


Figure 4.16: Example of manually delineated regions and generated regions: (a) Manually delineated regions for cortex and medulla on T2* map, (b) generated mask using erosion for cortex, (c) generated mask using erosion for medulla.

The manually delineated regions were created with high precision, ensuring that the selected pixels reliably included the intended anatomical structures. These regions typically reflect conservative selections made with high certainty. In contrast, the generated masks are derived from volumetric segmentations and tend to include a broader range of pixels, including areas with lower certainty near tissue boundaries, despite the application of erosion to minimize such effects.

Figure 4.17, indicate that there is no obvious correlation between predicted and reference ADC values across the training, validation, or test datasets, for both cortex and medulla. A substantial variability can be observed, with many data points falling outside a 5% difference margin.

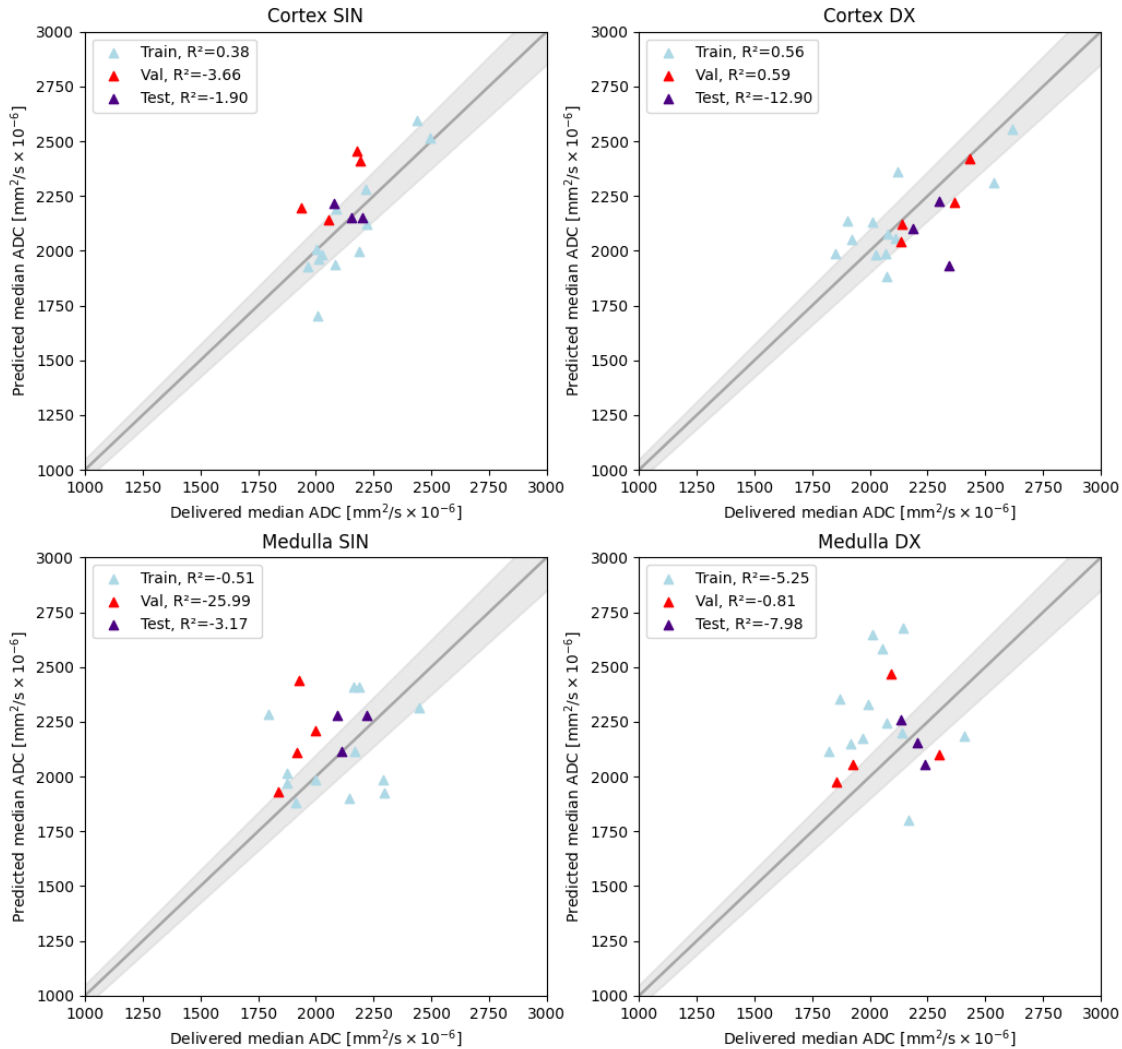


Figure 4.17: Correlation of predicted and delivered median ADC value for cortex and medulla. Model is a proposed multi-modal model (2D ResUNet). The identity line represents perfect correlation between the CNN-predicted and delivered values. Grey area corresponds to a 5% difference in predicted value.

Similar to the ADC median intensity quantification, the correlation plots for $T2^*$ values (Figure 4.18) show little to no correlation between predicted and reference values. However, in contrast to the ADC results, the $T2^*$ values are closer to the reference values. For the cortex, the majority of data points fall within the 5% difference margin, indicating better agreement between predicted and reference values despite the low correlation. In contrast, a clear trend of overestimation is observed for the medulla values, suggesting that the medulla masks are still not properly aligned with the underlying anatomy.

4. Results

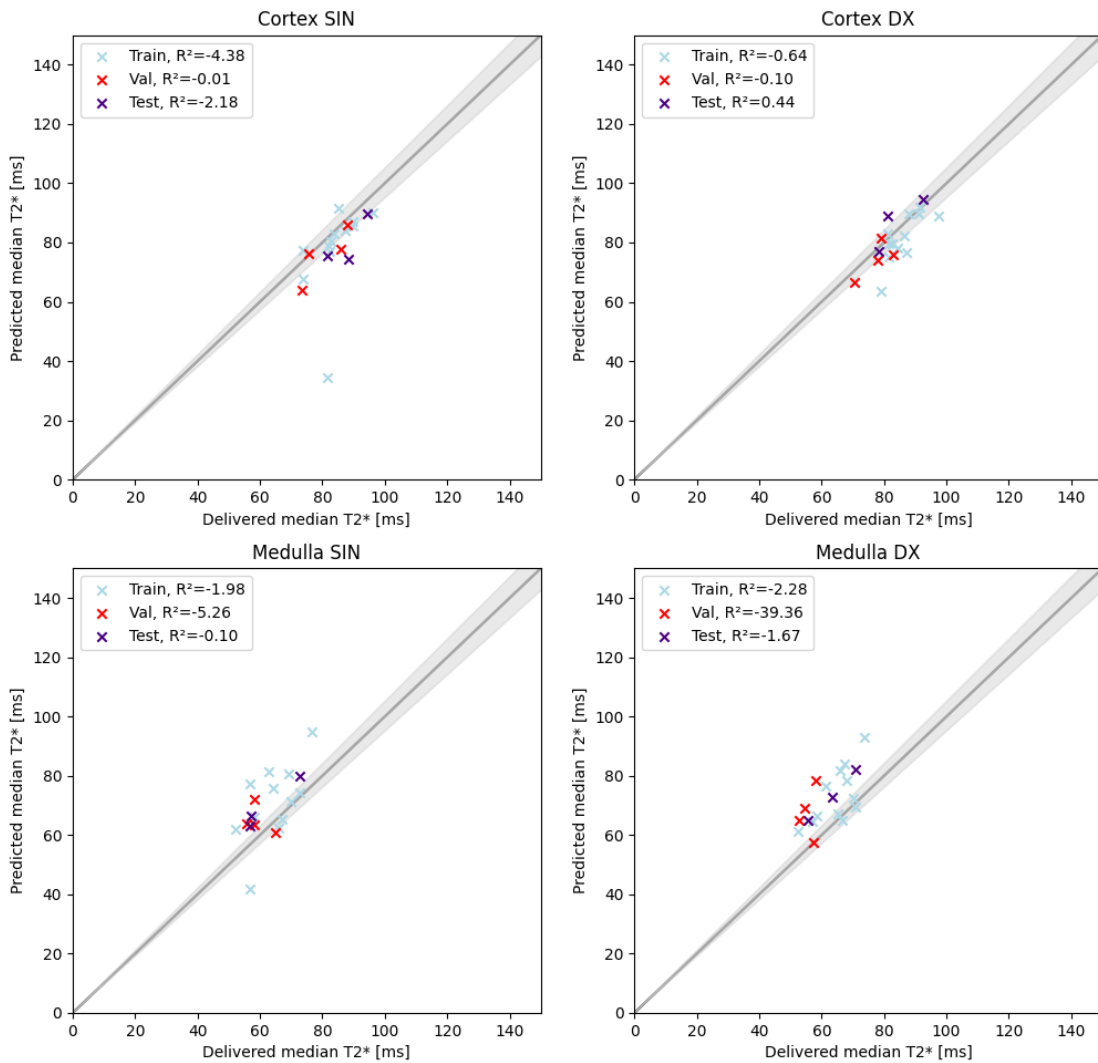


Figure 4.18: Correlation of predicted and delivered median T2star value for cortex and medulla. Model is a proposed multi-modal model (2D ResUNet). The identity line represents perfect correlation between the CNN-predicted and delivered values. Grey area corresponds to a 5% difference in predicted value.

Overall, the quantification of median intensity values using the multi-modal model shows limited agreement between predicted and reference values for both ADC and T2* maps. Although the T2* cortex values demonstrated better alignment with the delivered values, the overall correlation remained low. The observed deviations, particularly the consistent overestimation in the medulla, highlight persistent issues with segmentation accuracy and alignment that affect the reliability of intensity-based measurements.

5

Discussion

In this chapter, the results are evaluated with a focus on segmentation performance, model behavior, and potential clinical relevance. Key comparisons and limitations are also discussed.

5.1 Comparison of 2D and 3D Segmentation

The proposed 2D model demonstrates strong segmentation performance, accurately capturing fine anatomical structures like the cortex while also maintaining high segmentation accuracy for the parenchyma. In contrast, although the 3D model performs comparably well on parenchyma segmentation, it underperforms in delineating finer details such as the cortex. This is evident in the overlays, where the 2D model better captures the boundary between cortex and medulla. These observations are further supported by cross-validation results, where the 2D model consistently outperforms the 3D model. Although both models show similar standard deviation values across folds, indicating consistent performance across the dataset, the 2D approach proves more effective for this specific segmentation task.

In general, 3D models excel when there is a spatial relationship between slices in the volume, utilizing depth information to identify details between slices. However, in this study, true 3D spatial relationships are not present in the data, as image volumes are formed by stacking 2D slices acquired at different time points. This synthetic volume construction introduces challenges for a 3D network, as patient motion between acquisitions can disrupt spatial alignment. It is likely that the slice-wise acquisition does not provide any additional information, and that the 3D model is unlikely to benefit from the additional dimension, making the 2D model more effective in this case.

Furthermore, a 3D implementation introduces greater computational costs. Due to memory constraints, it often requires smaller input or batch sizes during training, which can negatively affect model convergence and generalization. This, combined with the limited number of image volumes available, likely impacted the 3D model's ability to learn meaningful features. In contrast, the 2D model takes advantage of a larger effective training set, as each 3D volume is decomposed into multiple slices at the input-stage. This not only increases the dataset size significantly, but also allows the model to focus on local contrast patterns within each slice. Inoue et al. [7] trained their 3D ResUnet on 100 image volumes and achieved a Dice score of 0.902 ± 0.034 for parenchyma segmentation. In contrast, the 3D model in this thesis

was constrained by a substantially smaller dataset of 37 image volumes, but would likely benefit a lot from a larger dataset. As a result, the 2D model becomes more effective in this setting, especially when volumetric data is limited.

5.2 The Effect of Multi-Channel Inputs

Following the decision to proceed with the 2D model due to its improved performance in capturing fine details such as CMD, the next step was to enhance its segmentation accuracy by integrating multi-channel inputs. By incorporating multiple input channels, each representing different contrast properties, the network could learn a more diverse and informative set of features. This proved especially useful for highlighting anatomical boundaries and tissue variations within the kidney. Different inversion times from the T1-MOLLI sequence produce varying contrast levels based on the magnetization recovery of tissues with different relaxation times. By presenting these images as separate input channels, the model was able to benefit from resulting contrast variations and use complementary information to improve cortico-medullary differentiation and boundary detection.

In the T1-MOLLI-based model, four TIs were selected based on their individual contributions to tissue contrast, detailed in 4.2 and Table 4.3. Lower TIs offered strong CMD contrast, which supported accurate segmentation of internal structures like the cortex and medulla, while higher TIs contributed to better boundary detection. When combined with the parametric T1 map, which provided high CMD contrast and yielded the best individual segmentation performance, the multi-channel setup resulted in robust segmentation for both the cortex and parenchyma.

It is important to note that increasing the number of input channels does not inherently lead to better performance and model accuracy. Rather, the effectiveness of multi-channel input depends on the contrast and complementary information of the selected input channels. Including TIs with poor tissue contrast or low image quality can introduce noise, reduce the model’s ability to distinguish meaningful features, and ultimately impact the overall segmentation performance. This is particularly important in regions with subtle intensity differences, such as the cortex and medulla, where reduced contrast may reflect underlying pathology, as low CMD is characteristic of CKD.

This highlights the importance of careful selection of input channels, which provide unique and meaningful information relevant to the segmentation task, rather than maximizing their number. In theory, this approach follows the basic idea behind how deep learning models learn features from multiple inputs. Increased input representations can improve generalization, but only when they add useful information without introducing unnecessary noise. Essentially, the effectiveness of a multi-channel approach depends more on the quality and relevance of the input channels than on their quantity.

5.3 The Effect of Multi-modal Integration

In the multi-modal integration approach, the concept of using multi-channel inputs was extended by incorporating different MRI modalities. Specifically, the model was designed to include DWI and T2* images with varying b-values and TEs. All available b-value images and TEs up to 30 ms were used, alongside the corresponding parametric maps derived from DWI, T2*, and T1-MOLLI sequences. While this approach is conceptually similar to the multi-channel strategy, integrating multiple modalities offers greater robustness and the potential for broader clinical applicability, as the model becomes capable of operating across different types of MRI data rather than being restricted to a single sequence.

However, one of the main challenges with the multi-modal approach was the absence of ground-truth segmentation labels for the DWI and T2* images. To overcome this, the segmentation masks generated from the T1-MOLLI sequence were used as reference labels. Although the datasets were registered to one another to the best extent possible, misalignment between the images and labels was still present, partly due to patient motion and sequence-specific acquisition differences. To reduce inconsistencies caused by inter-modality slice misalignment, a subset of visibly misaligned images was excluded from the training set. Additionally, slices from the MOLLI sequence were carefully matched to the corresponding DWI and T2* slices. While this approach helped minimize spatial mismatch, some residual misalignment remained. These differences were visually noticeable and likely contributed to the reduced performance of the multi-modal model, particularly in the absence of 3D information to support accurate inter-modality registration.

Despite these limitations, the model achieved reasonably good results for parenchyma, indicating the potential of a multi-modal segmentation approach. However, performance for the cortex was notably lower, likely due to limited CMD contrast in both the DWI and T2* images, as well as in the corresponding parametric maps. The insufficient tissue contrast made it challenging for the model to clearly distinguish the cortex, which subsequently affected the resulting delineation of the medulla. This, in turn, impacted the reliability of the quantification of parametric values, thereby limiting its potential to be used for clinical purposes.

5.4 Model Strengths and Limitations

While the segmentation models demonstrated strong overall performance, particularly in delineating the renal parenchyma and cortex, several recurring strengths and failure cases were observed across the dataset. A noticeable limitation was the undersegmentation in outer slices of the kidney volume. These regions, only containing small parts of the anterior or posterior surfaces of the kidney, proved to be challenging to interpret without spatial context. The 2D network, while outperforming its 3D counterpart in overall segmentation accuracy, lacks context from adjacent slices, which likely contributes to its difficulty in correctly identifying these outer regions. As a result, undersegmentation was frequently observed in these regions when compared to manual segmentations. This issue is likely caused by partial

volume effects, where boundaries become blurred due to mixing of tissues within a voxel, as well as other acquisition-related factors that make the boundaries of the ROI less well-defined.

In contrast, the models tends to oversegment in the renal pelvis area, potentially mistaking nearby tissues or vessels for parenchymal tissue. This misclassification suggests that while the models are sensitive to intensity patterns associated with kidney tissue, they sometimes lacks the contextual understanding needed to correctly differentiate between closely located anatomical structures. False positives were also observed in slices where no kidney tissue is present, typically at the edges of the scanned volume. In such challenging scenarios, manual image analysts benefit from reviewing adjacent slices or additional imaging sequences to guide accurate segmentation.

The segmentation of cortical regions presents additional challenges for the models. Particularly, they struggle with accuracy in inner regions surrounding the medulla and renal columns, where both false positives and false negatives are common due to limited contrast and the absence of well-defined anatomical boundaries. These limitations are seen in both single-modality and multi-modality approaches. Despite this, the models performs well in segmenting the outer cortex regions. The multi-label approach appears to contribute positively in this regard, as the outer boundary of the cortex is shared with the parenchyma. Training the models to segment both cortex and parenchyma simultaneously may encourage the learning of shared spatial and textural features, thereby improving boundary detection of both regions.

Nevertheless, the proposed multi-channel T1-MOLLI-based model maintained a high level of consistency across both easier and challenging cases. Even when false predictions occurred, they generally followed the same patterns, which suggests that the model could still be helpful by offering reliable starting points for segmenting images. Such suggestions could save time during manual workflows by allowing image analysts to focus on refining only the most complex or ambiguous regions. It is also important to recognize that the outer boundaries of the kidney are difficult to define, even for experienced analysts, and are often areas with high inter- and intra-observer variability due to differences in how various kidney regions are interpreted and outlined by analysts. A deep learning-based network, however, has the advantage of learning features from multiple anatomical aspects simultaneously, which could provide more consistent segmentation results and help reduce reliance on individual operator interpretations.

However, an important limitation of the proposed segmentation approach is its reliance on input channels with contrast characteristics similar to those seen during training. Results indicate that segmentation performance is highly dependent on the combination of input-channels provided, which may limit the network's generalizability to clinical applications where imaging protocols vary. Consequently, there is a risk of performance degradation when the input data differs significantly from the training set.

5.5 Quantifying Volume and Tissue Parameters

The results demonstrate that the proposed segmentation approach is capable of supporting the quantification of renal tissue parameters and volume, although with some notable limitations. The multi-channel model trained on T1-MOLLI images achieved a high correlation with ground truth volumes and showed relatively small differences in measurement. However, it tended to overestimate both parenchyma and cortex volumes. This overestimation is likely due to false positive predictions in the outer regions, where surrounding tissues were occasionally misclassified as kidney. Although these errors reduce quantitative accuracy, they are often visually apparent and can typically be identified and corrected by experienced image analysts.

The model also demonstrated promising capability in quantifying tissue parameters such as T1 values, particularly when segmentation accuracy was high. This is expected, as better segmentation enables more precise delineation of small and complex anatomical regions, thereby improving the accuracy of tissue parameter estimation. However, tissue values in parametric maps are highly sensitive to misalignments between the segmentation mask and the underlying image. Even minor registration errors can result in significant variations in the extracted values, which shows the importance of precise image alignment.

In clinical settings, analysts often limit segmentation to regions where tissue boundaries are clearly defined, avoiding areas of uncertainty. This practice differs from the model’s approach, which attempts to segment the entire structure. As a result of this, automatically generated segmentations may benefit from manual adjustment, particularly in regions where contrast is low or anatomical boundaries are unclear.

Despite these limitations, the model demonstrated consistent behavior and potential to support renal volume and tissue parameter estimation. When segmentation quality is high and the input images are well-aligned, the method can produce reliable and reproducible measurements, making it a potentially valuable tool in both clinical and research applications.

5.6 Comparison to Existing Works

While direct comparison to related studies is limited by differences in imaging modalities, segmentation strategies, and dataset characteristics, the proposed 2D ResUNet demonstrates performance that is in line with existing works, summarized in Table 5.1. On T1-MOLLI images, the proposed method achieves a Dice score for parenchyma segmentation (0.9089) and cortex (0.8552). These results are comparable to those achieved in other studies, even with the limited size and quality of the used dataset. It is also worth noting that many of the existing approaches rely on contrast-enhanced or higher-resolution input data, whereas this model demonstrates promising performance using standard clinical MRI sequences. This highlights the potential of a multi-channel approach for renal segmentation.

Table 5.1: Summary of related work on kidney segmentation.

Author	Approach	Results
Daniel et al. [8]	Used T2-weighted MRI from 60 subjects (30 healthy, 30 with CKD) and a 2D U-Net to segment left and right parenchyma for total kidney volume estimation.	Healthy: 0.94 (Dice), 0.88 (IoU) CKD: 0.92 (Dice), 0.86 (IoU) Vol diff %: 1.2 ± 16.2 mL
Bin Islam et al. [6]	ResNet18-Self-ONN-UNet++ architecture, a hybrid of ResUNet and U-Net++, used to segment whole kidney (parenchyma) from 100 T2-weighted MRI scans.	Dice: 0.9157 IoU: 0.8234
Klepaczko et al. [46]	35 CE-MRI exams used for a 3-branch CNN with a shared encoder for joint kidney classification and 2D segmentation of parenchyma and cortex.	Parenchyma IoU: 0.94 Cortex IoU: 0.76–0.92 Volume R ² : 0.92 (total) 0.90 (cortex)
Inoue et al. [7]	T1-weighted MRI from 170 subjects (healthy and non-healthy). 3D ResUNet to segment the kidney parenchyma, exploring both unilateral and bilateral segmentation strategies.	Dice of 0.902 ± 0.034 achieved using bilateral segmentation.
Korfiatis et al. [5]	Using 1930 contrast-enhanced CT exams, a modified 3D U-Net was developed to segment four kidney substructures: right cortex, left cortex, right medulla, and left medulla.	Dice: 0.94 (cortex) 0.90 (medulla)

5.7 Future Work

To ensure that the developed model can be effectively applied in real-world scenarios, future work should focus on improving its robustness and generalizability. One of the primary limitations of this study was the limited dataset size, which restricted the model’s learning capacity and made it difficult to draw concrete conclusions. Expanding the dataset with additional cases, including high-quality ground-truth segmentations, would allow for more comprehensive evaluation across both validation and test sets. With access to data from 3D acquisitions, further investigation of the 3D model would be feasible. Furthermore, incorporating an external test set would provide a more thorough assessment of the model’s ability to generalize beyond the training environment. This is particularly important for ensuring clinical applicability and robustness under varying acquisition conditions.

To improve consistency and quality of ground-truth data, additional manual quality control and exclusion of poorly aligned cases should be implemented. This will help ensure that only high-quality, well-aligned samples are used for training, which is critical for robust model performance. For multi-modal model development, future

work should focus on improving label-image alignment across modalities. This could be achieved either by acquiring a 3D reference acquisition to enable accurate inter-modality registration or by creating segmentation masks specifically for the DWI and T2* images, rather than relying on those derived from T1-MOLLI. Ensuring proper alignment between input images and corresponding labels is essential for accurate model performance and reliable predictions in multi-modal applications.

Moreover, further exploration of model architectures could provide performance improvements. While this study investigated CNNs like ResUNet and U-Net, alternative architectures could be explored to possibly further improve the outcome. Future work should also explore alternative fusion strategies within the network architecture to assess their effectiveness for kidney segmentation using multi-modal MRI. In this study, only early fusion was implemented by concatenating the input images across channels. Investigating intermediate or late fusion approaches could provide further insight into how different modalities contribute to segmentation performance. Additionally, an extensive evaluation of hyperparameters, such as learning rate, batch size, optimizer selection, number of filters, dropout rates, and data augmentation strategies could help identify a more optimal configuration. Automated techniques such as grid search or Bayesian optimization could be utilized to further improve model performance by efficiently fine-tuning multiple settings simultaneously.

6

Conclusion

In this thesis, a deep learning-based approach for automatic segmentation of kidneys regions in MRI were developed and evaluated, focusing on multi-channel and multi-modal inputs. The proposed 2D ResUNet demonstrated strong performance for segmenting the renal parenchyma and potential for cortex segmentation, particularly from T1-MOLLI images offering high corticomedullary contrast. Multi-channel inputs helped the model distinguish renal substructures, especially when channels offered complementary contrast, highlighting the benefit of using diverse channels for richer feature extraction. However, the model struggled with complex cases such as outer slices and low-contrast regions, where contextual information is needed. This suggest directions for future work focused on improving the model’s ability to capture and leverage spatial context.

A key challenge in the multi-modal setup was using T1-MOLLI-derived labels for DWI and T2*, which introduced misalignments between images and corresponding labels. These findings emphasize the importance of accurate image alignment and consistent labeling to enable reliable multi-modal segmentation. To properly assess the value of multi-modal input for renal segmentation, evaluation of the performance on properly aligned data is necessary.

Although manual corrections are still necessary, the model shows promise in providing reliable segmentation suggestions that can support manual analysis and reduce operator variability. It also showed potential to quantify renal volume and tissue-specific parametric values. However, the reliability of such quantifications remains highly dependent on segmentation quality and image alignment.

In conclusion, this thesis presents a deep learning framework for automatic kidney MRI segmentation and analysis. The findings highlight the potential of using deep learning-assisted tools for automating renal imaging workflows, which could help the diagnosis and monitoring of kidney disease. However, the limited dataset posed a key challenge, impacting models’ robustness and generalizability across scanners and imaging protocols. Future work should focus on expanding the dataset, and validating the model on external data to ensure clinical applicability.

Bibliography

- [1] A. Whaley-Connell, R. Nistala, and K. Chaudhary, “The importance of early identification of chronic kidney disease.,” *Missouri medicine*, vol. 108, no. 1, pp. 25–8, 2011, ISSN: 0026-6620.
- [2] O. Bane, E. Seeliger, E. Cox, *et al.*, “Renal MRI: From Nephron to NMR Signal,” *Journal of Magnetic Resonance Imaging*, vol. 58, no. 6, pp. 1660–1679, Dec. 2023, ISSN: 1053-1807. DOI: 10.1002/jmri.28828.
- [3] D. Zhao, W. Wang, T. Tang, Y.-Y. Zhang, and C. Yu, “Current progress in artificial intelligence-assisted medical image analysis for chronic kidney disease: A literature review,” *Computational and Structural Biotechnology Journal*, vol. 21, pp. 3315–3326, 2023, ISSN: 20010370. DOI: 10.1016/j.csbj.2023.05.029.
- [4] S. Will, P. Martirosian, C. Würslin, and F. Schick, “Automated segmentation and volumetric analysis of renal cortex, medulla, and pelvis based on non-contrast-enhanced T1- and T2-weighted MR images,” *Magnetic Resonance Materials in Physics, Biology and Medicine*, vol. 27, no. 5, pp. 445–454, Oct. 2014, ISSN: 0968-5243. DOI: 10.1007/s10334-014-0429-4.
- [5] P. Korfiatis, A. Denic, M. E. Edwards, *et al.*, “Automated Segmentation of Kidney Cortex and Medulla in CT Images: A Multisite Evaluation Study,” *Journal of the American Society of Nephrology*, vol. 33, no. 2, pp. 420–430, Feb. 2022, ISSN: 1046-6673. DOI: 10.1681/ASN.2021030404.
- [6] M. S. B. Islam, M. S. I. Sumon, R. Sarmun, E. H. Bhuiyan, and M. E. Chowdhury, “Classification and segmentation of kidney MRI images for chronic kidney disease detection,” *Computers and Electrical Engineering*, vol. 119, p. 109613, Nov. 2024, ISSN: 00457906. DOI: 10.1016/j.compeleceng.2024.109613.
- [7] K. Inoue, Y. Hara, K. Nagawa, *et al.*, “The utility of automatic segmentation of kidney MRI in chronic kidney disease using a 3D convolutional neural network,” *Scientific Reports*, vol. 13, no. 1, p. 17361, Oct. 2023, ISSN: 2045-2322. DOI: 10.1038/s41598-023-44539-z.
- [8] A. J. Daniel, C. E. Buchanan, T. Allcock, *et al.*, “Automated renal segmentation in healthy and chronic kidney disease subjects using a convolutional neural network,” *Magnetic Resonance in Medicine*, vol. 86, no. 2, pp. 1125–1136, Aug. 2021, ISSN: 0740-3194. DOI: 10.1002/mrm.28768.
- [9] S. Srinivasan, K. Durairaju, K. Deeba, S. K. Mathivanan, P. Karthikeyan, and M. A. Shah, “Multimodal Biomedical Image Segmentation using Multi-Dimensional U-Convolutional Neural Network,” *BMC Medical Imaging*, vol. 24, no. 1, p. 38, Feb. 2024, ISSN: 1471-2342. DOI: 10.1186/s12880-024-01197-5.

- [10] T. Zhou, S. Ruan, and S. Canu, “A review: Deep learning for medical image segmentation using multi-modality fusion,” *Array*, vol. 3-4, p. 100 004, Sep. 2019, ISSN: 25900056. DOI: 10.1016/j.array.2019.100004.
- [11] Q. He, N. Summerfield, M. Dong, and C. Glide-Hurst, “Modality-Agnostic Learning for Medical Image Segmentation Using Multi-Modality Self-Distillation,” in *2024 IEEE International Symposium on Biomedical Imaging (ISBI)*, IEEE, May 2024, pp. 1–5, ISBN: 979-8-3503-1333-8. DOI: 10.1109/ISBI56570.2024.10635881.
- [12] S. Ali, R. Khurram, K. u. Rehman, *et al.*, “An improved 3D U-Net-based deep learning system for brain tumor segmentation using multi-modal MRI,” *Multimedia Tools and Applications*, vol. 83, no. 37, pp. 85 027–85 046, May 2024, ISSN: 1573-7721. DOI: 10.1007/s11042-024-19406-2.
- [13] G. Chetty, M. Yamin, and M. White, “A low resource 3D U-Net based deep learning model for medical image analysis,” *International Journal of Information Technology*, vol. 14, no. 1, pp. 95–103, Feb. 2022, ISSN: 2511-2104. DOI: 10.1007/s41870-021-00850-4.
- [14] T.-Y. Su and Y.-H. Fang, “Automatic Liver and Spleen Segmentation with CT Images Using Multi-channel U-net Deep Learning Approach,” in 2020, pp. 33–41. DOI: 10.1007/978-3-030-30636-6{_}5.
- [15] H. G. Preuss, “Basics of Renal Anatomy and Physiology,” *Clinics in Laboratory Medicine*, vol. 13, no. 1, pp. 1–11, Mar. 1993, ISSN: 02722712. DOI: 10.1016/S0272-2712(18)30456-6.
- [16] D. Keshwani, Y. Kitamura, and Y. Li, “Computation of Total Kidney Volume from CT images in Autosomal Dominant Polycystic Kidney Disease using Multi-Task 3D Convolutional Neural Networks,” Sep. 2018.
- [17] I. A. Dekkers, A. de Boer, K. Sharma, *et al.*, “Consensus-based technical recommendations for clinical translation of renal T1 and T2 mapping MRI,” *Magnetic Resonance Materials in Physics, Biology and Medicine*, vol. 33, no. 1, pp. 163–176, Feb. 2020, ISSN: 0968-5243. DOI: 10.1007/s10334-019-00797-5.
- [18] F. G. Zöllner, E. Svarstad, A. Z. Munthe-Kaas, L. R. Schad, A. Lundervold, and J. Rørvik, “Assessment of Kidney Volumes From MRI: Acquisition and Segmentation Techniques,” *American Journal of Roentgenology*, vol. 199, no. 5, pp. 1060–1069, Nov. 2012, ISSN: 0361-803X. DOI: 10.2214/AJR.12.8657.
- [19] D. Nishimura, “Principles of Magnetic Resonance Imaging,” in 1990, ch. 5, pp. 58–59, ISBN: 5800118106336.
- [20] K. Chow and R. Thompson, “Myocardial T1 Mapping – Comparison of Techniques,” *Clinical Cardiology*, Jan. 2015. [Online]. Available: www.siemens.com/magnetom-world.
- [21] A. J. Taylor, M. Salerno, R. Dharmakumar, and M. Jerosch-Herold, “T1 Mapping,” *JACC: Cardiovascular Imaging*, vol. 9, no. 1, pp. 67–81, Jan. 2016, ISSN: 1936878X. DOI: 10.1016/j.jcmg.2015.11.005.
- [22] G. B. Chavhan, P. S. Babyn, B. Thomas, M. M. Shroff, and E. M. Haacke, “Principles, Techniques, and Applications of T2*-based MR Imaging and Its Special Applications,” *RadioGraphics*, vol. 29, no. 5, pp. 1433–1449, Sep. 2009, ISSN: 0271-5333. DOI: 10.1148/rg.295095034.

-
- [23] C.-C. Topriceanu, I. Pierce, J. C. Moon, and G. Captur, “T2 and T2* mapping and weighted imaging in cardiac MRI,” *Magnetic Resonance Imaging*, vol. 93, pp. 15–32, Nov. 2022, ISSN: 0730725X. DOI: 10.1016/j.mri.2022.07.012.
- [24] M. Pruijm, B. Milani, and M. Burnier, “Blood Oxygenation Level-Dependent MRI to Assess Renal Oxygenation in Renal Diseases: Progresses and Challenges,” *Frontiers in Physiology*, vol. 7, Jan. 2017, ISSN: 1664-042X. DOI: 10.3389/fphys.2016.00667.
- [25] V. Baliyan, C. J. Das, R. Sharma, and A. K. Gupta, “Diffusion weighted imaging: Technique and applications.,” *World journal of radiology*, vol. 8, no. 9, pp. 785–798, Sep. 2016, ISSN: 1949-8470. DOI: 10.4329/wjr.v8.i9.785.
- [26] A. Caroli, M. Schneider, I. Friedli, *et al.*, “Diffusion-weighted magnetic resonance imaging to assess diffuse renal pathology: a systematic review and statement paper,” *Nephrology Dialysis Transplantation*, vol. 33, no. suppl_2, pp. ii29–ii40, Sep. 2018, ISSN: 0931-0509. DOI: 10.1093/ndt/gfy163.
- [27] M. Franke, B. Baeßler, J. Vechtler, *et al.*, “Magnetic resonance T2 mapping and diffusion-weighted imaging for early detection of cystogenesis and response to therapy in a mouse model of polycystic kidney disease,” *Kidney International*, vol. 92, no. 6, pp. 1544–1554, Dec. 2017, ISSN: 00852538. DOI: 10.1016/j.kint.2017.05.024.
- [28] A. Stadler, W. Schima, A. Ba-Ssalamah, J. Kettenbach, and E. Eisenhuber, “Artifacts in body MR imaging: their appearance and how to eliminate them,” *European Radiology*, vol. 17, no. 5, pp. 1242–1255, Apr. 2007, ISSN: 0938-7994. DOI: 10.1007/s00330-006-0470-4. [Online]. Available: https://www.mriquestions.com/uploads/3/4/5/7/34572113/stadler_eur_rad_2007_body_artifacts_review.pdf.
- [29] H.-P. Chan, R. K. Samala, L. M. Hadjiiski, and C. Zhou, “Deep Learning in Medical Image Analysis,” in 2020, pp. 3–21. DOI: 10.1007/978-3-030-33128-3{_}1.
- [30] H. Liu, D. Hu, H. Li, and I. Oguz, “Medical Image Segmentation Using Deep Learning,” in 2023, pp. 391–434. DOI: 10.1007/978-1-0716-3195-9{_}13.
- [31] M. Thoma, “A Survey of Semantic Segmentation,” Feb. 2016.
- [32] I. Qureshi, J. Yan, Q. Abbas, *et al.*, “Medical image segmentation using deep semantic-based methods: A review of techniques, applications and emerging trends,” *Information Fusion*, vol. 90, pp. 316–352, Feb. 2023, ISSN: 15662535. DOI: 10.1016/j.inffus.2022.09.031.
- [33] Y. LeCun, Y. Bengio, and G. Hinton, “Deep learning,” *Nature*, vol. 521, no. 7553, pp. 436–444, May 2015, ISSN: 0028-0836. DOI: 10.1038/nature14539.
- [34] D. P. Kingma and J. Ba, “Adam: A Method for Stochastic Optimization,” Dec. 2014.
- [35] R. Yamashita, M. Nishio, R. K. G. Do, and K. Togashi, “Convolutional neural networks: an overview and application in radiology,” *Insights into Imaging*, vol. 9, no. 4, pp. 611–629, Aug. 2018, ISSN: 1869-4101. DOI: 10.1007/s13244-018-0639-9.
- [36] G. E. Hinton, N. Srivastava, A. Krizhevsky, I. Sutskever, and R. R. Salakhutdinov, “Improving neural networks by preventing co-adaptation of feature detectors,” Jul. 2012.

- [37] K. Sanjar, O. Bekhzod, J. Kim, J. Kim, A. Paul, and J. Kim, “Improved U-Net: Fully Convolutional Network Model for Skin-Lesion Segmentation,” *Applied Sciences*, vol. 10, no. 10, p. 3658, May 2020, ISSN: 2076-3417. DOI: 10.3390/app10103658.
- [38] O. Ronneberger, P. Fischer, and T. Brox, “U-Net: Convolutional Networks for Biomedical Image Segmentation,” May 2015.
- [39] R. Azad, E. K. Aghdam, A. Rauland, *et al.*, “Medical Image Segmentation Review: The success of U-Net,” Nov. 2022.
- [40] F. I. Diakogiannis, F. Waldner, P. Caccetta, and C. Wu, “ResUNet-a: A deep learning framework for semantic segmentation of remotely sensed data,” *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 162, pp. 94–114, Apr. 2020, ISSN: 09242716. DOI: 10.1016/j.isprsjprs.2020.01.013.
- [41] K. He, X. Zhang, S. Ren, and J. Sun, “Deep Residual Learning for Image Recognition,” Dec. 2015.
- [42] J. Ma, J. Chen, M. Ng, *et al.*, “Loss odyssey in medical image segmentation,” *Medical Image Analysis*, vol. 71, p. 102035, Jul. 2021, ISSN: 13618415. DOI: 10.1016/j.media.2021.102035.
- [43] F. Milletari, N. Navab, and S.-A. Ahmadi, “V-Net: Fully Convolutional Neural Networks for Volumetric Medical Image Segmentation,” Jun. 2016.
- [44] M. J. Cardoso, W. Li, R. Brown, *et al.*, “MONAI: An open-source framework for deep learning in healthcare,” Nov. 2022.
- [45] The MONAI Consortium, *Project MONAI - Transforms*, 2020. [Online]. Available: <https://docs.monai.io/en/1.3.1/transforms.html#monai.transforms.RemoveSmallObjects>.
- [46] A. Klepaczko, M. Majos, L. Stefańczyk, E. Eikefjord, and A. Lundervold, “Whole kidney and renal cortex segmentation in contrast-enhanced MRI using a joint classification and segmentation convolutional neural network,” *Biocybernetics and Biomedical Engineering*, vol. 42, no. 1, pp. 295–311, Jan. 2022, ISSN: 02085216. DOI: 10.1016/j.bbe.2022.02.002.

A

Appendix A

The Appendix contains representative examples of the medical image data used in this thesis. This includes MRI images from the T1-MOLLI, DWI and T2* mapping, acquired at varying acquisition parameters, as well as parametric maps from each of the sequences. Examples are provided to give a visual understanding of the characteristics and variability in contrast across each MRI sequence.

Figures A.1–A.3 illustrate example images from one patient acquired from multiple MRI sequences. Additionally, parametric maps derived from each imaging sequence are presented in Figure A.4.

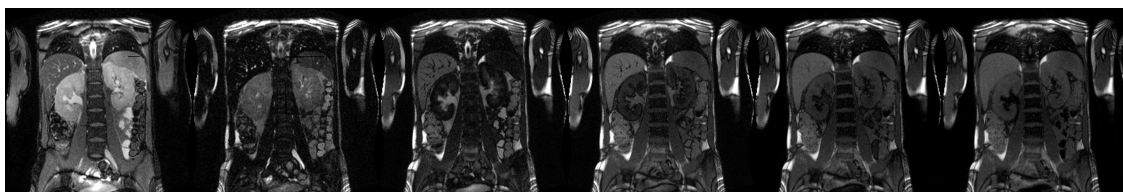


Figure A.1: Example images for T1-MOLLI images acquired with increasing inversion times, ranging from 174 ms to 2574 ms.

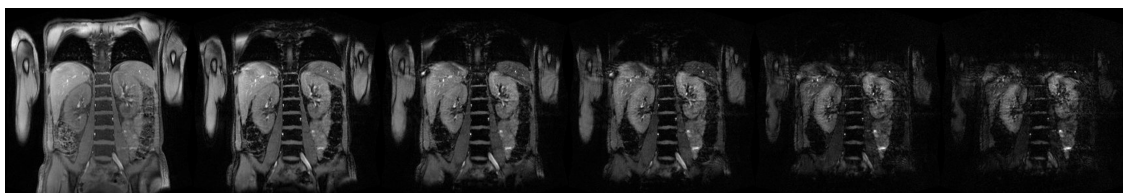


Figure A.2: Example images for T2*w images acquired with increasing echo times, ranging from 3 ms to 62 ms.

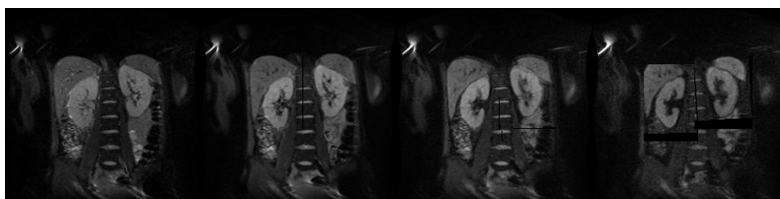


Figure A.3: Example images for DWI acquired with increasing b-values, ranging from 0 to 500 s/mm². Left to right: b=0, b=50, b=200, b=500.

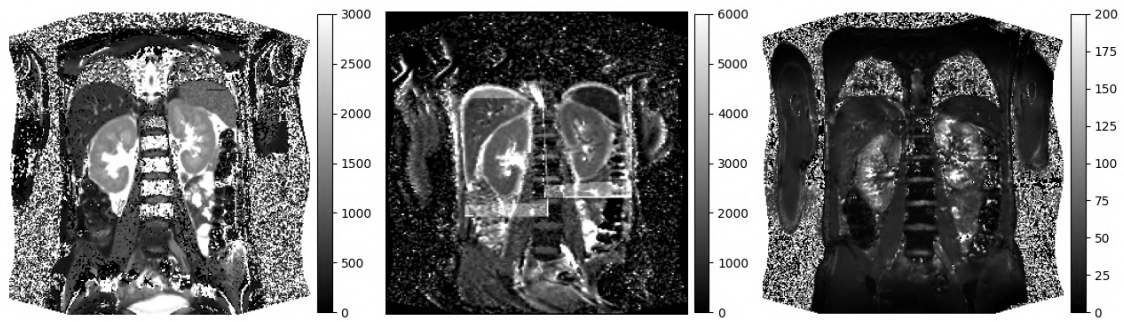


Figure A.4: Examples of parametric maps derived from different MRI sequences. From left to right: T1 map from T1-MOLLI, ADC map from DWI images, and T2* map from T2* mapping.

B

Appendix B

The Appendix contains tables showing R^2 values between predicted and actual T1 values for cortex and medulla regions, based on erosion using square structural elements of varying sizes. Results are reported for left (SIN) and right (DX) kidneys across the training, validation, and test sets.

Table B.1: R^2 correlation values between predicted and actual T1 values for SIN and DX cortex, using erosion with a square structural element of various sizes.

Size	Train		Validation		Test	
	SIN	DX	SIN	DX	SIN	DX
1×1	0.87	0.61	0.64	0.73	0.88	0.86
2×2	0.94	0.80	0.72	0.85	0.95	0.94
3×3	0.96	0.89	0.79	0.93	0.98	0.98
4×4	0.93	0.87	0.78	0.89	0.98	0.96

Table B.2: R^2 correlation values between predicted and actual T1 values for SIN and DX medulla, using erosion with a square structural element of various sizes.

Size	Train		Validation		Test	
	SIN	DX	SIN	DX	SIN	DX
1×1	0.77	0.56	0.75	0.53	0.80	0.35
2×2	0.78	0.48	0.68	0.32	0.77	0.36
3×3	0.66	-0.20	0.49	-0.39	0.49	0.04
4×4	0.32	-0.63	0.45	-0.54	0.52	-0.21

DEPARTMENT OF ELECTRICAL ENGINEERING (E2)
CHALMERS UNIVERSITY OF TECHNOLOGY
Gothenburg, Sweden
www.chalmers.se



CHALMERS
UNIVERSITY OF TECHNOLOGY