Title
Improving the Quantitative Interpretation of Multi-parametric MRI (mp-MRI) in Prostate Cancer

Author
Zhong, Xinran

Publication Date
2019

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA

Los Angeles

Improving the Quantitative Interpretation of Multi-parametric MRI (mp-MRI) in Prostate Cancer

A dissertation submitted in partial satisfaction of the

requirements for the degree Doctor of Philosophy

in Biomedical Physics

by

Xinran Zhong

2019

ABSTRACT OF THE DISSERTATION

Advancing Prostate Cancer Characterization via Improving the Quantitative Interpretation of Multi-parametric MRI (mp-MRI)

by

Xinran Zhong

Doctor of Philosophy in Biomedical Physics

University of California, Los Angeles, 2019

Professor Kyunghyun Sung, Chair

Purpose

Prostate cancer (PCa) is the second leading cause of cancer-related death in men in the United States. The accurate diagnosis of PCa is crucial for proper treatment decision. Although biopsy is still the gold standard for diagnosis, it is limited to low sensitivity and invasiveness. On the other hand, as a non-invasive imaging tool, multi-parametric MRI (mp-MRI) has excellent potential in PCa diagnosis such as detection and stratification of aggressiveness. The mp-MRI includes both anatomical and functional information to be able to provide a comprehensive characterization of the tissue. However, diagnosis with mp-MRI is limited to inconsistent and

qualitative interpretation. Clinically, the evaluation of mp-MRI is often through a standardized scoring system, PIRADS v2, which can lead to high inter- and intra-observer variability, and with a large amount of data for each case, the diagnosis process can be time-consuming. In order to get a more consistent quantitative evaluation, there are mainly two ways to utilize algorithms to help the diagnosis. The first one is creating quantitative biomarkers through mathematical models proposed based on assumption and understanding of physics and physiology, such as pharmacokinetic models for quantitative dynamic contrast-enhanced (DCE) MRI. The second one is using a machine learning technique to train a system with existing data to get diagnosis prediction on new data. The purpose of this work is to improve the quantitative interpretation of mp-MRI in PCa diagnosis regarding consistency and accuracy.

Methods

To evaluate existing B_1^+ estimation techniques to achieve a more consistent pre-contrast T_1 estimation for quantitative DCE- MRI, 21 volunteers were prospectively recruited and scanned twice on two 3T MRI scanners, resulting in 84 variable flip angle (VFA) T_1 exams. Two B_1^+ mapping techniques, including reference region variable flip angle (RR-VFA) and saturated turbo FLASH (satTFL), were used for B_1^+ correction, and T_1 maps with and without B_1^+ correction were tested for the intra-scanner repeatability and inter-scanner reproducibility. Volumetric regions of interest were drawn on the transition zone, peripheral zone of the prostate and the obturator internus left and right muscles in the corresponding slices. The average T_1 within each ROI for each scan was compared for both intra- and inter-scanner variability using concordance correlation coefficient and Bland-Altman plot.

To simplify B_1^+ compensation for quantitative DCE MRI in clinical and clinical research settings, an analytical B_1^+ correction method is proposed using a Taylor series approximation to

the steady-state spoiled gradient echo signal equation. The proposed approach only requires B_1^+ maps and uncorrected pharmacokinetic (PK) parameters as the input, and was evaluated using a prostate digital reference object (DRO) and 82 in-vivo prostate DCE-MRI cases. The approximated analytical correction was compared with the ground truth PK parameters in simulation, and compared with the reference numerical correction in in-vivo experiments, using percentage error as the metric.

To develop a deep transfer learning (DTL) based model to distinguish indolent lesions from clinically significant PCa lesions using multiparametric MRI, 140 patients with 3T mp-MRI and whole-mount histopathology (WMHP) were included as the study cohort with IRB approval. The DTL based model was trained on 169 lesions in 110 arbitrarily selected patients and tested on the remaining 47 lesions in 30 patients. We compared the DTL based model with the same deep learning (DL) model architecture trained from scratch and the classification based on PIRADS v2 score with a threshold of 4 using accuracy, sensitivity, specificity, and area under curve (AUC). Bootstrapping with 2000 resamples was performed to estimate the 95% confidence interval (CI) for AUC.

Results

Both RR-VFA-corrected T_1 and satTFL-corrected T_1 showed higher intra- and inter-scanner correlation (0.89/0.87 and 0.87/0.84 respectively) than VFA T_1 (0.84 and 0.74). Bland-Altman plots showed that VFA T_1 had a wider 95% limits of agreement and a larger range of T_1 for each tissue compared to T_1 with B_1^+ correction.

The prostate DRO results show that the proposed approach provides residual error less than 0.4% for both K^{trans} and v_e, compared to the ground truth. This noise-free residual error was smaller than the noise-induced error using the reference numerical correction, which had a

minimum error of 2.1±4.3% with baseline SNR of 234.5. For the 82 in-vivo cases, percentage error compared to the reference numerical correction method had a mean of 0.1% (95% central range of [0.0%, 0.2%]) across the prostate volume.

After training on 169 lesions in 110 patients, the AUC of discriminating indolent from clinically significant PCa lesions of the DTL based model, DL model without transfer learning and PIRADS v2 score > 4 were 0.726 (CI [0.575, 0.876]), 0.687 (CI [0.532, 0.843]) and 0.711 (CI [0.575, 0.847]), respectively in the testing set.

Conclusion

The application of B_1^+ correction (both RR-VFA and satTFL) to VFA T_1 results in more repeatable and reproducible T_1 estimation than VFA T_1. This can potentially provide improved quantification of the prostate DCE-MRI parameters.

The approximated analytical B_1^+ correction method provides comparable results with less than 0.3% error within 95% central range, compared to reference numerical B_1^+ correction. The proposed method is a practical solution for B_1^+ correction in prostate DCE-MRI due to its simple implementation.

The DTL based model achieved higher AUC compared to the DL model without transfer learning and PIRADS v2 score > 4 in discriminating clinically significant lesions in the testing set. The DeLong test indicated that the DTL based model achieved comparable AUC compared to the classification based on PIRADS v2 score ($p = 0.89$).

The dissertation of Xinran Zhong is approved.

Michael McNitt-Gray

Steven S. Raman

Dan Ruan

Peng Hu

Holden H. Wu

Kyunghyun Sung, Committee Chair

University of California, Los Angeles

2019

TABLE OF CONTENTS

List of Tables x

List of Figures xi

List of Equations xiv

Acknowledgements xv

Vita xvii

1 INTRODUCTION 1

1.1 Evaluation of B_1^+ Estimation Techniques 4

1.2 Simplified B_1^+ Estimation Technique 5

1.3 Deep Transfer Learning based Lesion Classification 6

2 REPEATABILITY AND REPRODUCIBILITY OF VARIABLE FLIP ANGLE T_1 QUANTIFICATION IN THE PROSTATE AT 3 TESLA 8

2.1 Introduction 8

2.2 Theory 9

2.2.1 Quantitative DCE MRI 9

2.2.2 Pre-contrast VFA T_1 Estimation Method 10

2.3 Methods 12

2.3.1 Image Analysis 14

2.3.2 Statistical Analysis 16

2.4 Results 18

2.5 Discussion 23

2.6 Conclusion 26

3 PROSTATE DCE-MRI WITH B_1^+ CORRECTION USING AN APPROXIMATED ANALYTICAL APPROACH 28

3.1 Introduction 28

3.2 Theory 30

3.2.1 Quantitative Analysis for prostate DCE-MRI 30

3.2.2 B_1^+ Correction: Reference Numerical Approach 31

3.2.3 B_1^+ Correction: Approximated Analytical Approach 32

3.3 Methods 34

3.3.1 Prostate Digital Reference Object (DRO) 34

3.3.2 In-vivo Prostate DCE-MRI Data *38*
3.4 RESULTS 40
3.4.1 Prostate DRO *40*
3.4.2 In-vivo Prostate DCE-MRI Data *44*
3.5 DISCUSSION 47
3.6 CONCLUSION 52
4 DEEP TRANSFER LEARNING-BASED PROSTATE CANCER CLASSIFICATION USING 3 TESLA MULTIPARAMETRIC MRI 54
4.1 INTRODUCTION 54
4.2 METHODS 55
4.2.1 Study Population and MR Imaging Technique *55*
4.2.2 General Workflow *57*
4.2.3 Pre-processing *58*
4.2.4 DTL based Model *59*
4.2.5 Model Evaluation and Comparison *60*
4.2.6 Statistical Analysis *63*
4.3 RESULTS 64
4.3.1 DTL based Model Evaluation *64*
4.3.2 DTL based Model Comparison with Other Models *65*
4.3.3 DTL based Model Comparison with PIRADS v2 Score Expert Reader Performance *66*
4.4 DISCUSSION 69
4.5 CONCLUSION 72
5 SUMMARY AND FUTURE STUDIES 73
5.1.1 B_1^+ Estimation in Quantitative DCE MRI *74*
5.1.2 B_1^+ Correction in quantitative DCE MRI *75*
5.1.3 Combination of B_1^+ Estimation and B_1^+ Correction for PCa Diagnosis *75*
5.1.4 Improvement for Deep Transfer Learning based PCa Classification *75*
5.1.5 Improvement for Lesion Classification in Prostate Cancer *77*
5.1.6 Combination of Quantitative DCE MRI and PCa Classification Model *77*
5.1.7 Summary *78*

6 APPENDICES 79
7 REFERENCES 84

LIST OF TABLES

Table 2-1 Summary of mean T_1, mean difference, 95% CI of the difference and RC after bootstrapping for each tissue repeatability and reproducibility. Metrics were averaged among each pair of comparisons (two pairs for intra-scanner and four pairs for inter-scanner) respectively. 23

Table 2-2 CCC between pairs of scans for intra-scanner repeatability and inter-scanner reproducibility. CCCs were averaged among each pair of comparisons (two pairs for intra-scanner and four pairs for inter-scanner) respectively. P-value from z-test of CCCs for repeatability and reproducibility comparison (after Bonferroni adjustment statistically significant difference was marked with *). 24

Table 3-1 Details of the DRO modification. Compared to QIBA DRO designed for brain DCE MRI, the proposed DRO modified the flip angles, repetition time and the size of the DRO. 35

Table 3-2 Summary of in-vivo results. The mean, standard deviation and 95% central range for K^{trans}, v_e, k, $E_{A,\text{ in-vivo}}$ for K^{trans} and v_e were listed. 47

Table 4-1 Summary for lesion distribution in training and testing data 61

Table 4-2 Performance summary on the testing set (n = 47) 65

LIST OF FIGURES

Figure 2-1 General workflow for quantitative DCE-MRI. Representative VFA images, Dynamic T_1 weighted images, contrast agent concentration curve for one pixel, quantitative DCE biomarkers K^{trans} and v_e map were also shown for each step. 12

Figure 2-2 Experiment design summary. Each volunteer was scanned four times on two 3.0 T scanners. The volunteers were repositioned between two scans on the same scanner, and the time interval between scans on two scanners varies from same-day to 70 days. Each scan consists of three different sequences and after post-processing, we obtained one T2W image, two B_1^+ maps and three T_1 maps to analyze. 13

Figure 2-3 Representative ROI placements on T2W-TSE images after registration. Volumetric ROIs were created for transition zone (TZ), peripheral zone (PZ) in the prostate and the obturator internus left and right muscles (left and right pelvic muscles) 15

Figure 2-4 Representative slices (apex, midgland and base) for one scan with $T_{1,\,VFA}$, $T_{1,\,satFTL}$ as well as $T_{1,\,RR\text{-}VFA}$ within ROIs overlaid on T2w-TSE images. The inconsistent $T_{1,\,VFA}$ estimation between left and right pelvic muscles became more consistent in $T_{1,\,satTFL}$, and $T_{1,\,RR\text{-}VFA}$. 18

Figure 2-5 Bland-Altman Plot for comparing mean T_1 within ROIs between two scans on Scanner 1 (a, c, e) and between Scan 1 on Scanner 1 and Scanner 2 (b, d, f). Each encoded color indicates one tissue. The mean T_1 for each tissue had a more condensed distribution for $T_{1,\,satTFL}$ (c, d) and $T_{1,\,RR\text{-}VFA}$ (e, f) compared to $T_{1,\,VFA}$ (a, b). 20

Figure 2-6 Linear regression plots for comparing mean T_1 within ROIs between two scans on Scanner 1 (a ,c ,e) and between Scan 1 on Scanner 1 and Scanner 2 (b, d, f). Each encoded color indicates one tissue. 21

Figure 3-1 Comparison between the two B correction approaches. Conventional correction method needs to repeat MRI modeling and PK modeling pixel by pixel with original T_1 images and dynamic T_1 weighted images, while approximated analytical correction method only requires the uncorrected PK parameters to perform the correction and simplifies the computation. 34

Figure 3-2 Summary of the simulation study design using DRO under various population-averaged AIFs (a) and Gaussian noise (b). The images in (a) are examples of corresponding K^{trans} and v_e maps in each step and DRO images. P represents PK parameters K^{trans} and v_e 37

Figure 3-3 A representative slice of ROI positioning for in-vivo prostate data. ROI was first drawn on contrast-enhanced images (a) and was copied to corresponding relative B_1^+ (k) map (b) $E_{B1,\,in\text{-}vivo}$ map (c) as well as $E_{A,in\text{-}vivo}$ for K^{trans} 39

Figure 3-4 Comparison between numerical correction method and approximated analytical correction method in simulation with k ranging from 0.7 to 1.3 (ground truth K^{trans}' = 0.05 min^{-1}, $v_e = 0.1$, $T_{10} = 1000$ ms) for K^{trans} (a) and v_e (b) and $E_{A,DRO}$ for K^{trans} and v_e (c). Two example areas around k of 1.1 and 1.3 are zoomed. The difference between blue and red curves indicates $E_{A,\,DRO}$ (also shown in (c)), and the difference of y-axis and 100% indicate $E_{B1,\,DRO}$. $E_{A,\,DRO}$ is negligible compared to $E_{B1,\,DRO.}$ 41

Figure 3-5 $E_{A,\,DRO}$ maps using three population-averaged AIFs for K^{trans} estimation (a-c) and for v_e estimation (d-f). The maximum residual error for K^{trans} is 0.2% and for v_e is 0.4%. 42

Figure 3-6 Comparison of correction residual percentage errors between two correction methods ($E_{N,DRO}$ and $E_{A,\,DRO}$) for K^{trans} maps (a) and v_e maps (b) with various levels of noise added. There are 100 Monte-Carlo simulations for each PK parameter combination. For each SNR level, noise-induced errors for 3000 pixels ($5 \times 6 \times 100$) excluding outliers were averaged. Across all simulated baseline SNRs, the residual error for both correction methods are comparable to each other. 43

Figure 3-7 $E_{N,\,DRO}$ averaged for each parameter (100 pixels) (a, d) and $E_{A,DRO}$ averaged for each parameter (b, e) for K^{trans} and v_e maps with baseline SNR of 38.5. The error patterns are similar between the two methods. The absolute value difference maps ($E_{N,DRO} - E_{A,DRO}$) averaged for each parameter (c and f) indicate outliers appear when v_e is low, where the fitting process is more sensitive to noise. 44

Figure 3-8 Linear regression and Bland-Altman plots for K^{trans} maps (a-b) and v_e maps (c-d) in DRO experiment with noise added. There are 100 Monte-Carlo simulations for each PK parameter combination. The corrected PK parameters from approximated analytical correction and numerical correction are highly comparable ($r^2 = 0.97$ for K^{trans} and $r^2 = 1.00$ for v_e) with baseline SNR of 38.5. 45

Figure 3-9 Ktrans maps after numerical B1+ correction method (a), before B1+ correction (b) and after approximated analytical correction method (c), B1+ induced error for Ktrans (d), correction residual error of Ktrans (e), ve maps after numerical B1+ correction method (f), before B1+ correction (g) and after approximated analytical correction method (h), B1+ induced error for ve (i), and correction residual error of ve (j) 46

Figure 3-10 Summary of average k (a), correction residual error $E_{A,\,in\text{-}vivo}$ (b) for 82 patients. Different scanners (Skyra and Trio) have slightly different k distribution as shown in (a), but all the average $E_{A,\,in\text{-}vivo}$ is smaller than 0.4%. 48

Figure 3-11 Extend Tofts model was simulated in the DRO with three v_p value, 0.001, 0.005 and 0.01. The results were evaluated using $E_{A,DRO}$ for K^{trans} (a-c) and v_e estimation (d-f). 52

Figure 4-1 Summary of the workflow. The input of the deep transfer learning model would be the image patches enclosing the lesion after pre-processing, and the out put would be the predicted probability that if a lesion is clinically significant. The label was based on the GS from whole-mount histopathology. 58

Figure 4-2 Example of how the image patches (yellow) was extracted based on the lesion contours (red). The intensity of image patches after pre-processing became more consistent between cases with and without endorectal coils. 59

Figure 4-3 Deep transfer learning structure. 60

Figure 4-4 Representative examples of the data augmentation. The left column shows the image patches before data augmentation, and the right column shows six examples of the image patches after data augmentation for T2 SPACE and ADC images 62

Figure 4-5 Representative cases for successful and failed prediction cases based on the DTL based model on clinically significant lesion (a and b) and indolent lesion (c and d). The corresponding GS from WMHP and PIRADS v2 score for each lesion are also shown under the images 64

Figure 4-6 Receiver operating characteristics curve (a) and precision recall curve (b) comparison between DTL, DTL_{T2}, DTL_{ADC}, DTL_{T2+ADC} and DL models on 47 testing lesions, validating the DTL based model architecture 67

Figure 4-7 Receiver operating characteristics curve (a) and precision recall curve (b) comparison between the proposed DTL based model and the expert reader PIRADS v2 score (n=47) 68

LIST OF EQUATIONS

Equation 2-1	11
Equation 2-2	11
Equation 3-1	30
Equation 3-4	31
Equation 3-5	31
Equation 6-1	79
Equation 6-2	80
Equation 6-3	80
Equation 6-4	80
Equation 6-5	81
Equation 6-6	81
Equation 6-7	81
Equation 6-8	81
Equation 6-9	82
Equation 6-10	82
Equation 6-11	82
Equation 6-12	82

ACKNOWLEDGEMENTS

I would like first to express my gratitude to my advisor, Dr. Kyunghyun Sung. Without him, I could not achieve what I have done so far. He is a great mentor not only providing insightful advice on research but also sharing his wisdom on communication, work-life balance and many other essential skills to get me prepared for my future career. He is always supportive and wants the best for every student. I feel lucky to have him as my advisor.

I would thank Dr. Peng Hu, Dr. Dan Ruan, Dr. Steven S. Raman, Dr. Michael McNitt-Gray and Dr. Holden H. Wu for being in my committee. You helped me to refine my projects and made sure I was on the right track towards my graduation. Particularly, I would like to thank Dr. Holden H. Wu who would provide enlightening comments on my experiment design, manuscripts and presentations, which helped me to present my work more clearly.

I would also like to thank my collaborators within and outside UCLA. Thanks to your help, our projects went smoothly. I want to express my special gratitude to Dr. Steven Raman and his clinical team, Dr. Sepideh Shakeri, Dr. Amirhossein Mohammadian Bajgiran and Dr. Sohrab Afshari Mirak, for continually providing professional clinical perspectives on my projects. Thank you, Dr. James Sayre, Dr. Grace Hyun J. Kim, Dr. Scalzo, Fabien, Dr. Dan Ruan for the helpful discussions related to statistics and machine learning. Moreover, I want to thank Dr. Krishna S. Nayak, Yannick Bliesener, Dr. Yi Guo, Dr. William Hsu, Dr. Sharon Qi, Dr. Dieter R. Enzemann, Teresa Chanlaw and Sarah H. Choi. It was a pleasure to work with you, and I gained a vast array of knowledge from each of you.

I am grateful for being in the family of the Physics and Biology in Medicine program. I made some good friends within the program. I want to thank all my friends for making my Ph.D.

life colorful and supporting me through stressful times. I would like to thank Dr. Michael McNitt-Gray, Reth Im and Alondra T. Correa Bautista for always providing help and making the program so warm and welcoming.

I want to thank all the past and current members in MRRL for all the inspiring discussions and supports. Specifically, I want to give thanks to Ruiming, Alibek, Thomas, Le, Andres, Jiahao, Haoxin, Felix, Ziwu, Yu, Chang, Jiaxin, Xinzhou, Zhaohuan, Tess, Caroline, Fadil and Yongkai. I also want to thank Heather for taking good care of the lab so that we could fit into the lab without difficulty. Thanks to all MR technologists for helping me during the scans.

Last and most importantly, I would like to express my sincere gratitude to my parents. They taught me how to become an independent and responsible person. They encourage me to try my best, and they are always there for me through good and bad times.

VITA

Education

2014 – 2019	University of California Los Angeles	Los Angeles, CA
	Ph.D., Biomedical Physics	
2010 – 2014	Tsinghua University	Beijing, China
	B.S., Biomedical Engineering	

Honors & Awards

2017	ISMRM Summa Cum Laude Award
2015, 2016, 2017	ISMRM Educational Stipend
2012	Academic Scholarship of Tsinghua University

First Author Peer Reviewed Publications

1. **Zhong X**, Martin T, Wu HH, Nayak K, Sung K. Prostate DCE-MRI with B_1^+ correction using an approximated analytical approach. Magnetic Resonance in Medicine. 2018;80;2525-2537. https://doi.org/10.1002/mrm.27232
2. **Zhong X**, Liu D, Sayre J, Wu HH, Sung K, Repeatability and Reproducibility of Variable Flip Angle T_1 Quantification in the Prostate at 3 Tesla. Journal of Magnetic Resonance Imaging. 2018. https://doi.org/10.1002/jmri.26596
3. **Zhong X**, Cao R, Shakeri S, Scalzo F, Lee Y, Enzmann DR, Wu HH, Raman SS, Sung K. Deep transfer learning-based prostate cancer classification using 3 Tesla multi-parametric MRI. Abdominal Radiology. 2018:1-10. https://doi.org/10.1007/s00261-018-1824-5

Co-Author Peer Reviewed Publications

4. Bliesener Y, **Zhong X**, Guo Y, Bosca R, Kirsch J, Boss M, Sung K, Nayak K, RF Transmit

Calibration for Dynamic Contrast-Enhanced MRI, Accepted by Medical Physics, early view

5. Cao R, **Zhong X**, Shakeri S, Afshari Mirak S, Mohammadian Bajgiran A, Enzmann D, Raman S, Sung K. Prostate Cancer Detection and Segmentation in Multi-parametric MRI via CNN and Conditional Random Field 2019 IEEE 16th International Symposium on Biomedical Imaging (ISBI 2019). IEEE

6. Cao R, Mohammadian Bajgiran A, Afshari Mirak S, Shakeri S, **Zhong X**, Enzmann D, Raman S, Sung K. Joint Prostate Cancer Detection and Gleason Score Prediction in mp-MRI via FocalNet IEEE Transactions on Medical Imaging (TMI), early access. http://doi.org/ 10.1109/TMI.2019.2901928.

Patent

7. Sung K, Hsu W, Shen S, **Zhong X**, 2018. Deep-Learning-Based Computerized Prostate Cancer Classification using a Hierarchical Classification Framework. App Serial No. 16/104,131, Non-provisional U.S. Patent Application filed on August 30, 2018

Selected Conference Abstracts: Oral Presentations

8. **Zhong X**, Li N, Sung K, Qi XS, Rectal Cancer Prognosis Prediction using Radiomics from Pre-treatment MRI, AAPM, SU-F-DBRB-7, July 2018

9. **Zhong X**, Wu H, Nayak K, Sung K, B_1^+ correction evaluation for Pharmacokinetic Modeling in prostate DCE MRI using Digital Reference Object, Proceedings ISMRM Twenty-Fifth Scientific Session, 0622, April 2017, Magna Cum Laude Merit Award

10. **Zhong X**, Sung K, Evaluation of Prostate Cancer Detection using Modified MR Dispersion Imaging, Proceedings ISMRM Twenty-Seventh Scientific Session, 0378, May 2019

1 INTRODUCTION

Prostate cancer (PCa) is among one of the most common cancers and is a leading cause of cancer-related death in men in the United States[1]. Detection and grading of PCa are important for patient prognosis and treatment since small low grade lesions may undergo observation, whereas larger, higher grade lesions would be treated aggressively with surgery or radiation[2]. For the past 25 years, the standard diagnostic method of PCa diagnosis has been elevation of prostate-specific antigen (PSA) followed by transrectal ultrasound (TRUS)-guided biopsy, which has resulted in decreased PCa mortality by 20%-30%[3]. However, this approach is associated with the underdetection of higher-grade (clinically significant) lesions and overdetection of low-grade (indolent) lesions due to sampling bias.[4] As a result, patients can receive overtreatment related morbidity such as incontinence and impotence [5,6] or repeated invasive biopsy assessment, which is costs for both patients and health care systems.[7]

On the other hand, multi-parametric (mp-MRI), mainly consists of diffusion-weighted imaging, dynamic contrast-enhanced (DCE) imaging as well as T_2 weighted imaging, is a great non-invasive diagnosis tool and has been widely used in PCa diagnosis[8]. 3T mp-MRI enables

detection of 50% of all PCa lesions and 80% of clinically significant lesions[9]. Moreover, studies have shown the diagnosis improvement with MRI guided biopsy compared to TRUS-guided systematic biopsy.[4,7] However, one of the main limitations of the broader applications of mp-MRI is its requirement for training and expertise for image interpretation by subspecialized radiologists, which may lead to high inter- and intra-reader variability in diagnosis[10], even with standardized guidelines such as Prostate Imaging Reporting and Data System (PIRADS v2)[11]. These guidelines which mainly evaluate qualitative parameters for T2, DWI and DCE enable more standardized and reproducible reporting of clinically significant PCa foci, but are still subject to limitations such as interobserver variability[12] and improved but limited correlation with final pathology[13]. Specifically, the reader's experience was proved to influence the inter-reader agreement and cancer detection rate.[14,15] Methods to standardize image interpretation would be desirable, and one solution is to establish quantitative models to automate the image interpretation procedure to help diagnosis.

One way to build the quantitative interpretation model is to develop quantitative MRI biomarkers based on prior knowledge of physics and physiology. Current clinical acquisition of mp-MRI in prostate is not quantitative since they are mainly based on so-called weighted images that the image intensity is affected by different factors including various tissue properties and scanning protocol. On the other hand, quantitative MRI image intensity is expected to represent certain physical or chemical variables that can be measured in physical units and compared between tissue regions and among subjects. In this way, the images can be used as a measuring tool instead of a camera. In other words, the quantification value should be accurate so that it reflects certain tissue character, and precise so that the changes in the measurements corresponds to tissue property changes. So quantitative MRI can help to achieve more consistent interpretation

for tissue characterization but can still be limited due to error in the acquisition and modeling. For example, volume transfer factor K^{trans} is a biomarker to estimate vascular permeability calculated from DCE MRI, which was proved useful in lesion detection, aggressive and treatment response.[16] However, B_1^+ inhomogeneity effect could lead to variation in K^{trans} estimation from scan to scan, reducing the estimation consistency.

Another way to build a quantitative model is to use machine learning technique. In general, a model is can be developed with optimization algorithms to explore the correlation between mp-MRI and histopathology based on known data, namely training data, in order to make predictions on the unknown data, namely testing data. Traditional machine learning models mainly consisted of two steps. Careful designed handcrafted features were extracted from images and then served as the input for the classifier. The performance of the model was highly dependent on the design of the feature extraction, which is less representative and needs expertise in feature engineering. In recent years, the development in deep learning models enables to explore highly nonlinear features from images directly and achieved better performance at the same time. Among them, convolutional neural networks (CNN) has become the leading model architecture for most image classification tasks. In PCa diagnosis, distinguishing clinically significant lesion from indolent lesion using mp-MRI is an important and challenging task, which can benefit from deep learning techniques.

The purpose of this dissertation was to improve the quantitative interpretation of mp-MRI by improving the consistency of quantitative DCE MRI biomarkers and developing deep learning-based lesion classification model using mp-MRI. The background and overview of each project are discussed in the following.

1.1 Evaluation of B_1^+ Estimation Techniques

Growth of PCa is reportedly associated with the development of a rich blood supply fed by a large network of immature, leaky blood vessels, and the density of this network was related to tumour grade and metastasis potential.[17] Therefore as one major component in mp-MRI, DCE-MRI has shown great promise for detection and grading of PCa[18–20] for it reveals tissue vasculature property. Conventional image analysis for DCE-MRI are typically based on subjective evaluation or qualitative analyses of signal uptakes, and such evaluations are limited by inter-observer variability and highly dependent on data acquisition[21]. To develop quantitative biomarkers, pharmacokinetic DCE MRI analysis was proposed for tissue vasculature evaluation[16,22] and became one of the most popular analysis models in DCE-MRI[23–28]. The general idea of quantitative DCE MRI was to estimate the contrast agent concentration curve based on the change of T_1 induced by contrast agent and then analyze the curve. However, quantitative DCE-MRI analysis results still suffer from high variability among different perfusion analysis tools and among different scanners[29].

Radiofrequency transmit field (B_1^+) inhomogeneity is one of the major challenges in quantitative DCE-MRI, for it varies the prescribed flip angles, thus reducing the accuracy of T_1 and contrast agent concentration curve estimation. The B_1^+ inhomogeneity becomes more severe with high field scanners ($\geq$ 3.0 T). An intra-subject B_1^+ variation from 30% - 50% over the chest was observed[30]. Additionally, an inter-subject average B_1^+ variation of 30% is observed in the prostate region[31]. Although various techniques were proposed to measure B_1^+ inhomogeneity, there are few applied in the clinic. One reason is that most techniques require a separate scan and may not be available on all systems. There is no clear clinical guideline on which technique to use to estimate B_1^+ inhomogeneity. Also, B_1^+ inhomogeneity problem in T_1 estimation in prostate at

3T has not been fully evaluated.

Chapter 2 describes the work to evaluate how different B_1^+ estimation techniques will influence the pre-contrast T_1 estimation regarding inter-scanner repeatability and intra-scanner reproducibility on 21 healthy volunteers on two 3T systems. It is important to evaluate the variation induced by B_1^+ inhomogeneity in prostate at 3T. Moreover, T_1 estimation consistency is a crucial evaluation metric for B_1^+ estimation. With this project, the B_1^+ inhomogeneity problem and several B_1^+ estimation techniques can be better understood.

1.2 Simplified B_1^+ Estimation Technique

Although multiple techniques have been proposed to estimate B_1^+ accurately[32–39] and two of them were evaluated on healthy volunteers in my previous work, the broad application of B_1^+ compensation in quantitative DCE-MRI is still hard to achieve. One reason is that the conventional numerical B_1^+ correction approach is demanding[40]. It requires repeating the pixel-by-pixel modeling process for 3D coverage and needs access to both raw images and complete modeling implementation, which can be hard in clinical settings. Therefore, a simplified analytical based correction approach can be beneficial for it can be more clinically feasible.

Chapter 3 describes the work on the development of the approximation based analytical B_1^+ correction approach, and the evaluation of this B_1^+ correction technique in comparison to the conventional numerical B_1^+ correction method using both simulation and in-vivo data. The proposed B_1^+ correction method only requires the B_1^+ maps and the uncorrected quantitative DCE MRI maps as the input with simplified computation. It is a promising technique to both understand the error propagation from B_1^+ inhomogeneity to quantitative DCE biomarkers and increase the feasibility of integrating B_1^+ compensation into clinical workflow.

1.3 Deep Transfer Learning based Lesion Classification

With a large number of images to process per case and the variety of scanning protocols and readers, a fast and consistent interpretation of mp-MRI can be challenging. Specifically, one crucial task in PCa diagnosis is to distinguish clinically significant lesions from indolent lesions. Gleason Score (GS) from histopathology analysis was shown to be a powerful indicator for patient outcome and is used to define clinically significant lesion. As described before, machine learning models can help identify clinically significant lesions automatically, reducing the reading time and diagnosis variability.[41] Many attempts have been made to develop reliable machine learning systems for this task[42–45]. However, most of them are utilizing handcrafted features. On the other hand, the deep neural networks draw considerable attention for they are capable of extracting multi-level abstraction from raw images directly with little human intervention. Deep learning has achieved great success in various computer vision tasks[46].

Although deep learning technique is promising, its transition to medical imaging domain is challenging. To train a reliable CNN, it often requires a large amount of labeled data to make sure the trained model is generalizable in prediction tasks. In the medical imaging domain, getting large numbers of high-quality data is difficult. First of all, organized patients' images are scarce and expensive to collect compared to natural images. Also, patient privacy issue makes it harder to share medical images, limiting data collection from multiple centers. Moreover, the labeling process of medical images requires great expertise and can be expensive, making it even harder to get large numbers of well-labeled data. For example, GS from biopsy samples is not as accurate as GS from whole-mount histopathology (WMHP) after prostatectomy because of the sampling bias, and labelling with the latter information improves the data quality. However, getting GS from WMHP further limits the patient population so this type of data is less available. On the other

hand, to overcome the data size requirement limitation in CNN, one technique called deep translation learning was proposed, which utilizes information from models trained in natural images or other different imaging domains to help to solve the problem in target image domain with fewer data available. Fine-tuning is a representative and powerful technique in this category[47] and can be used in the lesion classification model.

Chapter 4 describe the work on the development and evaluation of a deep transfer learning-based algorithm in distinguishing clinically significant lesions from indolent lesions using mp-MRI. The proposed method was compared with the deep learning model without transfer learning as well as PIRADS v2 score from experienced radiologists. The goal of this study is to propose a deep transfer learning-based model to distinguish indolent lesions from clinically significant lesions and to compare the proposed model with deep learning model without transfer learning and PIRADS v2 on 3T mp-MRI with WMHP validation.

2 Repeatability and Reproducibility of Variable Flip Angle T_1 Quantification in the Prostate at 3 Tesla

2.1 Introduction

Quantitative DCE-MRI is an important tool in the diagnosis of PCa due to its ability to reveal changes in tissue vascular physiology such as tissue perfusion and permeability [18,19]. Unlike qualitative or semi-quantitative analysis, the quantitative DCE-MRI is theoretically independent of patient- and scanner-specific variations but in practice can be highly influenced by errors in estimation of pre-contrast T_1 and arterial input function [48]. Variable flip angle (VFA) T_1 mapping is a common approach to measure pre-contrast T_1 [49,50] but is limited due to its sensitivity to transmit RF (B_1^+) field inhomogeneity, which becomes more severe at higher field strengths (≥ 3 Tesla). Several studies showed that the VFA T_1 mapping can be improved after compensating for B_1^+ inhomogeneity [51–53]. Since B_1^+ inhomogeneity mainly depends on the subject geometry and hardware, the B_1^+ field inhomogeneity should be measured by a subject-by-subject basis. Around 30% inter-subject B_1^+ variation has been noted in the prostate region at 3T [31,54]. Therefore, accurate and reproducible methods for T_1 mapping with B_1^+ correction should be important for prostate DCE-MRI.

Many techniques have been proposed to measure the B_1^+ inhomogeneity. These methods can be grouped into two main categories [55]: magnitude-based [33,34,36,37,56] and phase-based methods [39,55]. Although these techniques have been shown to provide accurate and fast B_1^+ mapping, the majority of these B_1^+ correction techniques employ a 2D multi-slice acquisition, leading to slice profile and position mismatches with the corresponding 3D VFA imaging for T_1 mapping. More importantly, all the proposed techniques require a separate B_1^+ mapping scan, which requires additional scan time and may not be available on all MRI scanner systems.

Recently, a simultaneous 3D T_1 and B_1^+ mapping technique, called reference-region variable flip angle (RR-VFA), was proposed [31,57]. In RR-VFA, it was assumed that the reference region (e.g., fatty tissue) has a known T_1 relaxation time, and the B_1^+ variation for the target tissue was interpolated based on the B_1^+ calculated from the reference region. The study demonstrated the feasibility of using RR-VFA in the breast [31] and prostate [57] without adding any additional scan time. The RR-VFA has the potential to avoid potential mismatches between B_1^+ mapping and actual imaging sequences and can be a better practical choice due to wide availabilities on many MRI systems. Therefore, further investigation of the repeatability and reproducibility of RR-VFA may increase the utilities of RR-VFA in various settings. In this study, we aim to evaluate the intra-scanner repeatability and inter-scanner reproducibility of prostate VFA T_1 estimation with and without B_1^+ inhomogeneity correction using either saturated turbo FLASH (satTFL) [37] or RR-VFA B_1^+ estimation at 3.0 T.

2.2 Theory

2.2.1 Quantitative DCE MRI

As described in Chapter 1.1, the foundation of DCE MRI is that paramagnetic contrast

agent injected in the vascular system will reduce T_1, and the perfusion can be estimated by observing the temporal variation of signal or T_1 in the tissue. The acquisition comprised of a set of T_1 weighted images, which is usually acquired with radiofrequency-spoiled gradient echo (SPGR) sequence, including baseline images before contrast injection and dynamic images after contrast injection within a few minutes. From those images, a signal intensity curve can be generated for each pixel. This signal intensity curve is related to contrast agent concentration variation, but the scaling factor is unknown and varies for different scanning protocols. Analysis performed on this curve is called semi-quantitative analysis, and the resulting biomarkers can be hardly compared between different scans.

To make the quantification more reliable, pre-contrast T_1 maps can also be estimated so that the contrast agent concentration curve can be generated based on the T_1 variation before and after contrast injection. This process is called MRI modeling for it is based on MRI physics. This curve is independent of scanning protocol and solely related to the tissue perfusion property. With the contrast agent concentration curve available, the pharmacokinetic analysis[58] can be performed on, and this analysis is also called quantitative DCE MRI. The quality of the quantitative DCE-MRI analysis highly relies on the accuracy of pre-contrast T_1 quantification.

2.2.2 Pre-contrast VFA T_1 Estimation Method

In prostate pre-contrast T_1 estimation, VFA is a commonly used. This technique requires multiple spoiled gradient-echo (SPGR) acquisitions with constant short repetition time and a set of flip angles. T_1 maps can be estimated based on the pixel-wise linear fitting on the acquired images. This method is also called Driven Equilibrium Single-Pulse Observation of T_1 (DESPOT1).[59] The quantification process is described in the following.

The signal intensity (S) of SPGR, ignoring T_2* decay, can be expressed as,

Equation 2-1

$$\boldsymbol{S = M_0 \frac{\sin\theta(1 - E_1)}{1 - E_1 \cos\theta}}$$

where M_0 is the equilibrium magnetization, θ is the flip angle, and $E_1 = e^{-TR/T_1}$.

VFA with SPGR can be used to generate T_{10} maps by employing a set of flip angles, $\alpha_i \in \{\alpha_1, \alpha_2, \cdots, \alpha_N\}$, with fixed repetition time (TR) and echo time (TE)[4950]. T_{10} can be calculated using a simple linear regression by substituting α_i and $E_{10} = e^{-TR/T_{10}}$ into Equation 2-1,

Equation 2-2

$$\boldsymbol{\frac{S(\alpha_i)}{\sin(\alpha_i)} = E_{10} \frac{S(\alpha_i)}{\tan(\alpha_i)} + M_0(1 - E_{10})}$$

Figure 2-1 shows the general workflow for quantitative DCE MRI analysis. It is demonstrated that if the actual flip angles deviate from the prescribed ones, the error will propagate to the T_1 estimation and then to the quantitative DCE biomarkers, so it is crucial to consider B_1^+ inhomogeneity effect in VFA T_1 quantification. In this chapter, the repeatability and reproducibility of T_1 maps with and without B_1^+ compensation will be evaluated.

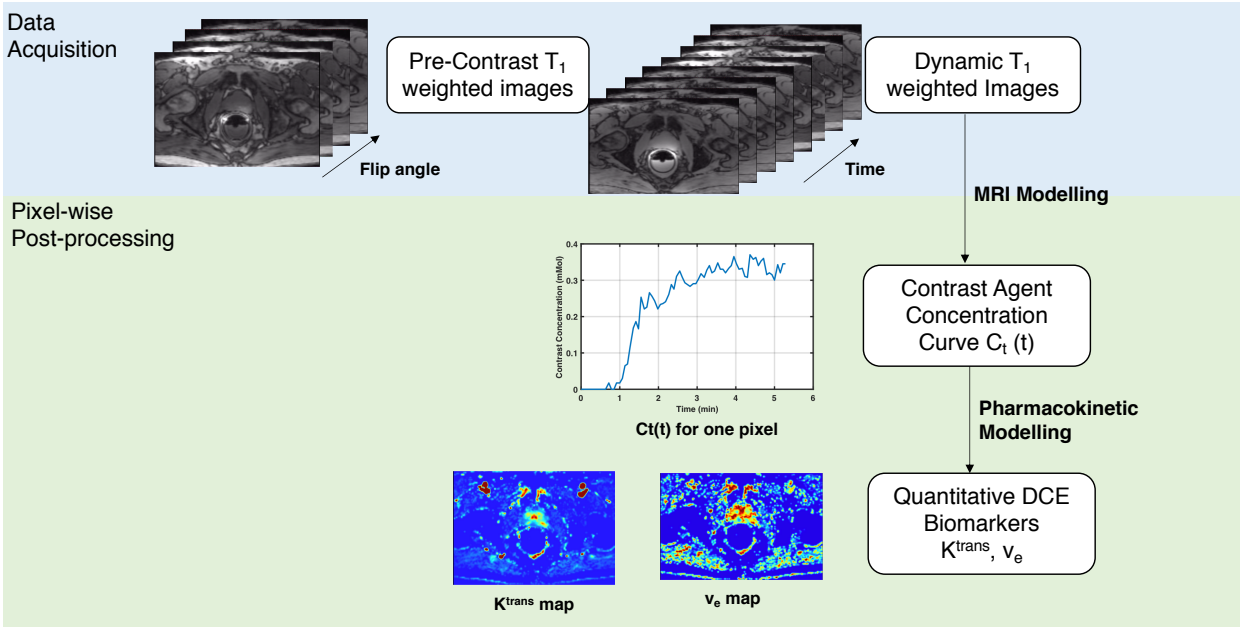

Figure 2-1 General workflow for quantitative DCE-MRI. Representative VFA images, Dynamic T_1 weighted images, contrast agent concentration curve for one pixel, quantitative DCE biomarkers K^{trans} and v_e map were also shown for each step.

2.3 Methods

Twenty-one male volunteers prospectively participated in the study, approved by the local institutional review board (IRB), for which all subjects provided written informed consent. The twenty-one subjects (27 ± 4 years old) were imaged twice in two different 3T MRI scanners (Skyra as "Scanner 1" and Prisma as "Scanner 2"; Siemens Healthcare, Erlangen, Germany) using the body coil for RF transmission and receive-only phased-array coil for signal reception, resulting in 84 scans. All subjects underwent MRI exams in a feet-first supine position, and the vendor-provided "TrueForm" RF transmission mode was active for both scanners for B_1^+ shimming [60].

The overall experiment setup for repeatability and reproducibility testing is summarized in Figure 2-2. For each scan, our protocol included a 2D T_2-weighted Turbo Spin Echo (T2w-TSE) sequence, a 2D satTFL-B_1^+ mapping sequence, and 3D VFA imaging with a dual echo bipolar readout. The axial multi-slice 2D T2w-TSE was used for an anatomical reference (TE/TR=101/4000 ms, FOV=200×200 mm^2, spatial resolution=0.6×0.6×3.0 mm^3, flip angle=160°, and total scan time = 3 min 38 s). The multi-slice 2D satTFL sequence was

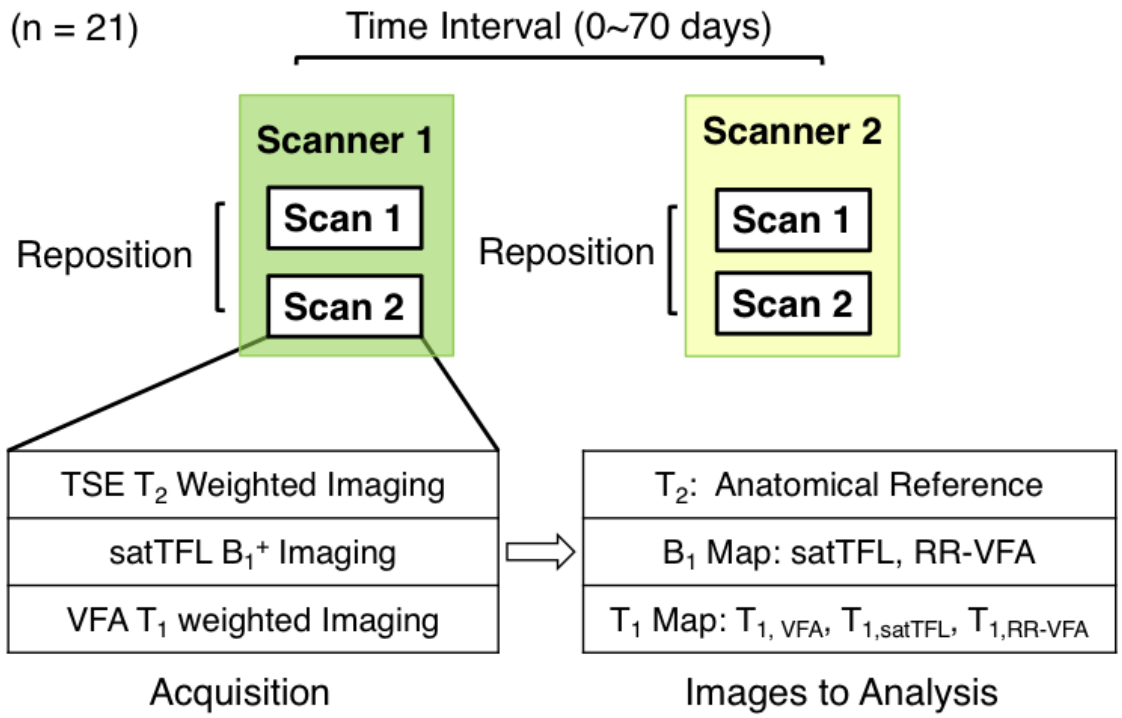

Figure 2-2 Experiment design summary. Each volunteer was scanned four times on two 3.0 T scanners. The volunteers were repositioned between two scans on the same scanner, and the time interval between scans on two scanners varies from same-day to 70 days. Each scan consists of three different sequences and after post-processing, we obtained one T2W image, two B_1^+ maps and three T_1 maps to analyze.

commercially available for all Siemens scanners, and the axial satTFL B_1^+ mapping sequence was used to cover the entire prostate (TE/TR =1.83/8000 ms, FOV=300 × 300 mm^2, spatial resolution=4.7×4.7×8.0 mm^3, flip angle = 80°, and total scan time = 20 s). The 3D VFA imaging was acquired axially and consisted of four flip angles, identical to our standard clinical prostate

DCE-MRI protocol. The two-point Dixon option was used to create fat- and water-only images (TEs=1.23/2.46 ms, TR=4.17 ms, FOV=260 × 260 mm^2, spatial resolution=1.6×1.6×3.6 mm^3, flip angles = 2°/5°/10°/15°, and total scan time = 4 × 35 s). The protocol was repeated twice on each scanner. After the first scan, the subject was taken off the scanner table, re-positioned on the scanner table, and scanned again. The time interval between scans on two different scanners varied from same-day to 70 days.

2.3.1 Image Analysis

The optimized prostate RR-VFA method [31] was used to simultaneously estimate both B_1^+ and T_1. The fatty reference tissue was identified in fat-only images using Otsu's method [61], and a population-based effective fat T_1 of 320 ms was used [31]. The B_1^+ inhomogeneity was initially calculated in the fatty tissue surrounding the prostate and then interpolated to the prostate, in terms of relative flip angle ($rFA_{RR\text{-}VFA}$ = obtained flip angle/prescribed flip angle×100 %). For comparison, the relative flip angle maps from satTFL were generated using the obtained flip angle divided by the prescribed flip angle (rFA_{satTFL} = obtained flip angle/prescribed flip angle×100 %).

Three T_1 maps were calculated based on VFA ($T_{1,VFA}$), VFA with B_1^+ correction using satTFL ($T_{1,satTFL}$), and RR-VFA ($T_{1,RR\text{-}VFA}$) for each scan. To account for different resolutions among the different sequences (T2w-TSE, VFA and satTFL), the B_1^+ and T_1 maps were linearly interpolated with respect to T2w-TSE images. All post-processing was performed using in-house scripts written in Matlab (Mathworks, Natick, Mass). Since the repeated MRI scans include re-positioning the subjects, we did rigid registration between four scans for the same volunteer using T2W images and registered corresponding B_1^+ and T_1 maps accordingly.

A genitourinary (GU) radiologist (interpreted over 2,000 prostate MR exams) reviewed the

T2w-TSE images of all volunteers and confirmed no visible abnormalities. The GU radiologist manually defined the regions of interest (ROIs) to cover the transition zone (TZ), peripheral zone (PZ) in the prostate and the obturator internus left and right muscles (left and right pelvic muscles) on the same slices as the prostate on T2w-TSE images. These latter muscle ROIs were selected to check the consistency by comparing T_1 values on contralateral sides. Four volumetric ROIs were

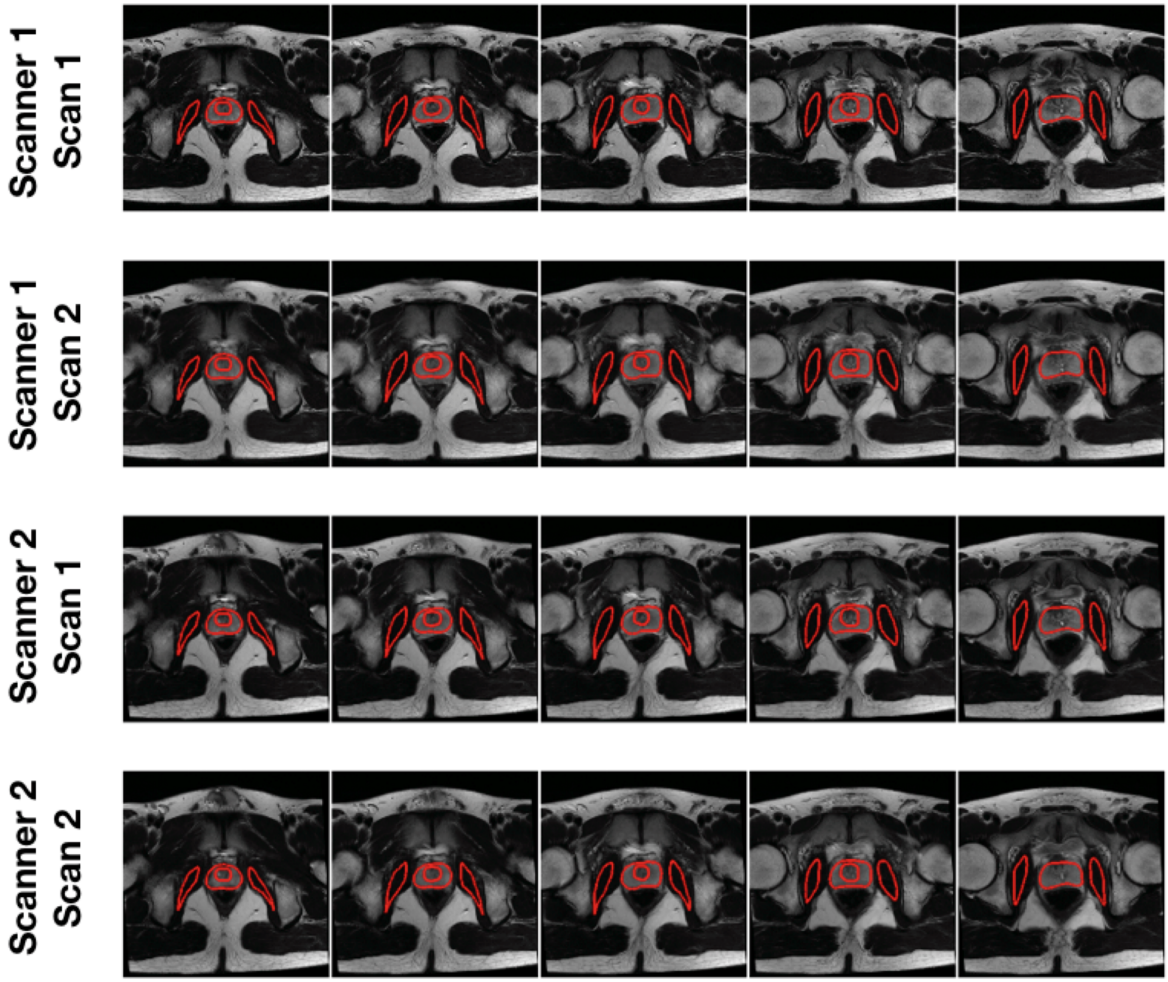

Figure 2-3 Representative ROI placements on T2W-TSE images after registration. Volumetric ROIs were created for transition zone (TZ), peripheral zone (PZ) in the prostate and the obturator internus left and right muscles (left and right pelvic muscles)

drawn for each volunteer, and the registered T2w-TSE images as well as the volumetric ROIs on one representative volunteer were shown in Figure 2-3. We also minimized partial volume effects

by excluding the tissue boundary. Each ROI was copied to the corresponding interpolated B_1^+ and T_1 maps. The average value within each ROI was computed for statistical analysis.

2.3.2 Statistical Analysis

A total of twenty-one subjects with 84 VFA T_1 scans were evaluated for the repeatability and reproducibility test. For each subject, there were two scans on two scanners, and the comparison was performed between scans both intra- and inter-scanners. The repeatability of VFA T_1 estimation with and without B_1^+ correction was evaluated by comparing T_1 values between two scans on the same scanner (intra-scanner), and the reproducibility was evaluated by comparing T_1 values between two scanners (inter-scanner). Three groups of T_1 estimation, including $T_{1,VFA}$, $T_{1,satTFL}$ and $T_{1,RR\text{-}VFA}$ were evaluated separately.

Statistical analysis of repeatability and reproducibility was performed based on Quantitative Imaging Biomarkers Alliance (QIBA)'s recommendation.[62] Lin's Concordance correlation coefficients (CCC) were mainly used to assess both repeatability and reproducibility. T_1 measurements from both prostate and muscles were evaluated using CCC between each pair of scans, and there were two intra-scanner comparisons and four inter-scanner comparisons in total. The intra- and inter-scanner CCC was then averaged respectively. Z-test with Bonferroni adjustment was used to compare the CCC between three pairs of T_1 maps ($\alpha = 0.05/3 = 0.017$). The null hypothesis for the z-test is two types of T_1 maps have equal CCCs for either repeatability or reproducibility comparison with the critical value $z_c = 1.96$ for two-sided comparison. Bland-Altman plots were also used to compare the agreement between scans with 95% confidence interval (CI) and limits of agreement (LOA). Linear regression plots were created to reveal the Due to limited sample size and non-Gaussian distribution, bootstrapping sampling with replacement (n = 1000) was used to calculate the mean T_1, repeatability coefficient (RC) and

reproducibility coefficient (RDC), mean and 95% confidence interval (CI) of the difference between two measurements. RC was defined as $2.77 \times \sigma_R$ and RDC was defined as $2.77 \times \sigma_{RD}$, where σ_R and σ_{RD} were the standard deviations of the difference between two measurements from repeatability test and reproducibility test respectively. Statistical analysis was performed using the in-house software in Matlab.

2.4 Results

One example scan with representative slices in apex, midgland and base were shown in Figure 2-4. The left and right pelvic muscles showed clearly different $T_{1,\,VFA}$ but the T_1 became more consistent in $T_{1,\,satTFL}$ and $T_{1,\,RR\text{-}VFA}$ after B_1^+ compensation.

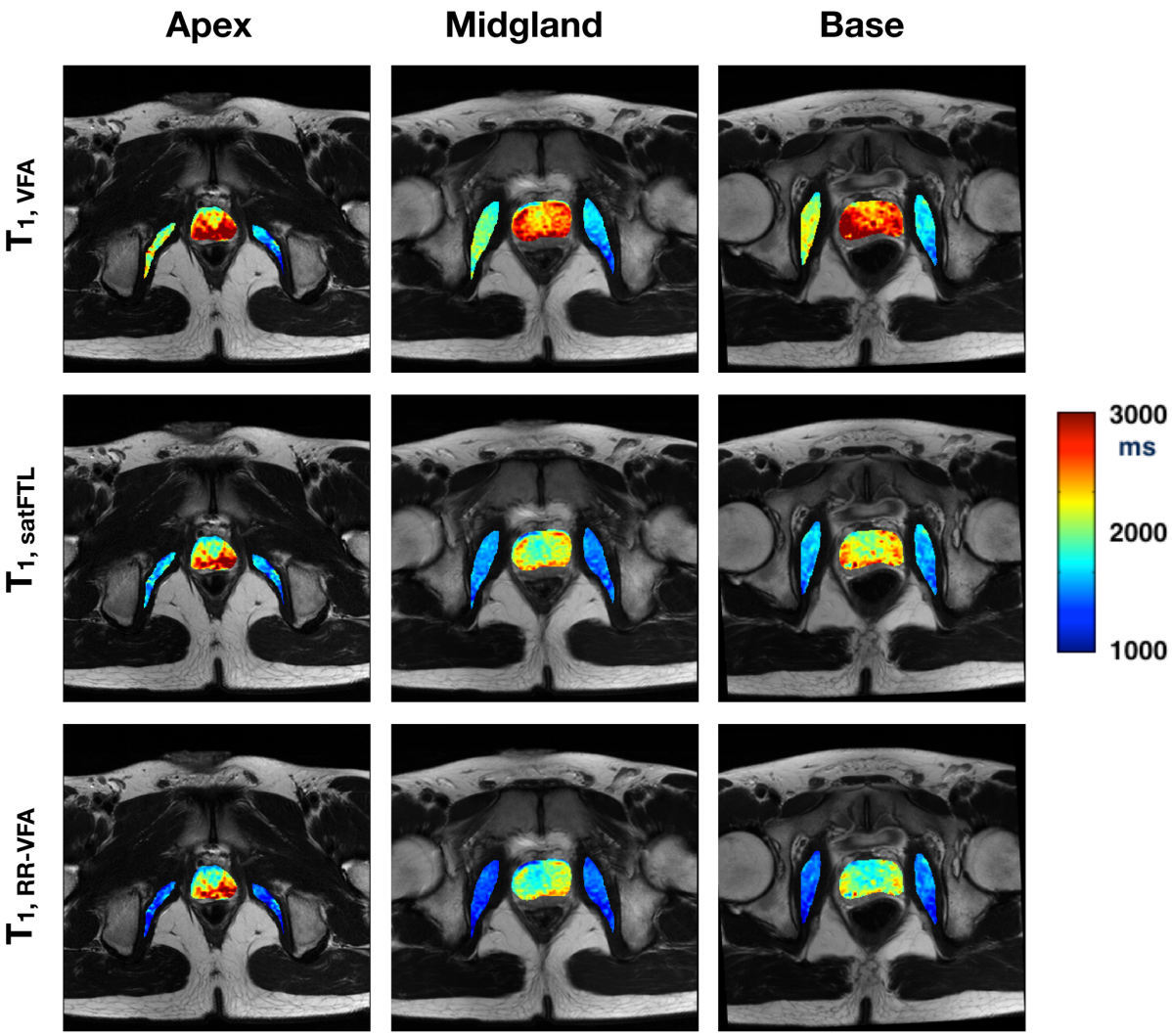

Figure 2-4 Representative slices (apex, midgland and base) for one scan with $T_{1,\,VFA}$, $T_{1,\,satFTL}$ as well as $T_{1,\,RR\text{-}VFA}$ within ROIs overlaid on T2w-TSE images. The inconsistent $T_{1,\,VFA}$ estimation between left and right pelvic muscles became more consistent in $T_{1,\,satTFL}$, and $T_{1,\,RR\text{-}VFA}$.

Figure 2-5 shows the Bland-Altman plots for intra-scanner (a, c, e) and inter-scanner comparison (b, d, f) of average T_1 values within ROIs. Figure 2-5 (a, c, e) showed the comparison

between two scans on Scanner 1. Note that $T_{1,\,VFA}$ had a broader range of mean T_1 for similar tissues among different volunteers (Figure 2-5a), and the T_1 value became much more consistent after compensating for B_1^+ variation using both B_1^+ mapping techniques (Figure 2-5c, e). The mean T_1 among volunteers converged to a more consistent value using both B_1^+ correction techniques, and the 95% LOA was narrower for $T_{1,satTFL}$ (510ms) and $T_{1,RR\text{-}VFA}$ (460ms) compared with $T_{1,VFA}$ (720ms). Figure 2-5 (b, d, f) showed the Bland-Altman plots for comparison between Scan 1 on Scanner 1 and Scan 1 on Scanner 2, and the conclusion is similar, $T_{1,satTFL}$ (LOA 680ms) and $T_{1,RR\text{-}VFA}$ (LOA 610ms) showed more consistent T_1 for each tissue compared to $T_{1,VFA}$ (LOA 1140ms).

Moreover, $T_{1,\text{RR-VFA}}$ achieved similar agreement with $T_{1,\text{satTFL}}$.

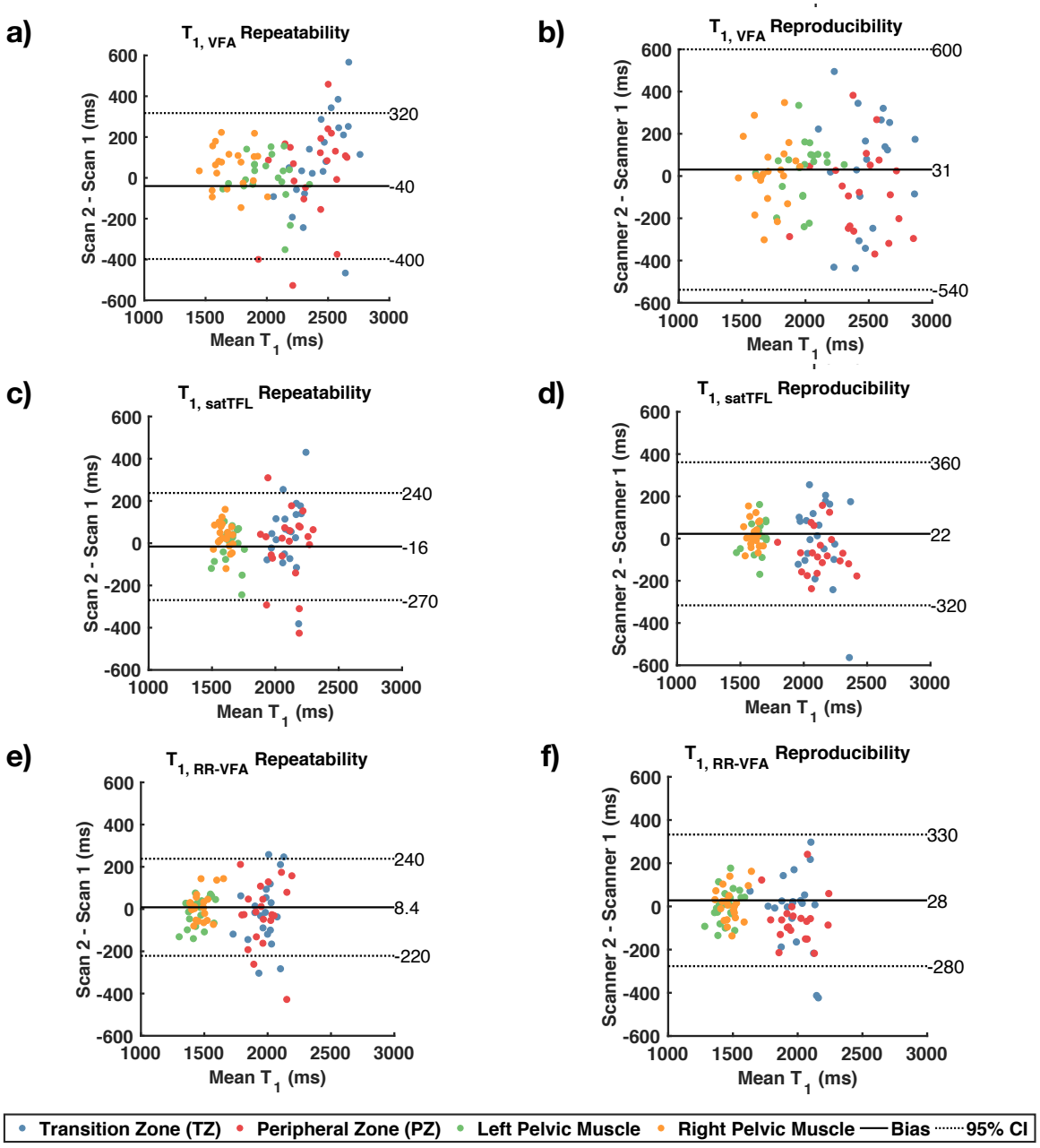

Figure 2-5 Bland-Altman Plot for comparing mean T_1 within ROIs between two scans on Scanner 1 (a, c, e) and between Scan 1 on Scanner 1 and Scanner 2 (b, d, f). Each encoded color indicates one tissue. The mean T_1 for each tissue had a more condensed distribution for $T_{1,\text{satTFL}}$ (c, d) and $T_{1,\text{RR-VFA}}$ (e, f) compared to $T_{1,\text{VFA}}$ (a, b).

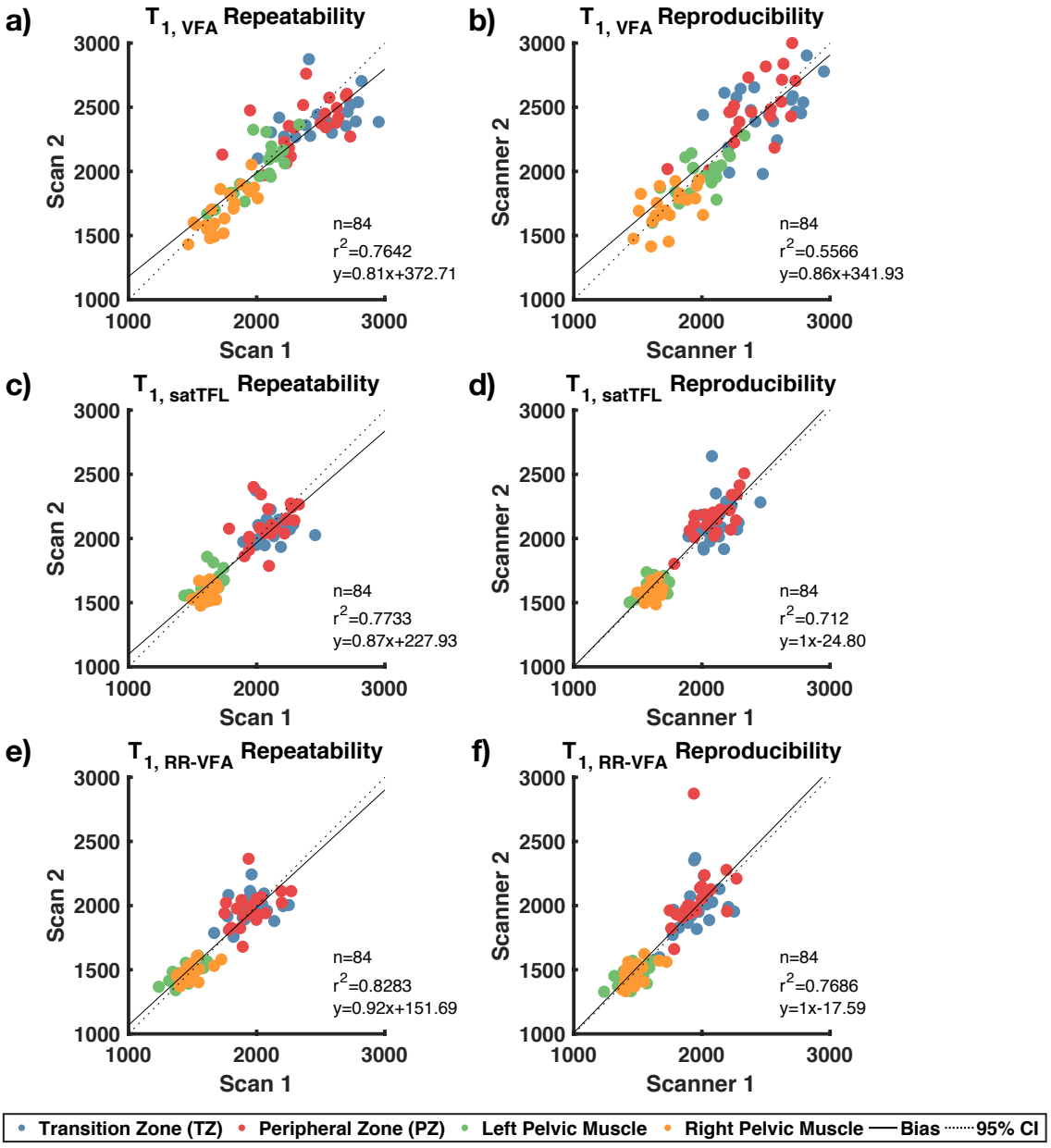

Figure 2-6 Linear regression plots for comparing mean T_1 within ROIs between two scans on Scanner 1 (a ,c ,e) and between Scan 1 on Scanner 1 and Scanner 2 (b, d, f). Each encoded color indicates one tissue.

Statistical analysis results are summarized in Table 2-2. Based on the averaged CCC evaluation of intra- and inter-scan pairs in Table 2-2, all correlations for each T_1 data were excellent. However, $T_{1,\text{RR-VFA}}$ gave a generally higher correlation (0.89 and 0.87) than $T_{1,\text{satTFL}}$ (0.87 and 0.83) and $T_{1,}$

$_{VFA}$ (0.84 and 0.74), showing higher T_1 correlation both intra- and inter- scanner. P-values from z-test of CCCs are also reported in Table 2-2. The reproducibility between $T_{1,VFA}$ and $T_{1,RR\text{-}VFA}$ showed statistically significant difference ($p = 0.015$). CCC of other pairs of comparison did not show significant difference due to our limited sample size. Table 2-1 contained the intra-scanner (left column) and inter-scanner (right column) comparison of mean T_1, mean difference, RC and RDC after bootstrapping for each tissue respectively. $T_{1,\,VFA}$ had inconsistent mean T_1 between left and right pelvic muscles, wider difference between measurements, higher RC and RDC compared to both corrected T_1.

Table 2-1 Summary of mean T_1, mean difference, 95% CI of the difference and RC after bootstrapping for each tissue repeatability and reproducibility. Metrics were averaged among each pair of comparisons (two pairs for intra-scanner and four pairs for inter-scanner) respectively.

		Repeatability			Reproducibility		
		Mean (ms)	Mean Difference (95% CI) (ms)	RC (ms)	Mean (ms)	Mean Difference (95% CI) (ms)	RDC (ms)
TZ	$T_{1,VFA}$	2461	-55 [-170, 61]	501	2462	47 [-96, 192]	616
	$T_{1,satTFL}$	2107	-34 [-113, 44]	339	2106	15 [-69, 101]	368
	$T_{1,RR\text{-}VFA}$	1978	1 [-75, 79]	329	1978	6 [-84, 97]	390
PZ	$T_{1,VFA}$	2448	-38 [-164, 93]	556	2448	145 [-15, 301]	675
	$T_{1,satTFL}$	2148	-11 [-105, 84]	405	2148	96 [-3, 191]	422
	$T_{1,RR\text{-}VFA}$	2004	-8 [-100, 89]	406	2004	75 [-8, 158]	357
Left Pelvic Muscle	$T_{1,VFA}$	1986	5 [-53, 65]	252	1987	-25 [-107, 58]	354
	$T_{1,satTFL}$	1638	4 [-29, 38]	145	1638	-13 [-51, 25]	164
	$T_{1,RR\text{-}VFA}$	1448	-1 [-29, 27]	122	1448	-12 [-43, 17]	131
Right Pelvic Muscle	$T_{1,VFA}$	1703	-44 [-101, 10]	238	1702	-11 [-75, 53]	277
	$T_{1,satTFL}$	1593	-27 [-58, 3]	131	1593	-11 [-40, 18]	124
	$T_{1,RR\text{-}VFA}$	1472	-16 [-43, 10]	112	1472	-12 [-44, 22]	141

2.5 Discussion

Repeatable and reproducible T_1 estimation is the foundation for accurate quantitative DCE

Table 2-2 CCC between pairs of scans for intra-scanner repeatability and inter-scanner reproducibility. CCCs were averaged among each pair of comparisons (two pairs for intra-scanner and four pairs for inter-scanner) respectively. P-value from z-test of CCCs for repeatability and reproducibility comparison (after Bonferroni adjustment statistically significant difference was marked with *).

Average CCC	Repeatability Comparison	Reproducibility Comparison
$T_{1,VFA}$	0.84	0.74
$T_{1,satTFL}$	0.87	0.83
$T_{1,RR\text{-}VFA}$	0.89	0.87
p-value	**Repeatability Comparison**	**Reproducibility Comparison**
$T_{1,VFA}$ vs. $T_{1,satTFL}$	0.478	0.131
$T_{1,VFA}$ vs. $T_{1,RR\text{-}VFA}$	0.201	0.015*
$T_{1,satTFL}$ vs. $T_{1,RR\text{-}VFA}$	0.569	0.358

MRI. In this study, we mainly compared the repeatability and reproducibility of T_1 estimation regarding different B_1^+ estimation techniques. When evaluating repeatability within each scanner, since the time interval between two scans is minimal, the main influence factor of the fluctuation is from the change of B_1^+ pattern resulting from the change of patient location relative to the isocenter. On the other hand, the inter-scanner reproducibility integrates more factors such as various scanning intervals, different hardware properties on different scanners as well as patient location relative to isocenter. All intra-scanner CCCs were higher than inter-scanner CCCs, which are as expected. In general, both $T_{1,satTFL}$ and $T_{1,RR\text{-}VFA}$ showed great repeatability and reproducibility under all of the conditions.

We mainly evaluated the variable flip angle T_1 estimation in male pelvis, and we expect

similar behavior in other organs. Quantitative DCE MRI is not only useful in PCa diagnosis. As minimal invasive treatment has been playing a more important role, highly repeatable and reproducible quantitative DCE MRI has great potential to help monitor the treatment response as well.

We also compared the B_1^+ estimation between two B_1^+ mapping techniques directly, and the linear regression plots of two methods showed good linear correlations with slopes slightly larger than one. This could be the main reason of different mean T_1 values between $T_{1,\,satTFL}$ and $T_{1,\,RR\text{-}VFA}$. We believe this may be due to the imperfect slice profile correction in satTFL as the technique is based on multi-slice 2D imaging and the bias was consistent across different scans and scanners. Further investigation may be needed to fully eliminate the inconsistencies between 2D and 3D imaging. In addition, 1000-1300ms prostate T_1 was reported in previous studies using inversion recovery approach, and we observed ~2000ms T_1 using VFA at 3.0T. The overestimation is expected due to both field difference and pulse sequence variation. For example, a 30% overestimation of white matter T_1 from VFA compared to inversion recovery was reported.[63]

In this study, we used satTFL B_1^+ as a reference to compare the repeatability and reproducibility because it is the only commercially available B_1^+ mapping technique on Siemens MRI systems. On the other hand, RR-VFA is widely available across scanners by different vendors as it does not require additional B_1^+ mapping sequences, and we expect the comparable repeatability and reproducibility would not be different for other scanners. Other advantages of using RR-VFA include no extra scan time and no inconsistencies between B_1^+ and T_1 mapping sequences (e.g., 2D/3D imaging, and in-plane and through-plane resolution). As a result, RR-VFA may provide a good practical solution to accurate pre-contrast T_1 estimation for prostate quantitative DCE MRI.

Our study has the following two major limitations: small cohort size and noise introduced by motion. The scan was performed on two scanners and repeated on two systems on twenty-one volunteers, resulting in 84 scans in total. Based on power analysis, to achieve 80% power, we need 390, 1250, 2000 volunteers respectively to detect the repeatability difference between $T_{1,VFA}$ vs. $T_{1,RR\text{-}VFA}$, $T_{1,VFA}$ vs. $T_{1,satTFL}$ and $T_{1,satTFL}$ vs. $T_{1,RR\text{-}VFA}$ with alpha = 0.05. Similarly, we need 110, 280, 750 volunteers respectively to detect the reproducibility difference between $T_{1,VFA}$ vs. $T_{1,RR\text{-}VFA}$, $T_{1,VFA}$ vs. $T_{1,satTFL}$ and $T_{1,satTFL}$ vs. $T_{1,RR\text{-}VFA}$ to achieve 80% power with alpha = 0.05. Even we could only detect statistically significant difference between the reproducibility between $T_{1,VFA}$ and $T_{1,RR\text{-}VFA}$ with current sample size, the conclusion is not expected to be changed with a larger cohort. Also, we observed several scans had motion-induced artifact in T_1 maps, especially on the edge. We excluded the edge when creating the contour to reduce the influence from motion, but the motion artifact could hardly be fully removed. This indicates that the repeatability and reproducibility was slightly underestimated. Nonetheless, the comparison between $T_{1,VFA}$, $T_{1,satTFL}$ and $T_{1,RR\text{-}VFA}$ was fair regardless motion-induced noise exist or not.

2.6 Conclusion

We showed improved repeatability and reproducibility of VFA T_1 estimation with compensating for B_1^+ variation in the prostate MRI at 3T. Both RR-VFA and satTFL B_1^+ mapping methods had comparable repeatability and reproducibility of T_1 estimation while RR-VFA does not require additional scan time nor a B_1^+ mapping sequence. The application of RR-VFA T_1 estimation has great potential to improve T_1 quantification resulting in more repeatable and reproducible prostate DCE-MRI quantification.

This work has been published as:

Zhong X, Liu D, Sayre J, Wu HH, Sung K, Repeatability and Reproducibility of Variable Flip Angle T1 Quantification in the Prostate at 3 Tesla. Journal of Magnetic Resonance Imaging. 2018. https://doi.org/10.1002/jmri.26596

3 PROSTATE DCE-MRI WITH B_1^+ CORRECTION USING AN APPROXIMATED ANALYTICAL APPROACH

3.1 Introduction

PCa is one of the leading causes of cancer deaths for men in the United States[1]. Biopsy is one of the current gold standards for diagnosing PCa, however, it is invasive and has a relatively low specificity[64]. Mp-MRI, which includes DCE-MRI is now widely used as a promising non-invasive technique for diagnosing PCa[18–20]. Conventional image analyses for DCE-MRI are typically based on qualitative analyses of signal uptake, where the subjective evaluation or qualitative analyses are limited by inter-observer variability and high dependence on data acquisition[21].

Quantitative DCE-MRI has shown great potential in tumor detection, staging, and treatment response evaluation[16,24,28]. Quantitative analysis of DCE-MRI usually requires modeling to generate contrast concentration curves in the tissue and then employs pharmacokinetic (PK) analysis to estimate parameters such as volume transfer constant (K^{trans}) and extravascular extracellular volume fraction (v_e)[65]. To calculate an accurate contrast agent concentration curve, pre-contrast T_1 (T_{10}) maps need to be estimated. A variable flip angle (VFA) method is commonly used for T_{10} estimation[49,50]. However, the VFA image acquisition is sensitive to flip angle variation

caused by transmit radiofrequency (B_1^+) field inhomogeneity[51,53,66].

Increased signal-to-noise (SNR) from 3T MRI systems can improve the quantification accuracy, thus becoming preferable for prostate mp-MRI[22]. However, B_1^+ field inhomogeneity becomes more severe with increased field strengths ($\geq$ 3.0T)[67]. If B_1^+ field is inhomogeneous, the spins within excitation can fail to achieve the exact flip angle as prescribed, thus reducing the accuracy of the quantitative analysis of DCE-MRI. Previous studies have shown an inter-subject B_1^+ variation of 32% in the prostate[31] at 3T, which can induce significant errors into the PK estimation. Di Giovanni et al. showed that the 55% overestimation of flip angle due to B_1^+ inhomogeneity could result in up to 66% underestimation for measured K^{trans} and 77% underestimation of v_e[68].

Various B_1^+ mapping techniques have been developed to enable B_1^+ compensation, including the double-angle method (DAM)[32], Block-Siegert (BS)[38], actual flip angle imaging (AFI)[36] and reference region variable flip angle (RR-VFA)[57]. However, even if B_1^+ maps are available, applying the B_1^+ correction to quantitative DCE-MRI analysis is sometimes difficult due to practical limitations when closed-form software is used. Using B_1^+-corrected flip angles, the B_1^+ correction requires a full numerical reprocessing of the entire DCE-MRI modeling, from signal intensity to PK parameters. This numerical reprocessing can be challenging especially when closed-source software is used for DCE-MRI analysis[40] and can be time-consuming due to the pixel-by-pixel reprocessing. Especially for clinical or clinical research settings, simple yet efficient B_1^+ correction approaches will be highly desirable since closed-form or commercial software is commonly used.

In this work, we present a simplified and practical approach that compensates for B_1^+ inhomogeneity in quantitative prostate DCE-MRI analysis. Our proposed approximated analytical

approach enables a simple and practical application of B_1^+ correction in quantitative DCE-MRI since it does not require full access to the entire DCE-MRI analysis and avoids repeated pixel-by-pixel PK parameter estimation. The accuracy of the approximated analytical approach was evaluated using numerical simulation and prostate-specific digital reference object (DRO). The approximated analytical approach was also compared with reference numerical correction[40] on 82 in-vivo 3T prostate DCE-MRI cases.

3.2 Theory

3.2.1 Quantitative Analysis for prostate DCE-MRI

As described in the theory session in Chapter 2, the pre-contrast T_1 (T_{10}) can be estimated using VFA method. Once T_{10} is estimated, the dynamic T_1-map, $T_1(t)$, generated by using SPGR sequence with the flip angle β, can be computed by using T_{10} and the normalized signal intensity $S(T_1(t))/S(T_{10})$, where $S(T_{10})$ is the pre-contrast baseline signal intensity and $S(T_1(t))$ is the dynamic contrast-enhanced signal intensity, as shown in Equation 3-1.

Equation 3-1

$$\Xi = \frac{\boldsymbol{S(T_1(t)) - S(T_{10})}}{\boldsymbol{S(T_{10})}} = \frac{\boldsymbol{(E_1(t) - E_{1,0})(cos\beta - 1)}}{\boldsymbol{(E_{1,0} - 1)(E_1(t)cos\beta - 1)}}$$

To determine the PK parameters, the tissue contrast agent concentration, C(t), needs to be calculated from $T_1(t)$ and T_{10}. C(t) is proportional to the change of longitudinal relaxation rate and can be computed by

Equation 3-2

$$\boldsymbol{C(t) = \frac{1}{r_1}\left(\frac{1}{T_1(t)} - \frac{1}{T_{10}}\right)}$$

where r_1 is the T_1 relaxivity related to the contrast agent. Once C(t) is estimated, PK modeling such as standard Tofts model[58] can be applied to estimate PK parameters (K^{trans} and v_e) using nonlinear curve fitting:

Equation 3-3

$$\boldsymbol{C(t) = K^{trans}\int_0^t C_p(\tau)e^{-\frac{K^{trans}}{v_e}(t-\tau)}d\tau}$$

where $C_p(t)$ is the contrast agent concentrations in the plasma (or arterial input function, AIF), K^{trans} is the rate constant from plasma to extravascular-extracellular space (EES), and v_e is the fractional volume constant of the EES. $C_p(t)$ can be either measured or pre-assumed[21].

When there exists B_1^+ inhomogeneity, the prescribed flip angles (α_i and β) are not the same as the actual flip angle, which causes errors in the measurement of PK parameters. B_1^+ estimation is needed to accurately determine α_i and β for each pixel.

3.2.2 B_1^+ Correction: Reference Numerical Approach

The conventional approach that compensates for B_1^+ inhomogeneity numerically reprocesses the whole quantitative DCE-MRI analysis with B_1^+-corrected flip angles[40]. For a given B_1^+ mapping technique, a pixel-by-pixel relative flip angle (defined as $k = \frac{\alpha_i'}{\alpha_i} = \frac{\beta'}{\beta}$, where α_i and β are the prescribed flip angles, and α_i' and β' are the actual flip angles), is determined. The B_1^+-corrected flip angle, assumed to be the actual flip angle, can be simply computed by multiplying k and the prescribed flip angle. The whole DCE-MRI analysis needs to be reprocessed for each pixel using the B_1^+-corrected flip angles. As a note, all variables with prime (') indicate B_1^+-

corrected variables. When α_i' is determined, then it is used in Equation 2-2 to estimate T_{10}', and β' has to be used in Equation 3-1 to estimate T_1'(t). Once T_{10}' and T_1'(t) are computed, both K^{trans}' and v_e' can be computed by $C'(t)$ from Equation 3-2 and Equation 3-3[40].

This numerical approach is well defined, but can be demanding since it requires full access to the MRI modeling, PK modeling and raw DCE-MRI images. Many commercial and closed-source software do not include pixel-by-pixel B_1^+ correction, nor can we modify the software to perform B_1^+ correction with the numerical approach. Moreover, even when the numerical approach is possible, the B_1^+ correction would need to repeat the pixel-by-pixel estimation of PK parameters, which can be time-consuming, especially for volumetric PK maps.

3.2.3 B_1^+ Correction: Approximated Analytical Approach

An analytical approach is desirable in many clinical and research settings as it allows direct derivation of B_1^+-corrected PK parameters. Analytical correction does not require full access to the DCE modeling, nor raw DCE-MRI images, and can enable a more practical B_1^+ correction process by only using B_1^+ maps and uncorrected PK maps as input. However, the full analytical expression of the B_1^+-corrected PK parameters is highly complicated to derive due to multiple nonlinear processes, as described before in Equation 2-2 to Equation 3-3. Here, we describe an approximated analytical approach to derive B_1^+-corrected PK parameters (K^{trans}' and v_e') by approximating the full analytical expression with certain assumptions. This approximated analytical approach will improve the utility of B_1^+ correction in DCE-MRI in various settings with minimal approximation error[69].

In the approximated analytical approach, we assume that the flip angles and TR/T_1 are small ($\alpha_i^3 \approx 0$, $\beta \approx 0$, $TR/T_{10} \approx 0$, and $TR/T_1(t) \approx 0$), and k is close to 1 ($k \approx 1$). Using a Taylor

series approximation on Equation 2-2 and Equation 3-1, T_{10}' and T_1'(t) can be simply expressed as $T'_{10} \approx \frac{1}{k^2} T_{10}$ and $T'_1(t) \approx \frac{1}{k^2} T_1(t)$. For simplicity, we used two flip angles for VFA process in the analytical derivation. Based on Equation 3-2, the corrected contrast agent concentration curve $C'(t)=k^2 C(t)$ can be derived. As a result, the B_1^+-corrected PK parameters can be approximated as $K^{trans\prime} \approx k^2 K^{trans}$ and $v'_e \approx k^2 v_e$ from Equation 3-3. The full derivation of the approximated approach can be found in Chapter 6. Using the approximated derivation of B_1^+-corrected PK parameters from uncorrected PK parameters allows for direct compensation for B_1^+ inhomogeneity without fully accessing MRI modeling and PK modeling because the relationship does not change regardless of corrected T_{10}, K^{trans} and v_e value. Figure 3-1 shows the comparison between the proposed approximate analytical correction method and the conventional numerical correction method. The conventional correction method requires the access to the images and MRI modeling and pharmacokinetic modeling while the approximated analytical correction method only requires the uncorrected K^{trans} and v_e maps as the input.

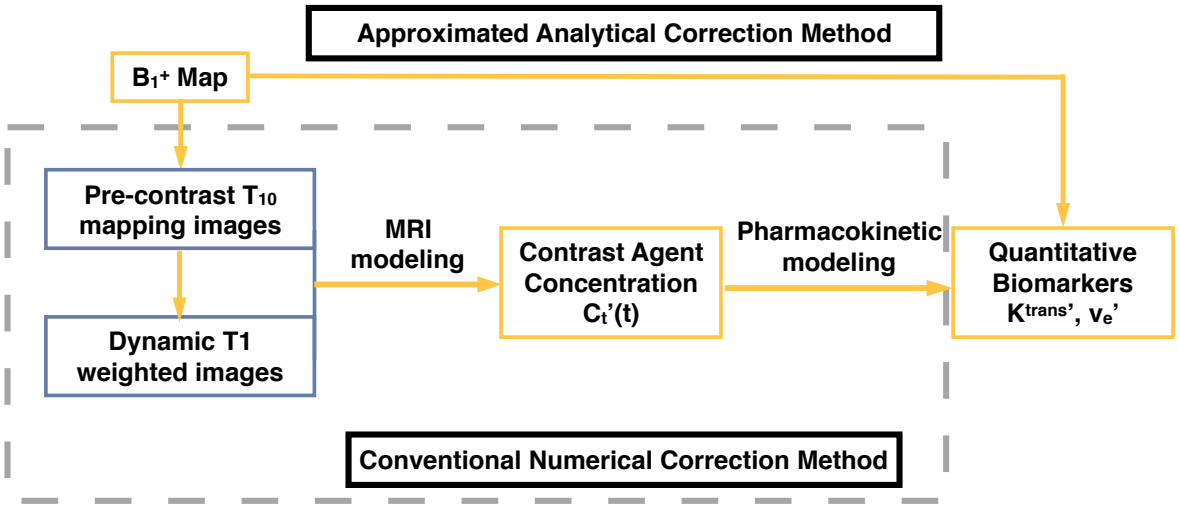

Figure 3-1 Comparison between the two B correction approaches. Conventional correction method needs to repeat MRI modeling and PK modeling pixel by pixel with original T_1 images and dynamic T_1 weighted images, while approximated analytical correction method only requires the uncorrected PK parameters to perform the correction and simplifies the computation.

3.3 Methods

Our approximated analytical approach relies on a set of assumptions, including small flip angles, small TR/T_1, and k close to one. We first evaluated the approximation by numerical simulation and DRO and then compared the approximated analytical approach with the conventional numerical correction using 82 in-vivo prostate DCE-MRI cases based on our standard clinical prostate DCE-MRI protocol.

3.3.1 Prostate Digital Reference Object (DRO)

We used the numerical simulation and prostate DRO[70–72] to carefully separate each source of errors (e.g. noise and B_1^+ inhomogeneity), providing a more systematic way to evaluate B_1^+ correction approaches. The DRO was composed of simulated grid-based MRI images with a set of pre-assumed PK parameters, including VFA images as well as dynamic images. We assumed a certain set of PK parameters, K^{trans} (ranged from 0.01 to 0.35 min^{-1}) and v_e (ranged from 0.01 to 0.5), defined as a ground truth set P_{nat}, and generated C(t) based on the set. The detailed sequence parameters are shown in Table 3-1, and are derived from our clinical prostate DCE-MRI protocol. We then created signals in both VFA imaging and dynamic T_1-weighted images based on actual flip angles. The B_1^+-induced uncorrected flip angles were created by applying various k (ranged from 0.7 to 1.3) to the actual flip angle, and these uncorrected flip angles were used in the following model fitting from signal intensity to pharmacokinetic parameters (Equation 2-2 to Equation 3-3). The calculated parameters were defined as uncorrected parameters P_{uncor}. Both the numerical and approximated analytical correction methods were applied on the simulated signals to generate the

corrected parameters ($P_{cor,N}$ and $P_{cor,A}$).

All signal simulations and fittings were done using Matlab (Mathworks, Inc., Natick, MA, USA), and the trust-region reflective algorithm[73] with lower bound of zero for K^{trans} and v_e was used in the nonlinear fitting process in PK modeling. In the following analysis, the percentage error relative to ground truth, $P_{nat,}$ was calculated as the evaluation metric. The B_1^+-induced error ($E_{B1,DRO}$) was defined as $\frac{|P_{uncor}-P_{nat}|}{P_{nat}} \times 100\%$ and the correction residual errors for $P_{cor,N}$ and $P_{cor,A}$ were defined by $E_{N,DRO} = \frac{|P_{cor,N}-P_{nat}|}{P_{nat}} \times 100\%$ and $E_{A,DRO} = \frac{|P_{cor,A}-P_{nat}|}{P_{nat}} \times 100\%$.

Table 3-1 Details of the DRO modification. Compared to QIBA DRO designed for brain DCE MRI, the proposed DRO modified the flip angles, repetition time and the size of the DRO.

	QIBA DRO	Prostate DRO
K^{trans}	0.01, 0.02, 0.05, 0.1, 0.2, 0.35 min^{-1}	
v_e	0.01, 0.05, 0.1, 0.2, 0.5	
Relative B_1^+ (k = θ'/θ)	1.2	
T_{10}	1000 ms	
VFA Flip Angle	3, 6, 9, 15, 24, 35°	2, 5, 10, 15°
Dynamic Flip Angle	25°	12°
Repetition Time	5 ms	4.17 ms
Patch Size	10×10 pixels	2×2 pixels

To assess the bias and variance of the percentage errors for B_1^+ correction under a certain prostate-like DCE-MRI conditions, we first calculated B_1^+-induced and correction residual errors within a realistic range for B_1^+ inhomogeneity in the prostate[31]. The numerical simulation included

100 points with uniformly distributed k between 0.7 – 1.3 for one representative combination of K^{trans} = 0.05 min^{-1}, v_e = 0.1 and T_{10} = 1000 ms. To further assess the B_1^+-induced errors with various K^{trans} and v_e, we created a prostate DRO, modified from the original DRO by Quantitative Imaging Biomarkers Alliance (QIBA)[74], using our clinical prostate DCE-MRI parameters (see Table 3-1). Other parameters of the prostate DRO are shown in Table 3-1. The DRO simulation, shown in Figure 3-2, was repeated by using three different widely-used population-based AIFs[29] (Parker[75], Weinmann[76] and Fritz-Hansen[77]) with the standard Tofts model.

Noise was added to both VFA images and dynamic images by $S' = \sqrt{(S + n_1)^2 + n_2^2}$, where S is the original signal intensity, and n_1 and n_2 are Gaussian noise with the mean 0 and standard deviation ranging from 5 to 150, resulting in a baseline SNR ranging from 7.8 to 234.5. PK maps with and without correction were calculated as shown in Figure 3-2 B). With each SNR, the process was repeated 25 times, resulting in 100 available samples for each K^{trans} and v_e combination. For fair comparison, estimation parameters (K^{trans} or v_e) larger than 1 were excluded as outliers[78]. $E_{N,DRO}$ and $E_{A,DRO}$ for each SNR from 3000 pixels (5×6×100), except for those outliers, were averaged to evaluate residual errors varying with SNRs. Here, $E_{N,\,DRO}$ provided an estimation of the error tolerance due to noise, compared with the approximation-induced error ($E_{A,}$

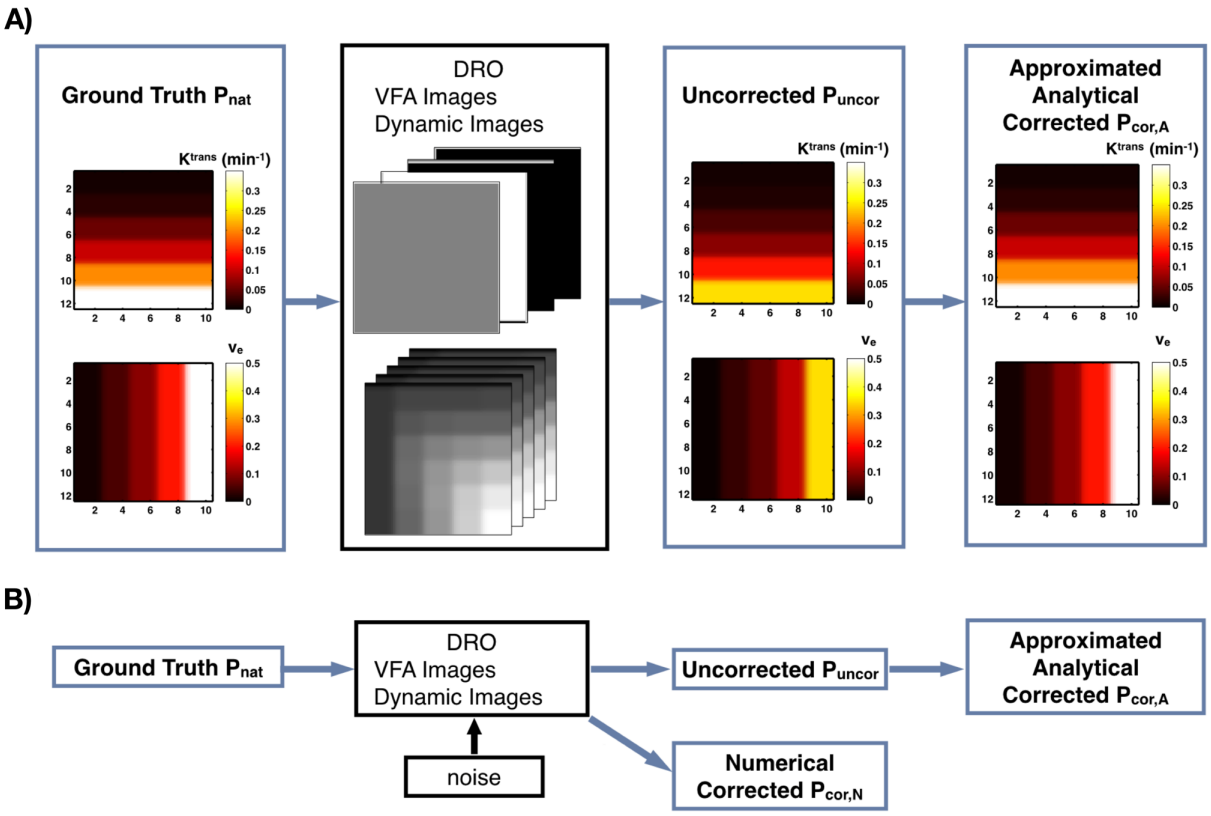

Figure 3-2 Summary of the simulation study design using DRO under various population-averaged AIFs (a) and Gaussian noise (b). The images in (a) are examples of corresponding K^{trans} and v_e maps in each step and DRO images. P represents PK parameters K^{trans} and v_e

$_{DRO}$ without noise). Moreover, $E_{N,\,DRO}$ and $E_{A,\,DRO}$ for each parameter combination were averaged separately to show the correction residual error distribution with P_{nat}. Linear regression and Bland-Altman plots were used to evaluate the correlation between PK parameters corrected by the two correction methods.

Similarly, to test the sensitivity of the two correction methods for k variation, we performed Monte-Carlo simulation with random Gaussian noise (zero mean and standard deviation ranging from 0.001 to 0.1) added in ground truth k = 1.2. The sensitivity was evaluated using $E_{A,DRO}$ and $E_{N,DRO}$.

3.3.2 In-vivo Prostate DCE-MRI Data

With the local Institutional Review Board (IRB) approval, 82 cases were used to evaluate the approximated analytical approach in-vivo. The 82 cases were acquired between June 2010 and September 2014 (age = 65.9 ± 6.9 years and mass = 81.9 ± 13.5 kg). All in-vivo DCE-MRI cases were performed on two 3T scanners (MAGNETOM Skyra and MAGNETOM Trio, Siemens Medical Systems), using a body array matrix and spine array coil. The 3D SPGR sequence was used in both VFA and dynamic imaging with a TR of 4.17ms. The slice thickness was 3.6mm, and the flip-angles used were 2, 5, 10, 15° for variable flip angle acquisition and 12° for dynamic acquisition. For most cases a matrix size of 160×160 with 20 slices was used, and those parameters varied slightly for other cases. For VFA imaging, a dual-echo bipolar readout (TE_1 = 1.23ms, TE_2 = 2.46ms) was used to generate the fat-only and water-only images using a two-point Dixon algorithm[79], and the B_1^+ maps were estimated using RR-VFA[31]. A single-dose injection of gadopentetate dimeglumine (Magnevist; Bayer, Wayne, NJ) contrast agent was administered to the patients at a dose of 0.1 mmol/kg through a peripheral vein at a rate of 2mL/sec using a mechanical injector, and dynamic images were acquired before, during, and after contrast injection. About 65 contrast-enhanced sets of images (temporal resolution of 4.3 sec) were acquired sequentially without delay between acquisitions with the total acquisition time of 5 min.

The standard Tofts model with Parker AIF[75] was used for the PK modeling. The fitting algorithm and constraint are the same as in DRO experiments. Prostate regions of interest (ROIs) were manually drawn on the five central slices in the contrast-enhanced images and were copied to other images such as B_1^+, B_1^+-corrected, and uncorrected PK maps. A representative example of the prostate ROI is shown in Figure 3-3. The evaluation was performed using percentage error with respect to the B_1^+-corrected parameters using the numerical approach. Specifically, the B_1^+-

induced error ($E_{B1,\text{ in-vivo}}$) was defined by $\frac{|P_{uncorr}-P_{cor,N}|}{P_{cor,N}} \times 100\%$, and the correction residual error

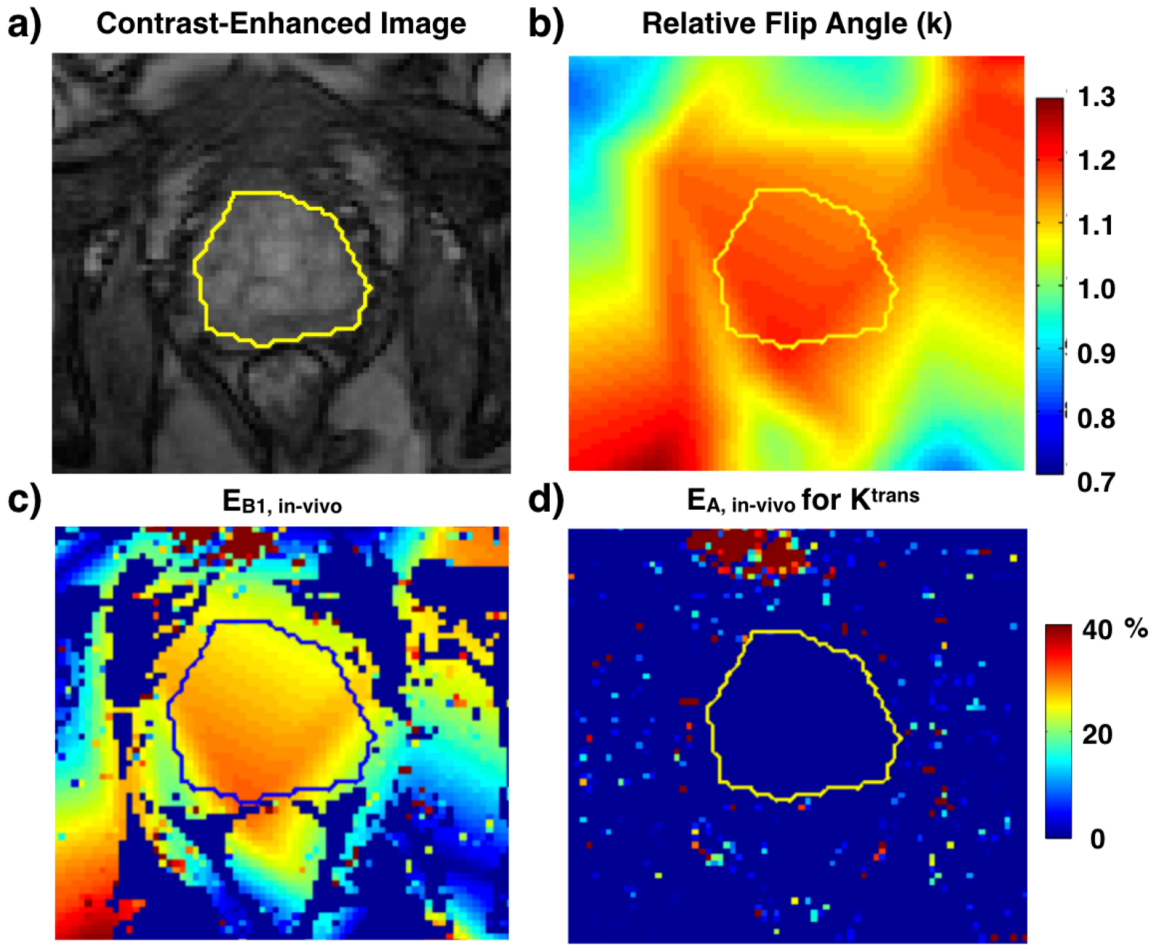

Figure 3-3 A representative slice of ROI positioning for in-vivo prostate data. ROI was first drawn on contrast-enhanced images (a) and was copied to corresponding relative B_1^+ (k) map (b) $E_{B1,\text{ in-vivo}}$ map (c) as well as $E_{A,\text{in-vivo}}$ for K^{trans}

($E_{A,\text{ in-vivo}}$) was defined by $\frac{|P_{cor,A}-P_{cor,N}|}{P_{cor,N}} \times 100\%$. The mean, standard deviation and 95% central range of all the voxels within the ROIs from all 82 patients were computed. An average k and the residual error for each patient's volumetric ROI from five central slices were also computed to evaluate the approximated analytical method among different cases. Any pixels with estimated v_e or K^{trans} larger than one were considered to be outliers[78] and therefore were excluded for all in-vivo experiments.

3.4 Results

3.4.1 Prostate DRO

Figure 3-4 shows that the numerical and approximated analytical methods are comparable in the numerical simulation with k variation. Within k range of 0.7 to 1.3, the maximum $E_{A,DRO}$ is less than 0.4% for K^{trans} and v_e. This is negligible compared to $E_{B1,\,DRO}$ (maximum of 104.1%). Figure 3-4 also describes how P_{uncor} deviates from the ground truth as k varies. For example, when k equals to 1.28, uncorrected K^{trans} and v_e underestimate around 40% of the true value. For all simulated points, $P_{cor,N}$ and P_{nat} are the same with a precision of 10^{-12} as expected, assuring the accuracy of the numerical correction method.

Despite K^{trans} and v_e variation, the $E_{A,\,DRO}$ is small and uniform for K^{trans} and v_e for all three AIFs based on the DRO simulation with a k of 1.2 (Figure 3-5). The maximum $E_{A,DRO}$ is 0.2% for K^{trans} estimation and 0.4% for v_e estimation, while the $E_{B1,\,DRO}$ of 30.7±0.1%. Overall, $E_{A,DRO}$ is almost negligible compared to $E_{B1,\,DRO}$ regardless of P_{nat} and the pre-assumed AIFs. By comparison, $E_{N,\,DRO}$ has a maximum of 0.2%. This further confirms the accuracy of the numerical correction method under the noise-free situation. These results are consistent with our expectation in the Theory section.

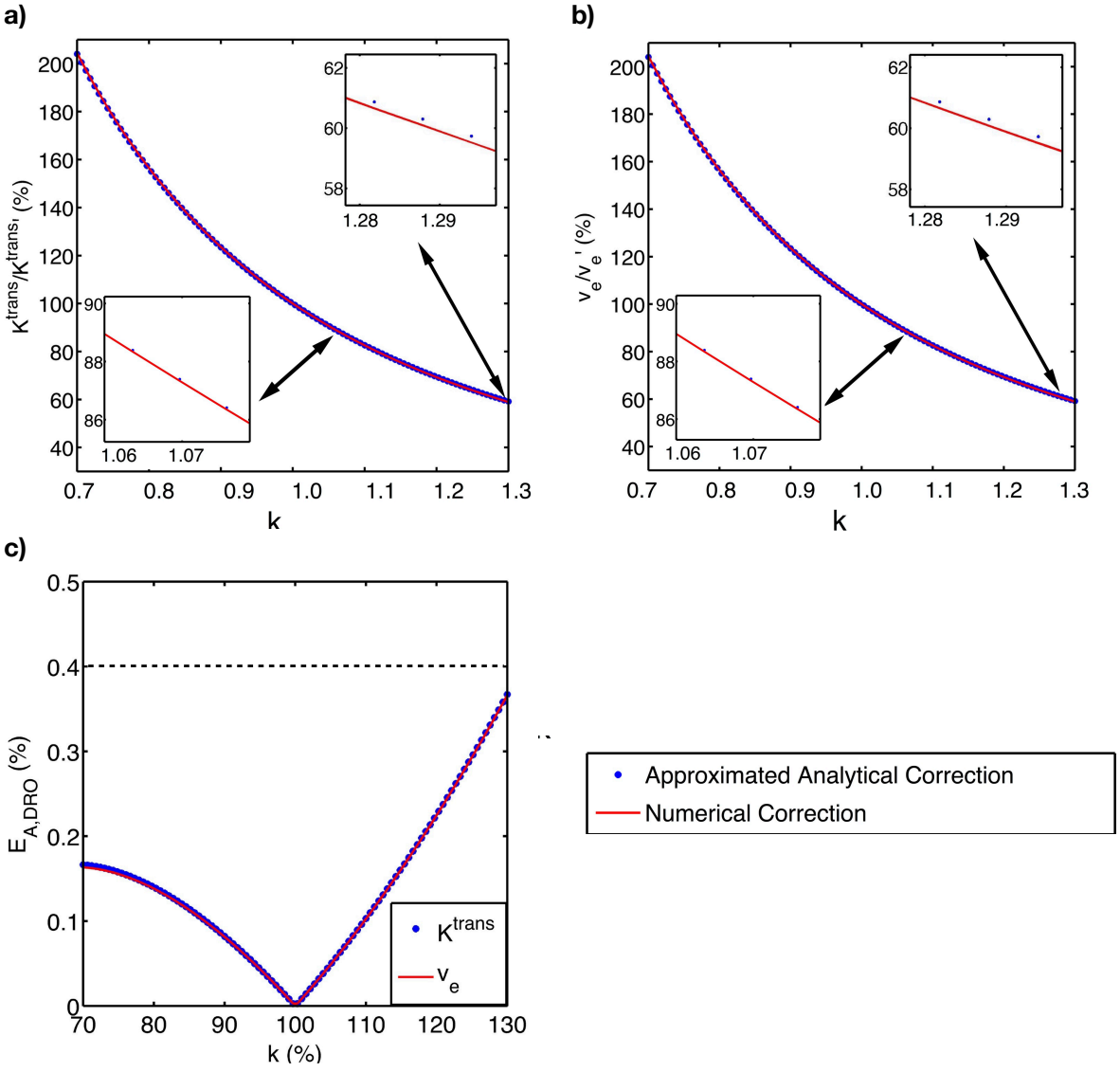

Figure 3-4 Comparison between numerical correction method and approximated analytical correction method in simulation with k ranging from 0.7 to 1.3 (ground truth $K^{trans'} = 0.05$ min^{-1}, $v_e = 0.1$, $T_{10} = 1000$ ms) for K^{trans} (a) and v_e (b) and $E_{A,DRO}$ for K^{trans} and v_e (c). Two example areas around k of 1.1 and 1.3 are zoomed. The difference between blue and red curves indicates $E_{A,DRO}$ (also shown in (c)), and the difference of y-axis and 100% indicate $E_{B1,DRO}$. $E_{A,DRO}$ is negligible compared to $E_{B1,DRO}$.

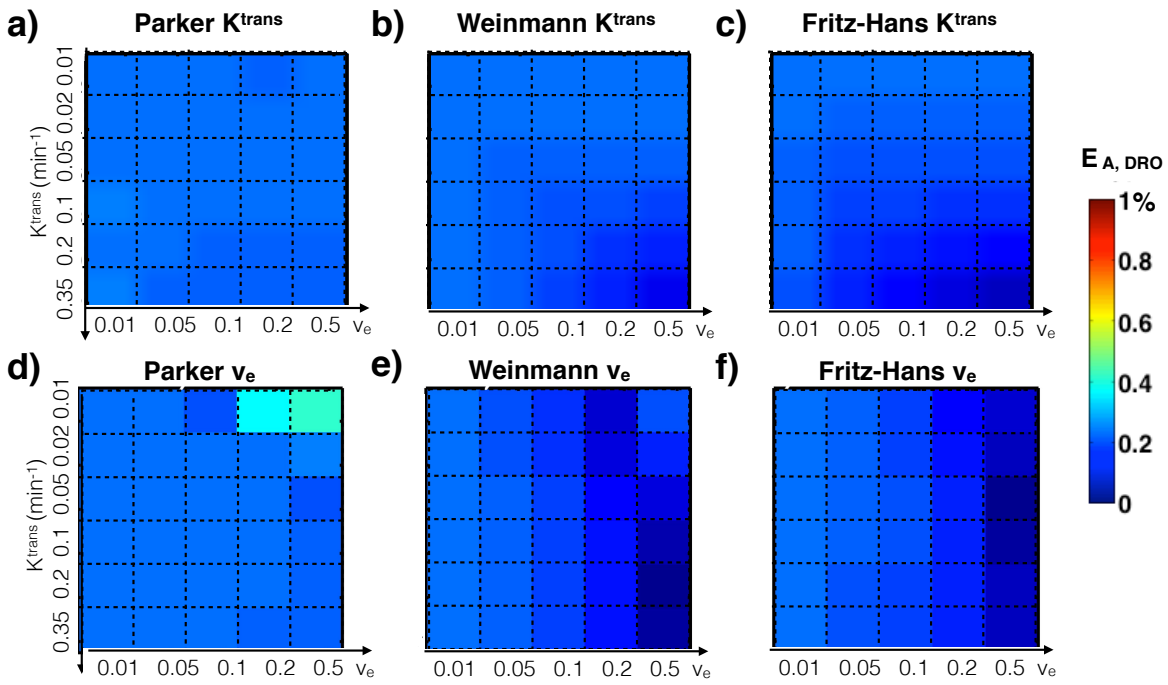

Figure 3-5 $E_{A,\,DRO}$ maps using three population-averaged AIFs for K^{trans} estimation (a-c) and for v_e estimation (d-f). The maximum residual error for K^{trans} is 0.2% and for v_e is 0.4%.

With various levels of noise added in DRO, a minimum $E_{N,\,DRO}$ of 2.1%±4.3% with baseline SNR of 234.5 was observed, which indicates the estimation uncertainty induced by noise. Based on previous simulation, the maximum $E_{A,\,DRO}$ without noise is 0.4%, which is much smaller than the minimum $E_{N,\,DRO}$ induced by minimal noise of standard deviation of 5. The overall residual errors against baseline SNR are shown in Figure 3-6. Across various baseline SNR (ranging from 7.8 to 234.5), the difference between $E_{N,\,DRO}$ and $E_{A,\,DRO}$ is minimal compared to $E_{N,\,DRO}$, which means that the approximation induced error is small compared to noise-induced error. For example, when baseline SNR is 41.2, the noise-induced error for numerical correction method is 21.9% for K^{trans} and 14.4% for v_e. However, the difference of the mean error of the analytical correction method from mean error of numerical correction method is 4.3% for K^{trans} and -0.1% for v_e. Those results indicate that under various noise levels, the analytical correction

method provides similar performance as numerical correction method.

More specifically, with an example baseline SNR of 38.5 when n_1 and n_2 had standard

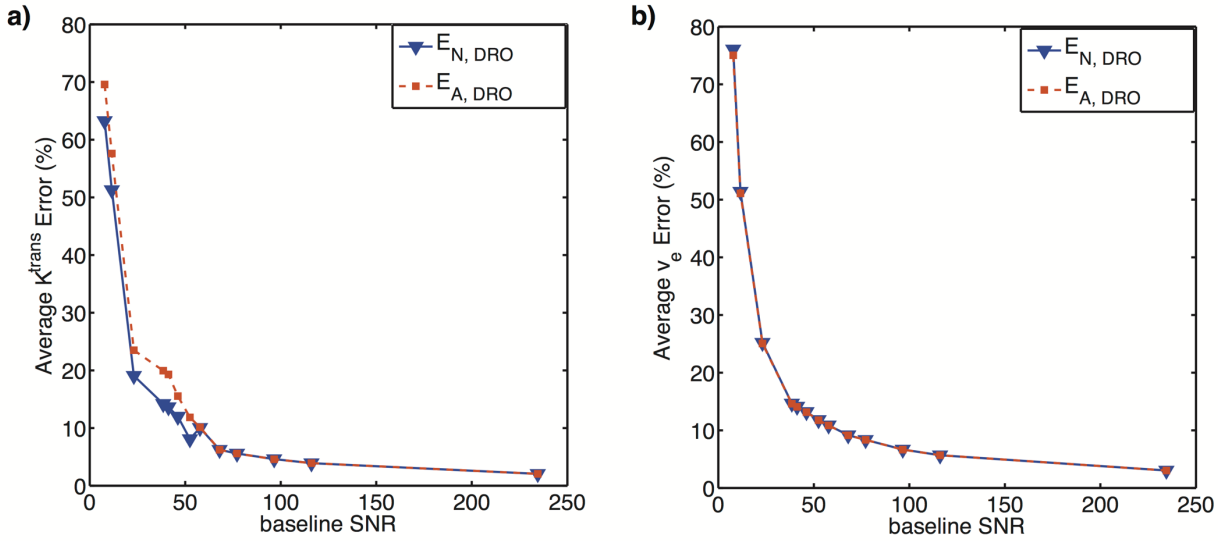

Figure 3-6 Comparison of correction residual percentage errors between two correction methods ($E_{N,DRO}$ and $E_{A, DRO}$) for K^{trans} maps (a) and v_e maps (b) with various levels of noise added. There are 100 Monte-Carlo simulations for each PK parameter combination. For each SNR level, noise-induced errors for 3000 pixels (5 × 6 × 100) excluding outliers were averaged. Across all simulated baseline SNRs, the residual error for both correction methods are comparable to each other.

deviation of 30, Figure 3-6 displays the average of all Monte-Carlo experiments for each PK parameter combination (100 pixels) excluding outliers. Figure 3-7 indicates that both methods provide robust estimation except in extreme PK parameters. The large errors occur when v_e is small (v_e = 0.01) for K^{trans} estimation. Those areas with large correction residual errors are mainly because the curve characteristic is more sensitive to noise under those circumstances. The difference maps between the two correction residual error maps on the right column confirms that the inconsistency exists only under extreme situations. Both correction methods are not reliable under extreme situations. Additionally, comparison between the corrected PK parameters from two methods, with added noise, was performed using linear regression and Bland-Altman plots (as shown in Figure 3-8). With 100 times Monte-Carlo simulation for each PK parameters combination, most K^{trans} and v_e values are highly comparable between two methods. Pearson

correlation results also show that the approximated analytical method is comparable to the conventional numerical correction methods ($r^2 = 0.97$ for K^{trans} and $r^2 = 1.00$ for v_e).

With noise added in ground truth k, the difference between average $E_{N,DRO}$ and $E_{A,DRO}$ based on the Monte-Carlo simulation over each PK parameter was smaller than 0.1%. For example, with the noise standard deviation of 0.01, both $E_{N,DRO}$ and $E_{A,DRO}$ were almost identical (4.1 ± 3.5%) for both K^{trans} and v_e. The results show that the robustness to k measurement accuracy for

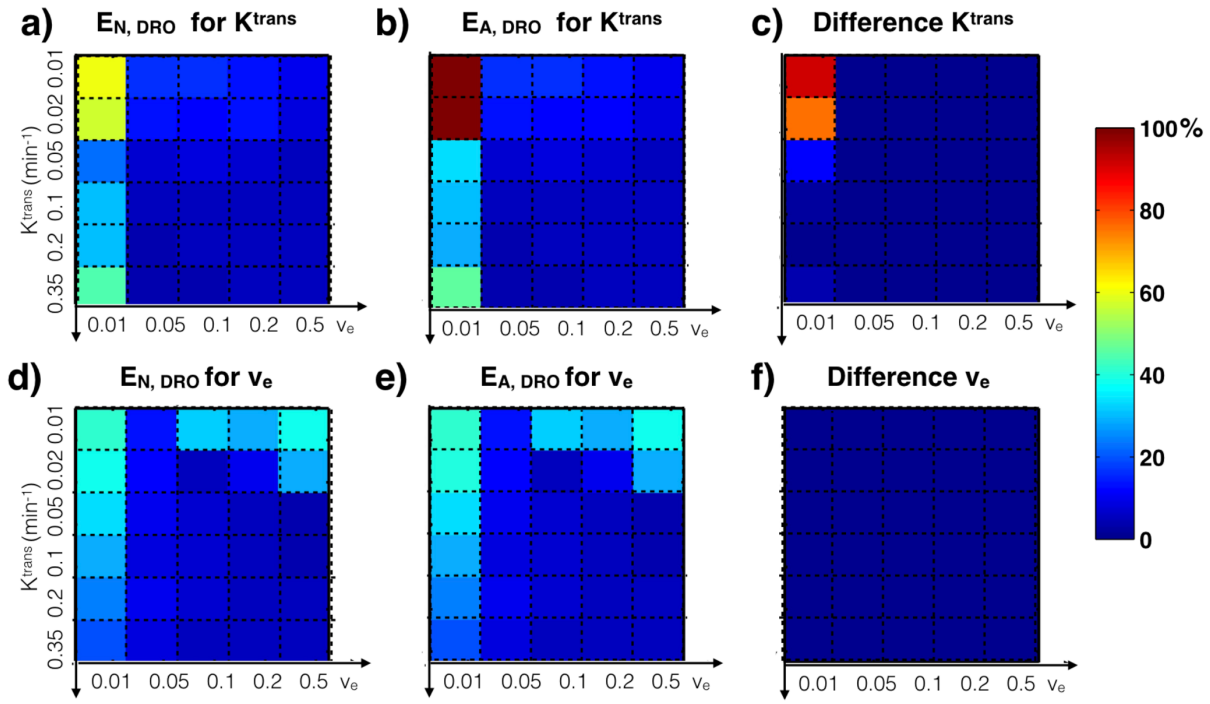

Figure 3-7 $E_{N,DRO}$ averaged for each parameter (100 pixels) (a, d) and $E_{A,DRO}$ averaged for each parameter (b, e) for K^{trans} and v_e maps with baseline SNR of 38.5. The error patterns are similar between the two methods. The absolute value difference maps ($E_{N,DRO} - E_{A,DRO}$) averaged for each parameter (c and f) indicate outliers appear when v_e is low, where the fitting process is more sensitive to noise.

those two correction methods is similar.

3.4.2 In-vivo Prostate DCE-MRI Data

Based on B_1^+ maps measured from 82 cases, mean k value from each subject gives a range

from 0.78 to 1.22 with projected 80% inter-subject B_1^+-induced error difference based on our

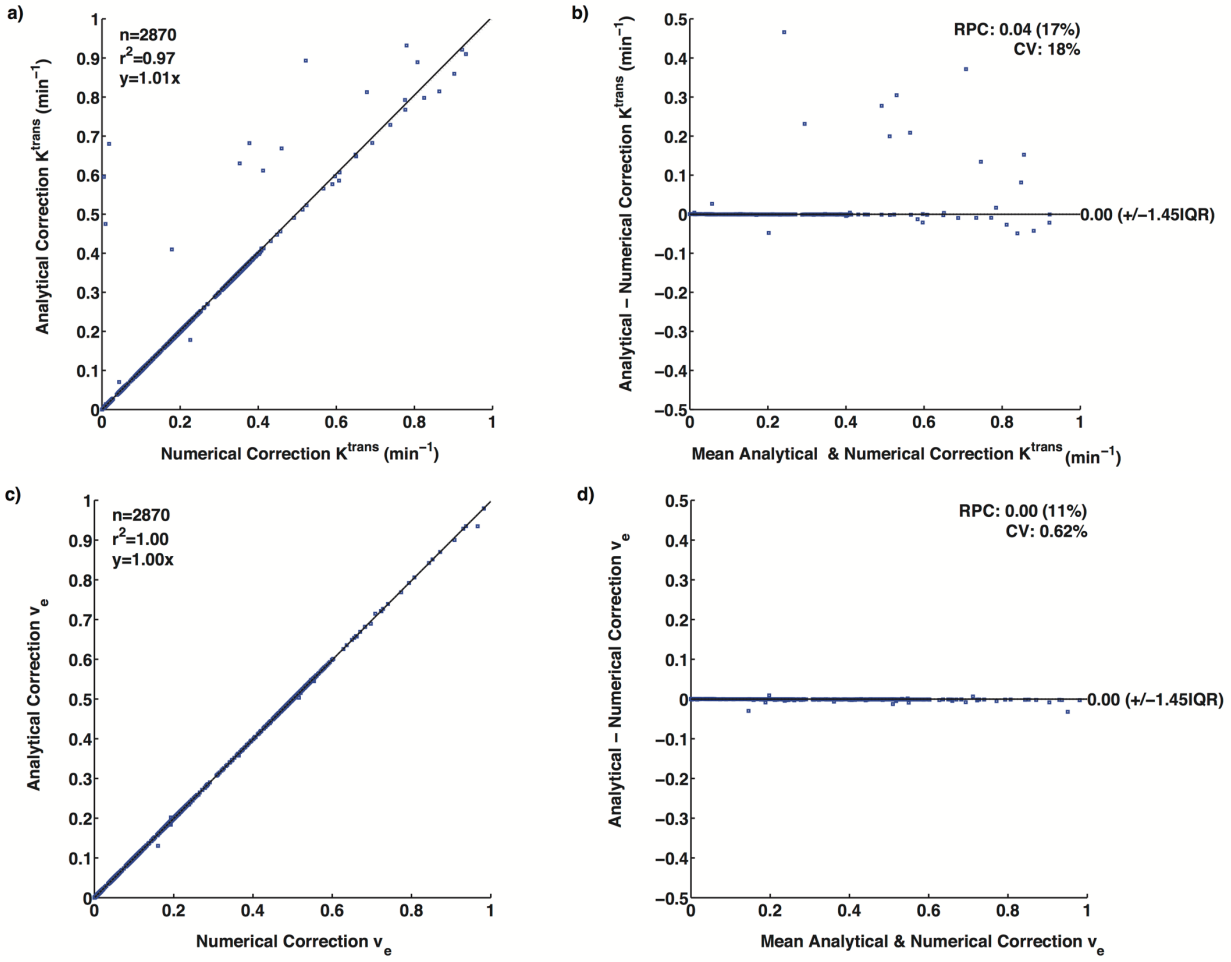

Figure 3-8 Linear regression and Bland-Altman plots for K^{trans} maps (a-b) and v_e maps (c-d) in DRO experiment with noise added. There are 100 Monte-Carlo simulations for each PK parameter combination. The corrected PK parameters from approximated analytical correction and numerical correction are highly comparable (r^2 = 0.97 for K^{trans} and r^2 = 1.00 for v_e) with baseline SNR of 38.5.

analytical theory, indicating the necessity for B_1^+ correction.

A representative K^{trans} and v_e comparison is shown in Figure 3-9. This figure shows $E_{A,\,in\text{-}vivo}$ (Figure 3-9e and Figure 3-9j) is small compared to $E_{B1,\,in\text{-}vivo}$ (Figure 3-9d and Figure 3-9i). A

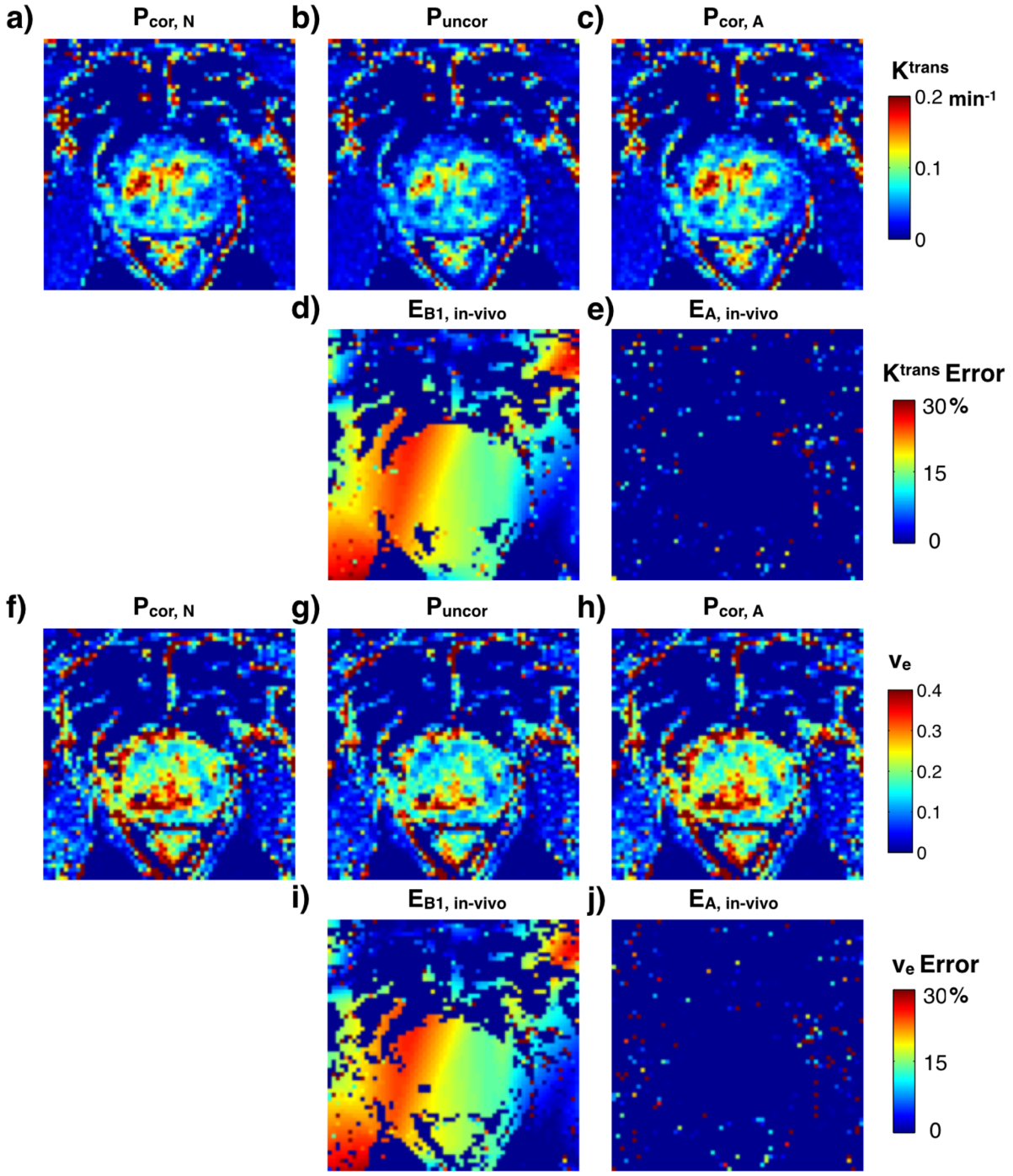

Figure 3-9 Ktrans maps after numerical B1+ correction method (a), before B1+ correction (b) and after approximated analytical correction method (c), B1+ induced error for Ktrans (d), correction residual error of Ktrans (e), ve maps after numerical B1+ correction method (f), before B1+ correction (g) and after approximated analytical correction method (h), B1+ induced error for ve (i), and correction residual error of ve (j)

summary of in-vivo measurements statistics for all 82 cases is shown in Table 3-2. With a wide

range of B_1^+ variation, residual correction error from analytical correction method for K^{trans} and v_e are 0.1±0.3% and 0.1±0.4%, which are minimal. Figure 3-10 summarizes the average k, $E_{A,\ in\text{-}vivo}$ for K^{trans} and v_e among different cases, and the cases are grouped with different scanners (3T Skyra and 3T Trio). The B_1^+ inhomogeneity patterns were significantly different between two scanners ($p < 0.01$), due to different B_1^+ shimming modes. The corresponding correction residual error shows good consistency with k, and all average $E_{A,\ in\text{-}vivo}$ is smaller than 0.4%.

Table 3-2 Summary of in-vivo results. The mean, standard deviation and 95% central range for K^{trans}, v_e, k, $E_{A,\ in\text{-}vivo}$ for K^{trans} and v_e were listed.

	Mean	Standard Deviation	95% Central Range
K^{trans} (min^{-1})	0.11	0.06	[0.03, 0.25]
v_e	0.26	0.13	[0.08, 0.60]
k	1.05	0.08	[0.83, 1.17]
$E_{A,\ in\text{-}vivo}$ for K^{trans} (%)	0.1	0.3	[0.0, 0.2]
$E_{A,\ in\text{-}vivo}$ for v_e (%)	0.1	0.4	[0.0, 0.2]

3.5 Discussion

In this work, a simple and practical B_1^+ correction for quantitative DCE-MRI analysis using an approximated analytical approach was proposed and evaluated. We performed a numerical simulation and a prostate DRO to evaluate the behavior of the approximated analytical method under a set of clinical imaging parameters and noise. The approximated analytical approach was also tested using 82 in-vivo prostate DCE-MRI cases by comparing it with the conventional numerical correction method. All the evaluations showed that the approximated analytical method provides comparable B_1^+ correction to the reference numerical method (less than 0.4% percentage

error) under the practical situation in prostate DCE-MRI.

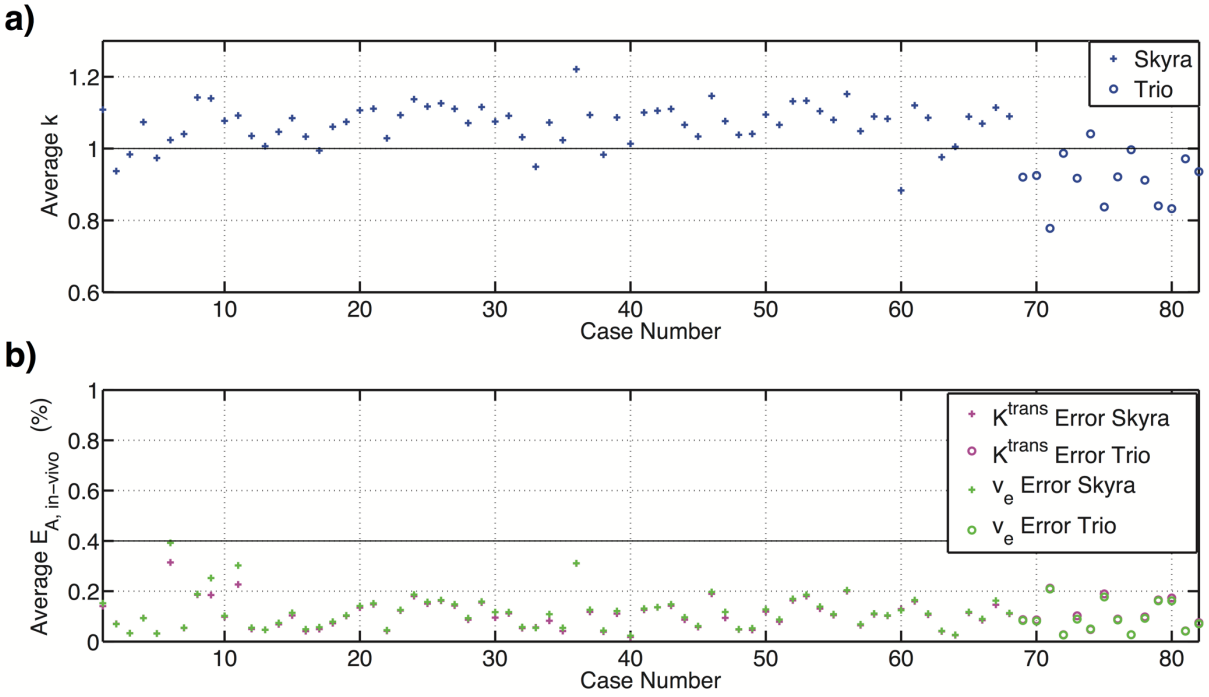

Figure 3-10 Summary of average k (a), correction residual error $E_{A, in-vivo}$ (b) for 82 patients. Different scanners (Skyra and Trio) have slightly different k distribution as shown in (a), but all the average $E_{A, in-vivo}$ is smaller than 0.4%.

The approximated analytical method will enable more practical solutions for B_1^+ correction in DCE-MRI because it does not need access to the full modeling implementation in quantitative DCE-MRI analyses and does not need the acquired images for T_{10} mapping and dynamic MRI. The approximated analytical method only requires B_1^+ maps and uncorrected PK parameter maps as input to estimate the corrected PK parameters. This makes the approximated analytical correction method more practical in clinical research environments, where the model implementation access may be limited. In addition, the approximated analytical correction method provides an easy implementation of B_1^+ correction and can potentially improve the computational efficiency, because for each voxel, the calculation becomes a simple multiplication instead of a series of fitting. For example, in our in-vivo analysis, the reference numerical correction method

took more than 3 hours for each case while the approximated analytical correction method required less than 0.01s using Matlab on the same computer. Although there exist computational acceleration techniques for the numerical correction method, such as parallelization and approximation[80], the approximated correction method can be a good alternative when such accelerations are not available.

The approximated analytical correction relies on three assumptions: small flip angles, small TR/T_1, and k close to one. The experiments in this paper employ our clinical protocol to evaluate the reliability of those assumptions. The first assumption (i.e. small flip angles), can be violated with increased flip angles and this could increase the approximation-induced error. However, in our experiment using the original QIBA DRO flip angles shown in Table 3-1ntio, the maximum percentage error in the simulation is only 0.2%, which is smaller than using our protocols (maximum = 0.4%). Considering those two protocols are similar to other studies[16,22,81], we expect the approximated analytical correction method generalizes well to other flip angle settings. The second assumption is that k is close to 1, and based on our simulation within the k range from 0.7 to 1.3, the correction residual error is smaller than 0.4%. Even when B_1^+ overestimation is 100%, the approximation-induced error is still smaller than the baseline defined in the noise DRO experiments (2.1%±4.3%). Considering the increasing trend of the residual error (difference between two curves) shown in Figure 3-4, we do not expect residual error lager than 1% within a practical B_1^+ range. Also, the derivation in the Appendix actually assumes $(1-k^2)\cdot TR$ is small, considering TR is usually a few milliseconds; the dependency on k is not strong. The last assumption is that TR/T_1 is small, and this is generally true for T1W imaging protocols. For example, in our clinical protocol, TR is 4.17 ms, and a T_1 value of 1579 ±42ms in the prostate region was reported[82].

Figure 3-6 shows how residual errors vary with baseline SNR in the DRO experiment. Because of the potential difference of our pre-assumed parameters in DRO compared to in-vivo, the scale of signal enhancement relative to baseline signal was not exactly the same as in-vivo data. Also, motion-induced errors may play a more important role than image noise for in-vivo data. We do not expect the residual error curve with baseline SNRs will be identical to in-vivo data, but we believe to observe a similar trend for the residual error from numerical correction method.

The noise in the DRO and in-vivo experiments caused outliers with relatively large correction residual error when v_e is low. This is because the fitting procedure is highly sensitive to noise under those circumstances. In the DRO experiments, the numerical correction method also gives $E_{N,\,DRO}$, as shown in Figure 3-7. For in-vivo experiments, the large noise may arise from rectal and bowel motion. We observed outliers near the edge of the prostate as shown in Figure 3-3d. Although we tried to avoid the boundary of the prostate during ROI positioning, due to the anatomy complexity of in-vivo cases, we still observed 0.09% of the pixels with $E_{A,\,in\text{-}vivo}$ lager than 1%. As shown in the DRO experiment, when large noise exists, neither of the fitting methods are reliable; therefore in our in-vivo evaluation, we reported 95% central range of the data to exclude the outliers.

In the numerical simulation and DRO experiments, we determined the percentage error relative to the ground truth, P_{nat}, to utilize the advantage of numerical simulation for error evaluation. We evaluated the percentage error relative to the numerical corrected $P_{cor,N}$ in in-vivo experiments for comparison because we do not have the ground truth, and $P_{cor,N}$ is proved to be a good estimation of the ground truth in simulation experiment in noise-free situation ($E_{N,DRO}$ less than 0.2%). With this in mind, we chose $P_{cor,N}$ as the reference in in-vivo experiments because we

want the evaluation in in-vivo experiment be more consistent with that of the simulation and DRO experiments. However, the DRO experiment also showed that if noise is present, $P_{cor,N}$ might deviate from the ground truth (as shown in Figure 3-7a). With baseline SNR of 38.5 and without taking outliers into evaluation, the numerical correction method will have an average of 14.4% $E_{N,DRO}$. This error will lead to inconsistent evaluation between the results of simulation and in-vivo experiments.

Our study included a few limitations. One limitation is the practical utility of the approximated correction is mainly limited to situations where closed-form or commercial software is used for quantitative DCE-MRI analysis. However, closed or commercial software is widely used in clinical prostate DCE-MRI[83], and to the best of our knowledge, most of them do not include the B_1^+ correction. Also, our approach can be practically useful when an in-house B_1^+ correction process is time-consuming, especially in clinical or clinical research settings. The second limitation is that we focused on the error propagation behavior under the standard Tofts model with population averaged AIFs, which are commonly used techniques in clinical prostate DCE-MRI[1629]. The error propagation analysis may need to be updated if other PK modeling settings is used for DCE-MRI quantification, including subject-based measured AIF[84] and/or the extended Tofts model[58]. For a subject-based measured AIF, our approximation method can be easily modified with minimal error if blood T_{10} is also measured (AIF1). The modified correction method becomes $\frac{K^{trans}{}'}{K^{trans}} \approx (\frac{k_t}{k_p})^2$, $\frac{v_e'}{v_e} \approx (\frac{k_t}{k_p})^2$, where k_t and k_p are k values in measured blood and tissue pixel or ROI. However, if pre-determined blood T_{10} is used for the subject-based measured AIF (AIF2), the modification becomes highly complicated, which may need to be further investigated in future. With DRO simulation with $k_t = 1.2$ and $k_p = 1.1$, $E_{A,DRO}$ using AIF1 is 0.4% for K^{trans} while $E_{A,DRO}$ using AIF2 is 87.6% for K^{trans}. Lastly, the difference between the standard and

extended Tofts models is generally small in the prostate due to the small contribution of v_p. Based

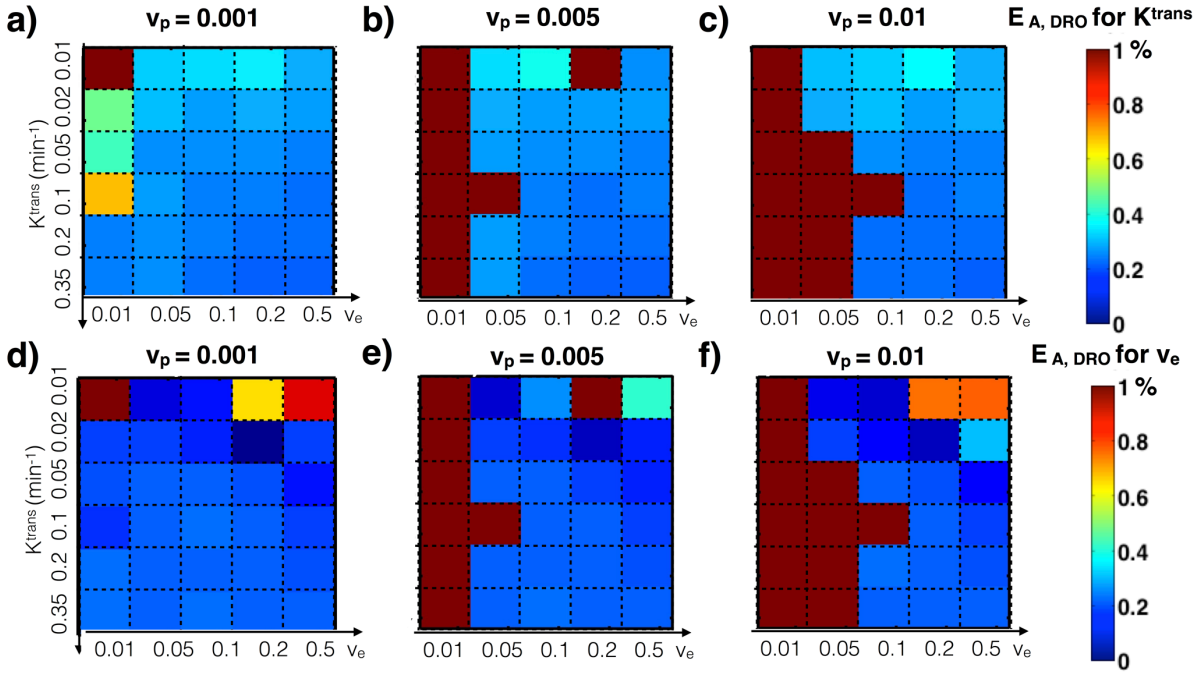

Figure 3-11 Extend Tofts model was simulated in the DRO with three v_p value, 0.001, 0.005 and 0.01. The results were evaluated using $E_{A,DRO}$ for K^{trans} (a-c) and v_e estimation (d-f).

on the in-vivo simulation (n=82), v_p within the prostate was 0.0026±0.0030, and the approximation errors with the extended Tofts model ($E_{A,\ in\text{-}vivo}$) were 0.1±0.1% and 0.1±0.6% for K^{trans} and v_e, similar to the ones with the standard Toft model. With higher v_p relative to v_e, the approximation induced error could be significant, and therefore the method is limited to organ with small v_p (see Figure 3-11 for the influence of v_p in extended Tofts model).

3.6 Conclusion

We have demonstrated the feasibility and accuracy of a simple approximated analytical B_1^+ correction approach for quantitative prostate DCE-MRI. This method only requires B_1^+ maps and uncorrected PK parameters as input to calculate corrected PK parameter maps. The approximated analytical method was evaluated by both numerical digital reference object and 82

in-vivo prostate DCE-MRI cases. In all cases, the approximated analytical method had very low approximation error (less than 0.3% correction residual error compared to conventional numerical correction within 95% central range). Most importantly, this B_1^+ correction method can be easily implemented in clinical workflow, and has the potential to improve the performance and reproducibility of clinical quantitative prostate DCE-MRI.

This work has been published as:

Zhong X, Martin T, Wu HH, Nayak K, Sung K. Prostate DCE-MRI with B1+ correction using an approximated analytical approach. Magnetic Resonance in Medicine 2018;80;2525-2537.

4 DEEP TRANSFER LEARNING-BASED PROSTATE CANCER CLASSIFICATION USING 3 TESLA MULTIPARAMETRIC MRI

4.1 Introduction

As described in Chapter 1, PCa is the most common solid organ malignancy and the second leading cause of cancer-related deaths in men in the United States[85]. The identification of clinically significant lesion is crucial for proper patient management. Gleason Score from histopathology has been proved to be a powerful prognostic factor. In Gleason Score, cell pattern can be described with a Gleason grade raging from 1-5, indicating the tissue aggressiveness, and Gleason Score is composed of two Gleason grades of the most and second most predominant cell pattern. The histological primary or secondary Gleason grade of 4 has generally been shown to be predictive of less favorable outcomes such as increased rates of cancer progression and mortality [86]. As a result, a lesion with GS > 7 is defined as a clinically significant lesion with higher rates of adverse outcomes [87].

Over the past decade, 3T mp-MRI, mainly consisting of 2D or 3D T_2 weighted (T2) imaging, high b-value DWI, and high temporal resolution DCE-MRI, has become the dominant

noninvasive diagnostic tool for diagnosing and grading PCa[8]. Current PIRADS v2 provides a standardized guideline to detect clinically significant lesion with high sensitivity but moderate specificity, so the detection accuracy is still not optimal. Moreover, as described in Chapter 1, the inter-observer diagnosis variety from PIRADS v2 can be improved with the help with automatic models.

Of currently available techniques, deep neural networks have demonstrated superior capabilities in non-medical imaging domains for extracting multi-level abstraction from raw images directly with little human intervention [46]. Application in medical imaging domains remains challenging in part due to large amount of labeled data required to train a reliable neural network. In addition, the labeling process is tedious, requires expertise and can be expensive. Another technique, deep transfer learning (DTL) using fine-tuned pre-trained convolutional neural network (CNN) alleviates the large labeled data requirement and has been successfully applied in medical domains [47].

In this study, we develop the DTL based PCa classification model and compare the classification performance of the DTL based model with deep learning (DL) model without transfer learning and standard PIRADS v2 radiologist interpreted score to distinguish clinically significant from indolent PCa lesions on a curated 3T mp-MRI dataset with whole-mount histopathology (WMHP) correlation.

4.2 Methods

4.2.1 Study Population and MR Imaging Technique

This retrospective study was approved by the institutional review board (IRB) and was compliant with the 1996 Health Insurance Portability and Accountability Act. Between December

2010 and June 2016, a standardized 3T mp-MRI protocol consisting of T2, DWI and DCE imaging was performed. The initial study cohort consisted of 154 patients scanned prior to planned robotic radical prostatectomy with biopsy-proven PCa. Out of 154 patients, 14 were excluded due to lack of pathology. The final study cohort comprised of 140 patients (age 43 to 80 years and weight 59.0 to 133.8 kg, and PSA = 7.9 ± 12.5 ng/dl) with 216 lesions identified on mp-MRI and correlated to thin section WMHP. 41% of the patients had MRI scans before undergoing MRI-targeted fusion biopsy (n = 58), and the rest had MRI scans after biopsy for surgical planning and local staging (n = 82). The scans were interpreted by one of three expert readers and each MR detected lesion was scored by PIRADS v2 components.

3T mp-MRI was performed on a variety of scanners (Trio, Verio, Prisma or Skyra, Siemens Healthineers, Erlangen Germany) using a pelvic phase-array coil with or without the endorectal coil. Each scan used a standard mp-MRI scanning protocol including 3D axial T2 images using Sampling Perfection with Application optimized Contrasts using different flip angle Evolution (SPACE) sequence, echo-planar imaging DWI sequence, and DCE images using Time-resolved angiography With Interleaved Stochastic Trajectories (TWIST). In this study, we used T2 SPACE images and apparent diffusion coefficient (ADC) images calculated from DWI. The echo time and repetition time of the T2 SPACE was 2200/200ms, and the echo train length was 88. With a 17cm FOV and matrix size of 256 x 230, we acquired and reconstructed T2 SPACE images with 0.66mm in-plane resolution and 1.5mm through plane resolution. For DWI acquisition, we used echo time and repetition time of 4800 and 80ms. With FOV of 21 x 26 cm and matrix of 94 x 160, DWI images were reconstructed with in-plane resolution of 1.6 mm and a slice thickness of 3.6mm. The ADC images were calculated from four b values 0, 100, 400 and 800 s/mm^2.

The ground-truth of this study was lesions detected by genitourinary (GU) pathologist on

post robotic-assisted laparoscopic prostatectomy WMHP, blinded to all MRI information. At a separate matching session, a team of one GU radiologist and one GU pathologist matched each previously reported lesion on 3T mp-MRI and each previously reported lesion on WMHP classifying MRI lesions as true or false positives. After matching, the T2 SPACE and ADC images were imported into a commercially available image processing program OsiriX (Pixmeo SARL, Bernex, Switzerland) and a region of interest (ROI) was drawn around each true and false positive lesion on T2 SPACE and ADC images. Each ROI was marked as either clinically significant lesion (GS $\geq$ 7) or indolent lesion (GS $\leq$ 6 or false positive) [87]. In total we marked 111 indolent lesions and 105 clinically significant lesions, consisting of 45 false positive lesions (21%), 66 GS 3+3 lesions (31%), 66 GS 3+4 lesions (31%), 23 GS 4+3 lesions (11%) and 16 GS > 7 lesions (7%). Representative lesion segmentation and corresponding label definition are shown in Figure 4-1.

4.2.2 General Workflow

Our general workflow is shown in Figure 4-1. The input data were T2 SPACE and ADC images with each lesion contoured on both sequences using OsiriX. After proper pre-processing, the image patches enclosing the lesion were generated as the input to the proposed DTL based model. A predicted probability of clinically significant PCa lesion was estimated through the model and was evaluated to determine the final prediction. The zone information of the prostate lesion (either peripheral zone or transition zone) was reviewed by the multidisciplinary team and

added to the DTL based model as a separate feature input.

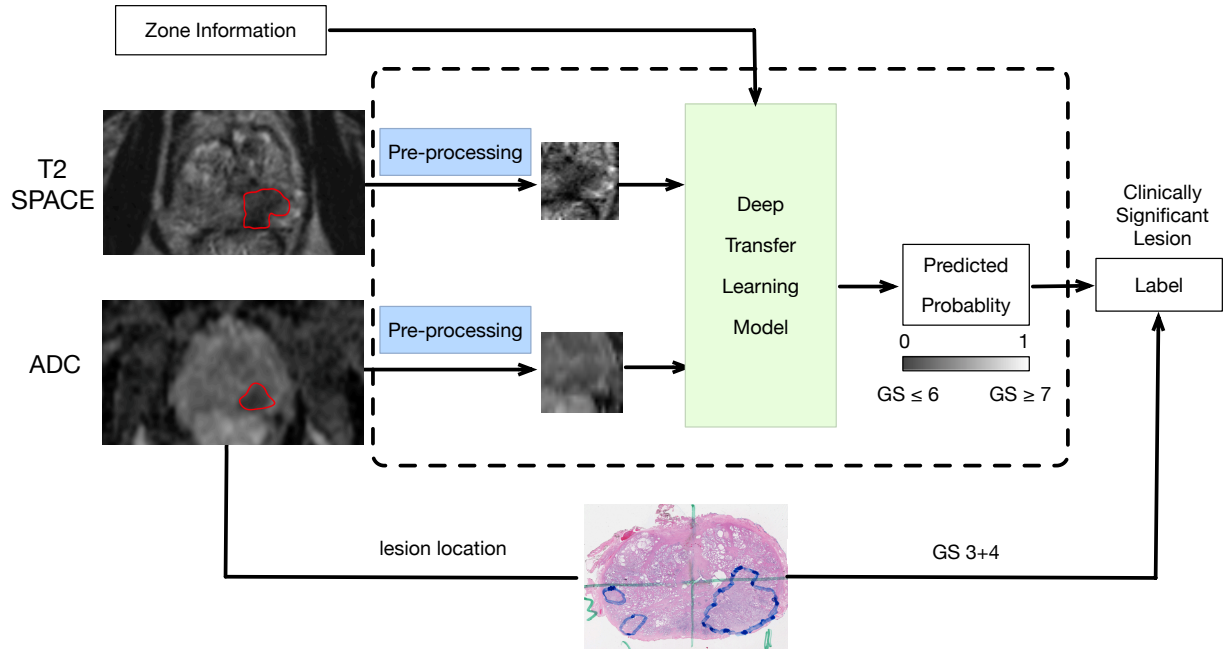

Figure 4-1 Summary of the workflow. The input of the deep transfer learning model would be the image patches enclosing the lesion after pre-processing, and the out put would be the predicted probability that if a lesion is clinically significant. The label was based on the GS from whole-mount histopathology.

4.2.3 Pre-processing

During pre-processing, the square image patches of T2 SPACE and ADC enclosing the lesion were created based on the lesion contour (shown in Figure 4-2). To prepare for further data augmentation, we cropped the image patches with a certain margin. Based on each lesion contour, a rectangle ROI was generated (orange box), and the ROI was expanded to a square ROI (yellow box) which has the same center and a side length of 1.4 times of the longer side of the rectangle ROI. The image patch was cropped based on the square ROI. The side lengths of cropped T2 SPACE image patches ranged from 17 pixels to 92 pixels with a mean of 35 pixels, and this T2 pixel size was recorded and fed into the DTL based model as a separate feature as well.

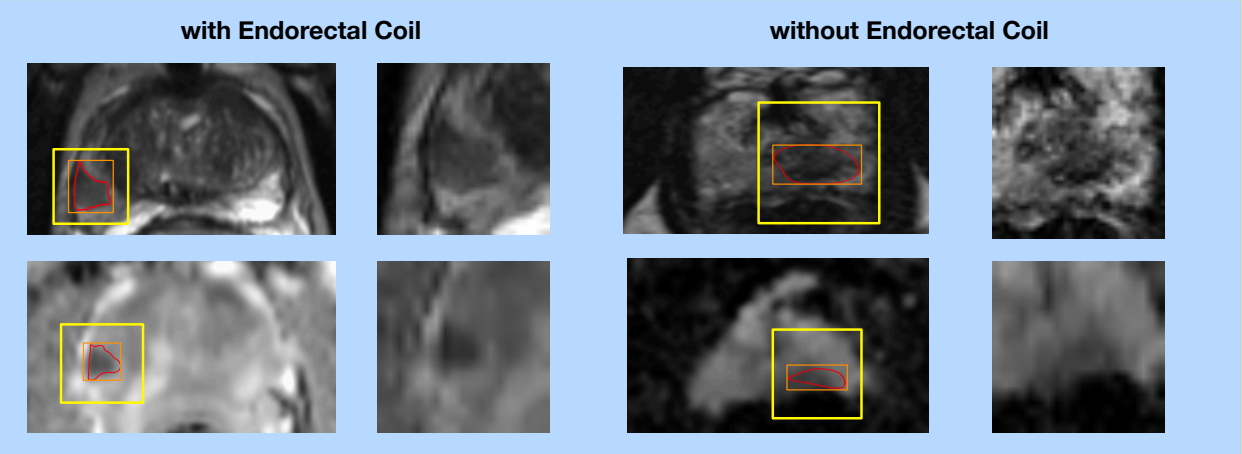

Figure 4-2 Example of how the image patches (yellow) was extracted based on the lesion contours (red). The intensity of image patches after pre-processing became more consistent between cases with and without endorectal coils.

Each image patch was normalized to pixel intensity of 0 to 255 before fed into the DTL based model. The study population included both scans with and without the endorectal coil, and an additional image normalization of the T2 SPACE images was implemented for patients with the endorectal coil to account for high signal variation at the interface of coil and tissue. The image normalization of T2 SPACE was based on the maximum value within a rough contour of normal prostate tissue defined on each patient. For quantitative ADC images, we set an empirical intensity upper bound of 4000 to filter extreme values before normalization to maintain the distribution of ADC values within and among cases. One representative example with and without the endorectal coil is shown in Figure 4-2. The T2 SPACE and ADC image patches were resized to 32 by 32 pixel before fed into CNNs.

4.2.4 DTL based Model

The DTL based model structure is described in Figure 4-3. We utilized ResNet [88], a state-of-the-art variant of CNN performing well on similar tasks [89]. The ResNet model we utilized consists of 19 convolutional layers (blue) and 9 building blocks (dark green). To fuse the two

ResNet models, features (f_{T2} and f_{ADC}) from the last average pooling layer (green) with two key features of T2 SPACE image patch size and the zone information of the lesion were concatenated. The two ResNet models were initialized with weights trained from CIFAR10 dataset [90], and the weights of the last convolutional layer of the two ResNet models together with the added fully connected layer were fine-tuned using back-propagation during the training process. In this study, we implemented the deep learning models using Caffe framework (University of California, AI Research, Berkeley CA) [91].

4.2.5 Model Evaluation and Comparison

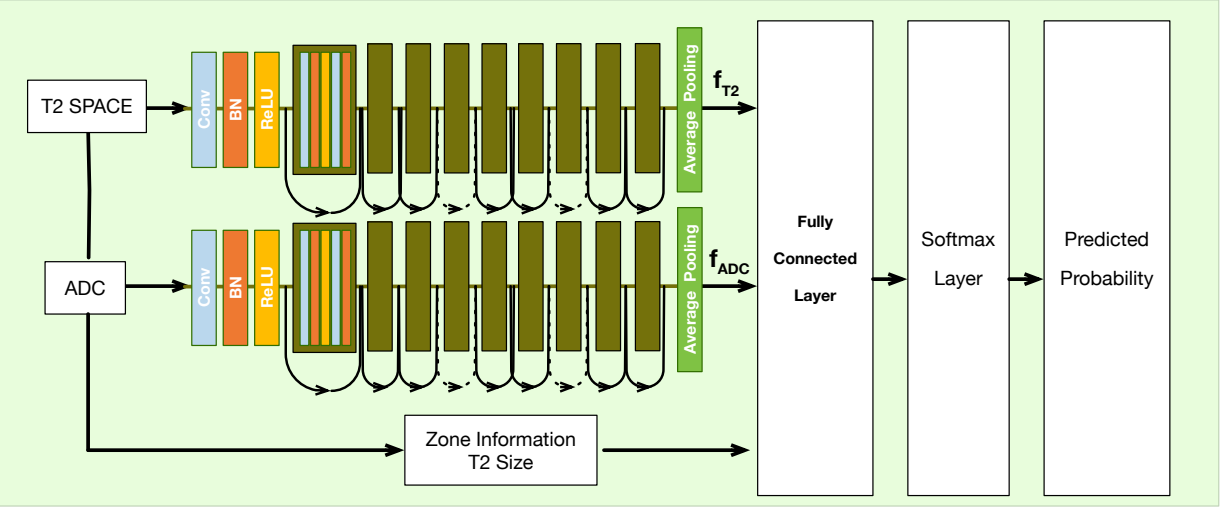

Figure 4-3 Deep transfer learning structure.

We selected 169 lesions (83 clinically significant and 86 indolent) in 110 patients as training set and the remaining 47 lesions (22 clinically significant and 25 indolent) in 30 patients as testing set from 105 clinically significant and 111 indolent lesions. The training and testing set data splitting was based per patient so that all lesions from the same patient would be in either the training or testing set. The lesion distribution for training and testing data was summarized in

Table 4-1.

Table 4-1 Summary for lesion distribution in training and testing data

Lesion Number	Training Set (Case Number)	Testing Set (Case Number) [CS Lesion/Indolent Lesion]
With Endorectal Coil	66 (45)	17 (12) [9/8]
Without Endorectal Coil	103 (65)	30 (18) [13/17]
Index Lesion	90	25 [20/5]
Non-index Lesion	79	22 [2/20]
Transition Zone (TZ)	40	11 [5/6]
Peripheral Zone (PZ)	129	36 [17/19]

As described by Shin et al, proper data augmentation on the training data with large variety is crucial for applying deep learning algorithm in medical image domain [92]. We applied the same random flipping, cropping and slightly shearing transformation to each T2 SPACE and ADC image pair to increase the variety of the training data and the robustness of the model. Random contrast adjustment was only applied to T2 SPACE images. An example of T2 SPACE and ADC images after data augmentation was shown in Figure 4-4. After data augmentation, 2574 clinically significant and 2580 indolent lesions were used as the input of the training process. The numbers of samples after data augmentation in both classes were almost equal so that the prediction bias during training introduced by class imbalance can be avoided. The DTL based model was trained

on the augmented training set and evaluated on the testing set.

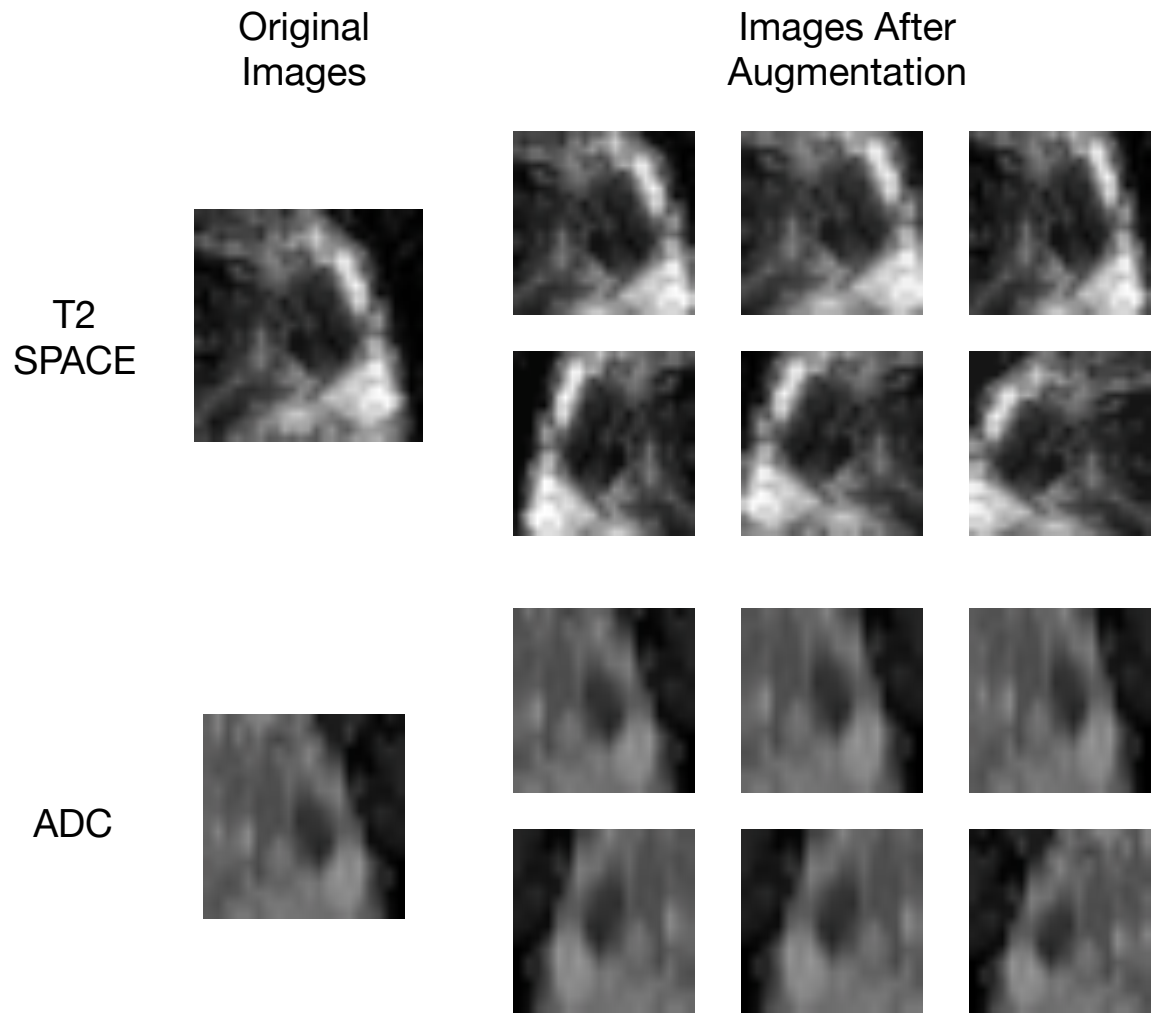

Figure 4-4 Representative examples of the data augmentation. The left column shows the image patches before data augmentation, and the right column shows six examples of the image patches after data augmentation for T2 SPACE and ADC images

Additionally, DTL based models with T2 SPACE images (DTL_{T2}) alone, ADC images (DTL_{ADC}) alone and combined T2 SPACE and ADC images (DTL_{T2+ADC}) without added features were individually evaluated to determine the contribution of each component of the model. In DTL_{T2} and DTL_{ADC} model, one ResNet model trained on CIFAR10 was fined-tuned, and the input to the fully connected layer was f_{T2} or f_{ADC} only. Likewise, DTL_{T2+ADC} model architecture was

similar to DTL based model architecture with only f_{T2} and f_{ADC} concatenated and fed to the fully connected layer. DL model without transfer learning was implemented with the same architecture as DTL based model but trained from scratch with MSRA initialization [93]. DL model was evaluated to illustrate the contribution of transfer learning. All these models were also trained on the augmented training set and evaluated on the testing set to validate our proposed DTL based model architecture.

The PIRADS v2 score for each lesion was extracted from the final radiology report interpreted by a subspecialized radiology fellow and an experienced prostate subspecialized genitourinary radiologist. We compared the performance of our DTL based model and the performance of expert reader PIRADS v2 score > 4 in the task of distinguishing clinically significant PCa lesions from indolent lesions on the same testing set. With a threshold of 4, lesion with PIRADS v2 score smaller than 4 was considered indolent lesion whereas lesion with PIRADS v2 score larger than or equal to 4 was considered clinically significant lesion. The PIRADS v2 score served as an expert reader baseline to compare with results from our proposed DTL based model.

4.2.6 Statistical Analysis

The classification performance was quantitatively evaluated by accuracy, sensitivity, specificity and area under curve (AUC) of receiver operating characteristics (ROC) curve using 47 testing lesions from 30 cases. ROC curves and precision-recall curves were also shown to give a more thorough description of the classification performance. The threshold to calculate accuracy, sensitivity and specificity was picked based on the best accuracy. Those analyses were performed on Matlab 2014a (MathWorks, Natick, MA). Bootstrapping with 2000 resamples was performed to estimate the 95% confidence interval (CI) for AUC and DeLong test was used to compare the

AUC of the DTL based model and other models as well as PIRADS v2 score. The AUC related analysis was implemented using pROC toolbox [94] in R package [95].

4.3 Results

4.3.1 DTL based Model Evaluation

With GS from WMHP as the ground truth, the DTL based model achieved an accuracy of 0.723 in distinguishing indolent from clinically significant PCa lesion. The corresponding sensitivity and specificity were 0.636 and 0.800 respectively. The AUC of the ROC curve was 0.726 (CI [0.575, 0.876]). Representative examples of successful and failed predictions of the DTL

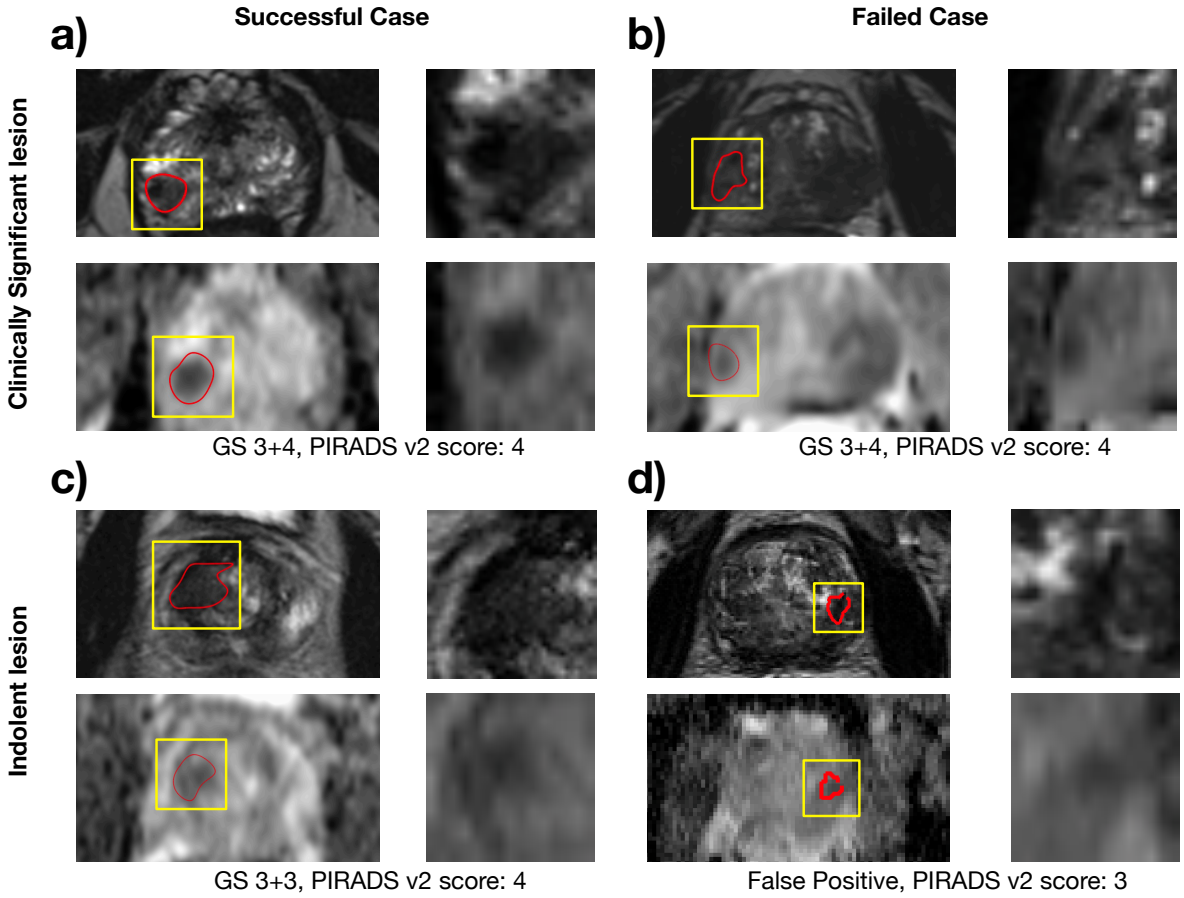

Figure 4-5 Representative cases for successful and failed prediction cases based on the DTL based model on clinically significant lesion (a and b) and indolent lesion (c and d). The corresponding GS from WMHP and PIRADS v2 score for each lesion are also shown under the images

based model are depicted in Figure 4-5. The false positive (Figure 4-5b) and false negative (Figure 4-5d) examples were all from scans with the endorectal coil, and the T2 SPACE contrast after pre-processing was suboptimal. Also, those examples illustrated how similar lesions appear on 3T mp-MRI and showed challenges for both algorithms and human to properly score the lesions.

4.3.2 DTL based Model Comparison with Other Models

To properly evaluate the contribution from T2 SPACE images, ADC images and two added features to the classification performance, we trained and compared the models with each image separately (DTL_{T2} and DTL_{ADC}) and the model with two images combined (DTL_{T2+ADC}). The experiments results are summarized in Table 4-2. The resulting ROC and precision-recall curve are shown in Figure 4-6. DTL_{T2} achieved the least accuracy of 0.617 among all the models, indicating currently T2 SPACE images provided less valuable information compare to ADC images regarding the classification task. DTL_{T2}, DTL_{ADC} and DTL_{T2+ADC} models all achieved a higher sensitivity of 0.773 compared to the DTL based model but with less specificity, which illustrated that the added zone information and T2 size features help to reduce false positive prediction, increasing the specificity. Overall, these experiment results validated the DTL based model structure design as DTL_{T2}, DTL_{ADC} and DTL_{T2+ADC} experiments generated less accurate prediction compared to DTL based model regarding accuracy and AUC. The p-values of DeLong

Table 4-2 Performance summary on the testing set (n = 47)

Performance	AUC (95% CI)	Accuracy	Sensitivity	Specificity
DTL_{T2}	0.580 (0.411-0.748)	0.617	0.773	0.480
DTL_{ADC}	0.620 (0.455-0.785)	0.638	0.773	0.520
DTL_{T2+ADC}	0.713 (0.563-0.863)	0.702	0.773	0.640
DL	0.687 (0.532-0.843)	0.702	0.636	0.760
DTL	**0.726 (0.575-0.876)**	**0.723**	0.636	**0.800**
PIRADS v2 score	0.711 (0.575-0.874)	0.660	**0.864**	0.480

test between DTL based model and DTL_{T2}, DTL_{ADC}, DTL_{T2+ADC} were 0.14, 0.30 and 0.74 respectively. Although the difference was not statistically significant due to the limited size for testing, the general trend indicated the relative significance of each component.

To evaluate the contribution of transfer learning, we also compared the DTL based model with the DL model without transfer learning using the same model structure with MSRA initialization. Our DTL based model achieved higher accuracy and AUC compared to non-transfer leaning DL (Table 4-2), which achieved an accuracy of 0.702 and AUC of 0.687 ($p = 0.70$). This experiment showed that transfer learning helped to improve the testing accuracy with a limited training set. Additionally, both DL model and DTL based model provided predictions with higher specificity and lower sensitivity compared to DTL_{T2}, DTL_{ADC} and DTL_{T2+ADC} models, confirming the two added feature could reduce false positive prediction. The corresponding ROC and precision-recall curves are shown in Figure 4-6.

4.3.3 DTL based Model Comparison with PIRADS v2 Score Expert Reader Performance

In detection of clinically significant from indolent PCa, the accuracy of the expert radiologist assigned PIRADS v2 score > 4 was 0.708 with AUC of 0.765 in all lesions (both training and testing set), and was 0.660 with AUC of 0.711 (CI [0.575, 0.874]) in the testing set.

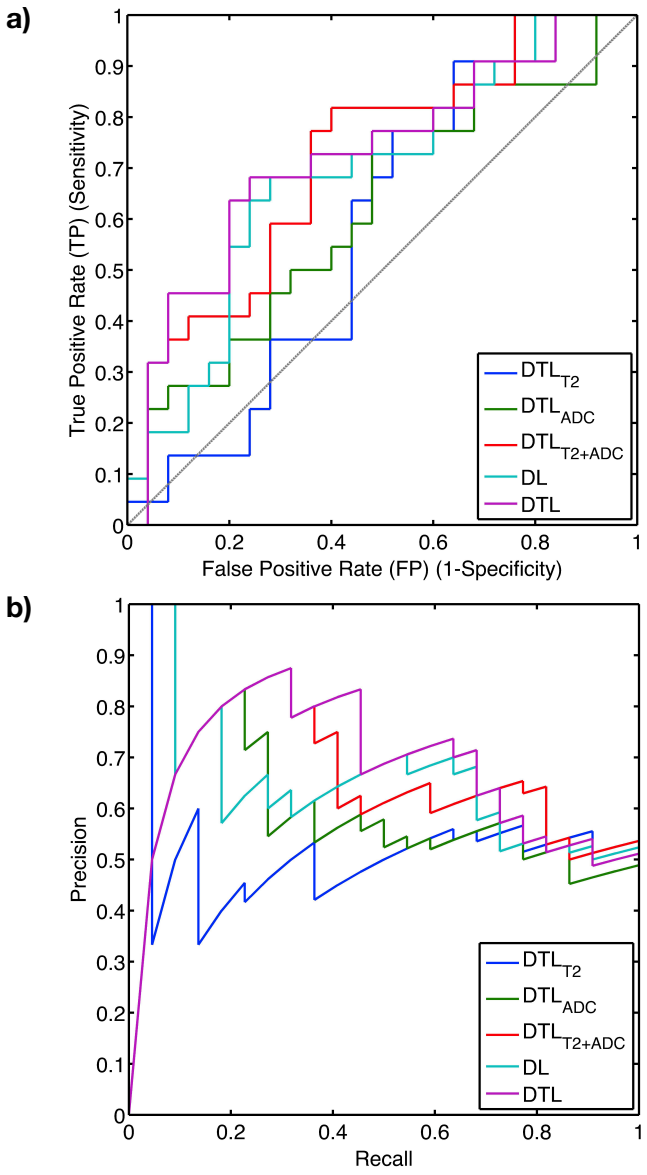

Figure 4-6 Receiver operating characteristics curve (a) and precision recall curve (b) comparison between DTL, DTL_{T2}, DTL_{ADC}, DTL_{T2+ADC} and DL models on 47 testing lesions, validating the DTL based model architecture

The performance with PIRADS v2 score > 4 is summarized in Table 4-2. DeLong test showed that the DTL based model generated comparable performance with PIR ADS v2 score ($p = 0.89$).

The ROC and precision-recall curve comparison between the DTL based model and PIRADS v2 score are shown in Figure 4-7. With estimation favoring specificity (left side of the curve), the proposed DTL based model outperformed PIRADS v2 score expert radiologist

interpretation. While with estimation favoring sensitivity (right side of the curve), the expert reader

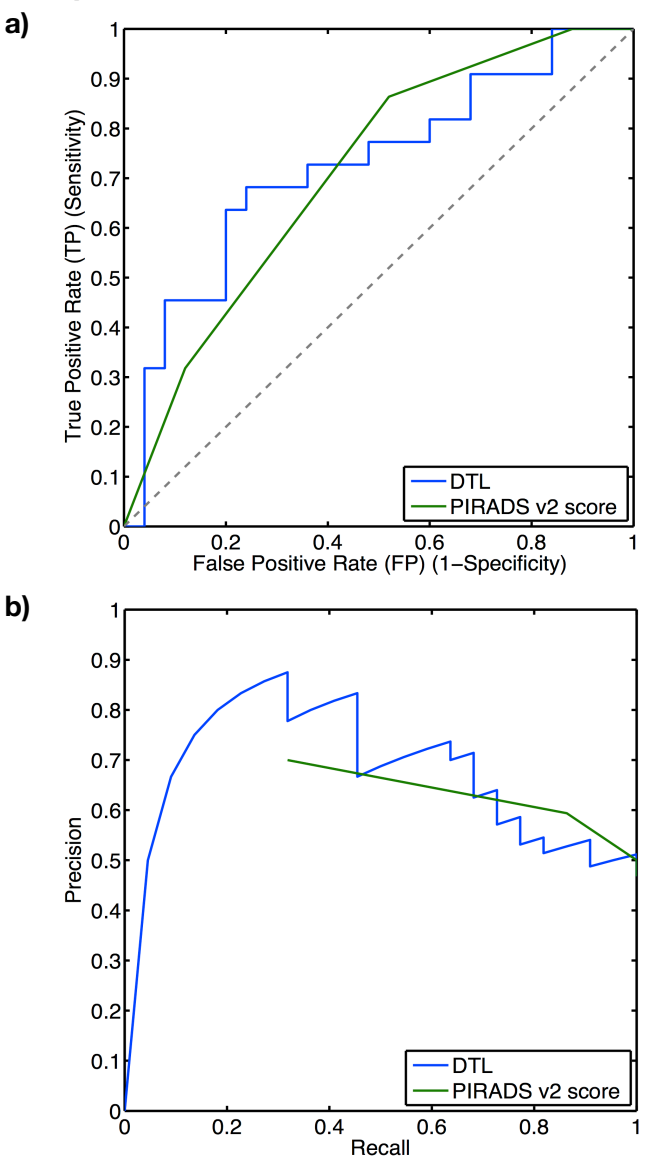

Figure 4-7 Receiver operating characteristics curve (a) and precision recall curve (b) comparison between the proposed DTL based model and the expert reader PIRADS v2 score (n=47)

PIRADS v2 score outperformed proposed DTL based model. This observation is consistent with performance in Table 4-2, where the best sensitivity and specificity were 0.636 and 0.800 for DTL based model and 0.864 and 0.480 for PIRADS v2 score. The proposed DTL based model tends to reduce the over-diagnosis with higher specificity compared to PIRADS v2 score.

4.4 Discussion

The DTL based model was developed and compared with the performance of a DL model without transfer learning and an expert radiologist detecting clinically significant PCa using the PIRADS v2 score to distinguish clinically significant from indolent PCa lesions using 3T mp-MRI with the highest available reference standard, histopathological grading on WMHP. Each discrete prostate lesion was first identified on 3T mp-MRI and contoured on OsiriX based on the appearance, and small image patches enclosing the lesion of T2 SPACE and ADC images were used as input to the DTL based model with labeling based on WMHP GS. The proposed DTL based model outperformed DL model without transfer learning and achieved comparable performance compared to PIRADS v2 score using single slice 2D image patches from T2W and ADC images on our 47 lesions testing data.

One advantage of our method is that it does not rely on the detailed contour defined by the radiologists, because the information used to generate the image patch was mainly the location and a rough size of the lesion. We further improved the robustness of the model to lesion ROI definition by adding random cropping into the data augmentation. In this way, the potential inconsistency of ROI definition from either lesion detection algorithms or radiologists will have little influence on the model performance.

A prior study has shown great potential to use DTL to distinguish clinically significant lesion[89] using publicly available prostate MRI data. Here, we have chosen to use the ResNet trained on CIFAR10 data to minimize potential overfitting and added T2 size and zone information as additional features to the DTL based model to improve the overall classification performance. Also, our model implementation and evaluation were based on labeling information referenced by the WMHP and included more practical situations, such as cases with and without the endorectal

coil.

Our study has compared the DTL based PCa classification to PIRADS v2 score. Although PIRADS v2 score was not originally designed for this task, studies have shown that lesions on 3T mp-MRI with higher PIRADS v2 score correlate with PCa lesions with higher GS on WMHP [96]. We picked a threshold four, which achieved the best accuracy for PIRADS v2 score in the testing set. All prostate mp-MRI cases were interpreted by expert genitourinary radiologists who have more than 10 years of experience with approximately 500 prostate MRI cases per year. We believe the PIRADS v2 scores in the study would be close to the upper limit of the prostate MRI interpretation, a relatively good approximation of human performance.

We distinguished clinically significant lesions from indolent lesions and false positive lesions rather than normal tissue, as shown in Figure 4-5. The task is challenging because differences on MRI images between these lesions are sometimes not visually apparent. Although there is no universally accepted definition of the clinically significant lesion, we used one of the most common definitions, GS > 6 [87], as our working definition of the clinically significant lesion to emphasize any differences between low- and intermediate/high-grade prostate lesions. We have also compared our proposed DTL based model with a traditional machine learning method as another baseline. We vectorized the pre-processed T2 SPACE and ADC images as the input and trained a random forest model [97] using the same training cohort before data augmentation. On the testing set, the random forest model provided a prediction with AUC of 0.66. This result confirmed the well-known challenges of not using detailed lesion contours and indicated that deep learning based methods can alleviate the requirements of handcrafted imaging features to achieve relatively satisfactory results.

The proposed DTL based model has achieved comparable results compared to PIRADS v2

score, but there still exist potential improvements in the performance, as the model has not fully utilized the imaging information from 3T mp-MRI. For example, the current input was from single 2D segmentation, and the feature extracted from 3D region of interest may provide a more comprehensive depiction of the tissue character. Also, we did not take DCE-MRI into consideration because DCE-MRI was not available for some of the cases. As the cases accumulate, we would only include cases with K^{trans} maps and add K^{trans} maps into our model. Another potential improvement is to apply registration between T2 SPACE and ADC images for each case so that the two images can be fed into the same model, reducing the model complexity and potential overfitting problem.

Our study has several limitations. One limitation of this study is the small sample size for testing because of the limited available labeled data. Even with the transfer learning, the model requires sufficient training data to produce a more generalized model, resulting in a limited testing set size. The evaluation on a larger testing cohort will be conducted in the future as we acquire more labeled data. With a larger testing set, a more statistically powerful evaluation of the DTL method can be made. Moreover, our current evaluation was performed on one random splitting cohort. Although it is commonly used in deep learning related papers due to computation limitation, it might give prediction with a certain bias. We included confidence interval of the AUC from bootstrapping resampling to give a more thorough evaluation. Another limitation of our study is that we included a manual segmentation of the prostate to assist the normalization for the cases with the endorectal coil (55 out of 140 cases were scanned with the endorectal coil). We also included contrast adjustment in data augmentation process to improve the robustness to T2 contrast. Although this method was proved to achieve similar contrast for all T2 SPACE images, the manual segmentation of the prostate could be burdensome, and a more systematic way to find

the normalization threshold is preferable and still under investigation. In the evaluation of DTL_{T2} and DTL_{ADC}, the features from T2 SPACE images contribute less than features from ADC in this classification task. This observation is consistent with radiologists' experience, but the performance of DTL_{T2} can be improved with well-designed pre-processing. Another limitation is that the system requires the lesion detection as the input to define an image patch. Although a prostate lesion detection and classification system is possible by combining other lesion detection systems [98] with our classification system, it may be preferable to design one integrated system that could complete the pixel-level lesion detection and classification task together.

4.5 Conclusion

We implemented and evaluated a deep transfer learning (DTL) based model to differentiate between clinically significant (GS $\geqslant$ 7) and indolent PCa lesions (GS < 6 and false positive) using 3T mp-MRI with WMHP correlation. The proposed DTL based model outperformed the DL based model without transfer learning, confirming the contribution of transfer learning. The DTL based model performance generated comparable performance to the expert reader PIRADS v2 score ($p = 0.89$), showing great potential to augment PCa for non-experts. This model would need to be validated in much larger datasets to further evaluate its clinical utility.

This work has been published as:

Zhong X, Cao R, Shakeri S, Scalzo F, Lee Y, Enzmann DR, Wu HH, Raman SS, Sung K. Deep transfer learning-based prostate cancer classification using 3 Tesla multi-parametric MRI. Abdom Radiol. 2018:1-10. https://doi.org/10.1007/s00261-018-1824-5

5 SUMMARY AND FUTURE STUDIES

In summary, I have evaluated two different B_1^+ estimation techniques regarding the repeatability and reproducibility of the pre-contrast T_1 estimation, and both B_1^+ corrected T_1 maps showed higher consistency for each tissue among different scans. I have developed and evaluated an approximated analytical B_1^+ correction technique for quantitative DCE MRI that can be easily implanted in the clinical workflow. By combining both RR-VFA B_1^+ estimation method and approximated analytical B^+ correction method, the B_1^+ compensation for quantitative DCE MRI is more clinically feasible and could potentially improve the consistency for quantitative DCE MRI biomarker. I have also developed and evaluated a deep transfer learning-based lesion classification system to distinguish clinically significant lesions from indolent lesions using mp-MRI, which could help to achieve a more consistent and accurate interpretation for mp-MRI together with PIRADS v2. Our proposed method achieved comparable performance to the PIRADS v2 score from experienced radiologists regrading accuracy and AUC.

All these projects have the potential to improve clinical diagnosis of prostate cancer and are in transition to clinical application. With the evaluation of B_1^+ estimation techniques, the evaluated RR-VFA B_1^+ estimation technique has been integrated into our routine clinical scan protocol and will be used to compensate B_1^+ in quantitative DCE-MRI. With simplified B_1^+ correction method development, this method can be easily applied in clinic and is especially useful with closed-form post-processing software used. In UCLA radiology department there are more

than 1,500 prostate DCE-MRI cases per year using the commercial closed-form software (DynaCAD, In Vivo) for more than six years to generate quantitative parametric maps, the proposed approximated analytical B_1^+ correction provides a valuable alternative solution to perform feasible B_1^+ compensation. With the deep transfer learning based prostate cancer classification model developed and evaluated, it has the potential to provide a consistent tool to help radiologists identify false positive lesions and avoid overtreatment. This model needs to be further improved and evaluated before applied in clinic workflow.

5.1.1 B_1^+ Estimation in Quantitative DCE MRI

Repeatability and reproducibility are crucial for quantitative biomarker evaluation since quantitative biomarkers are supposed to give consistent results for the same tissue. As described in Chapter 2, I evaluated the repeatability and reproducibility of three T_1 maps with or without B_1^+ correction on ten healthy volunteers on two 3T Siemens scanners. Further study is needed to expand the evaluation to more vendors such as GE and Phillips and more B_1^+ estimation techniques such as Bloch-Siegert. Also, based on the feasibility, we evaluated the B_1^+ estimation techniques using pre-contrast T_1 on healthy volunteers. Although this indicates the potential benefit for quantitative DCE analysis, a more rigorous evaluation can be made in the future. For example, compared to healthy volunteers, patients tend to have higher BMI and might have slightly different B_1^+ property. In addition, pre-contrast T_1 is only part of the quantification in quantitative DCE analysis. To further validate each B_1^+ estimation technique, experiments should be done to measure the repeatability and reproducibility for quantitative DCE MRI biomarkers on patients in the future.

5.1.2 B_1^+ Correction in quantitative DCE MRI

Our proposed approximated analytical B_1^+ correction method was proved to simplify the B_1^+ correction process for quantitative DCE MRI in prostate. However, as described in Chapter 3, the error propagation analysis was based on the Tofts model and population-averaged AIF. I have done some experiments on extended Tofts model and patient-specific AIF (discussion in Chapter 3), and further analysis is needed to get a thorough understanding for the error propagation for other popular models such as the dispersion model.

5.1.3 Combination of B_1^+ Estimation and B_1^+ Correction for PCa Diagnosis

With the RR-VFA B_1^+ estimation and approximated analytical B_1^+ estimation technique validated, the next step would be to integrate those two components into the clinical workflow, and test if there is any clinical benefit for B_1^+ compensation in quantitative DCE MRI. We have collected around 80 patient scans with RR-VFA B_1^+ maps, and I am working on the statistical analysis to test if there is any improvement for lesion delineation using B_1^+ corrected K^{trans} and v_e maps. The hypothesis is that by integrating B_1^+ compensation, K^{trans} and v_e would enable better separation between normal tissue and cancerous tissue as well as between clinically significant lesions and indolent lesions.

5.1.4 Improvement for Deep Transfer Learning based PCa Classification

The deep transfer learning-based lesion classification model showed superior performance compared to alternative model structures and comparable performance to PIRADS v2 from experienced radiologists. Further study direction could be to explore the full potential of this technique by increasing the data size and improving pre-processing.

Although the requirement for training data size was partly solved by introducing deep

transfer learning and data augmentation, they are still not as good as getting more training data from patients. For one thing, natural images and MRI images have many differences besides certain common properties. Moreover, the samples generated from data augmentation are less various and realistic compared to real data. Overall, the performance of the model is expected to increase with more training data. Furthermore, although I have provided the lesion statistics regarding if the lesion is scanned with endorectal coil, is index lesion or is in PZ in Table 4-1, because of our limited testing size, I didn't evaluate the performance on separate lesion groups. With patient data accumulates, testing data can also be expanded to provide a more reliable and specific evaluation for each lesion type and lesion location. In summary, increasing the data size helps with both training and testing of the model.

At the same time, as mentioned in Chapter 4, current pre-processing is not optimal. Specifically, the signal-inhomogeneity induced by the endorectal coil can be addressed by more automatic and consistent way. For example, Cao et al. utilized the signal intensity in bladder to normalize the T2 intensity. On the other hand, if we have enough training data for patients with and without endorectal coil, we could train two different models respectively. In summary, the performance of this lesion classification system can be improved in the future by increasing case number and improving pre-processing.

Another limitation of the current lesion classification system is that it still requires the input of the lesion location and rough lesion size. It is more desirable to develop or integrate a lesion detection algorithm such as proposed by Tsehay et al.[99] to generate the information so that the whole mp-MRI interpretation was automatic. The evaluation combining those two systems can be done in the future. Also, further assessment of how the system would improve the consistency and the accuracy of the diagnosis is to be studied. For instance, one could compare the inter-observer

and intra-observer variability and prediction accuracy with and without knowing the prediction results from the lesion classification system.

5.1.5 Improvement for Lesion Classification in Prostate Cancer

It is challenging to collect high-quality and large number of labeled data in the medical imaging domain, and I chose deep transfer learning to overcome this limitation. On the other hand, we have large number of unlabelled data available. For example, we have more mp-MRI data from screening scans, which can be used by the deep learning model to understand normal prostate tissue. Another future direction would be to combine supervised model with unsupervised model to fully utilize available data.

Additionally, there is no universally accepted definition of clinically significant lesion, and I picked the definition of GS > 6 from whole-mount histopathology. This definition is commonly used and a Gleason grade of 4 has been shown to be related to less favorable patient outcome. In addition, Gleason Score from whole-mount histopathology was proved to be a more powerful indicator for patient outcome, so the current label is valuable for diagnosis. However, Gleason Score is also limited by inter-observer variability[100] and alternative labels related to patient outcome can also be considered to train the model to aid diagnosis. For example, mutations in DNA-repair genes was proved to be related to metastatic prostate cancer[101] and can be served as a great label for prostate cancer classification task.

5.1.6 Combination of Quantitative DCE MRI and PCa Classification Model

Another limitation of the current lesion classification systems is that quantitative DCE MRI was not included. This is mainly because the current clinical data contains inconsistent quantitative DCE MRI data for patients and has no B_1^+ compensation. It would be valuable to explore if adding

DCE MRI would improve the lesion classification performance and if adding B_1^+ compensation would further improve the classification performance. It is a great way to utilize mp-MRI information by combining both quantitative biomarkers and deep learning algorithms.

5.1.7 Summary

Having clinically applicable and reliable B_1^+ estimation and correction techniques for quantitative DCE MRI and deep transfer learning-based clinically significant lesion classification system improved the quantitative interpretation of mp-MRI. B_1^+ estimation and compensation improved the consistency of quantitative biomarkers while lesion classification systems utilized the mp-MRI images more consistently compared to qualitative image interpretation. Those techniques have the potential to provide better diagnosis and patients comfort.

6 APPENDICES

In VFA process, the fitting procedure is simplified to only two flip angles $\alpha_1$and α_2. The measured signals with the two flip angles S_1 and S_2 are given as constant, namely $S_1 = S_1$' and $S_2 = S_2$'. Based on Equation 2-1, in actual situation, $S_i = \frac{M_0(1-E'_{10})sink\alpha_i}{1-E'_{10}cosk\alpha_i}$ for both signals S_1 and S_2.

First, we want to find the relationship between corrected E_{10}' and uncorrected E_{10}. The corrected E_{10}' can be expressed as a function of uncorrected E_{10} as follows:

Equation 6-1

$$E_{10} = \frac{\frac{S_2}{sin\alpha_2} - \frac{S_1}{sin\alpha_1}}{\frac{S_2}{tan\alpha_2} - \frac{S_1}{tan\alpha_1}}$$

$$= \frac{(1 - E'_{10}cosk\alpha_1)sink\alpha_2 sin\alpha_1 - (1 - E'_{10}cosk\alpha_2)sink\alpha_1 sin\alpha_2}{(1 - E'_{10}cosk\alpha_1)sink\alpha_2 sin\alpha_1 cos\alpha_2 - (1 - E'_{10}cosk\alpha_2)sink\alpha_1 sin\alpha_2 cos\alpha_1}$$

Equation 6-1 is the analytical form of the linear regression using two flip angles. When using more than two flip angles, the fitting estimation will be the same using any two flip angles without noise since it is an over-determined problem.

With the assumptions that flip angles α_i in rad are close to zero $({\alpha_i}^3 \approx 0)$, based on Taylor Series, we could get that $sin\alpha_i \approx \alpha_i, cos\alpha_i \approx 1 - \frac{{\alpha_i}^2}{2}$ for both flip angles $\alpha_1$and α_2. By substitute those equations into Equation 6-1, we could get a simplified version of the relationship.

Equation 6-2

$$\boldsymbol{E_{10} \approx 1 - \frac{1 - E_{10}{}'}{1 - E'_{10} + k^2 E_{10}{}'}}$$

Then based on the assumption that $\frac{\mathrm{TR}}{\mathrm{T_1}} \approx 0$ and Taylor Series $\ln(\mathrm{x}) \approx \mathrm{x} - 1$, it can be derived that $E_{10} \approx 1 - \frac{TR}{T_{10}}$ and $E_{10}{}' \approx 1 - \frac{TR}{T_{10}{}'}$. By substituting these equations into Equation 6-2, we could get the relationship between corrected T_{10}' and uncorrected T_{10}

Equation 6-3

$$\boldsymbol{T_{10} \approx (1 - k^2)TR + k^2 T_{10}{}'}$$

With a small TR (0.004s in our protocol) and k close to 1, the first term on the right side of Equation 6-3 is close to zero. The relationship can be further simplified as

Equation 6-4

$$\boldsymbol{\frac{T_{10}{}'}{T_{10}} \approx \frac{1}{k^2}}$$

In the process of dynamic T_1 quantification, based on $\frac{S(t)}{S_0} = \frac{(1-E_1(t))(1-E_{10}cos\beta)}{(1-E_{10})(1-E_1(t)cos\beta)}$, $T_1(t)$ at each time point is calculated from T_{10} and the ratio between signal at baseline S_0 and signal at the corresponding time point S(t). In this process, the ratios between signals are given as constant, as expressed in

Equation 6-5

$$\frac{\boldsymbol{S'(t)}}{\boldsymbol{S'_0}} = \frac{\boldsymbol{S(t)}}{\boldsymbol{S_0}}$$

After substituting Equation 2-1 into Equation 6-6 for all four signals and reformatting, we could get the following equation

Equation 6-6

$$\frac{[\mathbf{1} - \boldsymbol{E_1(t)}][\mathbf{1} - \boldsymbol{E'_1(t)cosk\beta}]}{[\mathbf{1} - \boldsymbol{E_1}'\boldsymbol{(t)}][\mathbf{1} - \boldsymbol{E_1(t)cos\beta}]} = \frac{[\mathbf{1} - \boldsymbol{E_{10}}][\mathbf{1} - \boldsymbol{E'_{10}cosk\beta}]}{[\mathbf{1} - \boldsymbol{E_{10}}'][\mathbf{1} - \boldsymbol{E_{10}cos\beta}]}$$

Based on the similar assumptions of small flip angle β, we could get $sin\beta \approx \beta, cos\beta \approx 1 - \frac{\beta^2}{2}$. By substituting those equations and Equation 6-2 into Equation 6-7 we could prove that

Equation 6-7

$$\frac{[\mathbf{1} - \boldsymbol{E_1(t)}][\mathbf{1} - \boldsymbol{E'_1(t)cosk\beta}]}{[\mathbf{1} - \boldsymbol{E_1}'\boldsymbol{(t)}][\mathbf{1} - \boldsymbol{E_1(t)cos\beta}]} \approx \mathbf{1}$$

Then similarly, based on assumption of $\frac{\mathrm{TR}}{\mathrm{T}_1(t)} \approx 0$ and $\beta \approx 0$, Equation 6-8 could be further simplified as

Equation 6-8

$$\boldsymbol{E_1(t)} \approx \mathbf{1} - \frac{\mathbf{1} - \boldsymbol{E_1}'\boldsymbol{(t)}}{\mathbf{1} - \boldsymbol{E'_1(t)} + \boldsymbol{k^2 E_1}'\boldsymbol{(t)}}$$

We can find the similarity between Equation 6-8 and Equation 6-2. Correspondingly as shown in Equation 6-4, with assumptions of small TR and k^2 close to 1, we could further simplify Equation 6-8 to the following Equation:

Equation 6-9

$$\frac{\boldsymbol{T_1}'(\boldsymbol{t})}{\boldsymbol{T_1}(\boldsymbol{t})} \approx \frac{\mathbf{1}}{\boldsymbol{k^2}}$$

Now we got the relationship of corrected and uncorrected T_1 values before and after contrast agent, and from the linearity of contrast agent shown in Equation 3-2, it could be easily derived that the relationship between corrected and uncorrected tissue contrast agent concentration C_t'(t) and C_t(t):

Equation 6-10

$$\frac{\boldsymbol{C_t}'(\boldsymbol{t})}{\boldsymbol{C_t}(\boldsymbol{t})} \approx \boldsymbol{k^2}$$

As the last step, according to Equation 3-3 of standard Tofts model[102], with a fixed C_p(t) which is not influenced by B_1^+, ratio between uncorrected and corrected K^{trans} is equal to the ratio of tissue contrast agent concentration.

Equation 6-11

$$\frac{K^{trans\prime}}{K^{trans}} = \frac{C_t{}'(t)}{C_t(t)} \approx k^2$$

As the integration following K^{trans} in Tofts model is not related to B_1^+, namely $\frac{K^{trans,}}{v_e{}'} = \frac{K^{trans}}{v_e}$, the ratio between v_e would be

Equation 6-12

$$\frac{v_e{}'}{v_e} = \frac{K^{trans\prime}}{K^{trans}} \approx k^2$$

In summary, based on three basic assumptions during the whole acquisition process: 1) small flip angle, 2) small TR and T_1 ratio, 3) k is close to 1, under standard Tofts model and population-averaged AIF condition, relationships between corrected and uncorrected PK parameters can be simplified to $\frac{K^{trans}{}'}{K^{trans}} \approx k^2$, $\frac{v_e{}'}{v_e} \approx k^2$.

7 REFERENCES

1. Siegel RL, Miller KD, Jemal A. Cancer statistics, 2016. *CA Cancer J Clin*. 2016;66(1):7-30. doi:10.3322/caac.21332

2. Ahmed HU, Akin O, Coleman JA, et al. Transatlantic Consensus Group on active surveillance and focal therapy for prostate cancer. *BJU Int*. 2012;109(11):1636-1647. doi:10.1111/j.1464-410X.2011.10633.x

3. Schröder FH, Hugosson J, Roobol MJ, et al. Screening and Prostate-Cancer Mortality in a Randomized European Study. *N Engl J Med*. 2009;360(13):1320-1328. doi:10.1056/NEJMoa0810084

4. Ahmed HU, El-Shater Bosaily A, Brown LC, et al. Diagnostic accuracy of multi-parametric MRI and TRUS biopsy in prostate cancer (PROMIS): a paired validating confirmatory study. *Lancet*. 2017;389(10071):815-822. doi:10.1016/S0140-6736(16)32401-1

5. Caster JM, Falchook AD, Hendrix LH, Chen RC. Risk of Pathologic Upgrading or Locally Advanced Disease in Early Prostate Cancer Patients Based on Biopsy Gleason Score and PSA: A Population-Based Study of Modern Patients. *Int J Radiat Oncol Biol Phys*. 2015;92(2):244-251. doi:10.1016/j.ijrobp.2015.01.051

6. Cohen MS, Hanley RS, Kurteva T, et al. Comparing the Gleason Prostate Biopsy and Gleason Prostatectomy Grading System: The Lahey Clinic Medical Center Experience and an International Meta-Analysis. *Eur Urol*. 2008;54(2):371-381. doi:10.1016/j.eururo.2008.03.049

7. Kasivisvanathan V, Rannikko AS, Borghi M, et al. MRI-Targeted or Standard Biopsy for Prostate-Cancer Diagnosis. *N Engl J Med*. 2018;378(19):1767-1777. doi:10.1056/NEJMoa1801993

8. Hoeks CMA, Barentsz JO, Hambrock T, et al. Prostate cancer: multiparametric MR imaging for detection, localization, and staging. *Radiology*. 2011;261(1):46-66. doi:10.1148/radiol.11091822

9. Tan N, Margolis DJ, Lu DY, et al. Characteristics of detected and missed prostate cancer foci on 3-T multiparametric MRI using an endorectal coil correlated with whole-mount thin-section histopathology. *Am J Roentgenol*. 2015;205(1):W87-W92. doi:10.2214/AJR.14.13285

10. Litjens GJS, Barentsz JO, Karssemeijer N, Huisman HJ. Clinical evaluation of a computer-aided diagnosis system for determining cancer aggressiveness in prostate MRI. *Eur Radiol*. 2015:3187-3199. doi:10.1007/s00330-015-3743-y

11. ACM. Pi-Rads Prostate Imaging - Reporting and Data System. *Am Coll Radiol*. 2015. http://www.acr.org/~/media/ACR/Documents/PDF/QualitySafety/Resources/PIRADS/PIRADS V2.pdf.

12. Giannarini G, Girometti R, Sioletic S, et al. Inter-reader agreement of Prostate Imaging Reporting and Data System version 2 in detecting prostate cancer on 3 Tesla multiparametric MRI: A prospective study on patients referred to radical prostatectomy. *Eur Urol Suppl*. 2018;17(2):e893. doi:10.1016/s1569-9056(18)31456-8

13. Zhang L, Tang M, Chen S, Lei X, Zhang X, Huan Y. A meta-analysis of use of Prostate Imaging Reporting and Data System Version 2 (PI-RADS V2) with multiparametric MR

imaging for the detection of prostate cancer. *Eur Radiol*. 2017;27(12):5204-5214. doi:10.1007/s00330-017-4843-7

14. Girometti R, Giannarini G, Greco F, et al. Interreader agreement of PI-RADS v. 2 in assessing prostate cancer with multiparametric MRI: A study using whole-mount histology as the standard of reference. *J Magn Reson Imaging*. 2019;49(2):546-555. doi:10.1002/jmri.26220

15. Stolk TT, de Jong IJ, Kwee TC, et al. False positives in PIRADS (V2) 3, 4, and 5 lesions: relationship with reader experience and zonal location. *Abdom Radiol*. 2019;44(3):1044-1051. doi:10.1007/s00261-019-01919-2

16. Franiel T, Hamm B, Hricak H. Dynamic contrast-enhanced magnetic resonance imaging and pharmacokinetic models in prostate cancer. *Eur Radiol*. 2011;21(3):616-626. doi:10.1007/s00330-010-2037-7

17. Ren J, Huan Y, Wang H, et al. Dynamic contrast-enhanced MRI of benign prostatic hyperplasia and prostatic carcinoma: correlation with angiogenesis. *Clin Radiol*. 2008;63(2):153-159. doi:10.1016/J.CRAD.2007.07.023

18. Isebaert S, De Keyzer F, Haustermans K, et al. Evaluation of semi-quantitative dynamic contrast-enhanced MRI parameters for prostate cancer in correlation to whole-mount histopathology. *Eur J Radiol*. 2012;81(3):e217-e222. doi:10.1016/j.ejrad.2011.01.107

19. Alonzi R, Padhani AR, Allen C. Dynamic contrast enhanced MRI in prostate cancer. *Eur J Radiol*. 2007;63(3):335-350. doi:10.1016/j.ejrad.2007.06.028

20. Kamrava M, Kishan AU, Margolis DJ, et al. Multiparametric magnetic resonance imaging for prostate cancer improves Gleason score assessment in favorable risk prostate cancer.

Pract Radiat Oncol. 2015;5(6):411-416. doi:10.1016/j.prro.2015.04.006

21. Khalifa F, Soliman A, El-Baz A, et al. Models and methods for analyzing DCE-MRI: A review. *Med Phys*. 2014;41(12):124301. doi:10.1118/1.4898202

22. Ocak I, Bernardo M, Metzger G, et al. Dynamic Contrast-Enhanced MRI of Prostate Cancer at 3 T: A Study of Pharmacokinetic Parameters. *Am J Roentgenol*. 2007;189(4):W192-W201. doi:10.2214/AJR.06.1329

23. Chung MP, Margolis D, Mesko S, Wang J, Kupelian P, Kamrava M. Correlation of quantitative diffusion-weighted and dynamic contrast-enhanced MRI parameters with prognostic factors in prostate cancer. 2014;58:588-594. doi:10.1111/1754-9485.12230

24. Vos EK, Litjens GJS, Kobus T, et al. Assessment of Prostate Cancer Aggressiveness Using Dynamic Contrast-enhanced Magnetic Resonance Imaging at 3 T. *Eur Urol*. 2013;64(3):448-455. doi:10.1016/j.eururo.2013.05.045

25. Li C, Chen M, Li S, Zhao X, Zhang C. Detection of prostate cancer in peripheral zone : comparison of MR diffusion tensor imaging , quantitative dynamic contrast-enhanced MRI , and the two techniques combined at 3 . 0 T. 2014;55(2):239-247. doi:10.1177/0284185113494978

26. Langer DL, van der Kwast TH, Evans AJ, Trachtenberg J, Wilson BC, Haider MA. Prostate cancer detection with multi-parametric MRI: Logistic regression analysis of quantitative T2, diffusion-weighted imaging, and dynamic contrast-enhanced MRI. *J Magn Reson Imaging*. 2009;30(2):327-334. doi:10.1002/jmri.21824

27. Akin O, Gultekin DH, Vargas HA, et al. Incremental value of diffusion weighted and dynamic contrast enhanced MRI in the detection of locally recurrent prostate cancer after

radiation treatment : preliminary results. 2011:1970-1978. doi:10.1007/s00330-011-2130-6

28. Fennessy FM, McKay RR, Beard CJ, Taplin M-E, Tempany CM. Dynamic Contrast-Enhanced Magnetic Resonance Imaging in Prostate Cancer Clinical Trials: Potential Roles and Possible Pitfalls. *Transl Oncol*. 2014;7(1):120-129. doi:10.1593/tlo.13922

29. Azahaf M, Haberley M, Betrouni N, et al. Impact of arterial input function selection on the accuracy of dynamic contrast-enhanced MRI quantitative analysis for the diagnosis of clinically significant prostate cancer. *J Magn Reson Imaging*. 2016;43(3):737-749. doi:10.1002/jmri.25034

30. Sung K, Nayak KS. Measurement and characterization of RF nonuniformity over the heart at 3T using body coil transmission. *J Magn Reson Imaging*. 2008;27(3):643-648. doi:10.1002/jmri.21253

31. Rangwala NA, Dregely I, Wu HH, Sung K. Optimization and evaluation of reference region variable flip angle (RR-VFA) B1+ and T 1 Mapping in the Prostate at 3T. *J Magn Reson Imaging*. 2017;45(3):751-760. doi:10.1002/jmri.25410

32. Cunningham CH, Pauly JM, Nayak KS. Saturated double-angle method for rapidB1+ mapping. *Magn Reson Med*. 2006;55(6):1326-1333. doi:10.1002/mrm.20896

33. Insko EK, Bolinger L. Mapping of the Radiofrequency Field. *J Magn Reson Ser A*. 1993;103(1):82-85. doi:10.1006/jmra.1993.1133

34. Akoka S, Franconi F, Seguin F, Le Pape A. Radiofrequency map of an NMR coil by imaging. *Magn Reson Imaging*. 1993;11(3):437-441. doi:10.1016/0730-725X(93)90078-R

35. Jiru F, Klose U. Fast 3D radiofrequency field mapping using echo-planar imaging. *Magn Reson Med*. 2006;56(6):1375-1379. doi:10.1002/mrm.21083

36. Yarnykh VL. Actual flip-angle imaging in the pulsed steady state: A method for rapid three-dimensional mapping of the transmitted radiofrequency field. *Magn Reson Med*. 2007;57(1):192-200. doi:10.1002/mrm.21120

37. Chung S, Kim D, Breton E, Axel L. Rapid B1+ mapping using a preconditioning RF pulse with turboFLASH readout. *Magn Reson Med*. 2010;64(2):439-446. doi:10.1002/mrm.22423

38. Sacolick LI, Wiesinger F, Hancu I, Vogel MW. B 1 mapping by Bloch-Siegert shift. *Magn Reson Med*. 2010;63(5):1315-1322. doi:10.1002/mrm.22357

39. Morrell GR. A phase-sensitive method of flip angle mapping. *Magn Reson Med*. 2008;60(4):889-894. doi:10.1002/mrm.21729

40. Bedair R, Graves MJ, Patterson AJ, et al. Effect of Radiofrequency Transmit Field Correction on Quantitative Dynamic Contrast- enhanced MR Imaging of the. 2016;000(0):1-10.

41. Rampun A, Tiddeman B, Zwiggelaar R, Malcolm P. Computer aided diagnosis of prostate cancer: A texton based approach. *Med Phys*. 2016;43(10):5412-5425. doi:10.1118/1.4962031

42. Fehr D, Veeraraghavan H, Wibmer A, et al. Automatic classification of prostate cancer Gleason scores from multiparametric magnetic resonance images. *Proc Natl Acad Sci*. 2015:201505935. doi:10.1073/pnas.1505935112

43. Wang S, Burtt K, Turkbey B, Choyke P, Summers RM. Computer aided-diagnosis of prostate cancer on multiparametric MRI: a technical review of current research. *Biomed Res Int*. 2014;2014:789561. doi:10.1155/2014/789561

44. Moradi M, Salcudean SE, Chang SD, et al. Multiparametric MRI maps for detection and grading of dominant prostate tumors. *J Magn Reson Imaging*. 2012;35(6):1403-1413. doi:10.1002/jmri.23540

45. Citak-Er F, Vural M, Acar O, Esen T, Onay A, Ozturk-Isik E. Final Gleason Score Prediction Using Discriminant Analysis and Support Vector Machine Based on Preoperative Multiparametric MR Imaging of Prostate Cancer at 3T. *Biomed Res Int*. 2014;2014:1-9. doi:10.1155/2014/690787

46. LeCun Y, Bengio Y, Hinton G. Deep learning. *Nature*. 2015;521(7553):436-444. http://dx.doi.org/10.1038/nature14539.

47. Tajbakhsh N, Shin JY, Gurudu SR, et al. Convolutional Neural Networks for Medical Image Analysis: Full Training or Fine Tuning? *IEEE Trans Med Imaging*. 2016;35(5):1299-1312. doi:10.1109/TMI.2016.2535302

48. Huang W, Chen Y, Fedorov A, et al. The Impact of Arterial Input Function Determination Variations on Prostate Dynamic Contrast-Enhanced Magnetic Resonance Imaging Pharmacokinetic Modeling: A Multicenter Data Analysis Challenge. *Tomography*. 2016;2(1):56-66. doi:10.18383/j.tom.2015.00184

49. Gupta RK. A new look at the method of variable nutation angle for the measurement of spin-lattice relaxation times using fourier transform NMR. *J Magn Reson*. 1977;25(1):231-235. doi:10.1016/0022-2364(77)90138-X

50. Homer J, Beevers MS. Driven-equilibrium single-pulse observation of T1 relaxation. A reevaluation of a rapid “new” method for determining NMR spin-lattice relaxation times. *J Magn Reson*. 1985;63(2):287-297. doi:10.1016/0022-2364(85)90318-X

51. Treier R, Steingoetter A, Fried M, Schwizer W, Boesiger P. Optimized and combined T1 and B1 mapping technique for fast and accurateT1 quantification in contrast-enhanced abdominal MRI. *Magn Reson Med*. 2007;57(3):568-576. doi:10.1002/mrm.21177

52. Sung K, Daniel BL, Hargreaves BA. Transmit B1+ field inhomogeneity and T1 estimation errors in breast DCE-MRI at 3 tesla. *J Magn Reson Imaging*. 2013;38(2):454-459. doi:10.1002/jmri.23996

53. Tsai W-C, Kao K-J, Chang K-M, et al. B1 Field Correction of T1 Estimation Should Be Considered for Breast Dynamic Contrast-enhanced MR Imaging Even at 1.5 T. *Radiology*. 2017;282(1):55-62. doi:10.1148/radiol.2016160062

54. Zhong X, Martin T, Wu HH, Nayak KS, Sung K. Prostate DCE-MRI with B1+ correction using an approximated analytical approach. *Magn Reson Med*. 2018;00(March):1-13. doi:10.1002/mrm.27232

55. Sacolick LI, Wiesinger F, Hancu I, Vogel MW. B1 mapping by Bloch-Siegert shift. *Magn Reson Med*. 2010;63(5):1315-1322.

56. Nehrke K, Bornert P. DREAM-a novel approach for robust, ultrafast, multislice B1 mapping. *Magn Reson Med*. 2012;68(5):1517-1526. doi:10.1002/mrm.24158

57. Sung K, Saranathan M, Daniel BL, Hargreaves BA. Simultaneous T 1 and B 1 + Mapping Using Reference Region Variable Flip Angle Imaging. *Magn Reson Med*. 2013;70(4):954-961. doi:10.1002/mrm.24904

58. Tofts PS. Modeling tracer kinetics in dynamic Gd-DTPA MR imaging. *J Magn Reson Imaging*. 1997;7(1):91-101. doi:10.1002/jmri.1880070113

59. Deoni SCL, Rutt BK, Peters TM. Rapid combinedT1 andT2 mapping using gradient

recalled acquisition in the steady state. *Magn Reson Med*. 2003;49(3):515-526. doi:10.1002/mrm.10407

60. Panagiotelis I, Blasche M. TrueForm RF Design. http://clinical-mri.com/wp-content/uploads/software_hardware_updates/Reprint_TrueForm_Flash_final.pdf.

61. Otsu N. A Threshold Selection Method from Gray-Level Histograms. *IEEE Trans Syst Man Cybern*. 1979;9(1):62-66. doi:10.1109/TSMC.1979.4310076

62. Obuchowski NA, Reeves AP, Huang EP, et al. Quantitative imaging biomarkers: A review of statistical methods for computer algorithm comparisons. *Stat Methods Med Res*. 2015;24(1):68-106. doi:10.1177/0962280214537390

63. Stikov N, Boudreau M, Levesque IR, Tardif CL, Barral JK, Pike GB. On the accuracy of T1mapping: Searching for common ground. *Magn Reson Med*. 2015;73(2):514-522. doi:10.1002/mrm.25135

64. Litjens G, Debats O, Barentsz J, Karssemeijer N, Huisman H. Computer-aided detection of prostate cancer in MRI. *IEEE Trans Med Imaging*. 2014;33(5):1083-1092. doi:10.1109/TMI.2014.2303821

65. Tofts P. T1-weighted DCE Imaging Concepts: Modelling, Acquisition and Analysis. *Signal*. 2010;500(450):400.

66. Subashi E, Choudhury KR, Johnson GA. An analysis of the uncertainty and bias in DCE-MRI measurements using the spoiled gradient-recalled echo pulse sequence. *Med Phys*. 2014;41(3):032301-1-8. doi:10.1118/1.4865790

67. Soher BJ, Dale BM, Merkle EM. A Review of MR Physics: 3T versus 1.5T. *Magn Reson Imaging Clin N Am*. 2007;15(3):277-290. doi:10.1016/j.mric.2007.06.002

68. Di Giovanni P, Azlan CA, Ahearn TS, Semple SI, Gilbert FJ, Redpath TW. The accuracy of pharmacokinetic parameter measurement in DCE-MRI of the breast at 3 T. *Phys Med Biol*. 2010;55(1):121-132. doi:10.1088/0031-9155/55/1/008

69. Zhong X, Rangwala N, Raman S, Margolis D, Wu H, Sung K. B1+ Inhomogeneity Correction for Estimation of Pharmacokinetic Parameters through an Approximation Approach. *Proc 24th Sci Meet Int Soc Magn Reson Med*. 2016:2491.

70. Bosca RJ, Jackson EF. Creating an anthropomorphic digital MR phantom—an extensible tool for comparing and evaluating quantitative imaging algorithms. *Phys Med Biol*. 2016;61(2):974-982. doi:10.1088/0031-9155/61/2/974

71. Mehrtash A, Gupta SN, Shanbhag D, et al. Bolus arrival time and its effect on tissue characterization with dynamic contrast-enhanced magnetic resonance imaging. *J Med Imaging*. 2016;3(1):014503. doi:10.1117/1.JMI.3.1.014503

72. Zhong X, Wu H, Nayak K, Sung K. Evaluation of Approximation Method for B1+ Correction using Digital Reference Object in Prostate DCE-MRI. *Proc 25th Sci Meet Int Soc Magn Reson Med*. 2017:622.

73. Coleman TF, Li Y. An Interior Trust Region Approach for Nonlinear Minimization Subject to Bounds. *SIAM J Optim*. 1996;6(2):418-445. doi:10.1137/0806023

74. QIBA Content | Daniel P. Barboriak Lab. https://sites.duke.edu/dblab/qibacontent/.

75. Parker GJM, Roberts C, Macdonald A, et al. Experimentally-derived functional form for a population-averaged high-temporal-resolution arterial input function for dynamic contrast-enhanced MRI. *Magn Reson Med*. 2006;56(5):993-1000. doi:10.1002/mrm.21066

76. Weinmann HJ, Laniado M, Mützel W. Pharmacokinetics of GdDTPA/dimeglumine after

intravenous injection into healthy volunteers. *Physiol Chem Phys Med NMR*. 1984;16(2):167-172. http://www.ncbi.nlm.nih.gov/pubmed/6505043.

77. Fritz-Hansen T, Rostrup E, Larsson HBW, Søndergaard L, Ring P, Henriksen O. Measurement of the arterial concentration of Gd-DTPA using MRI: A step toward quantitative perfusion imaging. *Magn Reson Med*. 1996;36(2):225-231. doi:10.1002/mrm.1910360209

78. Peng Y, Jiang Y, Yang C, et al. Quantitative Analysis of Multiparametric Prostate MR Images: Differentiation between Prostate Cancer and Normal Tissue and Correlation with Gleason Score—A Computer-aided Diagnosis Development Study. *Radiology*. 2013;267(3):787-796. doi:10.1148/radiol.13121454

79. Ma J. Breath-hold water and fat imaging using a dual-echo two-point dixon technique with an efficient and robust phase-correction algorithm. *Magn Reson Med*. 2004;52(2):415-419. doi:10.1002/mrm.20146

80. Orton MR, D'Arcy JA, Walker-Samuel S, et al. Computationally efficient vascular input function models for quantitative kinetic modelling using DCE-MRI. *Phys Med Biol*. 2008;53(5):1225-1239. doi:10.1088/0031-9155/53/5/005

81. Portalez D, Mozer P, Cornud F, et al. Validation of the European Society of Urogenital Radiology Scoring System for Prostate Cancer Diagnosis on Multiparametric Magnetic Resonance Imaging in a Cohort of Repeat Biopsy Patients. *Eur Urol*. 2012;62(6):986-996. doi:10.1016/j.eururo.2012.06.044

82. de Bazelaire CMJ, Duhamel GD, Rofsky NM, Alsop DC. MR Imaging Relaxation Times of Abdominal and Pelvic Tissues Measured in Vivo at 3.0 T: Preliminary Results. *Radiology*.

2004;230(3):652-659. doi:10.1148/radiol.2303021331

83. Chao S-L, Metens T, Lemort M. TumourMetrics : a comprehensive clinical solution for the standardization of DCE-MRI analysis in research and routine use. *Quant Imaging Med Surg*. 2017;7(5):1-15. doi:10.21037/qims.2017.09.02

84. Port RE, Knopp M V., Brix G. Dynamic contrast-enhanced MRI using Gd-DTPA: Interindividual variability of the arterial input function and consequences for the assessment of kinetics in tumors. *Magn Reson Med*. 2001;45(6):1030-1038. doi:10.1002/mrm.1137

85. Siegel RL, Miller KD, Jemal A. Cancer statistics, 2018. *CA Cancer J Clin*. 2018;68(1):7-30. doi:10.3322/caac.21442

86. Lavery HJ, Droller MJ. Do gleason patterns 3 and 4 prostate cancer represent separate disease states? *J Urol*. 2012;188(5):1667-1675. doi:10.1016/j.juro.2012.07.055

87. Felker ER, Raman SS, Margolis DJ, et al. Risk Stratification Among Men With Prostate Imaging Reporting and Data System version 2 Category 3 Transition Zone Lesions: Is Biopsy Always Necessary? *Am J Roentgenol*. 2017;209(6):1272-1277. doi:10.2214/AJR.17.18008

88. Wu S, Zhong S, Liu Y. Deep residual learning for image steganalysis. *Multimedia Tools and Applications*. http://arxiv.org/abs/1512.03385. Published December 10, 2017. Accessed August 18, 2017.

89. Le MH, Chen J, Wang L, et al. Automated diagnosis of prostate cancer in multi-parametric MRI based on multimodal convolutional neural networks. *Phys Med Biol*. 2017;62(16):6497-6514. doi:10.1088/1361-6560/aa7731

90. Krizhevsky A. Learning Multiple Layers of Features from Tiny Images. 2009.

https://www.cs.toronto.edu/~kriz/learning-features-2009-TR.pdf. Accessed March 27, 2018.

91. Jia Y, Shelhamer E, Donahue J, et al. Caffe: Convolutional Architecture for Fast Feature Embedding. June 2014. http://arxiv.org/abs/1408.5093. Accessed April 1, 2018.

92. Shin H-C, Roth HR, Gao M, et al. Deep Convolutional Neural Networks for Computer-Aided Detection: CNN Architectures, Dataset Characteristics and Transfer Learning. February 2016. http://arxiv.org/abs/1602.03409. Accessed March 28, 2018.

93. He K, Zhang X, Ren S, Sun J. Delving Deep into Rectifiers: Surpassing Human-Level Performance on ImageNet Classification. February 2015. http://arxiv.org/abs/1502.01852. Accessed April 27, 2018.

94. Robin X, Turck N, Hainard A, et al. pROC: an open-source package for R and S+ to analyze and compare ROC curves. *BMC Bioinformatics*. 2011;12(1):77. doi:10.1186/1471-2105-12-77

95. R Core Team. R: A language and environment for statistical computing. *R Found Stat Comput Vienna, Austria*. 2014;(2015):2014. doi:10.1017/CBO9781107415324.004

96. Zhao C, Gao G, Fang D, et al. The efficiency of multiparametric magnetic resonance imaging (mpMRI) using PI-RADS Version 2 in the diagnosis of clinically significant prostate cancer. 2016. doi:10.1016/j.clinimag.2016.04.010

97. Breiman L. Random forests. *Mach Learn*. 2001;45(1):5-32. doi:10.1023/A:1010933404324

98. Ishioka J, Matsuoka Y, Uehara S, et al. Computer-aided diagnosis of prostate cancer on magnetic resonance imaging using a convolutional neural network algorithm. *BJU Int*. 2018;122(3):411-417. doi:10.1111/bju.14397

99. Tsehay YK, Lay NS, Roth HR, et al. Convolutional neural network based deep-learning architecture for prostate cancer detection on multiparametric magnetic resonance images. In: Armato SG, Petrick NA, eds. Vol 10134. International Society for Optics and Photonics; 2017:1013405. doi:10.1117/12.2254423

100. Griffiths DFR, Melia J, McWilliam LJ, et al. A study of Gleason score interpretation in different groups of UK pathologists; techniques for improving reproducibility. *Histopathology*. 2006;48(6):655-662. doi:10.1111/j.1365-2559.2006.02394.x

101. Pritchard CC, Mateo J, Walsh MF, et al. Inherited DNA-Repair Gene Mutations in Men with Metastatic Prostate Cancer. *N Engl J Med*. 2016;375(5):443-453. doi:10.1056/NEJMoa1603144

102. Tofts PS, Kermode AG. Measurement of the blood-brain barrier permeability and leakage space using dynamic MR imaging. 1. Fundamental concepts. *Magn Reson Med*. 1991;17(2):357-367. doi:10.1002/mrm.1910170208