THE UNIVERSITY OF CHICAGO

MRI RADIOMICS FOR IMAGING GENOMICS, RISK CLASSIFICATION, AND PREDICTION OF DISEASE PROGRESSION IN AUTOSOMAL DOMINANT POLYCYSTIC KIDNEY DISEASE

A DISSERTATION SUBMITTED TO
THE FACULTY OF THE DIVISION OF THE BIOLOGICAL SCIENCES
AND THE PRITZKER SCHOOL OF MEDICINE
IN CANDIDACY FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

COMMITTEE ON MEDICAL PHYSICS

BY
LINNEA ELLEN KREMER

CHICAGO, ILLINOIS
JUNE 2024

TABLE OF CONTENTS

LIST OF FIGURES v

LIST OF TABLES xiv

ACKNOWLEDGMENTS xix

ABSTRACT xxi

1 INTRODUCTION 1
1.1 Autosomal Dominant Polycystic Kidney Disease (ADPKD) 1
1.1.1 Risk-stratification 1
1.2 Medical Imaging in ADPKD 3
1.2.1 Magnetic resonance imaging (MRI) 4
1.2.2 Non-cystic kidney parenchyma characterization 7
1.2.3 Magnetic resonance fingerprinting (MRF) 7
1.3 Artificial Intelligence in Medical Imaging 8
1.3.1 Radiomics 9
1.3.2 Radiomics in ADPKD 12
1.4 Research objectives and scope of work 14

2 MRI RADIOMICS ANALYSIS OF NON-CYSTIC KIDNEY PARENCHYMA TO CLASSIFY PKD1 AND PKD2 GENOTYPES 18
2.1 Introduction 18
2.2 Methods 19
2.2.1 Databases for feature reproducibility and genotype classification . . . 19
2.2.2 MR image pre-processing 22
2.2.3 Feature extraction 23
2.2.4 Feature reproducibility 26
2.2.5 Feature selection and classification 27
2.3 Results 27
2.3.1 Reproducibility of radiomic features across pre-processing 27
2.3.2 *PKD1* vs. *PKD2* classification using radiomic features 30
2.3.3 *PKD1* vs. *PKD2* using clinical and radiomic features 42
2.4 Discussion 45
2.5 Conclusion 50

3 MRI RADIOMICS ANALYSIS OF NON-CYSTIC KIDNEY PARENCHYMA TO DIFFERENTIATE AMONG MAYO IMAGING CLASSIFICATION CLASSES . 51
3.1 Introduction 51
3.2 Methods 53
3.2.1 Database 53
3.2.2 U-Net for segmentation 54

3.2.3 MR image pre-processing and feature extraction . . . 57
3.2.4 Correlation of radiomic features with kidney size . . . 57
3.2.5 Feature selection and classification . . . 58
3.2.6 MIC 1B (intermediate-risk) vs. MIC 1C (high-risk) classification . . . 59
3.3 Results . . . 60
3.3.1 Low- and intermediate-risk MIC (1A-1B) vs. high-risk MIC (1C-1E) . . . 60
3.3.2 Impact of feature correlation with kidney size across pre-processing . . . 72
3.3.3 MIC 1B (intermediate-risk) vs. MIC 1C (high-risk) classification . . . 76
3.4 Discussion . . . 80
3.5 Conclusion . . . 82

4 TEMPORAL ASSESSMENT OF MRI RADIOMIC FEATURES TO PREDICT KIDNEY FUNCTION DECLINE IN PATIENTS WITH ADPKD . . . 84
4.1 Introduction . . . 84
4.2 Methods . . . 85
4.2.1 Database . . . 85
4.2.2 MR image selection, segmentation, and pre-processing . . . 87
4.2.3 Feature extraction and predictive model . . . 87
4.3 Results . . . 92
4.3.1 Prediction to ≥CKD Stage 3A and ≥30% reduction in eGFR . . . 92
4.3.2 Radiomic features at timepoints . . . 99
4.4 Discussion . . . 117
4.5 Conclusion . . . 121

5 MAGNETIC RESONANCE FINGERPRINTING RADIOMICS ANALYSIS IN HEALTHY CONTROLS AND ADPKD PATIENTS: A PILOT STUDY . . . 122
5.1 Introduction . . . 122
5.2 Methods . . . 124
5.2.1 Database . . . 124
5.2.2 MRF acquisition and imaging parameters . . . 126
5.2.3 Segmentation . . . 127
5.2.4 Feature extraction and classification . . . 128
5.3 Results . . . 130
5.3.1 Segmentation comparisons . . . 130
5.3.2 Mean T1 and T2 . . . 135
5.3.3 Classification of ADPKD patients vs. a healthy cohort . . . 146
5.4 Discussion . . . 151
5.5 Conclusion . . . 154

6 SUMMARY AND FUTURE DIRECTIONS . . . 155

REFERENCES . . . 161

LIST OF FIGURES

1.1 Magnetic resonance imaging (MRI) of a coronal T1-weighted scan acquired using 3D VIBE/FMPSPGR/LAVA, without fat saturation, and 3 mm slice thickness (left), and coronal T2-weighted scan acquired using SSFSE/HASTE, fat saturation, and a 3 mm fixed slice thickness (right) from the same patient. Note the hyperintense hemorrhagic (blood-filled) cyst in the T1-weighted MR image and the hypointense signal of the same cyst in the T2-weighted MR image. The images shown here have cysts both in the kidneys and liver. 5
1.2 Flowchart of the general radiomics image processing scheme for computing radiomics features. Adapted with permission from Zwanenburg et al. [39]. 10
1.3 An example of fixed bin number gray-level discretization using 8, 16, 32, 64, 128, and 256 gray-level bins. Gray-level discretization is a pre-processing step before feature extraction. The image is a coronal T2-weighted fat-saturated magnetic resonance image of an autosomal dominant polycystic kidney patient with varying number of gray-level bins. 11
1.4 Dissertation outline of aims. 16

2.1 From left to right: (a) representative MRI slice, (b) the result of kidney segmentation, and (c) the result of cyst segmentation. These segmented regions were used for feature extraction for a Mayo Imaging Classification (MIC) 1B *PKD1* patient (top) and a MIC 1B *PKD2* patient (bottom). 25
2.2 Established MRI pre-processing pipeline for investigation on radiomic feature reproducibility and subsequent classification. 26
2.3 Intra-class correlation coefficient (ICC) scores for radiomic features (poor, moderate, good-excellent) across MRI normalizations. Radiomic features extracted from the non-cystic kidney parenchyma were sensitive to pre-processing parameters, with varying reproducibility depending on the parameter. The percentage of features with good-to-excellent ICC scores ranged from 14%-58%, increasing as the number of gray levels available for feature extraction increased. 28
2.4 Percentage (%) of total radiomic features for pairwise comparisons of MRI normalizations and radiomic reproducibility categorized as poor, moderate, and good-to-excellent. 30
2.5 Box and whisker plot of the MRI signal intensity range in arbitrary units (a.u.) across HALTA-PKD images for all patient data and normalization methods after downsampling. Images obtained were from 7 different sites. The box represents the interquartile range, with the central line indicating the median value; the whiskers extend to the minimum and maximum values, while outliers are represented as individual data points. 31
2.6 Area under the receiver operating characteristic curve (AUC) values in classifying genotype using radiomic features extracted from the non-cystic kidney parenchyma using fixed bin size (FBS) discretization. The dotted line at an AUC of 0.5 is random guessing. 34

2.7 Area under the receiver operating characteristic curve (AUC) values in classifying genotype using radiomic features extracted from the non-cystic kidney parenchyma using fixed bin number (FBN) discretization. The dotted line at an AUC of 0.5 is random guessing. 36
2.8 Area under the receiver operating characteristic curve (AUC) values in classifying genotype using radiomic features extracted from the entire kidney using fixed bin size (FBS) discretization. The dotted line at an AUC of 0.5 is random guessing. 38
2.9 Area under the receiver operating characteristic curve (AUC) values in classifying genotype using radiomic features extracted from the entire kidney using fixed bin number (FBN) discretization. The dotted line at an AUC of 0.5 is random guessing. 40
2.10 Box and whisker plot of *PKD1* and *PKD2* estimated glomerular filtration rate (eGFR) values, showing a statistically significant difference using the Wilcoxon rank sum test ($p<0.05$) [68]. The median eGFR [min eGFR, max eGFR] of *PKD1* and *PKD2* was 81.22 [37.01, 141.8] and 85.23 [57.85, 134.3], respectively. The box represents the interquartile range, with the central line indicating the median value; the whiskers extend to the minimum and maximum values, while outliers are represented as individual data points. 43
2.11 Scatter plot showing the relationship between estimated glomerular filtration rate (eGFR) and height-corrected total kidney volume (htTKV) for *PKD1* (shown in red) and *PKD2* (shown in blue). Additionally, linear regression lines representing the Pearson correlation coefficient (r) between eGFR and htTKV are plotted separately for *PKD1* (red) and *PKD2* (blue). 44
2.12 Fitted proper binormal receiver operating characteristic (ROC) curves for genotype classification using eGFR, radiomics, and combined clinical-radiomics models using radiomic features from the entire kidney or non-cystic kidney parenchyma. The legend gives the area under the receiver operating characteristic curve (AUC) along with the 95% confidence interval (CI) of AUC for each classifier. 45

3.1 Coronal MR images of the kidneys of an 18-year-old Mayo Imaging Classification (MIC) class 1A patient (top left), 34-year-old MIC class 1B patient (top right), 17-year-old MIC class 1C patient (bottom left), 23-year-old MIC class 1D patient (bottom center), and 43-year-old MIC class 1E patient (bottom right). MIC classes (based on age and height-corrected total kidney volume (htTKV)) range from low risk (1A) to high risk (1E) for kidney function decline. 52
3.2 U-net architecture (example for 32x32 pixels in the lowest resolution). Each blue box corresponds to a multi-channel feature map. The number of channels is denoted on top of the box. The x-y size is provided at the lower left edge of the box. White boxes represent copied feature maps. The arrows denote the different operations. Reprinted from Ronnebeger, et al. [74]. 54
3.3 Pipeline using the U-Net models (4 in total) to segment left kidney and left non-cystic kidney parenchyma (the same process was done for the right kidney and right non-cystic kidney parenchyma), first starting with the original T2-weighted fat saturated (T2W-FS) MR image. 56

3.4 Box and whisker plot of the MRI signal intensity range in arbitrary units (a.u.) across HALTA-PKD MIC patient (1A-1E) images and normalization methods after downsampling. The non-cystic kidney parenchyma (left) had extreme outliers (n=4) with signal intensities ranging between 4901-13540 using the original image; similarly using the original image, the entire kidney range (right) had extreme outliers (n=5) with ranges between 4596-21805. The box represents the interquartile range, with the central line indicating the median value; the whiskers extend to the minimum and maximum values, while outliers are represented as individual data points. The extreme outliers mentioned are not included in the figure. 61
3.5 Area under the receiver operating characteristic curve (AUC) values in classifying Mayo Imaging Classification (MIC) using radiomic features extracted from the non-cystic kidney parenchyma using fixed bin size (FBS) discretization. The dotted line at an AUC of 0.5 is random guessing. 64
3.6 Area under the receiver operating characteristic curve (AUC) values in classifying Mayo Imaging Classification (MIC) using radiomic features extracted from the non-cystic kidney parenchyma using fixed bin number (FBN) discretization. The dotted line at an AUC of 0.5 is random guessing. 66
3.7 Area under the receiver operating characteristic curve (AUC) values in classifying Mayo Imaging Classification (MIC) using radiomic features extracted from the entire kidney using fixed bin size (FBS) discretization. The dotted line at an AUC of 0.5 is random guessing. 68
3.8 Area under the receiver operating characteristic curve (AUC) values in classifying Mayo Imaging Classification (MIC) using radiomic features extracted from the entire kidney using fixed bin number (FBN) discretization. The dotted line at an AUC of 0.5 is random guessing. 70
3.9 Box and whisker plot of the distribution of kidney size from the representative coronal section (i.e., number of voxels) of low/intermediate- and high-risk Mayo Imaging Classification (MIC) patients, showing a statistically significant difference using the Wilcoxon rank sum test ($p < 0.05$). The box represents the interquartile range, with the central line indicating the median value; the whiskers extend to the minimum and maximum values, while outliers are represented as individual data points. 72
3.10 GLCM joint entropy Spearman rank correlation values ranging from -0.10-0.91 using FBS discretization and 0.29-0.95 using FBN discretization. The red dashed lines represent the moderate ($0.5 < \rho < 0.9$) and high correlation ($\rho > 0.9$) thresholds. 73
3.11 GLDM dependence non-uniformity Spearman rank correlation values ranging from 0.60-0.97 using FBS discretization and 0.88-0.98 using FBN discretization. The red dashed lines represent the moderate ($0.5 < \rho < 0.9$) and high correlation ($\rho > 0.9$) thresholds. 74

3.12 GLRLM run-length non-uniformity Spearman rank correlation values ranging from 0.31-1.0 using FBS discretization and 0.77-1.0 using FBN discretization. The red dashed lines represent the moderate ($0.5 < \rho < 0.9$) and high correlation ($\rho > 0.9$) thresholds. 74
3.13 GLSZM size-zone non-uniformity Spearman rank correlation values ranging from 0.16-0.98 using FBS discretization and 0.49-0.98 using FBN discretization. The red dashed lines represent the moderate ($0.5 < \rho < 0.9$) and high correlation ($\rho > 0.9$) thresholds. 75
3.14 Heatmaps of the number of patients for MIC 1B (left) and MIC 1C (right) with their baseline chronic kidney disease (CKD) stage (x-axis) and CKD stage at 60-months follow-up (y-axis). The color intensity represents the number of patients, with darker shades indicating a higher patient count. 77
3.15 Fuzzy c-means (FCM) clustering using clinical features (age, sex, and eGFR). For simplicity, the figure shows the age on the x-axis and eGFR on the y-axis. The cluster center for 1B is (33.74, 104.71) and for 1C is (41.24, 74.75). The scatter plot shows the cluster membership as well as the true class label of each patient (n=271). 79

4.1 Image selection and feature extraction pipeline used to predict kidney function decline to ≥CKD stage 3A or ≥30% reduction in eGFR at 60 months using radiomic features extracted from MR images at baseline (t_{baseline}), 24 months ($t_{\text{24-month}}$), and 48 months ($t_{\text{48-month}}$). Features were extracted from the 60-month timepoint ($t_{\text{60-month}}$) but were not used in the logistic regression (LR) prediction models. 91
4.2 Area under the receiver operating characteristic curve (AUC) values for the prediction of progression to ≥CKD stage 3A using radiomic features from the non-cystic kidney parenchyma (blue) and the entire kidney (orange). The dotted line at an AUC of 0.5 is random guessing. There were 11 radiomics models investigated using baseline and follow-up timepoints: (1) baseline (B); (2) 24-month delta radiomics (Δ24); (3) 24-month timepoint (24); (4) 48-month delta radiomics (Δ48); (5) 48-month timepoint (48); (6) combined baseline and 24-month delta radiomics (B+Δ24); (7) combined baseline and 24-month radiomics (B+24); (8) combined baseline and 48-month delta radiomics (B+Δ48); (9) combined baseline and 48-month timepoint (B+48); (10) combined baseline, 24-month delta radiomics, and 48-month delta radiomics (B+Δ24+Δ48); and (11) combined baseline 24-month timepoint, and 48-month timepoint(B+24+48). 94

4.3 Area under the receiver operating characteristic curve (AUC) values for prediction to a ≥30% reduction in eGFR using radiomic features from the non-cystic kidney parenchyma (blue) and the entire kidney (orange). The dotted line at an AUC of 0.5 is random guessing. There were 11 radiomics models investigated using baseline and follow-up timepoints: (1) baseline (B); (2) 24-month delta radiomics (Δ24); (3) 24-month timepoint (24); (4) 48-month delta radiomics (Δ48); (5) 48-month timepoint (48); (6) combined baseline and 24-month delta radiomics (B+Δ24); (7) combined baseline and 24-month radiomics (B+24); (8) combined baseline and 48-month delta radiomics (B+Δ48); (9) combined baseline and 48-month timepoint (B+48); (10) combined baseline, 24-month delta radiomics, and 48-month delta radiomics (B+Δ24+Δ48); and (11) combined baseline 24-month timepoint, and 48-month timepoint(B+24+48). 96
4.4 Histogram counts of radiomic features extracted from the non-cystic kidney for prediction of progression to ≥CKD stage 3A. The radiomic features were selected in all single timepoint models of baseline, 24-month, and 48-month timepoints. . 100
4.5 Histogram counts of radiomic features extracted from the entire kidney for prediction of progression to ≥CKD stage 3A. The radiomic features were selected in all single timepoint models of baseline, 24-month, and 48-month timepoints. . . 101
4.6 Scatter plots of radiomic feature GLSZM SZNN from the non-cystic kidney used for the prediction of progression to ≥CKD stage 3A at 24-month, 48-month, and 60-month timepoints, and its corresponding delta-feature from baseline at these timepoints. 103
4.7 Histograms of radiomic feature values of GLSZM SZNN from the non-cystic kidney for the prediction of progression to ≥CKD stage 3A histogram at baseline, 24-month, 48-month, and 60-month timepoints. GLSZM SZNN measures the variability of size zones with a lower value indicating more homogeneity among zone size volumes (i.e., number of connected voxels that share the same gray-level intensity) in an image. 104
4.8 Scatter plots of radiomic feature NGTDM strength from the entire kidney used for the prediction of progression to ≥CKD stage 3A at 24-month, 48-month, and 60-month timepoints, and its corresponding delta-feature from baseline at these timepoints. 106
4.9 Histograms of radiomic feature values of NGTDM strength from the entire kidney for the prediction of progression to ≥CKD stage 3A histogram at baseline, 24-month, 48-month, and 60-month timepoints. NGTDM strength measures primitives in an image with high values indicating a slow change in intensity with more large coarse differences in gray-level intensities. 107
4.10 Histogram counts of radiomic features extracted from the non-cystic kidney for prediction to a ≥30% reduction in eGFR. The radiomic features were selected in all single timepoint models of baseline, 24-month, and 48-month timepoints. . . 109
4.11 Histogram counts of radiomic features extracted from the entire kidney for prediction to a ≥30% reduction in eGFR. The radiomic features were selected in all single timepoint models of baseline, 24-month, and 48-month timepoints. 110

4.12 Scatter plots of radiomic feature GLCM IMC2 from the non-cystic kidney used for the prediction to a ≥30% reduction in eGFR at 24-month, 48-month, and 60-month timepoints, and its corresponding delta-feature from baseline at these timepoints. 112
4.13 Histograms of radiomic feature values of GLCM IMC2 from the non-cystic kidney for the prediction to a ≥30% reduction in eGFR at baseline, 24-month, 48-month, and 60-month timepoints. GLCM IMC2 assesses the correlation between the probability distribution of pixel pairs and describes the complexity of texture, with a value approaching one representing two fully dependent and uniform distributions. 113
4.14 Scatter plots of radiomic feature GLCM joint energy from the entire kidney used for the prediction to a ≥30% reduction in eGFR at 24-month, 48-month, and 60-month timepoints, and its corresponding delta-feature from baseline at these timepoints. 115
4.15 Histograms of radiomic feature values of GLCM joint energy from the entire kidney for the prediction to a ≥30% reduction in eGFR at baseline, 24-month, 48-month, and 60-month timepoints. GLCM joint energy is a measure of homogeneous patterns in the image, with a greater energy having more instances of intensity value pairs that neighbor each other at higher frequencies in the image. 116

5.1 T1 (top) and T2 (bottom) MRF maps of three ADPKD patients ranging from the lowest cyst burden (left) to the highest cyst burden (right). The images shown are cropped to the kidneys from the original image and are in units of milliseconds (ms). 125
5.2 Schematic of the MRF acquisition and postprocessing methodology. The kidney MRF acquisition combines 12 acquisition segments with each having a magnetization preparation (T1, T2, or no preparation), fat suppression (FS), and 144 fast imaging with steady-state free precession (FISP) MRF imaging readouts. The flip angle pattern over the 1728 total MRF data points is shown at the bottom. Vertical dashed lines within the flip angle profile indicate the 12 different segments. Reprinted with permission from MacAskill, et al. [27]. 127
5.3 Box and whisker plots of Dice similarity coefficient (DSC) between two segmenters using segmentations of all available images of the left and right kidney. The images show the red and blue corresponding to the segmentations, and purple is the overlap between the two segmentations. The box represents the interquartile range, with the central line indicating the median value; the whiskers extend to the minimum and maximum values, while outliers are represented as individual data points. 132
5.4 Box and whisker plots of Dice similarity coefficient (DSC) between two segmenters using three central sections of the left and right kidney. The images show the red and blue corresponding to the segmentation, and purple is the overlap. The box represents the interquartile range, with the central line indicating the median value; the whiskers extend to the minimum and maximum values, while outliers are represented as individual data points. 133

5.5 The fuzzy c-means (FCM) cyst removal from the entire kidney from an ADPKD patient with a high total cyst burden (37.99%) (top), an ADPKD patient with a low total cyst burden (1.98%) (middle), and a healthy control (bottom). Note that the FCM cyst removal method segmented the kidney medulla and cortex of the ADPKD patient with a low cyst burden and the healthy control. The colors represent the number of clusters (n=3) in the FCM algorithm: background (red), kidney (green), and cyst (blue). 134
5.6 Box and whisker plots of the adaptive manual T1 threshold value chosen on a per image basis for each patient. The box and whisker plots of each patient are images that presented macroscopic cysts and subsequently used for cyst removal. The red dashed line represents the fixed threshold value of 2160 ms for cyst removal. The box represents the interquartile range, with the central line indicating the median value; the whiskers extend to the minimum and maximum values, while outliers are represented as individual data points. 135
5.7 Box and whisker plots comparing the mean T1 values obtained from the entire kidney and non-cystic (NC) kidney parenchyma of the ADPKD cohort, using all available images and central kidney sections, alongside the healthy controls' kidneys. The cyst removal methods from the ADPKD kidney were (1) a fixed threshold of 2160 ms from T1 maps, (2) an adaptive manual threshold on a per-image basis for each patient from their T1 maps, and (3) a fuzzy c-means (FCM) clustering algorithm. The box represents the interquartile range, with the central line indicating the median value; the whiskers extend to the minimum and maximum values, while outliers are represented as individual data points. . 137
5.8 Box and whisker plots comparing the mean T2 values obtained from the entire kidney and non-cystic (NC) kidney parenchyma of the ADPKD cohort, using all available images and central kidney sections, alongside the healthy controls' kidneys. The cyst removal methods from the ADPKD kidney were (1) a fixed threshold of 2160 ms from T1 maps, (2) an adaptive manual threshold on a per-image basis for each patient from their T1 maps, and (3) a fuzzy c-means (FCM) clustering algorithm. The box represents the interquartile range, with the central line indicating the median value; the whiskers extend to the minimum and maximum values, while outliers are represented as individual data points. . 138
5.9 Mean T2 versus T1 of the entire kidney and non-cystic kidney parenchyma compared with healthy controls using all the images from the ADPKD cohort. Each red data point (kidney) corresponds with a green data point (non-cystic kidney parenchyma) of the same patient. 139
5.10 Mean T2 versus T1 of the entire kidney and non-cystic kidney parenchyma compared with healthy controls using the central kidney sections from the ADPKD cohort. Each red data point (kidney) corresponds with a green data point (non-cystic kidney parenchyma) of the same patient. 140
5.11 Mean T1 (left) and mean T2 (right) of the non-cystic kidney parenchyma versus the entire kidney using all the images from the ADPKD cohort, shown with the fitted Pearson line and Pearson correlation coefficient. 142

5.12 Mean T1 (left) and mean T2 (right) of the non-cystic kidney parenchyma versus the entire kidney using the central kidney sections from the ADPKD cohort, shown with the fitted Pearson line and Pearson correlation coefficient. 142
5.13 Mean T1 of the entire kidney (left) and non-cystic kidney parenchyma (right) with respect to cyst volumes using all images from the ADPKD cohort, shown with the fitted Pearson line and Pearson correlation coefficient. The legend contains the lowest to highest cyst fractions values. 143
5.14 Mean T1 of the entire kidney (left) and non-cystic kidney parenchyma (right) with respect to cyst volumes using the central kidney sections from the ADPKD cohort, shown with the fitted Pearson line and Pearson correlation coefficient. The legend contains the lowest to highest cyst fraction values. 143
5.15 Mean T2 of the entire kidney (left) and non-cystic kidney parenchyma (right) with respect to cyst volumes using all images from the ADPKD cohort, shown with the fitted Pearson line and Pearson correlation coefficient. The legend contains the lowest to highest cyst fraction values. 144
5.16 Mean T2 of the entire kidney (left) and non-cystic kidney parenchyma (right) with respect to cyst volumes using the central kidney sections from the ADPKD cohort, shown with the fitted Pearson line and Pearson correlation coefficient. The legend contains the lowest to highest cyst fraction values. 144
5.17 Height-corrected total kidney volume (htTKV) with respect to cyst volumes, shown with the fitted Pearson line and Pearson correlation coefficient. 14 of the 15 patients had htTKV data available. The legend contains the lowest to highest cyst fraction values. 145
5.18 Body surface area-total kidney volume (BSA-TKV) with respect to cyst volumes, shown with the fitted Pearson line and Pearson correlation coefficient. 14 of the 15 patients had BSA-TKV data available. The legend contains the lowest to highest cyst fraction values. 145
5.19 The radiomic features extracted from the entire kidney of T1 maps that were chosen during feature selection in merging radiomic features for classification. The counts refer to the number of times the radiomic feature was selected across training partitions of the leave-one-out approach. 148
5.20 The radiomic features extracted from the non-cystic kidney parenchyma of T1 maps that were chosen during feature selection in merging radiomic features for classification across all segmentation methods. The counts refer to the number of times the radiomic feature was selected across training partitions of the leave-one-out approach. 148
5.21 Histogram of radiomic features extracted from the non-cystic kidney parenchyma of the ADPKD cohort versus healthy controls using T1 maps (adaptive manual threshold cyst removal). GLCM autocorrelation measures the coarseness of texture (top) and entropy captures the randomness of signal intensities (bottom). 149

5.22 The radiomic features extracted from the entire kidney of T2 maps that were chosen during feature selection in merging radiomic features for classification. The counts refer to the number of times the radiomic feature was selected across training partitions of the leave-one-out approach. 150
5.23 The radiomic features extracted from the non-cystic kidney parenchyma of T2 maps that were chosen during feature selection in merging radiomic features for classification across all segmentation methods. The counts refer to the number of times the radiomic feature was selected across training partitions of the leave-one-out approach. 150

LIST OF TABLES

2.1 Consortium for Radiologic Imaging Studies of Polycystic Kidney Disease (CRISP) *PKD1* and *PKD2* subject characteristics of age and height-corrected total kidney volume (htTKV) with respect to Mayo Imaging Classification (MIC) class. . . . 20

2.2 HALT subject characteristics for classification. A $p < 0.05$ was significant using Wilcoxon rank sum test. Asterisks indicate statistically significant differences. . 21

2.3 Mean intra-class correlation coefficient (ICC) scores for radiomic features across feature families using up-sampling and down-sampling methods. 29

2.4 Area under the receiver operating characteristic curve (AUC) along with the 95% confidence interval (CI) of AUC for features extracted from the non-cystic kidney parenchyma using fixed bin size (FBS) discretization. 33

2.5 Area under the receiver operating characteristic curve (AUC) along with the 95% confidence interval (CI) of AUC for features extracted from the non-cystic kidney parenchyma using fixed bin number (FBN) discretization. 35

2.6 Area under the receiver operating characteristic curve (AUC) along with the 95% confidence interval (CI) of AUC for features extracted from the entire kidney parenchyma using fixed bin size (FBS) discretization. 37

2.7 Area under the receiver operating characteristic curve (AUC) along with the 95% confidence interval (CI) of AUC for features extracted from the entire kidney parenchyma using fixed bin number (FBN) discretization. 39

2.8 The radiomic features chosen in the training phase for any partition in the 5-fold repeated cross-validation (rCV). For the non-cystic kidney parenchyma, the representative pre-processing parameters of z-score normalization, pixel resampling to 2.0x2.0 mm, and fixed bin size (FBS) discretization using 64 gray levels was employed. For the entire kidney, the representative pre-processing parameters of psoas muscle normalization, pixel resampling to 1.0x1.0 mm, and FBS discretization using 64 gray levels was employed. 41

2.9 Area under the receiver operating characteristic curve (AUC) along with the 95% confidence interval (CI) of AUC for genotype classification using estimated glomerular filtration rate (eGFR), radiomic features, and combined clinical-radiomics model. 43

3.1 HALT subject characteristics for the differentiation among Mayo Imaging Classification (MIC) 1A-1E. 53

3.2 U-Net model performance. The images of the Consortium for Radiologic Imaging Studies of Polycystic Kidney Disease (CRISP) dataset (n=14 patients) were used as an external test set in addition to the test set from the HALT dataset (n=35 patients). 57

3.3 Area under the receiver operating characteristic curve (AUC) along with the 95% confidence interval (CI) of AUC for features extracted from the non-cystic kidney parenchyma using fixed bin size (FBS) discretization for Mayo Imaging Classification (MIC) 1A-1B vs. MIC 1C-1E classification. 63

3.4 Area under the receiver operating characteristic curve (AUC) along with the 95% confidence interval (CI) of AUC for features extracted from the non-cystic kidney parenchyma using fixed bin number (FBN) discretization for Mayo Imaging Classification (MIC) 1A-1B vs. MIC 1C-1E classification. 65
3.5 Area under the receiver operating characteristic curve (AUC) along with the 95% confidence interval (CI) of AUC for features extracted from the entire kidney parenchyma using fixed bin size (FBS) discretization for Mayo Imaging Classification (MIC) 1A-1B vs. MIC 1C-1E classification. 67
3.6 Area under the receiver operating characteristic curve (AUC) along with the 95% confidence interval (CI) of AUC for features extracted from the entire kidney parenchyma using fixed bin number (FBN) discretization for Mayo Imaging Classification (MIC) 1A-1B vs. MIC 1C-1E classification. 69
3.7 The radiomic features chosen during the training phase of LASSO feature selection across the 5-fold repeated cross-validation (rCV) in classifying low/intermediate- and high-risk MIC class using a representative pre-processing method (psoas muscle normalization, pixel resampling to 2.0x2.0 mm, and fixed bin number (FBN) discretization using 32 gray levels). 71
3.8 Radiomic features that exhibited moderate ($0.5 < \rho < 0.9$) Spearman rank correlation values with number of voxels across pre-processing parameters. 75
3.9 Area under the receiver operating characteristic curve (AUC) along with the 95% confidence interval (CI) of AUC in the task of classifying patients between low/intermediate- vs. high-risk Mayo Imaging Classification (MIC) classes using radiomic features with respect to their correlation with number of voxels (i.e., kidney size). Asterisks denote significance after accounting for multiple comparisons using Bonferroni-Holm corrections. 76
3.10 HALT subject characteristics for Mayo Imaging Classification (MIC) 1B vs. MIC 1C classification. Asterisks indicate statistically significant differences ($p < 0.05$). 77
3.11 Area under the receiver operating characteristic curve (AUC) along with the 95% confidence interval (CI) of AUC for 1B vs. 1C classification using age, sex, estimated glomerular filtration rate (eGFR), radiomic features, and combined clinical-radiomics model. The combined clinical features incorporate age, sex, and eGFR, but do not include height-corrected total kidney volume (htTKV). Although Mayo Imaging Classification (MIC) classification considers htTKV and age, htTKV was investigated as a standalone feature. 78
3.12 Area under the receiver operating characteristic curve (AUC) along with the 95% confidence interval (CI) of AUC for fuzzy c-means (FCM) clustering memberships of Mayo Imaging Classification (MIC) 1B and MIC 1C patients using age, sex, estimated glomerular filtration rate (eGFR), radiomic features, and combined clinical-radiomics model. 79

4.1 Subject characteristics based on baseline age, baseline sex, baseline estimated glomerular filtration rate (eGFR), and baseline height-corrected total kidney volume (htTKV) for 281 ADPKD patients. The chronic kidney disease (CKD) stage is the patient stage at 60-months follow-up. 86

4.2 A Wilcoxon rank sum test between the subject characteristics at baseline for those who progressed or did not progress to ≥ CKD stage 3A and between those who experienced or did not experience a ≥30% reduction in eGFR ($p < 0.05$ were significant). Asterisks indicate statistically significant differences. The "≥30% reduction in eGFR" cohort contains patients in CKD stages 1-5. 87
4.3 Area under the receiver operating characteristic curve (AUC) along with the 95% confidence interval (CI) of AUC for radiomics-based models to predict progression to ≥CKD stage 3A at 60-month timepoint using single timepoints, delta-radiomics features, and combined timepoints using radiomic features extracted from the non-cystic kidney parenchyma and entire kidney. Asterisks indicate statistically significant differences from baseline model performance (after Bonferroni-Holm correction for 10 multiple comparisons). 93
4.4 Area under the receiver operating characteristic curve (AUC) along with the 95% confidence interval (CI) of AUC for radiomics-based models to predict progression to ≥30% reduction in eGFR at 60-month timepoint using single timepoints, delta-radiomics features, and combined timepoints using radiomic features extracted from the non-cystic kidney parenchyma and entire kidney. Asterisks indicate statistically significant differences from baseline model performance (after Bonferroni-Holm correction for 10 multiple comparisons). 95
4.5 Area under the receiver operating characteristic curve (AUC) along with the 95% confidence interval (CI) of AUC for the clinical and combined clinical-radiomics models to predict kidney function decline at the 60-month timepoint. Asterisks on AUC results indicate performance statistically significant to combined clinical model performance. Asterisks indicate statistically significant differences from baseline model performance (after Bonferroni-Holm correction for 5 multiple comparisons). 98
4.6 The Spearman rank correlation coefficients between the baseline and timepoint feature and between the baseline and delta feature of GLSZM SZNN, used in baseline, 24-month timepoint, and 48-month timepoint models for prediction of progression to ≥CKD stage 3A using the non-cystic kidney for feature extraction. 102
4.7 The p-values obtained from the Wilcoxon rank sum test using radiomic feature GLSZM SZNN from the non-cystic kidney for prediction of progression to ≥CKD stage 3A. A corrected $p < 0.05$ was significant for comparisons with baseline feature values, and asterisks denote statistical significance after multiple comparison corrections. 102
4.8 The Spearman rank correlation coefficient was computed between the baseline and timepoint feature and between the baseline and delta feature of NGTDM strength, used in baseline, 24-month timepoint, and 48-month timepoint models for prediction of progression to ≥CKD stage 3A using the entire kidney for feature extraction. 105

4.9 The p-values obtained from the Wilcoxon rank sum test using radiomic feature NGTDM strength from the entire kidney for prediction of progression to ≥CKD stage 3A. A corrected $p < 0.05$ was significant for comparisons with baseline feature values, and asterisks denote statistical significance after multiple comparison corrections. 105
4.10 The Spearman rank correlation coefficient was computed between the baseline and timepoint feature and between the baseline and delta feature of GLCM IMC2, used in baseline, 24-month timepoint, and 48-month timepoint models for prediction to a ≥30% reduction in eGFR using the non-cystic kidney for feature extraction. 111
4.11 The p-values obtained from the Wilcoxon rank sum test using radiomic feature GLCM IMC2 from the non-cystic kidney for prediction to a ≥30% reduction in eGFR. A corrected $p < 0.05$ was significant for comparisons with baseline feature values, and asterisks denote statistical significance after multiple comparison corrections. 111
4.12 The Spearman rank correlation coefficient was computed between the baseline and timepoint feature and between the baseline and delta feature of GLCM joint energy, used in baseline, 24-month timepoint, and 48-month timepoint models for prediction to a ≥30% reduction in eGFR using the entire kidney for feature extraction. 114
4.13 The p-values obtained from the Wilcoxon rank sum test using radiomic feature GLCM joint energy from the entire kidney for prediction to a ≥30% reduction in eGFR. A corrected $p < 0.05$ was significant for comparisons with baseline feature values, and asterisks denote statistical significance after multiple comparison corrections. 114

5.1 ADPKD subject characteristics (n=14) based on mean age, sex, mean BSA-TKV, and mean htTKV. There was one patient whose clinical information was not completed at the time of this study. 125
5.2 Mean Dice similarity coefficient (DSC) [95% CI] of manual kidney segmentations and subsequent non-cystic kidney parenchyma segmentations between two different segmenters. 131
5.3 Mean Dice similarity coefficient (DSC) [95% CI] of cyst removal methods to obtain the non-cystic kidney parenchyma masks using kidney segmentations from one segmenter (the University of Chicago). 131
5.4 Mean T1 and T2 [95% CI] of manual kidney segmentations and non-cystic kidney segmentations using the cyst removal methods. A $p < 0.05$ was significant from the healthy control cohort using Wilcoxon rank sum test. Asterisks indicate statistically significant differences (after Bonferroni-Holm correction for 8 multiple comparisons). 136

5.5 Area under the receiver operating characteristic curve (AUC) values for differentiating ADPKD children and young adults from a healthy adult cohort using radiomic features from the non-cystic kidney parenchyma and entire kidney from T1 and T2 quantitative MRF maps. The mean T1 and mean T2 values and the top-two performing radiomic features were used for classification. 147

ACKNOWLEDGMENTS

I would like to give glory, honor, and praise to my Lord and Savior Jesus Christ: "For in Him we live and move and have our being" (Acts 17:28).

Thank you to my advisors, Sam Armato and Arlene Chapman, whose support and scientific guidance was critical to the dissertation work and my training as a scientist. I dedicate my doctorate success to their leadership over the last years.

Thank you to my committee members, Maryellen Giger and Steffen Sammet, that added necessary and insightful knowledge for this dissertation to come to fruition.

I would like to thank the Armato Lab members: Roger Engelmann, Feng Li, Mena Shenouda, and Adam Starkey. I would specifically like to thank my labmate, Mena, who provided both scientific and theological perspective in the office.

Thank you to the members of the Chapman Lab; I would like to specifically mention Bill Ding, who completed the image matching for Aim 3.

Thank you to the GPMP faculty and students; I would like to specifically thank my GPMP class Natalie Baughan, Julian Bertini, Hadley DeBrosse, Mira Liu, and Mena Shenouda. In addition, thank you to Sean Foxley's guidance in the beginning of my dissertation work and his expertise in MRI.

Thank you to the undergraduate researchers who joined me during their summer breaks: Natalie Perri, Eliza Sorber, Boris Fosso, Lucy Groothuis, John Trevino, and Madison Orlins.

Thank you to the Chicago KUH Forward TL1 Training Program that provided two years of funding, numerous events around Chicago, and the ability to connect to young researchers. Thank you to the Hiroshi and Yuri Sugiyama Fund for Graduate Studies in the Biological Sciences Division that provided funding of the last two quarters of my PhD.

Thank you to our collaborators at Case Western Reserve University: Katherine Dell, Chris Flask, and Madison Kretzler; their insight in magnetic resonance fingerprinting and excitement for research was refreshing to witness.

I have been a part of deeply enriching communities at the University of Chicago beyond the GPMP, which have shaped me as a believer, a woman, a scientist, a coach, and a friend. Thank you to Amy Reifert, Marlon McKenzie, Rachael Grotts, and all of the fantastic young women on the University of Chicago Women's Soccer Team. Thank you to Graduate Christian Fellowship that provided a spiritual solace on campus. Thank you to Augustana Lutheran Church and Campus Ministry that provided a church home and family for me.

Thank you to the Gamma Alpha Co-Op that provided a safe and affordable home, and the ability to meet people from different academic and personal backgrounds.

Thank you to my physics professors at Roanoke College, Matthew Fleenor and Rama Balasubramanian, who encouraged me to study physics and supported me in research.

Thank you to Susan Francis at the University of Nottingham, who led my master's research and helped my application process at the University of Chicago.

I would like to give thanks for my friends on campus and beyond. Thank you to my girlfriends I met here at the University of Chicago: Andrea Bryant, Nikita Mehta, and Nan Zhou. Their friendships during the good and hard times in my studies will forever be cherished and appreciated. Thank you to my long-time friends from Cincinnati and Roanoke College: Kimberly Janitz, Tessa Tubbs, Abby Finan, Abbie Guarino, Kirstin Figel, Caity Ashley, Grace Young, and others.

Thank you to my immediate family members, whose support is unwavering: Kristin and Chris, Dylan and Jess, Kai and Jojo. Thank you, Kristin (mom), for encouraging me to speak with Maryellen Giger about medical physics, which ultimately led me to apply to the PhD program. Thank you to my grandparents, Bob and Joanna, whose marriage, faith, and friendship have shaped me to be the woman I am today in Christ. Thank you, Bob (grandpa), for your intellectual curiosity and first paving the University of Chicago campus with your footsteps that I too walked 60 years later. Thank you to the rest of the Benne and Kremer family.

ABSTRACT

The overall goal of research was to analyze magnetic resonance imaging (MRI) radiomic features of patients with autosomal dominant polycystic kidney disease (ADPKD) for genotype and risk-stratified classification, temporal assessment of radiomic features for the prediction of kidney function decline, and the analysis of radiomic-based differences from magnetic resonance fingerprinting (MRF)-acquired T1 and T2 maps between a healthy and an ADPKD cohort.

Radiomics, the extraction and analysis of quantitative image features from medical images, has shown the ability to predict a patient's future kidney function decline. However, there has been limited research utilizing radiomics to address other research questions related to ADPKD. Longitudinal clinical studies have shown differences in disease progression and patient outcomes in genetic variants in ADPKD, specifically *PKD1* and *PKD2*, with respect to cyst burden and age at onset of end-stage kidney disease (ESKD). In addition, risk-stratified groups are clinically used to monitor disease progression using height-corrected total kidney volume (htTKV) and age, assigning to patients a Mayo Imaging Classification (MIC) class ranging from low-risk to high-risk of increase in htTKV growth and estimated glomerular filtration rate (eGFR) decline. Delta radiomics, the assessment of temporal change in image features, of longitudinal imaging studies will be used to investigate the additive power of temporal imaging data in predicting future kidney function decline versus baseline texture alone. Radiomic features provided additive power in classification of genetic variants, distinguished among risk-stratified MIC classes, captured cystogenesis, and predicted patient outcomes.

A key component of the radiomics workflow before feature extraction is image preprocessing, which is important for T1-weighted and T2-weighted MRI images due to arbitrary signal intensities. Therefore, MR image normalization has been understood to be an important step before feature extraction to standardize pixel intensities in an image and

improve feature robustness. In addition to normalization, pixel resampling and gray-level discretization are important steps in harmonizing images that come from different sites or timepoints. ADPKD radiomic studies differ in pre-processing methods, and the impact of these steps on radiomic features extracted from the kidney has not been investigated. This work investigated the effect of pre-processing on radiomic features of ADPKD patients. The results indicate that pre-processing parameters influenced the radiomic features extracted from kidney MR images and their subsequent classification.

Quantitative MRI has allowed for anatomic and numeric image analyses, but acquisition time can be long and susceptible to motion artifacts. Magnetic resonance fingerprinting (MRF) allows for repeatable and simultaneous acquisitions (e.g., T1 and T2 quantitative maps) in a single breath hold. Radiomic features were extracted from MRF-acquired T1 and T2 maps of ADPKD children and young adults and a healthy cohort. Radiomic features extracted from MRF-derived quantitative maps distinguished ADPKD patients from a healthy cohort using features extracted from both the entire kidney (including cysts) and non-cystic kidney parenchyma. This work advances non-invasive imaging biomarkers for characterizing the non-cystic kidney parenchyma, ultimately facilitating the identification of ADPKD children at high risk for disease progression.

Specific aims of this thesis were (1) to investigate MRI radiomics analysis of non-cystic kidney parenchyma to classify *PKD1* and *PKD2* genotypes, (2) to use MRI radiomics analysis of non-cystic kidney parenchyma to differentiate among MIC 1A through 1E patients, (3) temporal assessment of MRI radiomic features to predict kidney function decline in patients with ADPKD, and (4) to investigate MRF radiomics analysis in healthy controls and ADPKD patients using T1 and T2 quantitative maps.

CHAPTER 1

INTRODUCTION

1.1 Autosomal Dominant Polycystic Kidney Disease (ADPKD)

Autosomal dominant polycystic kidney disease (ADPKD) is the most common hereditary kidney disorder and is responsible for 10% of patients with end-stage kidney disease (ESKD) under the age of 65 [1, 2]. ADPKD results in gradual enlargement of the kidneys due to cyst growth over decades prior to decline in kidney function and kidney failure [3]. It is common for patients with ADPKD to have hypertension, acute and chronic pain, gross haematuria, cyst infection, and nephrolithiasis and also manifestations beyond the kidney such as hepatic and pancreatic cysts, intracranial aneurysms, abdominal hernias, and cardiac valvular lesions [4]. The Consortium for Radiologic Imaging Studies of Polycystic Kidney Disease (CRISP) was a major longitudinal imaging study of patients with ADPKD, and the results showed that total kidney volume (TKV) predicts the risk of developing kidney function decline in patients with ADPKD [3]; the CRISP study also showed that cyst expansion accounts for increased TKV and that increases in kidney size associate with loss of kidney function, with TKV and total cyst volume (TCV) increasing exponentially. ADPKD is predominantly caused by mutations in two genes, *PKD1* and *PKD2*, that make up over 90% of cases [5]. There is currently no cure for ADPKD, and treatments are limited; the US Food and Drug Administration approved tolvaptan, a vasopressin V2 receptor antagonist, as the first drug treatment in 2018 to slow kidney function decline in adults at risk of rapidly progressing ADPKD [6].

1.1.1 Risk-stratification

Due to the long asymptomatic phase and late age of onset of kidney dysfunction, typical clinical outcome measures (e.g., time to ESKD, doubling of serum creatinine, and 30%

reduction in glomerular filtration rate) are ineffective for identifying high-risk individuals [2]. The long asymptomatic phase is due to hyperfiltration by surviving nephrons that mask the true loss of functioning nephrons resulting from an increase in TKV due to the presence and growth of cysts [7]. The Mayo Imaging Classification (MIC) provides a measure of the rate of height-corrected TKV (htTKV) growth and is a validated risk stratification imaging biomarker for progression to ESKD; MIC has been validated in major clinical studies [8, 9, 10]. MIC uses htTKV and age for classification and is validated for patients between 15-80 years of age; MIC aims to optimize patient selection for enrollment into clinical trials and for available treatments [11]. Currently, there is no risk-stratification for children with ADPKD under 15 years of age. MIC assumes exponential kidney growth, and typical classes range from 1A and 1B (lowest and intermediate risk, respectively) to 1C, 1D, and 1E (highest risk). However, MIC is a simple model used for patients with typical presentations of ADPKD and does not always predict kidney function decline with patients who have a few large or exophytic cysts. While MIC risk stratification was developed independent of genotype, *PKD1* disease severity is significantly different from *PKD2* disease severity, with a difference in age of ESKD onset of over 20 years. The CRISP study demonstrated larger increases in TKV in those with a *PKD1* genotype [3]. *PKD2* patients demonstrated smaller TKV with 40% fewer kidney cysts, consistent with the known delayed onset of hypertension, longer time to ESKD, and a 10-year longer life expectancy [3, 12]. Lavu et al. [13] completed a univariate analysis and found that sex, genotype, baseline MIC, baseline eGFR, and baseline body mass index (BMI) were associated with both a 50% decline in eGFR or ESKD. ADPKD phenotype is heterogeneous across patients, and there is a need to understand characteristics associated with disease severity to help better predict patient outcomes.

1.2 Medical Imaging in ADPKD

Radiologic imaging is crucial for the assessment of ADPKD and risk stratification for progression to ESKD. Medical images are used to measure TKV, and the predominant methods to measure TKV are manual segmentation of whole kidneys and stereology, which uses grid points over the kidney but is not readily available [14]; while obtaining accurate and precise measurements of TKV is important, these methods are time consuming and differ in complexity due to the heterogeneous ADPKD phenotype. An ellipsoid equation (using principal diameters) has been used to quicken the task of TKV measurement by choosing the largest representative mid-slice of the kidney using three imaging views to measure the length, width, and depth of the kidney [15, 16]:

$$TKV(mL) = \frac{\pi}{6} \cdot (L \cdot W \cdot D), \tag{1.1}$$

$$htTKV\left(\frac{mL}{m}\right) = \frac{TKV(mL)}{\text{height}(m)}. \tag{1.2}$$

Computed tomography (CT), magnetic resonance imaging (MRI), and ultrasound (US) are clinically used for measuring TKV and quantifying kidney disease progression in ADPKD. TKV is most-often measured using MRI, which is advantageous because it is a non-ionizing imaging modality with superior soft-tissue contrast and in this case without the need for contrast agents [16]. The accuracy of the ellipsoid equation for TKV measurements using CT and MRI was found to be 87% and US was 21%, and the ellipsoid method for measuring TKV has shown varying reproducibility and variability compared with manual TKV measurements; however, previous research showed that using the ellipsoid equation still accurately determines an ADPKD patient MIC class [16, 17].

1.2.1 Magnetic resonance imaging (MRI)

Nuclear magnetic resonance has been fundamental in its applications to physics, biology, and medicine. MRI uses five major components that impact image quality and clinical utility: (1) main magnet (typical field strength of magnet between 0.5-3.0 Tesla), (2) radio-frequency (RF) transmitting and receiving coils and amplifiers, (3) magnetic field gradient coils and amplifiers, (4) electronic racks controlling gradient and RF waveform generation and signal acquisition, and (5) the computer [18]; MRI can generate images of the human body without ionizing radiation.

The majority of MR imaging is based on the nucleus of hydrogen or the single proton, which is abundant in the human body. The randomly oriented proton spins align parallel to the applied main magnetic field of the scanner and produce a bulk magnetization that precesses about the main magnetic field. RF pulses are applied perpendicular to the main magnetic field at the Larmor frequency, i.e., the precessional frequency of protons, and tilts the synchronous precession of spins away from the longitudinal magnetic field to the transverse plane. Transverse magnetization is generated by an applied alternating magnetic field, and induces a very small signal in sensitive coils tuned to the Larmor frequency. The MR signal formed is spatially encoded using magnetic field gradients, which produce a non-uniform magnetic field using small perturbations to the overall main magnetic field, and the Larmor frequency at the spins location. The raw data is represented in k-space, the spatial frequency domain of the signal, and the spatial frequencies correspond with the gradients in the spatial domain. The center of k-space defines the contrast and basic structures in the MR image and the periphery of k-space provides fine details. The inverse Fourier transform of the signal in k-space produces the MR image.

The weighting of an image is complex but is dependent on the time for the longitudinal magnetization vector of the magnetic moment of protons to return to alignment with the main magnetic field (T1-weighted (T1W)), the time for the transverse magnetization to

decay (T2-weighted (T2W)), and density of protons in the tissue (proton-density-weighted). MRI pulse sequences consist of RF pulses and gradients for selective slice excitation and phase and frequency encoding of the object's signal in k-space. The repetition time (TR) is time between successive gradients, and the time between the selective excitation pulse and the peak of the MRI signal is the echo time (TE); the timing parameters of TR and TE makes the contrast of an image, and the choice of TR is used to control T1W of the image and the choice of TE is used to control the T2W of the image [19].

Qualitative MR images have different pulse-sequences with contrasts that describe how energy is absorbed and dissipated within the body and returned to the RF coil. However, it is important to note that the gray levels or signal intensities within T1W and T2W MR images are not direct quantitative measurements of tissue properties, hence they are referred to as qualitative MR images. The MR image intensity, depicted on a gray-scale, is determined by the signal emitted from the hydrogen nuclei. T1W and T2W images are used in the imaging of ADPKD patients to measure htTKV and evaluate cyst number (Figure 1.1).

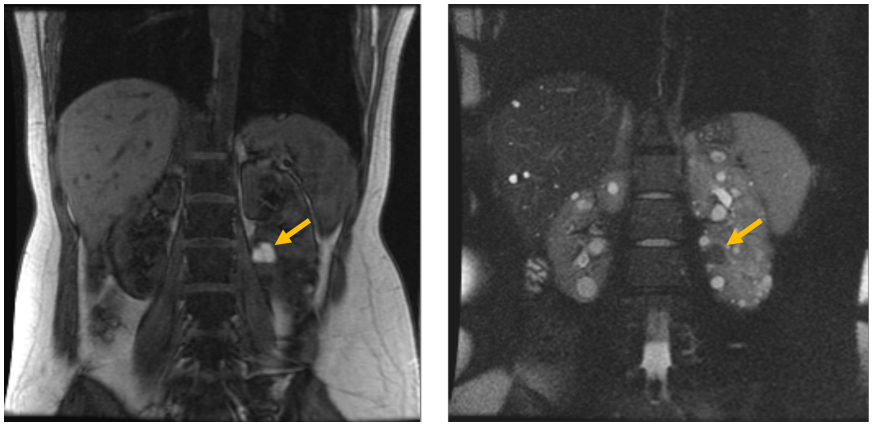

Figure 1.1: Magnetic resonance imaging (MRI) of a coronal T1-weighted scan acquired using 3D VIBE/FMPSPGR/LAVA, without fat saturation, and 3 mm slice thickness (left), and coronal T2-weighted scan acquired using SSFSE/HASTE, fat saturation, and a 3 mm fixed slice thickness (right) from the same patient. Note the hyperintense hemorrhagic (blood-filled) cyst in the T1-weighted MR image and the hypointense signal of the same cyst in the T2-weighted MR image. The images shown here have cysts both in the kidneys and liver.

Quantitative MRI captures physiological and functional characteristics of tissues, including T1 maps, T2 maps, diffusion-weighted imaging (DWI), arterial-spin labeling (ASL), and phase-contrast (PC): T1 and T2 maps provide tissue-specific relaxation times, DWI provides information on tissue microstructure using the apparent diffusion coefficient (ADC) of water molecules (i.e., measure of magnitude of diffusion within tissue using b-values of changing gradient amplitudes for the degree of diffusion-weighting), fractional anisotropy (FA) utilizes diffusion tensors obtained from DWI sequences to quantify the degree of directional preference in molecular displacement by diffusion (0 represents isotropic diffusion, with no preferred direction, while 1 signifies restricted anisotropic diffusion, typically occurring along a single axis), ASL measures tissue perfusion using magnetically labelled water protons in the blood, and PC measures the blood flow in the renal arteries [14]. Research utilizing T2 quantitative mapping has demonstrated a significant increase in T2 values in early-stage ADPKD patients (TKV <300 mL), with T2 values distinguishing them from healthy volunteers when TKV could not, and the potential for T2 values to identify cystogenesis from the very early disease stages [20]. Initial DWI research in ADPKD has revealed abnormally increased ADC values and reduced FA in ADPKD patients with normal kidney function compared to healthy subjects [21]; FA values were significantly lower in ADPKD patients than in healthy controls, indicating a reduced anisotropy (directionality), which may be a potential biomarker of micro-structural damage caused by cyst growth [21].

Utilizing qualitative and quantitative MRI in kidney imaging (multi-parametric MRI) has the ability to capture blood flow and tissue perfusion, anatomic structures, and tissue properties such as inflammation and fibrosis. In summary, kidney MRI has the ability to provide a non-invasive, radiation-free method to assess morphology, function, and microstructure [22]. For ADPKD patients, MRI biomarkers may help assess disease progression when changes are indiscernible using clinical markers such as eGFR.

1.2.2 Non-cystic kidney parenchyma characterization

The non-cystic kidney parenchyma (excluding the presence of macroscopic cysts) has been suspected to be a potential biomarker of disease progression and stratification. In ADPKD, increases in size of cysts did not associate with further kidney function decline, suggesting that changes occurring on a smaller-scale are resulting in greater loss of kidney function [14]. MRI could be used to further noninvasive markers of microscopic kidney tissue changes during cystogenesis and also assess effects of tubular atrophy and interstitial fibrosis alongside cyst development. Caroli et al. [23] used DWI to identify and characterize the non-cystic kidney tissue component, showing its higher diffusivity, lower pseudo-diffusion, and lower flowing fraction relative to healthy tissue; in addition, diffusivity significantly differed by MIC class from both the whole kidney and non-cystic tissue [23]. With respect to T2 mapping, T2 values have a higher correlation with the kidney cyst fraction than htTKV [24]. In summary, characterizing the non-cystic kidney parenchyma may be used for early monitoring and prediction of disease progression; however, the quantitative methods by which to characterize the non-cystic kidney parenchyma is an active area of research. The work presented in this thesis used data extracted from non-cystic kidney parenchyma, but it is important to note that non-cystic kidney parenchyma is not synonymous with healthy kidney tissue.

1.2.3 Magnetic resonance fingerprinting (MRF)

Limitations of conventional quantitative MRI are that data is acquired sequentially on single parameters, have long scan times, and are scanner dependent [25]. Magnetic resonance fingerprinting (MRF) is the simultaneous non-invasive quantification of tissues and is a novel approach with respect to data acquisition, post-processing, to visualization [25]. MRF uses pseudorandom acquisition parameters (i.e, flip angle and phase of radio frequency pulses, repetition time, echo time, sampling patterns) to cause tissues to have unique signal evo-

lution known as "fingerprints," which differs from conventional methods of repeated, serial acquisition of data. Acquisition schemes for MRF have been developed for brain, cardiac, and kidney, and specific MRF parameters may differ across pulse sequences to best exploit differential sensitivity of materials of interest [26]. An important step of this process to create quantitative maps is the pattern-recognition algorithm to match these unique fingerprints to a predefined dictionary of predicted signal evolutions using the Bloch equations from combinations of materials and system-related parameters. Upon identifying the best match to the unique fingerprint from the dictionary, the quantitative parameters of interest (e.g., T1, T2) are provided on a per-voxel basis.

A rapid kidney MRF method for acquiring quantitative T1 and T2 maps at 3.0 T in a single breath hold has been established and used for imaging patients with autosomal recessive polycystic kidney disease (ARPKD) [27]; the work presented investigated the first MRF-acquired T1 and T2 maps of ADPKD patients.

1.3 Artificial Intelligence in Medical Imaging

Artificial intelligence (AI) is a broad term that involves the mimicking of human intelligence using computers or machines, and its application in medical imaging includes computer-aided diagnosis (CAD) systems such as computer-aided detection (CADe) and diagnosis (CADx) [28, 29]. CAD systems are to aid in the interpretation and final decision-making process by a radiologist, and the first major breakthrough of CAD was in detecting lesions in mammography and chest radiographs [29, 30]. Within the CAD system pipeline, quantitative image analysis is used to extract mathematical descriptors, commonly known as radiomics, to capture characteristics of disease; the final output of the CAD system includes quantitative descriptions of the radiologists' task (e.g., lesion vs. non-lesion).

Segmentation of the region of interest (ROI) (e.g., a diseased organ or a lesion) is an important step in the CAD pipeline. AI applications in the task of segmentation in ADPKD

research have been utilized for the kidney, total cyst volume, and other organs such as the liver and spleen from imaging modalities such as CT, MRI, and US [31, 32, 33, 34, 35, 36]. A previous study implemented an AI algorithm in the clinic for kidney segmentation from MR images to compute TKV, and the results show that the AI-measured TKV obtained high levels of agreement with manual segmentations [34]. Shifting this segmentation task to computers would reduce the time required (typically 60 to 90 minutes) to segment entire kidneys and cysts to only a few minutes [34]. From automated kidney and cyst segmentation, additional quantitative measurements that have the potential to complement and enhance prognostication alongside TKV include total parenchymal volume, total cyst number, cyst surface area, and cyst parenchyma surface area [36].

1.3.1 Radiomics

Radiomics transforms medical images into mineable, quantitative data through the calculation of image features that range from simple, first-order signal intensity statistics to more complex spatial relationships of signal intensities, such as gray-level co-occurrence matrix (GLCM) texture features, that differ in mathematical computational complexity. Radiomics is an integral component of quantitative image analysis within the CAD pipeline, tasked with generating quantitative descriptors and mathematical features (e.g., size, shape, texture, and morphology) aimed at enhancing the diagnostic, prognostic, and predictive accuracy of the CAD system [37, 38].

The radiomics workflow involves image acquisition, reconstruction, segmentation, image pre-processing, feature computation, and statistical modeling [39]. Image pre-processing is a step before feature extraction and includes denoising images, interpolation to isotropic voxel spacing, and discretization (quantization) of image intensities inside the ROI (Figure 1.2) [39].

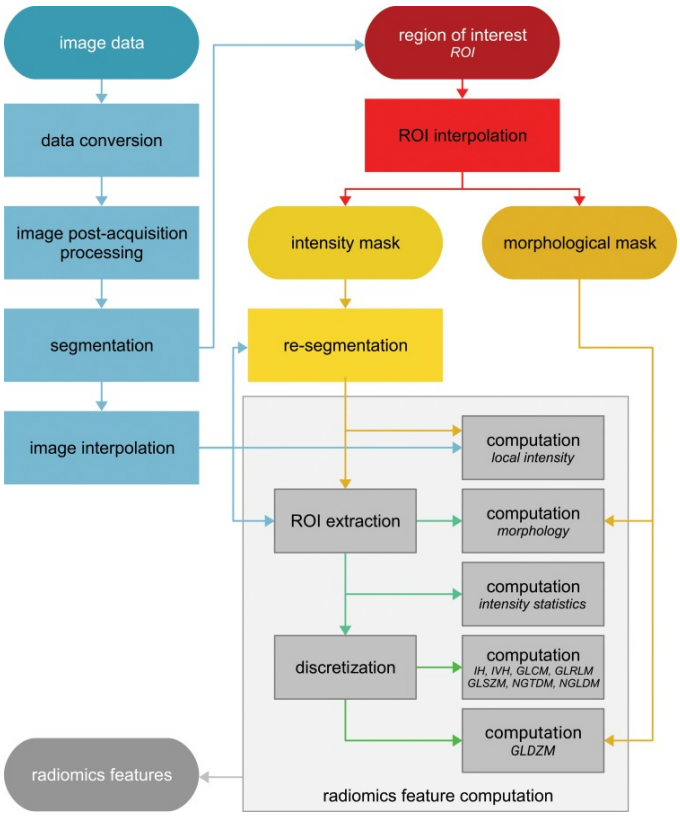

Figure 1.2: Flowchart of the general radiomics image processing scheme for computing radiomics features. Adapted with permission from Zwanenburg et al. [39].

Normalization and pre-processing are necessary steps when extracting features from qualitative signal intensities from T1W and T2W MR images, and these steps have been shown to improve feature robustness and repeatability [40]. There are two predominant discretization methods, "fixed bin number" (FBN) and "fixed bin size" (FBS). FBN is a relative discretization that considers the maximum and minimum gray level in the ROI using a specified bin count and has an adaptive normalization in its implementation (Figure 1.3). FBS uses the raw gray level in the ROI with a specified bin width. Currently, there is no consensus on which discretization is the best for MRI, but literature suggests using a bin width that has 8-130 bin counts for feature extraction [41]; binary discretization is a method that discretizes gray levels in powers of two and divides the number of gray levels in an image based on the

number of bits used to represent each pixel's intensity value. The ability to reproduce and validate published studies has proven difficult due to factors such as pre-processing and utilizing images from different sites and scanners [39]; although the impact of pre-processing is currently an active area of research in radiomics, there are a number of previous studies that do not state pre-processing parameters in their work. A limitation in the clinical use of radiomics is standardization of image pre-processing, which has inspired collaborations to standardize imaging biomarkers and metrics such as the radiomics quality score (RQS) that breaks down radiomics studies in five phases: data selection, medical imaging, feature extraction, exploratory analysis, and modelling [38, 39, 42].

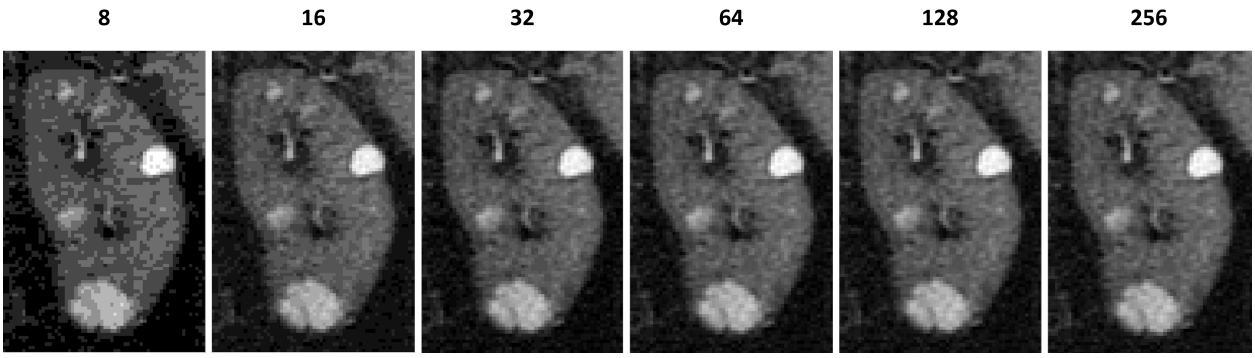

Figure 1.3: An example of fixed bin number gray-level discretization using 8, 16, 32, 64, 128, and 256 gray-level bins. Gray-level discretization is a pre-processing step before feature extraction. The image is a coronal T2-weighted fat-saturated magnetic resonance image of an autosomal dominant polycystic kidney patient with varying number of gray-level bins.

Radiomic feature extraction has been utilized for various applications such as screening for disease, predicting tumor prognosis, and associations with gene expression patterns and imaging phenotypes (i.e., imaging genomics) [38, 43, 44, 45, 46]. Radiomic studies can take advantage of multiple imaging timepoints versus one imaging timepoint alone for a given task; the use of features across multiple imaging timepoints has been shown to improve prediction tasks [47, 48]. Delta radiomics quantifies feature variation of longitudinal image

data, which has shown additive power in assessing response to therapy and patient survival and can also improve overall diagnosis and monitoring of disease [38, 49, 50, 51].

1.3.2 Radiomics in ADPKD

Radiomic features provide quantitative, objective measures of kidney texture that have clinical utility in aiding clinical decision support systems for radiologists in a non-invasive, personalized manner. Radiomics analyses of MRI and CT images to evaluate patients with ADPKD has only recently been explored, and promising studies have shown the additive power of texture in predicting future kidney function decline and classifying chronic kidney disease (CKD) stage in ADPKD patients [52, 53, 54, 55].

The first published report of applying radiomics in ADPKD was by Kline et al. [52], and the study extracted radiomic features (first-order entropy, GLCM correlation, and GLCM energy) from T2W fat-saturated (T2W-FS) MR images from 122 patients who had eGFR values >70 mL/min/1.73m^2 at baseline and assessed the ability to predict kidney function decline at eight-year follow-up; when incorporating texture with a traditional model with baseline age, htTKV, and eGFR, the AUCs increased for CKD stage 3A from 0.86 to 0.94, for CKD stage 3B from 0.90 to 0.96, and for a 30% or more reduction in eGFR from 0.75 to 0.85. Cong et al. [53] extracted radiomic features from T2W-FS MR images to classify ADPKD patients in CKD 1-2 versus CKD stage 3A or greater, and investigated the classification performance using a clinical model, image model, and fused clinical-image model across different machine learning classifiers; the combined clinical-image model outperformed the image and clinical models with an AUC value of 0.89 [0.83, 0.95]. Li et al. [54] extracted radiomic features from T1W fat suppression (T1W-FS) and T2W-FS MR images to classify CKD stage 1-2 versus CKD stage 3A or greater using radiomic features and combined radiomics with clinical factors (gender, hypertension, urinary albumin, ellipsoid formula for TKV measurement, urinary white blood cells, urinary red blood cells, maximum cyst length,

and maximum bleeding cyst length); the results report an increase in AUC value from using the radiomics model (AUC=0.75) to the combined radiomics-clinical model (AUC=0.84) [54]. This was the first work to combine image texture features from multi-parametric MRI for the task of classifying CKD stage of patients with ADPKD. Xie et al. [55] examined 340 ADPKD non-contrast CT images and assessed the predictive value of renal parenchyma volume (RPV) (excluding cysts) and a combined model of RPV and radiomic features extracted from the RPV at a five-year follow-up; at the five-year follow-up, the non-impairment group and the impairment group were defined as experiencing a reduction of 30% in baseline eGFR. The AUC value of RPV alone was 0.752 in prediction to a 30% reduction in eGFR [55]; incorporating RPV with radiomic features extracted from the RPV improved the predictive power of a 30% reduction in eGFR versus radiomic features alone, with an AUC value increasing from 0.849 to 0.902 [55].

Radiomics analyses of MRI images to evaluate patients with ADPKD has only recently been explored, and texture-based differences between PKD genotypes and among risk-stratified MIC patients have not been established. These texture differences have potential to non-invasively classify patient genotype, predict disease progression, and monitor treatment plans and patient outcomes. The majority of ADPKD is due to mutations in the *PKD1* and *PKD2* genes, which associate with different patient and kidney outcomes including the age of onset of ESKD. Establishing radiomic differences between PKD genotypes could help better identify gene-specific texture features aiding in proper diagnosis and risk stratification. Although *PKD1* individuals demonstrate larger kidneys with more cysts relative to age-matched *PKD2* individuals, they exhibit similar rates of increase in TKV and cyst growth and 40% fewer cysts [2]. These findings suggest that microscopic cysts not yet detectable by traditional imaging techniques may account for differences in htTKV between *PKD1* and *PKD2* individuals. Radiomics can examine non-cystic kidney compartments (potentially indicative of microcysts) to determine genotype. To our knowledge, there has

been no research evaluating radiomic feature differences between *PKD1* and *PKD2* patients in the non-cystic kidney parenchyma.

In addition, MIC, an imaging risk-stratification tool for ESKD, consists of groups ranging from slow-progressing disease (1A-1B) to fast-progressing disease (1C-1E) with respect to the rate of increase in htTKV growth. MIC class associates with gender, mutation strength, and mean age at ESKD, where MIC 1E patients have a mean age of onset ESKD of 43.4±7 years versus 65±6.8 years for MIC 1B patients [56]. Although *PKD1* patients are more frequently found to be in the MIC 1C-1E groups, there is significant overlap in MIC class between *PKD1* and *PKD2* patients. Assessing differences among MIC classes using radiomic features extracted from the non-cystic kidney parenchyma from MR images has yet to be investigated; such features may associate with MIC independent of genotype and provide significant prognostic information.

Previous ADPKD radiomic studies differed in their normalization methods before feature extraction, with little mention of additional pre-processing steps: normalization with respect to cerebral spinal fluid, z-score normalization, or image normalization method was not stated [52, 53, 54, 55]. The lack of pre-processing standardization makes the reproducibility and validation of kidney radiomics studies difficult. The effect of pre-processing on radiomic features and ultimate subsequent classification is a research topic of substantial interest and importance.

1.4 Research objectives and scope of work

The purpose of this dissertation was to investigate texture-based MRI biomarkers of ADPKD (Figure 1.4). We developed radiomics-based quantitative models for genotype classification, evaluation of texture-based differences of risk-stratified MIC classes, temporal assessment of radiomic features for the prediction of kidney function decline of ADPKD patients, and

assessment of texture differences of healthy and ADPKD cohorts based on MRF-acquired quantitative maps. The specific aims of this thesis were to:

Aim 1- Investigate MRI radiomics analysis of non-cystic kidney parenchyma to classify *PKD1* and *PKD2* genotypes (Chapter 2)

Aim 2- Use MRI radiomics analysis of non-cystic kidney parenchyma to differentiate among Mayo Imaging Classifications (MIC) 1A through 1E (Chapter 3)

Aim 3- Assess temporal change in radiomic features and the additive power for the prediction of kidney function decline in ADPKD patients (Chapter 4)

Aim 4- Investigate MRF radiomics analysis in healthy controls and ADPKD patients (Chapter 5)

Radiomics analysis—the extraction of quantitative image features—can non-invasively characterize disease and has shown predictive utility for ADPKD patient future decline in kidney function. Although promising, current studies have not established texture-based differences between PKD genotypes (*PKD1* and *PKD2*) or among risk-stratified MIC groups (1A-1E) that capture disease progression or patient outcomes. In addition, the non-cystic kidney parenchyma (which may undergo microscopic cystogenesis) has been suggested as a potential imaging biomarker. There is a need to investigate texture-based differences in the non-cystic kidney parenchyma between genotypes and among risk-stratified groups because features may quantify potential prognostic information. In addition, we examined how pre-processing parameters affect radiomic features and subsequent performance in classifying genotype and MIC groups.

Medical imaging is clinically indicated for ADPKD patients to monitor disease progression using htTKV. Delta radiomics was used to assess temporal imaging data to investigate the additive power in prediction models of kidney function decline at 60-months follow-up. Radiomic features of the non-cystic kidney parenchyma and the entire kidney were calculated

to determine whether texture features using machine learning provide additional information regarding disease progression and prediction of patient prognosis. We hypothesize that kidney texture features, that are extracted from longitudinal imaging studies, can accurately predict disease progression.

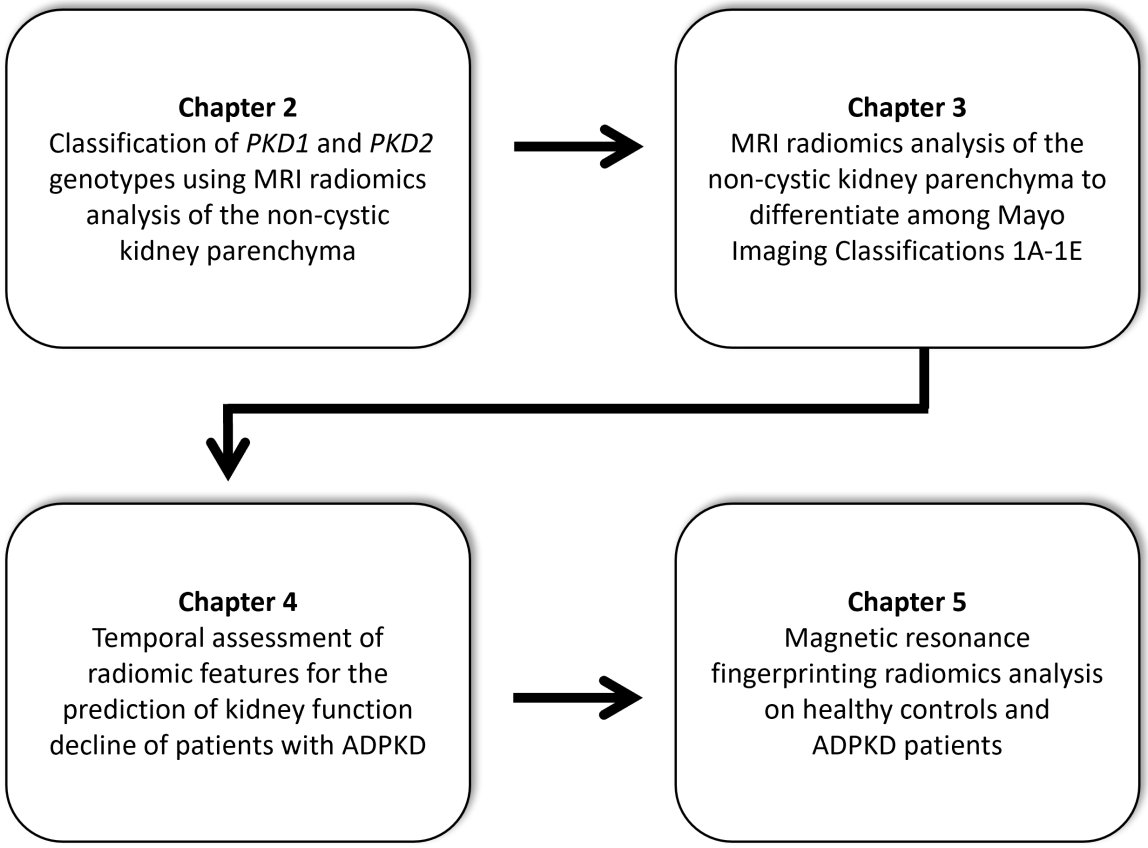

Figure 1.4: Dissertation outline of aims.

Clinical utility of MRI radiomics is limited due to a lack of standardized pre-processing steps prior to feature extraction. Specifically, T1W and T2W MR images require normalization or intensity standardization because image units are arbitrary. Quantitative MRI, specifically MRF, allows for direct quantitation of an image and provides absolute and repeatable measurements from tissues that are capable of characterizing underlying pathology. We investigated radiomic features from MRF-acquired T1 and T2 maps that differentiate

between healthy controls and children and young adults with ADPKD. This work advances non-invasive imaging biomarkers for characterizing the non-cystic kidney parenchyma using radiomic features, ultimately facilitating the identification of ADPKD children at high risk for disease progression.

Aim 1 classified ADPKD genotype and quantified texture differences that have yet to be shown in ADPKD radiomic studies. Aim 2 differentiated among risk-stratified MIC classes (independent of genotype) and identified features associated with risk of ESKD. Aim 3 utilized temporal imaging data and delta radiomics to assess whether radiomic features from 24-month and 48-month timepoints provide additive power versus baseline texture alone in predicting kidney function decline of patients with ADPKD. Aim 4 extracted features from MRF-acquired T1 and T2 maps to investigate texture-based differences in healthy and ADPKD cohorts.

CHAPTER 2

MRI RADIOMICS ANALYSIS OF NON-CYSTIC KIDNEY PARENCHYMA TO CLASSIFY PKD1 AND PKD2 GENOTYPES

2.1 Introduction

Autosomal dominant polycystic kidney disease (ADPKD) is the most common hereditary kidney disorder and is responsible for 10% of patients with end-stage kidney disease (ESKD) under the age of 65 [1, 2]. ADPKD is predominantly due to mutations in two genes, *PKD1* (85%) and *PKD2* (15%). ADPKD results in gradual enlargement of the kidneys due to cyst growth and enlargement over decades prior to decline in kidney function and kidney failure. Increases in kidney size associate with loss of kidney function in ADPKD and height-corrected total kidney volume (htTKV) is an imaging prognostic biomarker approved by the U.S. Food and Drug Administration. Typically, *PKD1* patients have greater htTKV than *PKD2* patients, and while *PKD1* and *PKD2* kidneys increase in size at the same rate, *PKD1* patients have 40% more detectable kidney cysts [2]. These findings indicate that the rate of cyst formation may account for the differences in htTKV observed between PKD genotypes. This is consistent with the observation that *PKD2* patients typically start dialysis 20 years later and live 10 years longer than *PKD1* patients. Understanding the nature of alterations in the non-cystic parenchyma in *PKD1* and *PKD2* patients may help to understand genetic differences in disease progression in ADPKD.

Radiomic features could capture textural alterations in the non-cystic compartments of the kidney due to differences in cystogenesis or tissue response to injury and add value to established kidney size differences. Although the power of radiomics enhances the understanding of a disease, the field lacks a standardized approach to extracting features. Imaging and radiomics workflows involve image acquisition, reconstruction, segmentation, image pre-

processing, feature computation, and statistical modeling [39]. Image pre-processing is an important part of the radiomics workflow to harmonize images before feature extraction and includes operations such as image resizing, pixel resampling, gray-level discretization, and filtering, which all have downstream effects on feature values [38, 57]. Qualitative MRI sequences (e.g., T1-weighted and T2-weighted) require a normalization or signal intensity standardization process for inter- and intra-patient radiomic comparisons due to arbitrary signal intensities. In ADPKD imaging studies, T2-weighted fat saturated (T2W-FS) MR images are used due to their superior cyst-to-parenchyma contrast; however, previous ADPKD radiomic studies differed in their pre-processing, specifically MR normalization and gray-level discretization methods [52, 53, 54]. Disease-specific radiomic studies, such as brain and prostate cancer, have investigated the effect of pre-processing on radiomic features; therefore, this work explores the effect of pre-processing parameters on MRI radiomic features specific to ADPKD.

The purpose of this work was two-fold: (1) to assess the impact of MRI pre-processing on radiomic features and (2) to evaluate the ability of features extracted from the non-cystic kidney parenchyma to classify *PKD1* and *PKD2* variants of ADPKD.

2.2 Methods

2.2.1 Databases for feature reproducibility and genotype classification

This work analyzed images that were obtained from two previously completed prospective imaging studies: the Consortium for Radiologic Imaging Studies of Polycystic Kidney Disease (CRISP) study, a longitudinal study of cyst and kidney growth in a large cohort of patients with ADPKD, and the HALT Progression of Polycystic Kidney Disease (HALTA-PKD) randomized clinical trial (NCT00283686). T2W-FS MR images were acquired with the same scanning protocol on 1.5T scanners using single-shot fast spin echo/half-Fourier acquisition

single-shot turbo spin echo imaging (SSFSE/HASTE) with fat saturation and a 3-mm fixed slice thickness. A single representative 2D MR images of the left and right kidney were manually chosen based on the coronal MR image that maximized the longitudinal length for each kidney individually.

The CRISP study was first used to evaluate the effect of pre-processing on radiomic features. T2W-FS MR images from 15 subjects (7 with the *PKD1* genotype and 8 with the *PKD2* genotype) were analyzed. The images had been acquired between 2001-2002 from a single clinical site and were matched for age, sex, and htTKV. Image matrix sizes for all CRISP images were 256x256 pixels, with pixel sizes ranging from 1.17-1.37 mm. Table 2.1 shows the subject characteristics of age, htTKV, and estimated glomerular filtration rate (eGFR).

Table 2.1: Consortium for Radiologic Imaging Studies of Polycystic Kidney Disease (CRISP) *PKD1* and *PKD2* subject characteristics of age and height-corrected total kidney volume (htTKV) with respect to Mayo Imaging Classification (MIC) class.

Clinical	*PKD1*	*PKD2*
No. Patients	7	8
Mean age ±SD		
Male (n=7)	23 ±7.79	27 ±6.80
Female (n=8)	25 ±9.42	24 ±7.76
Mean htTKV ±SD		
MIC 1A (n=0)	-	-
MIC 1B (n=8)	358.12 ±55.94	283.28 ±67.45
MIC 1C (n=4)	274.71 ±2.34	358.23 ±74.86
MIC 1D (n=1)	320.09	-
MIC 1E (n=2)	451.74	428.29

The HALTA-PKD randomized clinical trial was then subsequently used for genotype classification. A dataset of 136 age-, gender-, and Mayo Imaging Classification (MIC) matched baseline MR images was analyzed [11]. Due to the low prevalence of *PKD2* to *PKD1*, the HALTA-PKD dataset had 68 *PKD2* patients who had measured htTKV, baseline MR images, and the ability to match with 68 *PKD1* patients for age, gender, and MIC. The MR images had been acquired from 7 different sites between 2006-2009. All HALT MR image matrix sizes ranged from 256x256-560x560 pixels, and pixel sizes ranged from 0.63-1.8 mm. Table 2.2 shows the subject characteristics.

Table 2.2: HALT subject characteristics for classification. A $p < 0.05$ was significant using Wilcoxon rank sum test. Asterisks indicate statistically significant differences.

Clinical	*PKD1*	*PKD2*	p-value
No. Patients	68	68	
Mean age ±SD			
Male (n=64)	40 ±7.40	41 ±6.86	0.81
Female (n=72)	40 ±7.85	41 ±7.77	0.61
Mean htTKV ±SD			
MIC 1A (n=10)	257.61 ±41.28	231.38 ±37.34	0.31
MIC 1B (n=60)	408.66 ±90.41	380.98 ±91.28	0.28
MIC 1C (n=48)	651.07 ±198.56	681.08 ±247.64	0.78
MIC 1D (n=14)	1075.01 ±258.18	1064.88 ±287.25	0.90
MIC 1E (n=4)	1453.34 ±22.68	1627.36 ±349.44	1.0
Mean eGFR ±SD	80.91±16.55	87.01 ±15.48	0.01*

2.2.2 MR image pre-processing

For MRI pulse sequences with arbitrary signal intensities, normalization is employed to standardize images from different sites and scanners, mitigating large variations in signal intensities. The image normalization methods chosen were (1) z-score method and (2) reference-tissue normalization. Z-score normalization uses the mean and standard deviation of the entire image to normalize signal intensities:

$$X_{z-score} = \frac{X_i - \mu}{\sigma}\,, \tag{2.1}$$

where the μ represents the mean pixel intensity of the entire image, σ represents the standard deviation of pixel intensities of the entire image, X_i represents the pixel value of the pixel of interest, and $X_{z-score}$ is the corresponding normalized pixel value. Reference-tissue normalization uses a healthy tissue to standardize the gray levels in an image. The reference-tissue method chosen transformed a region of interest (ROI) mean and standard deviation extracted from the psoas muscle in each image to have a mean of 100 and a standard deviation of 10:

$$X_{reference-tissue} = \frac{X_i - \mu_{ROI}}{\sigma_{ROI}}\,, \tag{2.2}$$

$$X_{reference-tissue,transformed} = (X_{reference-tissue} * \mu_{ROI,all}) + \sigma_{ROI,all}. \tag{2.3}$$

$X_{reference-tissue}$ is the z-score normalization of X_i, the pixel value of the pixel of interest, using the mean pixel intensity of the psoas muscle, μ_{ROI}, and the standard deviation of the ROI, σ_{ROI}. $X_{reference-tissue,transformed}$ is the scaled and shifted pixel intensity of the pixel of interest after transforming the ROI to have a mean, $\mu_{ROI,all}$, of 100 and standard

deviation, $\sigma_{ROI,all}$, of 10. Muscle is hypointense on T2-weighted MR images and has been used in previous studies to standardize MR signal intensity [41, 58, 59]. The original images were also used for feature extraction without any normalization applied.

All images were resized to 256x256 for feature extraction using nearest-neighbor interpolation. Additionally, pixel sizes were harmonized by up-sampling to 1.0x1.0 mm and down-sampling to 2.0x2.0 mm using nearest-neighbor interpolation. According to the Imaging Biomarker Standardization Initiative (IBSI), it is not known whether up-sampling or down-sampling schemes are preferable [39].

3D Slicer, an open-source image-analysis software package, was used for segmentation of the kidney and cysts [60]. Kidneys were segmented manually, and cysts were semi-automatically segmented and removed from the resultant kidney segmentations to obtain pixel intensities from the non-cystic kidney parenchyma.

2.2.3 Feature extraction

Pyradiomics [61] was used for feature extraction of the images. Features were extracted from the non-cystic kidney parenchyma and the entire kidney parenchyma. Ninety-three features were extracted per kidney: first-order (18), gray-level co-occurrence matrix (GLCM) (24), gray-level run length matrix (GLRLM) (16), gray-level size zone matrix (GLSZM) (16), neighboring gray-tone difference matrix (NGTDM) (5), and gray-level dependence matrix (GLDM) (14). One of the discretization methods used in this work was fixed bin size (FBS) as implemented in Pyradiomics [61]:

$$X_{fbs,i} = \lfloor \frac{X_{gl,i}}{W} \rfloor - \lfloor \frac{min(X_{gl})}{W} \rfloor + 1\,. \tag{2.4}$$

$X_{gl,i}$ and $X_{fbs,i}$ are the gray level of voxel i before and after discretization, respectively, with bins equally spaced from 0 and $\lfloor \frac{min(X_{gl})}{W} \rfloor + 1$ ensuring that the minimum gray level

for feature extraction is always 1 after discretization. To analyze the impact of normalization on radiomic features from images with different ranges of signal intensities, bin widths (W) were calculated using mean signal intensity range of patient T2W-FS MR images:

$$W = \frac{Range_{mean}}{N_b}\,, \tag{2.5}$$

where the $Range_{mean}$ is the mean gray-level range across all patient MR images and N_b is the number of gray levels or bins for feature calculation. The second discretization method used was fixed bin number (FBN) as implemented in Pyradiomics [61]:

$$X_{b,i} = \begin{cases} \left\lfloor N_b \frac{(X_{\mathrm{gl},i} - \min(X_{\mathrm{gl}}))}{\max(X_{\mathrm{gl}}) - \min(X_{\mathrm{gl}})} \right\rfloor + 1 & \text{for } X_{\mathrm{gl},i} < \max(X_{\mathrm{gl}}) \\ N_b & \text{for } X_{\mathrm{gl},i} = \max(X_{\mathrm{gl}}) \end{cases}. \tag{2.6}$$

N_b is the number of bins used, $X_{gl,i}$ is the gray level of the pixel of interest, and the maximum and minimum X_{gl} is the maximum and minimum gray level of the pixels in the ROI. $X_{b,i}$ is the pixel value after discretization.

In this work six gray-level bins were analyzed for the non-cystic kidney parenchyma and entire kidney: 8, 16, 32, 64, 128, and 256. A difference between Pyradiomics implementation of FBS and IBSI is that Pyradiomics ensures that the minimum gray level starts at 1, while IBSI starts at the minimum gray-level intensity in the ROI [39, 61]. Gray-level discretization is used to suppress noise while retaining important biological variation in the ROI and making the feature calculation time more efficient [39]. An accepted range in the literature for the total number of bins for discretization is between 8-128, and this range is frequently used in studies that have previously investigated the effect of gray-level discretization on MRI radiomic feature repeatability and reproducibility in other disease cohorts [40, 62, 63]. Currently, there is no consensus on the best approach to discretizing gray levels of MR images that have arbitrary signal intensities. For example, IBSI [39] suggests using the FBN method,

but Pyradiomics [61] suggests normalizing MR images and using an FBS approach. FBN is a relative discretizaton approach that normalizes gray levels to the maximum and minimum of the ROI, and using this method on qualitative MR images to extract second-order texture features without any normalization has been implemented in previous studies [40]. The repeatability and reproducibility of radiomic features using FBN and FBS methods have been investigated in other disease cohorts, with FBS producing more reproducible features [63].

Figure 2.1 are example images of the original T2W-FS MR image, subsequent result of applying the kidney mask to the original image, and finally the result of removing cysts for feature extraction from the non-cystic kidney parenchyma. The pre-processing pipeline presented in this chapter is shown in Figure 2.2.

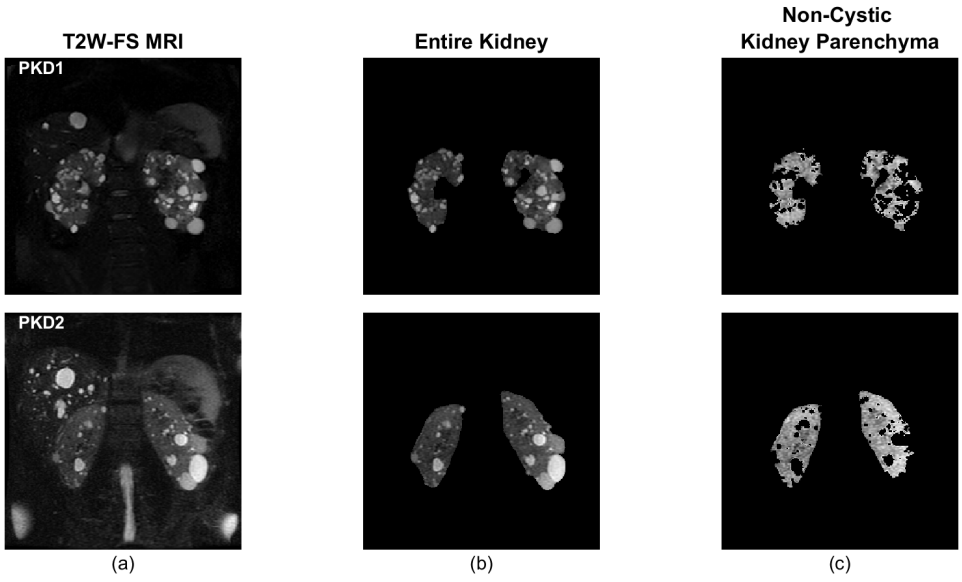

Figure 2.1: From left to right: (a) representative MRI slice, (b) the result of kidney segmentation, and (c) the result of cyst segmentation. These segmented regions were used for feature extraction for a Mayo Imaging Classification (MIC) 1B *PKD1* patient (top) and a MIC 1B *PKD2* patient (bottom).

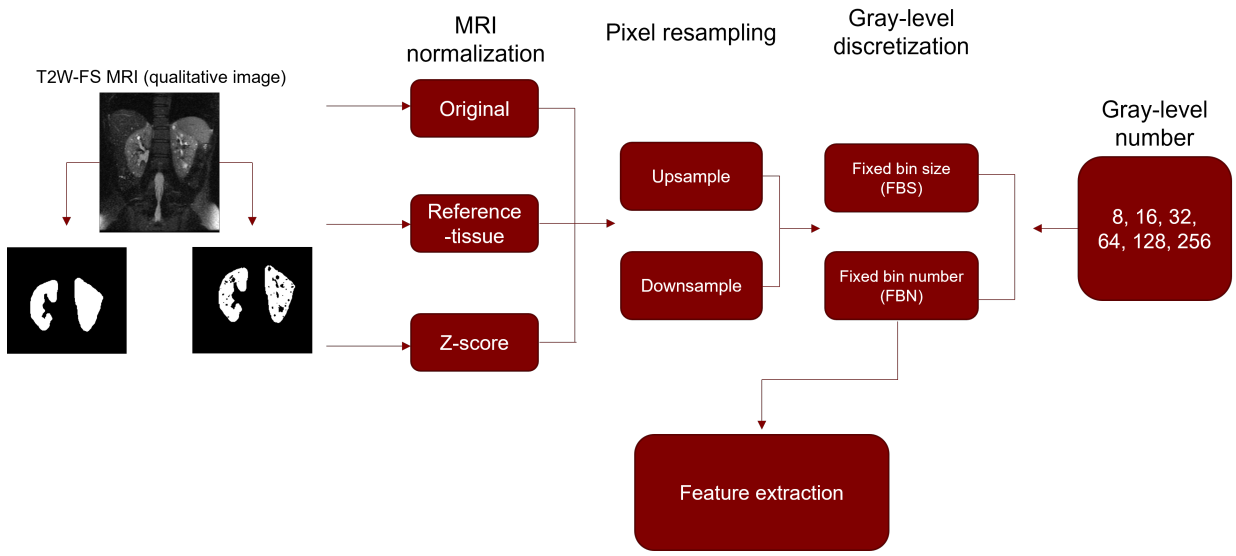

Figure 2.2: Established MRI pre-processing pipeline for investigation on radiomic feature reproducibility and subsequent classification.

2.2.4 Feature reproducibility

To evaluate the reproducibility of features across normalization and gray levels for up-sampling and down-sampling schemes, the intra-class correlation coefficient (ICC) was calculated for each radiomic feature extracted from the images of the CRISP dataset. ICC were calculated only on radiomic features using FBS discretization because FBN introduces a normalization effect (based on the minimum and maximum pixel intensity of the ROI) in its implementation, therefore the impact of normalization on the radiomic features would not be captured using ICC. The ICC metric has been used in previous radiomics literature to evaluate radiomic feature reproducibility based on test-retest, intrarater, and interrater analyses [40, 64, 65]. This statistical metric combines information about the degree of correlation and agreement between measurements. A two-way mixed effects, consistency, single-rater model was used:

$$ICC = \frac{MS_R - MS_E}{MS_R + (k-1)MS_E}, \tag{2.7}$$

where MS_R is the mean square for observations, MS_E is the mean square for error, and k represents the “raters of interest," which are the MR normalization methods used (z-score, psoas muscle, and the original image). ICC can take values between 0 and 1, with values closer to 1 representing stronger reproducibility; currently, there is no standard ICC value for “acceptable" reproducibility of radiomic features, and this determination differs across the radiomics literature. ICC values were calculated in MATLAB, and according to Koo and Li [66], values less than 0.5 indicate poor reproducibility, 0.5-0.75 indicate moderate reproducibility, 0.75-0.9 indicate good reproducibility, and greater than 0.9 indicate excellent reproducibility [67]. In this work, good and excellent reproducibility were combined so that ICC values on the range [0.75-1.0] were classified as good-excellent reproducibility.

2.2.5 Feature selection and classification

A logistic-regression classifier using 5-fold cross validation was utilized on the HALT dataset only to classify genotype. For each training partition, the top-10 performing radiomic features (Pearson correlation threshold of 0.7 with any other selected feature) from the 93 total extracted features were determined using the area under the receiver operating characteristic curve (AUC). This process used a repeated cross-validation (rCV) of 10 to account for variance in 5-fold cross-validation.

2.3 Results

2.3.1 Reproducibility of radiomic features across pre-processing

Figure 2.3 shows the variability of ICC scores of CRISP-derived radiomic features across normalizations and gray-level discretization using FBS discretization for both up-sampling and down-sampling schemes. As the bin counts increase (smaller bin widths and an increase in gray levels for feature computation), feature reproducibility across normalizations

also increases. Up-sampling and down-sampling methods under 64 bins yielded poor reproducibility for over 50% of the total features calculated. For both resampling methods, the largest increase in the number of features with good-excellent reproducibility was over 20% from 32 gray levels to 64 gray levels. Table 2.3 shows the mean ICC values across feature families; an increase in reproducibility is shown with increasing gray levels, the highest being features calculated with 256 gray levels. Increasing the gray levels of the ROI resulted in larger mean ICC values across feature families except for first-order features.

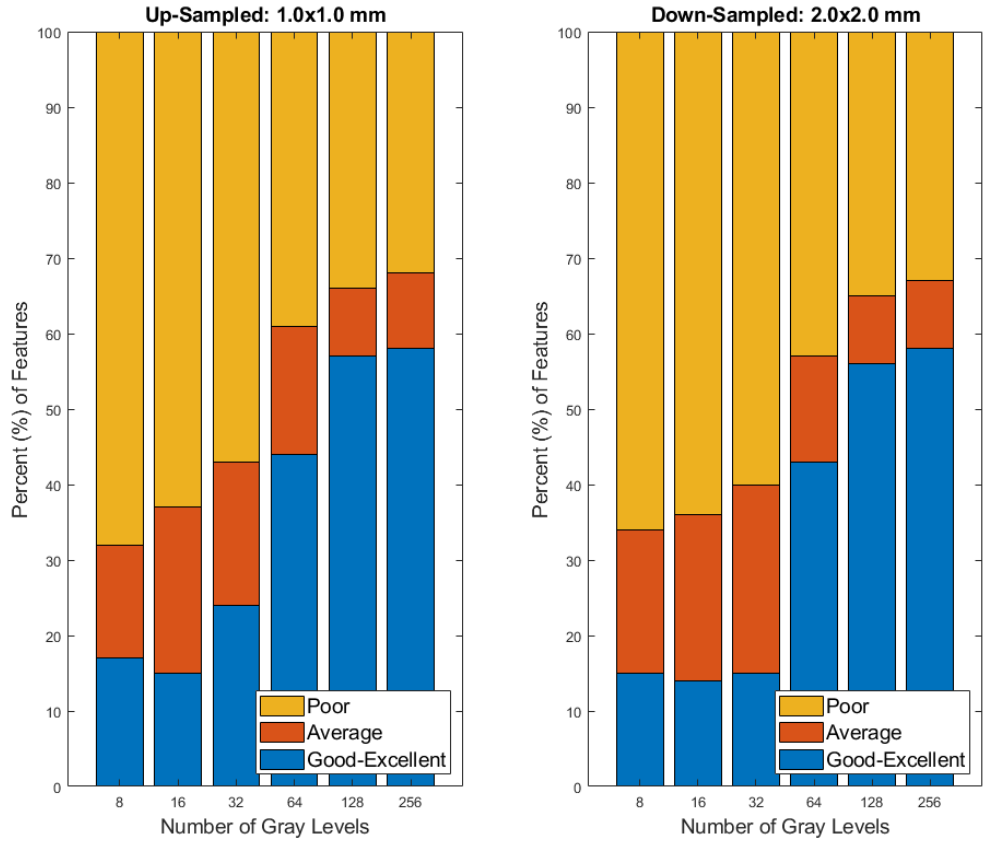

Figure 2.3: Intra-class correlation coefficient (ICC) scores for radiomic features (poor, moderate, good-excellent) across MRI normalizations. Radiomic features extracted from the non-cystic kidney parenchyma were sensitive to pre-processing parameters, with varying reproducibility depending on the parameter. The percentage of features with good-to-excellent ICC scores ranged from 14%-58%, increasing as the number of gray levels available for feature extraction increased.

Table 2.3: Mean intra-class correlation coefficient (ICC) scores for radiomic features across feature families using up-sampling and down-sampling methods.

Number of Gray Levels	8		16		32		64		128		256	
	1.0	2.0	1.0	2.0	1.0	2.0	1.0	2.0	1.0	2.0	1.0	2.0
First-order (n=18)	0.33	0.33	0.33	0.33	0.33	0.33	0.36	0.36	0.38	0.38	0.38	0.38
GLCM (n=24)	0.60	0.61	0.62	0.61	0.62	0.60	0.65	0.63	0.72	0.72	0.77	0.76
GLDM (n=14)	0.55	0.59	0.58	0.59	0.65	0.64	0.76	0.75	0.84	0.84	0.86	0.85
GLRLM (n=16)	0.58	0.60	0.56	0.56	0.63	0.60	0.78	0.75	0.85	0.84	0.87	0.86
GLSZM (n=16)	0.56	0.56	0.50	0.58	0.66	0.60	0.77	0.74	0.85	0.85	0.87	0.87
NGTDM (n=5)	0.54	0.58	0.56	0.58	0.55	0.54	0.60	0.62	0.76	0.77	0.80	0.81

There were 7 features that exhibited good-excellent (ICC $>$ 0.75) reproducibility across all gray levels, up-sampling and down-sampling schemes, and MRI normalization: first-order skewness and kurtosis, GLCM inverse difference matrix normalized (IDMN) and inverse difference normalized (IDN), GLDM dependence non-uniformity, GLSZM gray-level non uniformity, and NGTDM coarseness. Two features, GLCM maximum correlation coefficient (MCC) and GLSZM small area emphasis, were reproducible across all gray levels for up-sampling only.

Figure 2.4 contains the ICC values that were observed using pair-wise comparisons of MRI normalization. ICC values of radiomic features using z-score normalization and psoas muscle normalization had larger percentages of features with poor reproducibility, ranging from 32%-68%, versus original and z-score, 15%-42%, and the original image and psoas muscle normalization, 19%-41%. A similar trend for all pair-wise comparisons in Table 2.3 was that increasing the number of gray levels increased the percentage of features with good-excellent reproducibility across normalization pair-wise comparisons. In general, there were a larger percentage of radiomic features with poor ICC values when comparing between normalization methods more so than comparing the original image and normalization method (z-score and psoas muscle normalization), suggesting that the original signal intensities are

greatly impacted depending on the normalization method used and that different methods may result in varying signal intensities that impact downstream feature extraction.

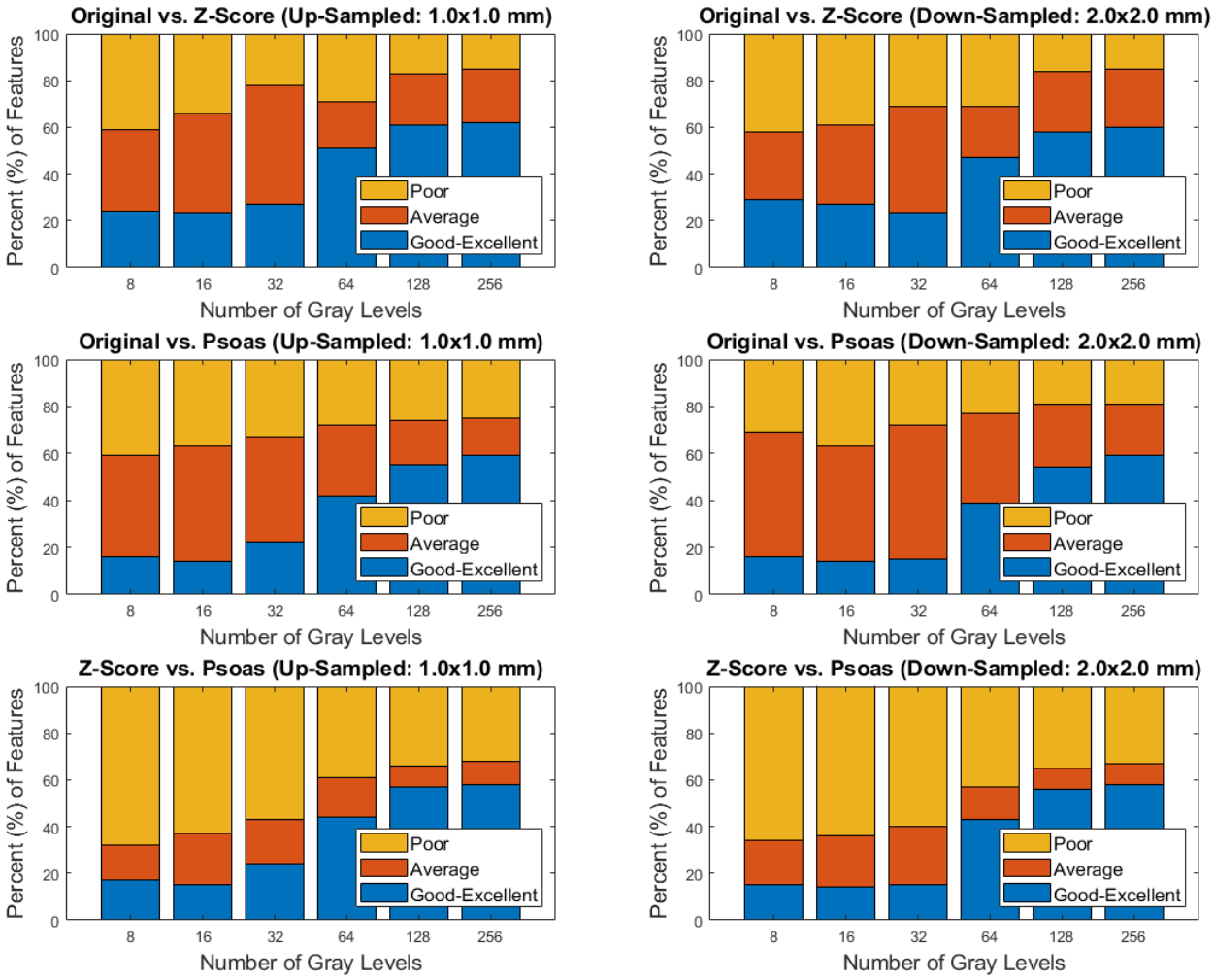

Figure 2.4: Percentage (%) of total radiomic features for pairwise comparisons of MRI normalizations and radiomic reproducibility categorized as poor, moderate, and good-to-excellent.

2.3.2 PKD1 vs. PKD2 classification using radiomic features

Figure 2.5 shows the MRI signal intensity range for all patient data after normalization and downsampling of pixel size; the non-cystic kidney parenchyma has a lower range of signal intensity than the entire kidney parenchyma due to the exclusion of cysts.

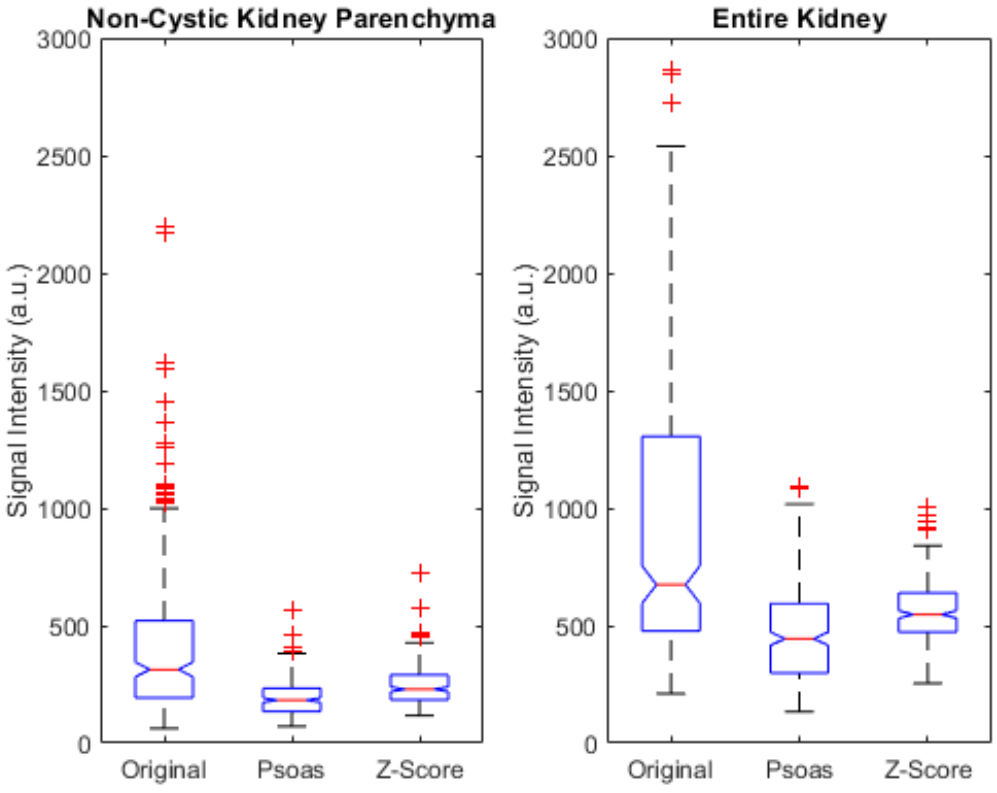

Figure 2.5: Box and whisker plot of the MRI signal intensity range in arbitrary units (a.u.) across HALTA-PKD images for all patient data and normalization methods after down-sampling. Images obtained were from 7 different sites. The box represents the interquartile range, with the central line indicating the median value; the whiskers extend to the minimum and maximum values, while outliers are represented as individual data points.

Tables 2.4-2.5 show the AUC values across pre-processing parameters using the original image, z-score normalization, and psoas muscle normalization using features extracted from the non-cystic kidney parenchyma. The range of AUC values was between 0.47-0.68. Across normalizations, the range of AUC values for the original image, z-score normalization, and psoas muscle normalization were 0.50-0.62, 0.47-0.68, and 0.50-0.61, respectively. Across normalizations, FBS discretization, and up-sampling only, the range of AUC values for the original image, z-score normalization, and psoas muscle normalization were 0.51-0.58, 0.47-0.55, and 0.51-0.59, respectively; the range of AUC values for the down-sampling method

for the original image, z-score normalization, and psoas muscle normalization were 0.55-0.62, 0.56-0.68, and 0.56-0.61, respectively. Across normalizations, FBN discretization, and up-sampling only, the range of AUC values for the original image, z-score normalization, and psoas muscle normalization were 0.52-0.61, 0.51-0.58, and 0.50-0.60, respectively; the range of AUC values for the down-sampling method for the original image, z-score normalization, and psoas muscle normalization were 0.50-0.58, 0.53-0.59, and 0.53-0.59, respectively.

Table 2.6-2.7 shows the AUC values across pre-processing parameters using the original image, z-score normalization, and psoas muscle normalization using features extracted from the entire kidney parenchyma. The range of AUC values was between 0.56-0.73. Across normalizations, the range of AUC values for the original image, z-score normalization, and psoas muscle normalization were 0.56-0.68, 0.57-0.68, and 0.57-0.73, respectively. Across normalizations, FBS discretization, and up-sampling only, the range of AUC values for the original image, z-score normalization, and psoas muscle normalization were 0.58-0.63, 0.58-0.68, and 0.68-0.73, respectively; the range of AUC values for the down-sampling method for the original image, z-score normalization, and psoas muscle normalization were 0.56-0.62, 0.58-0.64, and 0.65-0.70, respectively. Across normalizations, FBN discretization, and up-sampling only, the range of AUC values for the original image, z-score normalization, and psoas muscle normalization were 0.57-0.68, 0.57-0.68, and 0.57-0.68, respectively; the range of AUC values for the down-sampling method for the original image, z-score normalization, and psoas muscle normalization were 0.61-0.64, 0.61-0.65, and 0.61-0.65, respectively.

Table 2.4: Area under the receiver operating characteristic curve (AUC) along with the 95% confidence interval (CI) of AUC for features extracted from the non-cystic kidney parenchyma using fixed bin size (FBS) discretization.

	Number of Gray Levels					
Normalization	8	16	32	64	128	256
Original						
1.0x1.0	0.52	0.58	0.55	0.51	0.54	0.54
	[0.49,0.55]	[0.54,0.61]	[0.51,0.58]	[0.48,0.54]	[0.51,0.56]	[0.51,0.55]
2.0x2.0	0.62	0.60	0.55	0.57	0.58	0.55
	[0.59,0.65]	[0.57,0.63]	[0.52,0.58]	[0.54,0.60]	[0.55,0.61]	[0.52,0.58]
Psoas						
1.0x1.0	0.59	0.57	0.53	0.52	0.55	0.51
	[0.56,0.62]	[0.54,0.60]	[0.50,0.56]	[0.49,0.55]	[0.52,0.58]	[0.48,0.54]
2.0x2.0	0.60	0.56	0.61	0.60	0.58	0.58
	[0.57,0.63]	[0.52,0.59]	[0.58,0.64]	[0.57,0.63]	[0.56,0.61]	[0.55,0.61]
Z-score						
1.0x1.0	0.54	0.55	0.53	0.47	0.50	0.49
	[0.51,0.57]	[0.52,0.58]	[0.50,0.56]	[0.44,0.50]	[0.47,0.53]	[0.46,0.52]
2.0x2.0	0.57	0.56	0.61	0.68	0.65	0.56
	[0.54,0.60]	[0.53,0.59]	[0.58,0.64]	[0.65,0.71]	[0.62,0.68]	[0.53,0.60]

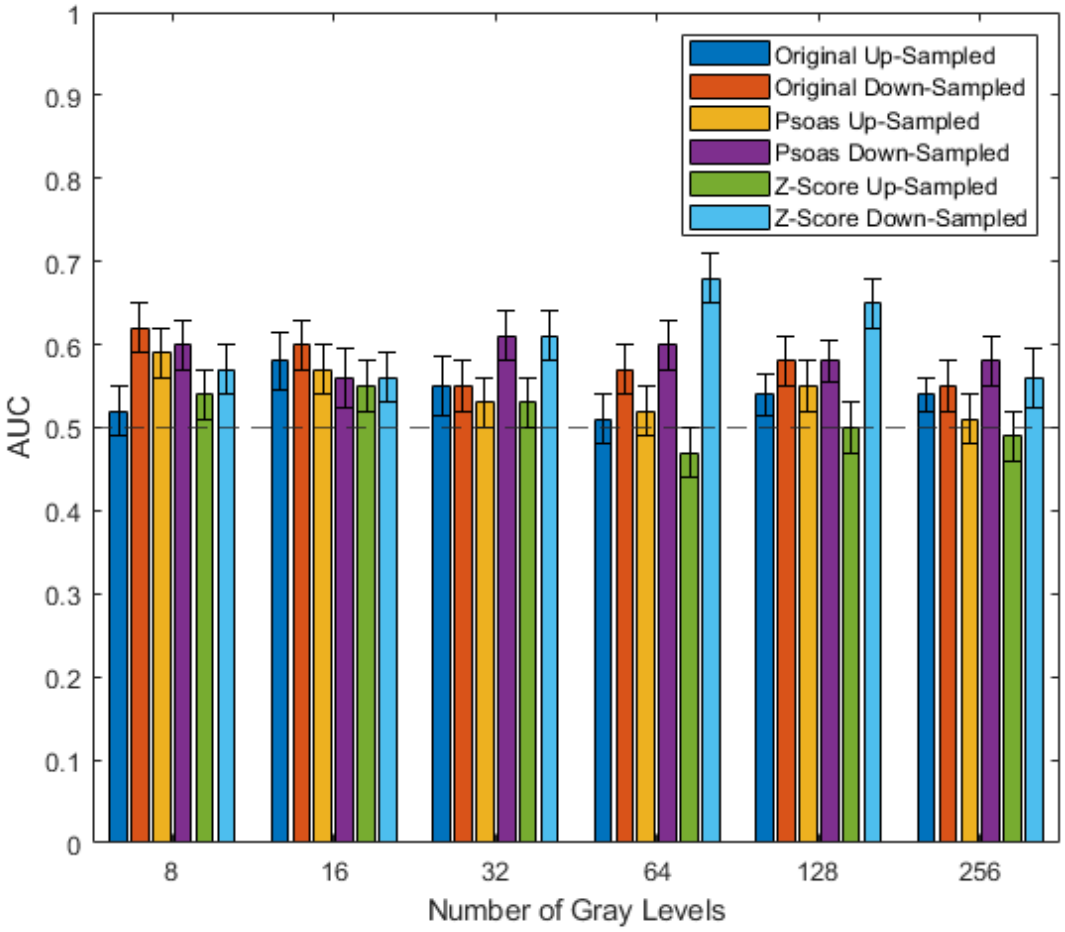

Figure 2.6: Area under the receiver operating characteristic curve (AUC) values in classifying genotype using radiomic features extracted from the non-cystic kidney parenchyma using fixed bin size (FBS) discretization. The dotted line at an AUC of 0.5 is random guessing.

Table 2.5: Area under the receiver operating characteristic curve (AUC) along with the 95% confidence interval (CI) of AUC for features extracted from the non-cystic kidney parenchyma using fixed bin number (FBN) discretization.

	Number of Gray Levels					
Normalization	8	16	32	64	128	256
Original						
1.0x1.0	0.61	0.58	0.53	0.57	0.54	0.52
	[0.58,0.63]	[0.55,0.61]	[0.50,0.56]	[0.54,0.60]	[0.51,0.57]	[0.49,0.55]
2.0x2.0	0.54	0.55	0.54	0.58	0.50	0.53
	[0.51,0.57]	[0.52,0.58]	[0.50,0.57]	[0.55,0.61]	[0.47,0.53]	[0.50,0.56]
Psoas						
1.0x1.0	0.60	0.54	0.52	0.57	0.55	0.50
	[0.57,0.63]	[0.51,0.57]	[0.49,0.55]	[0.54,0.60]	[0.52,0.58]	[0.47,0.53]
2.0x2.0	0.53	0.53	0.57	0.57	0.57	0.59
	[0.49,0.56]	[0.51,0.57]	[0.54,0.60]	[0.54,0.60]	[0.54,0.60]	[0.56,0.62]
Z-score						
1.0x1.0	0.58	0.55	0.51	0.57	0.56	0.52
	[0.55,0.61]	[0.52,0.58]	[0.47,0.54]	[0.54,0.60]	[0.53,0.59]	[0.48,0.54]
2.0x2.0	0.56	0.58	0.58	0.59	0.53	0.55
	[0.53,0.60]	[0.54,0.61]	[0.55,0.61]	[0.56,0.62]	[0.51,0.57]	[0.52,0.58]

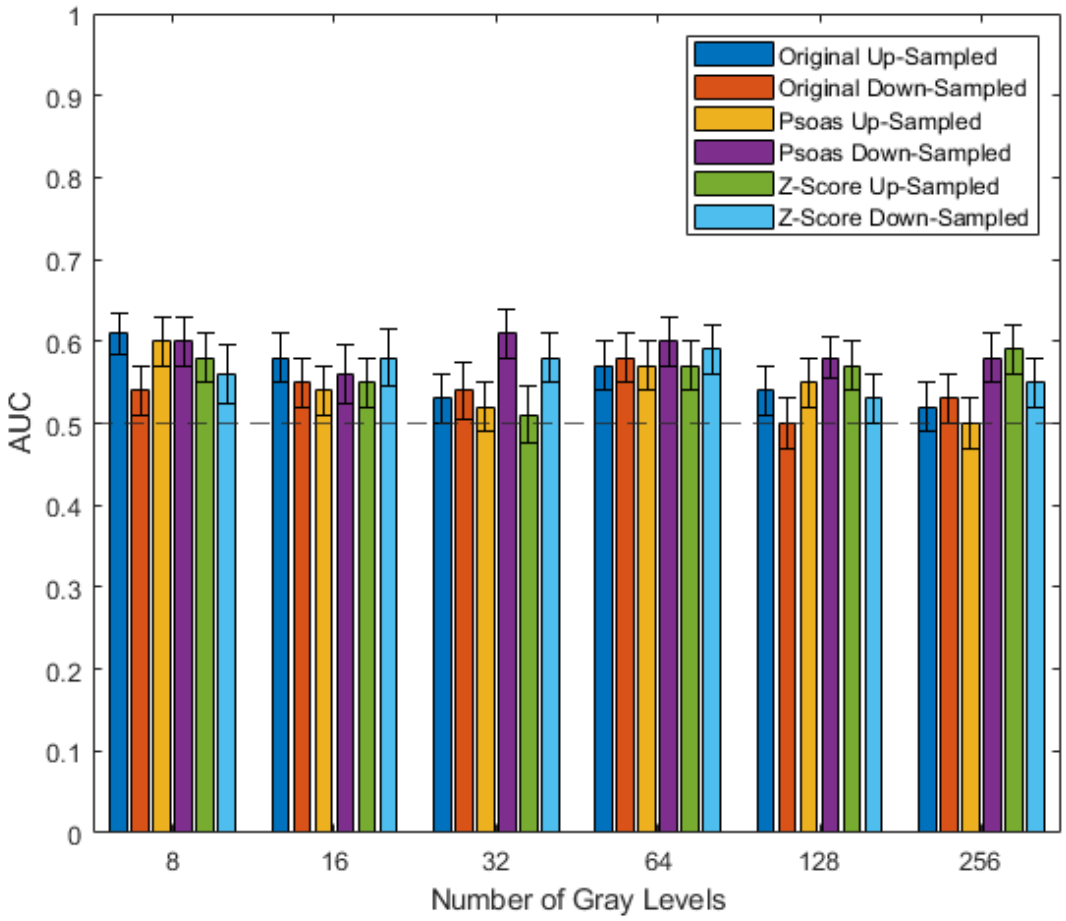

Figure 2.7: Area under the receiver operating characteristic curve (AUC) values in classifying genotype using radiomic features extracted from the non-cystic kidney parenchyma using fixed bin number (FBN) discretization. The dotted line at an AUC of 0.5 is random guessing.

Table 2.6: Area under the receiver operating characteristic curve (AUC) along with the 95% confidence interval (CI) of AUC for features extracted from the entire kidney parenchyma using fixed bin size (FBS) discretization.

	Number of Gray Levels					
Normalization	8	16	32	64	128	256
Original						
1.0x1.0	0.63	0.61	0.60	0.58	0.58	0.59
	[0.60,0.66]	[0.58,0.64]	[0.57,0.63]	[0.55,0.61]	[0.55,0.61]	[0.56,0.62]
2.0x2.0	0.62	0.62	0.56	0.58	0.57	0.58
	[0.59,0.64]	[0.59,0.65]	[0.53,0.59]	[0.55,0.61]	[0.54,0.60]	[0.54,0.61]
Psoas						
1.0x1.0	0.73	0.70	0.68	0.71	0.71	0.69
	[0.70,0.76]	[0.67,0.72]	[0.65,0.70]	[0.68,0.74]	[0.68,0.73]	[0.66,0.72]
2.0x2.0	0.70	0.65	0.70	0.67	0.67	0.65
	[0.68,0.73]	[0.62,0.68]	[0.67,0.73]	[0.64,0.70]	[0.64,0.69]	[0.62,0.68]
Z-score						
1.0x1.0	0.68	0.65	0.58	0.63	0.66	0.65
	[0.65,0.70]	[0.62,0.67]	[0.55,0.61]	[0.60,0.66]	[0.63,0.69]	[0.62,0.68]
2.0x2.0	0.62	0.64	0.59	0.61	0.62	0.58
	[0.59,0.65]	[0.61,0.67]	[0.56,0.63]	[0.58,0.64]	[0.59,0.65]	[0.55,0.61]

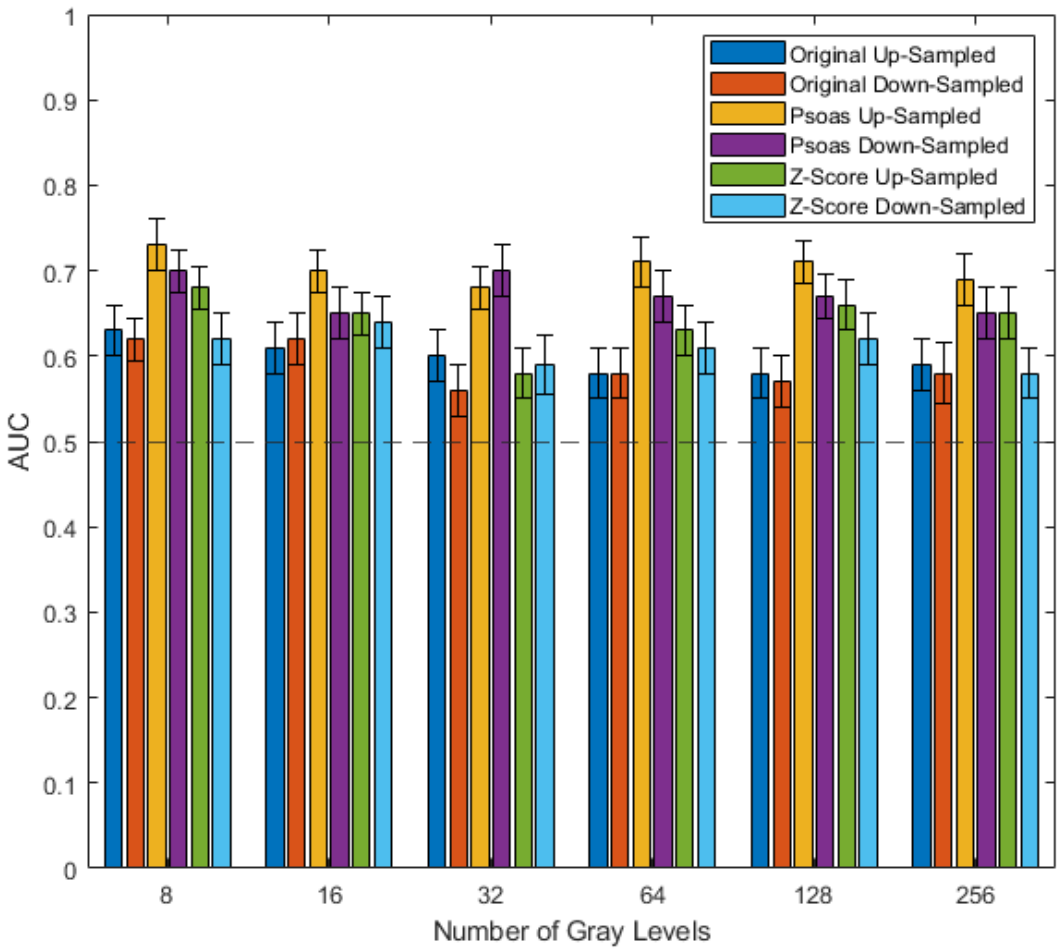

Figure 2.8: Area under the receiver operating characteristic curve (AUC) values in classifying genotype using radiomic features extracted from the entire kidney using fixed bin size (FBS) discretization. The dotted line at an AUC of 0.5 is random guessing.

Table 2.7: Area under the receiver operating characteristic curve (AUC) along with the 95% confidence interval (CI) of AUC for features extracted from the entire kidney parenchyma using fixed bin number (FBN) discretization.

	Number of Gray Levels					
Normalization	8	16	32	64	128	256
Original						
1.0x1.0	0.68	0.66	0.57	0.60	0.62	0.63
	[0.65,0.71]	[0.63,0.69]	[0.54,0.60]	[0.57,0.63]	[0.59,0.65]	[0.61,0.66]
2.0x2.0	0.64	0.62	0.62	0.61	0.61	0.62
	[0.60,0.66]	[0.59,0.64]	[0.59,0.65]	[0.58,0.64]	[0.58,0.64]	[0.58,0.65]
Psoas						
1.0x1.0	0.68	0.66	0.57	0.62	0.61	0.68
	[0.65,0.71]	[0.63,0.69]	[0.54,0.60]	[0.59,0.65]	[0.58,0.64]	[0.65,0.71]
2.0x2.0	0.64	0.62	0.64	0.61	0.65	0.65
	[0.62,0.67]	[0.59,0.65]	[0.61,0.67]	[0.58,0.65]	[0.61,0.67]	[0.62,0.67]
Z-score						
1.0x1.0	0.68	0.65	0.57	0.61	0.62	0.65
	[0.65,0.71]	[0.62,0.68]	[0.54,0.60]	[0.58,0.64]	[0.59,0.65]	[0.61,0.67]
2.0x2.0	0.65	0.62	0.63	0.62	0.63	0.61
	[0.62,0.68]	[0.59,0.64]	[0.61,0.66]	[0.58,0.65]	[0.60,0.66]	[0.58,0.64]

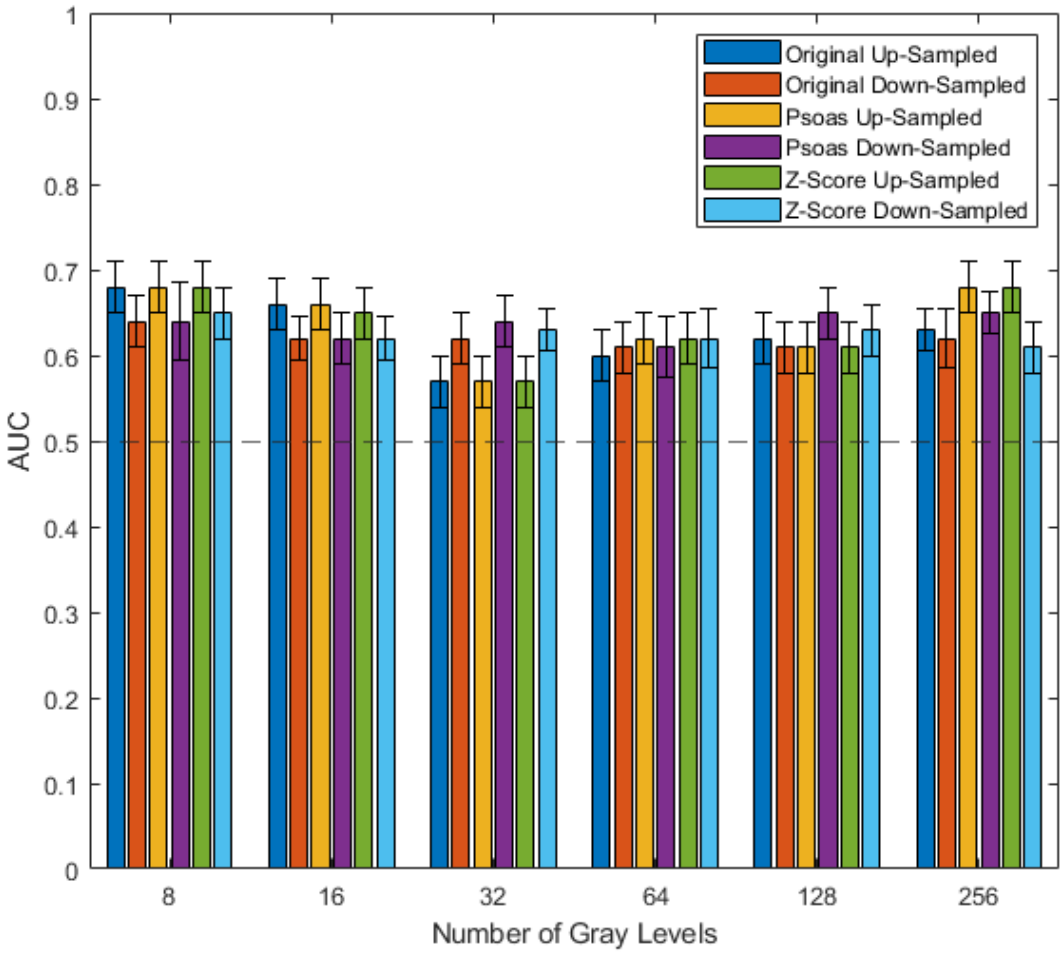

Figure 2.9: Area under the receiver operating characteristic curve (AUC) values in classifying genotype using radiomic features extracted from the entire kidney using fixed bin number (FBN) discretization. The dotted line at an AUC of 0.5 is random guessing.

The top-performing radiomic features selected across training partitions of the 5-fold rCV in the non-cystic kidney parenchyma and the entire kidney classification were explored of a representative pre-processing parameter (Table 2.8); z-score normalization, pixel resampling to 2.0x2.0 mm, and FBS discretization using 64 gray levels was the representative pre-processing method chosen for the non-cystic kidney parenchyma, and psoas muscle normalization, pixel resampling to 1.0x1.0 mm, and FBS discretization using 64 gray levels was

Table 2.8: The radiomic features chosen in the training phase for any partition in the 5-fold repeated cross-validation (rCV). For the non-cystic kidney parenchyma, the representative pre-processing parameters of z-score normalization, pixel resampling to 2.0x2.0 mm, and fixed bin size (FBS) discretization using 64 gray levels was employed. For the entire kidney, the representative pre-processing parameters of psoas muscle normalization, pixel resampling to 1.0x1.0 mm, and FBS discretization using 64 gray levels was employed.

Radiomic features	Entire kidney (n=17)	Union (n=20)	Non-cystic kidney (n=16)
First-order	Robust Mean Absolute Deviation	Energy, Interquartile Range, Kurtosis, Median, Minimum	10th Percentile, 90th Percentile, Maximum, Skewness, Total Energy, Uniformity
GLCM	Inverse Difference, Inverse Difference Moment, Inverse Variance	Cluster Prominence, Cluster Shade, Inverse Difference Moment Normalized, Inverse Difference Normalized, Maximal Correlation Coefficient, Maximum Probability, Sum Squares	Informational Measure of Correlation 1, Informational Measure of Correlation 2
GLSZM	Large Area High Gray-Level Emphasis, Small Area High Gray-Level Emphasis, Zone Entropy, Zone Percentage, Zone Variance	Gray-Level Non-Uniformity, Low Gray-Level Zone Emphasis, Size-Zone Non-Uniformity, Small Area Low Gray-Level Emphasis	Large Area Low Gray-Level Emphasis, Size Zone Non-Uniformity Normalized, Small Area Emphasis
GLRLM	Run Length Non-Uniformity, Run Percentage		
NGTDM	Coarseness, Strength	Busyness	Complexity, Contrast
GLDM	Dependence Non-Uniformity Normalized, Gray-Level Non-Uniformity, Small Dependence Emphasis, Small Dependence High Gray-Level Emphasis	Dependence Entropy, Dependence Non-Uniformity, Large Dependence Low Gray-Level Emphasis	Large Dependence High Gray-Level Emphasis, Low Gray-Level Emphasis, Small Dependence Low Gray-Level Emphasis

the representative pre-processing method chosen for features from the entire kidney. For feature extraction and classification from the non-cystic kidney parenchyma, there was a total of 36 features selected in the training partitions of the rCV; from the entire kidney, there was a total of 37 features selected in the training partitions of the rCV. There were 20 radiomic features that were used in both classification schemes, features from the non-cystic kidney parenchyma and entire kidney, and those are found in the Union column of Table 2.8.

2.3.3 PKD1 vs. PKD2 using clinical and radiomic features

Since the cohorts were matched for age-, gender-, and MIC class, eGFR was explored as a standalone clinical feature for classification of genotype. Between the cohorts, there was a statistically significant difference with respect to eGFR, a measure of kidney function (Figure 2.10); the mean (standard deviation) of *PKD1* was 80.91 (16.55) and *PKD2* was 87.01 (15.48), showing *PKD2* having a higher, preserved kidney function. Figure 2.11 shows the calculated Pearson correlation coefficient between eGFR and htTKV for *PKD1* and *PKD2*, with *PKD2* having a more negative correlation coefficient between eGFR and htTKV. When using eGFR as a standalone feature for genotype classification, the AUC was 0.61, compared to the highest AUC value using radiomic features of the non-cystic kidney parenchyma, 0.68, and the entire kidney parenchyma, 0.73. Combining eGFR with radiomics, the AUC performance using radiomic features extracted from the entire kidney was 0.74 and 0.69 using radiomic features extracted from the non-cystic kidney parenchyma, both resulting in minimally higher performance than radiomics alone.

Table 2.9: Area under the receiver operating characteristic curve (AUC) along with the 95% confidence interval (CI) of AUC for genotype classification using estimated glomerular filtration rate (eGFR), radiomic features, and combined clinical-radiomics model.

Features	AUC [95% CI]
eGFR	0.61 [0.58, 0.64]
Radiomics model	
Entire kidney	0.73 [0.70,0.76]
Non-cystic kidney	0.68 [0.65,0.71]
Clinical and radiomics model	
Entire kidney+eGFR	0.74 [0.71,0.76]
Non-cystic kidney+eGFR	0.69 [0.67,0.72]

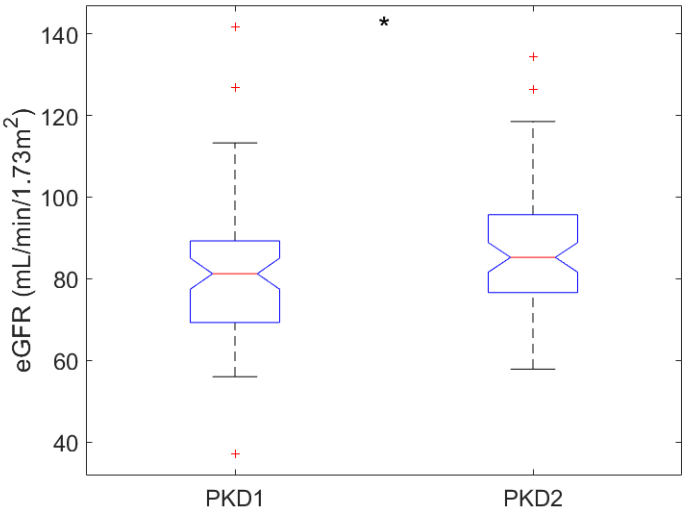

Figure 2.10: Box and whisker plot of *PKD1* and *PKD2* estimated glomerular filtration rate (eGFR) values, showing a statistically significant difference using the Wilcoxon rank sum test ($p<0.05$) [68]. The median eGFR [min eGFR, max eGFR] of *PKD1* and *PKD2* was 81.22 [37.01, 141.8] and 85.23 [57.85, 134.3], respectively. The box represents the interquartile range, with the central line indicating the median value; the whiskers extend to the minimum and maximum values, while outliers are represented as individual data points.

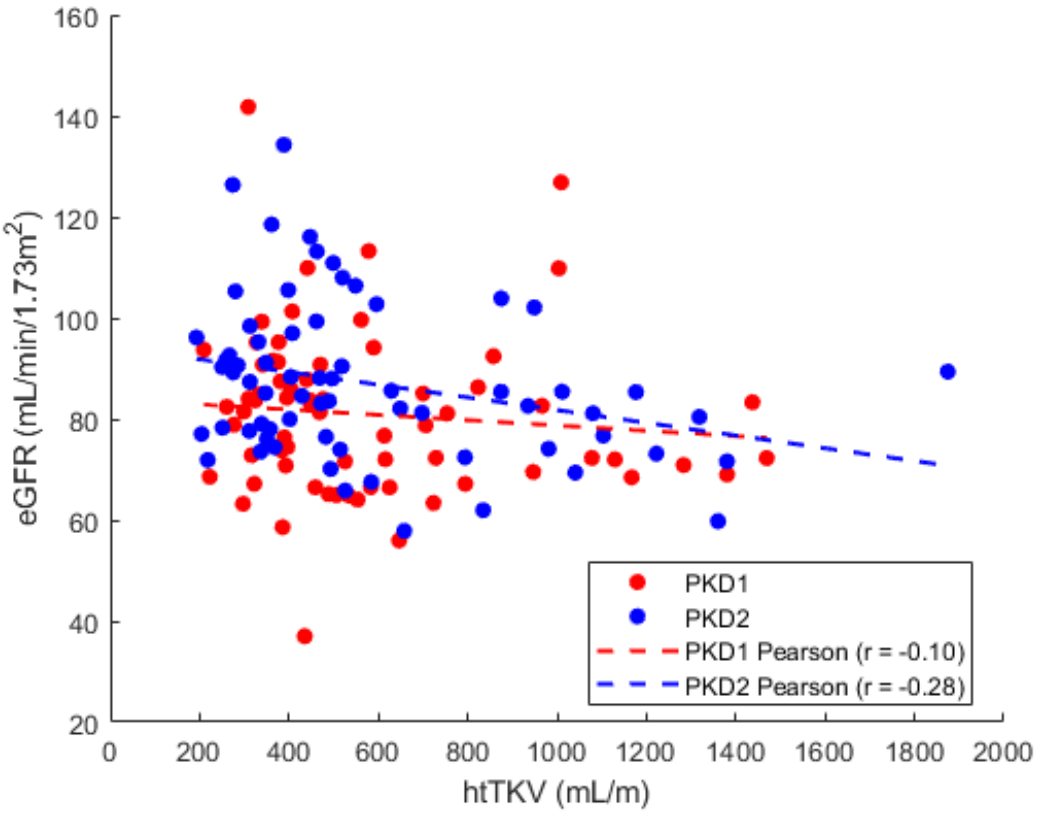

Figure 2.11: Scatter plot showing the relationship between estimated glomerular filtration rate (eGFR) and height-corrected total kidney volume (htTKV) for *PKD1* (shown in red) and *PKD2* (shown in blue). Additionally, linear regression lines representing the Pearson correlation coefficient (r) between eGFR and htTKV are plotted separately for *PKD1* (red) and *PKD2* (blue).

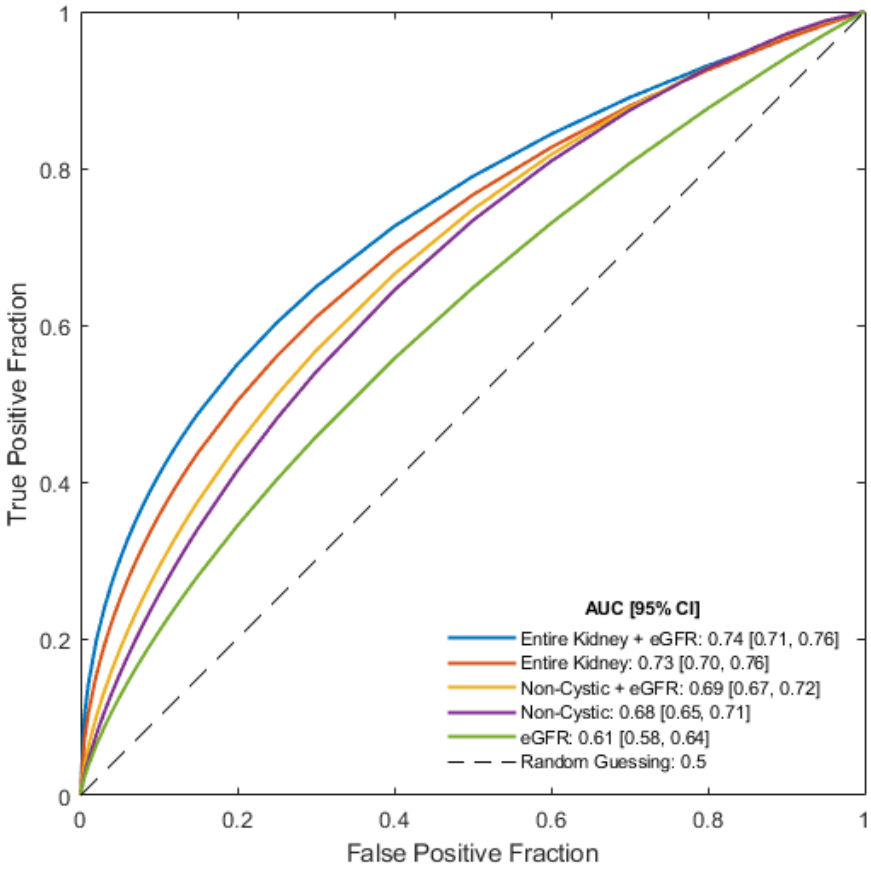

Figure 2.12: Fitted proper binormal receiver operating characteristic (ROC) curves for genotype classification using eGFR, radiomics, and combined clinical-radiomics models using radiomic features from the entire kidney or non-cystic kidney parenchyma. The legend gives the area under the receiver operating characteristic curve (AUC) along with the 95% confidence interval (CI) of AUC for each classifier.

2.4 Discussion

This work examined pre-processing of MR images and its effect on the reproducibility of radiomic features extracted from the non-cystic kidney parenchyma from patients with ADPKD. The ability of these radiomic features to classify patients with *PKD1* or *PKD2* genetic mutations was explored using both the non-cystic kidney parenchyma and the entire kidney. The percentage of radiomic features extracted with good-to-excellent feature repro-

ducibility ranged from 17-58% and 15-58% across the number of gray levels used for feature calculation for up-sampling and down-sampling, respectively. In examining the classification performance of patients with either the *PKD1* or *PKD2* genetic variant across MR normalizations and pre-processing parameters, the AUC values ranged between 0.47-0.68 and 0.56-0.73 for the non-cystic kidney parenchyma and entire kidney, respectively.

This study is the first to examine the non-cystic kidney parenchyma in MR images of ADPKD patients. Xie et al. [55] recently used radiomic features extracted from the renal parenchyma volume (RPV) on computed tomography (CT) to predict kidney function decline. The results of this study showed improved predictive power using texture features rather than TKV or RPV alone, suggesting that the non-cystic kidney parenchyma has prognostic power of future kidney function decline in ADPKD [55]. In the work presented in this chapter, two image normalization methods, z-score normalization and reference-tissue normalization using the psoas muscle, and the original image were chosen for feature extraction across different gray-level discretization and pixel resampling schemes. In the MRI ADPKD radiomics literature, pre-processing parameters are either heterogeneous or not defined, and the effect of pre-processing parameters on radiomic features extracted from the non-cystic kidney parenchyma has not been appreciated.

There was an overall trend of increasing mean ICC scores for first-order, GLCM, GLDM, GLRLM, GLSZM, and NGTDM features, and the percentage of total features with good-to-excellent feature reproducibility when the number of gray levels was increased in the ROI. In decreasing the number of gray levels for feature extraction, the impact of normalization on the inherent pixel information was evident, showing that the chosen normalization method dramatically affects the original pixel intensity information and downstream feature calculation for all feature families. A pair-wise comparison of z-score and psoas muscle normalization obtained poor radiomic feature reproducibility for more than 50% of the radiomic features using 32 gray levels or fewer; the lowest percentage of radiomic features with poor repro-

ducibility obtained was 32% when discretizing the non-cystic kidney parenchyma gray levels to 256. Carré et al. [40] investigated the impact of the MR normalization method and gray-level discretization on radiomic feature stability from post-contrast 3D axial T1-weighted and axial T2-weighted fluid attenuation inversion recovery images 1 month apart and found that a higher number of bins was associated with a higher number of robust features for both sequences. In the present study, among the 93 features extracted across gray levels and resampling schemes there were 7 features that exhibited good-excellent feature ICC scores; the up-sampling scheme resulted in 2 additional features, for a total of 9 features, that exhibited good-to-excellent reproducibility. Among these features, GLSZM gray-level non-uniformity and NGTDM coarseness have been identified in previous radiomics studies as being correlated with ROI size or voxel number [59, 69]. The effect of pre-processing on radiomic features from qualitative MR images has been investigated in the application of brain and prostate diseases [40, 59, 63, 70, 71]. Conclusions from these studies differ with respect to optimal pre-processing parameters, showing the unique parameters for a given task, body site, and MR sequence. For example, in the task of brain cancer assessment, radiomics studies have used z-score normalization of MR images and discretizing to 32 gray levels when using FBS discretization and first- and second-order radiomic features, while prostate cancer radiomics literature has used between 36-42 gray levels with FBS discretization and z-score normalization, histogram matching, or reference-tissue normalization [40, 71]. The results of this study show that radiomics features extracted from the non-cystic kidney parenchyma using different MR normalization and gray levels for discretization has an effect on radiomic feature reproducibility. These results add to the literature in other disease cohorts investigating the effect of MR pre-processing on radiomic features.

Extending these pre-processing parameters to a multi-site data set for the clinical task of genotype classification, the highest AUC values across the number of gray levels for discretization were 0.68, 0.61, and 0.62 for z-score normalization, psoas muscle normalization, and the

original image, respectively, for features extracted from the non-cystic kidney parenchyma. In addition to the non-cystic kidney parenchyma, the entire kidney was also used for feature extraction. Among the pre-processing parameters, the highest AUC values were 0.68, 0.73, and 0.63 for z-score normalization, psoas muscle normalization, and the original image, respectively, for features extracted from the entire kidney parenchyma. Table 2.5 shows the effect of normalization on the signal range across patients. The non-cystic kidney parenchyma has an overall lower range of signal intensities compared with the entire kidney including the cysts, which tend to be of higher signal intensity and, therefore, with a larger range of values for discretization. Normalization of the MR images tended to improve AUC values using non-cystic kidney parenchyma and entire kidney radiomic features for classification. Z-score normalization uses the mean and standard deviation of the entire image and includes signal intensities from cysts, while psoas muscle normalization used the mean and standard deviation of the psoas-muscle that is hypointense in T2W-FS images. Z-score normalization using the downsampling method and 64 gray levels for discretization resulted in the highest AUC value for classification using features from the non-cystic kidney parenchyma. When including textural information from cysts, the psoas muscle normalization resulted in the highest AUC values across gray levels and pixel resampling methods. There was not one pre-processing method that optimized the classification performance using the non-cystic kidney parenchyma and the entire kidney. The AUC performance across radiomic models and combined clinical-radiomic models using eGFR and features extracted from the non-cystic kidney parenchyma and the entire kidney were between 0.68-0.74; the AUC in classifying genotype using eGFR alone was 0.61, showing moderate increase in performance and the added potential of texture features in capturing differences in genotype beyond a traditional clinical measurement.

This study has a few limitations. With respect to pre-processing, there are additional normalization methods such as histogram-matching and gray-level discretization methods.

The results from this study raise an important trade-off in radiomics research of optimal pre-processing parameters for a given task and the effect on feature reproducibility across pre-processing parameters, as optimizing the pre-processing parameters to obtain the highest feature reproducibility does not necessarily lead to the best classification.

With respect to the patient data investigated, the impact of mutation type either in functional domains or in mutation strength was not taken into account in this analysis when matching for age, gender, and MIC class. *PKD1* patient outcomes differ for truncating and non-truncating mutations, and this variability may impact genotype classification [13]. With respect to patient data, the number of available patients using the HALT dataset was limited as the prevalence of *PKD2* is lower than that of *PKD1*. The number of *PKD2* subjects available to study was small, and a larger sample size would potentially confirm texture differences with greater precision. In controlling for htTKV, this allowed for minimal differences in cyst burden between *PKD1* and *PKD2* to impact the results. However, the purpose here was to evaluate regions of the kidney without the influence of differences in cyst burden. This study provided preliminary information about the potential impact of finding textural changes that associate with genotype in ADPKD.

Future work will include incorporating more image slices, as this study included one largest, representative, coronal slice of the kidney for analysis. Removing the cysts from the kidney is a tedious process and would benefit from an automated approach. Recent work has utilized semantic instance segmentation to segment kidneys and cysts within the kidney, which may prove useful in providing a fast calculation of entire kidney and non-cystic kidney parenchyma texture [72]. Utilizing multi-slice MR texture data of ADPKD kidneys may provide more information of kidney texture rather than a single imaging slice. Additionally, other MR imaging sequences such as T1-weighted, diffusion-weighted imaging, and quantitative maps could be used to analyze textural differences in ADPKD genotype with respect to mutation type and strength. Retrospective ADPKD MR datasets provide a

wealth of imaging data to explore the use of radiomic features and provide valuable insight in the prediction of future kidney function decline, texture differences in genetic mutation, and assessing the longitudinal change in texture over time. However, this work emphasizes the need for ADPKD radiomics studies to state the pre-processing parameters used for feature extraction as they may have a dramatic effect on subsequent classification.

2.5 Conclusion

This study investigated the impact of pre-processing on radiomic features extracted from the non-cystic kidney parenchyma on T2W-FS MR images of patients with ADPKD. Radiomic features extracted from the non-cystic kidney parenchyma were sensitive to MRI normalization, and the results show that feature reproducibility across MRI normalization is dependent on the number of gray levels available for discretization. Classification performance in distinguishing *PKD1* and *PKD2* varied with respect to the pre-processing parameters. This work revealed that there are texture features indicative of genotype expression in ADPKD in both the non-cystic and entire kidney parenchyma regions. Additionally, there were preferred pre-processing parameters when using features extracted from the non-cystic kidney parenchyma compared with those from the entire kidney, and there was not one method that optimized features extracted from either region. The results of this study show the importance of reporting pre-processing parameters used for feature extraction, as these affect subsequent classification of ADPKD genotype.

CHAPTER 3

MRI RADIOMICS ANALYSIS OF NON-CYSTIC KIDNEY PARENCHYMA TO DIFFERENTIATE AMONG MAYO IMAGING CLASSIFICATION CLASSES

3.1 Introduction

Mayo Imaging Classification (MIC) provides an estimate of annual growth rate of height-corrected total kidney volume (htTKV) and serves as a prognostic biomarker for the prediction of kidney function decline of patients with autosomal dominant polycystic kidney disease (ADPKD) [11, 73]. MIC is currently the best prediction model for selecting rapid progressors of patients with ADPKD and is used for patient enrollment into clinical trials. The rate of increase in htTKV provided by MIC is also used to predict the time to end-stage kidney disease (ESKD) or decline in estimated glomerular filtration rate (eGFR). Patients with low/intermediate MIC classes have extended, preserved kidney function compared with high MIC classes, which demonstrate steeper rates of kidney function decline [13]. The estimated yearly kidney growth rate for each class is <1.5% (1A), 1.5%–3% (1B), 3%–4.5% (1C), 4.5%–6% (1D), or >6% (1E) [11]. Patients in the 1C-1E MIC classes have rapidly progressing disease and are deemed most likely to benefit from therapy [11]; MIC 1A is considered to be low risk for kidney function decline, while MIC 1B patients are intermediate risk and should be re-evaluated yearly for a more accurate risk of progression [11]. Lavu et al. [13] found onset of ESKD at 45.1 years (1E), 55.6 years (1D), 62.8 years (1C), and 71.2 (1B) years across MIC classes 1B-1E, and less than 20% of MIC 1A patients reaching ESKD. While MIC is a strong predictor for kidney function decline, there remains a need to identify biomarkers that can quantify ADPKD-related kidney tissue damage at earlier stages, thereby enhancing disease prognostication [14].

Radiomic features may provide additional power in stratifying patients with ADPKD early on providing power in risk-assessment models to identify those patients that are fast progressors to kidney function decline, ultimately leading to ESKD. Specifically, characterizing the non-cystic kidney tissue using radiomic features may associate and differentiate among MIC classes.

Therefore, the work presented in this chapter was two-fold: (1) to determine whether radiomic features can differentiate between low/intermediate- and high-risk MIC classes in both the cystic and non-cystic components of the kidney and (2) investigate the effect of pre-processing on subsequent classification.

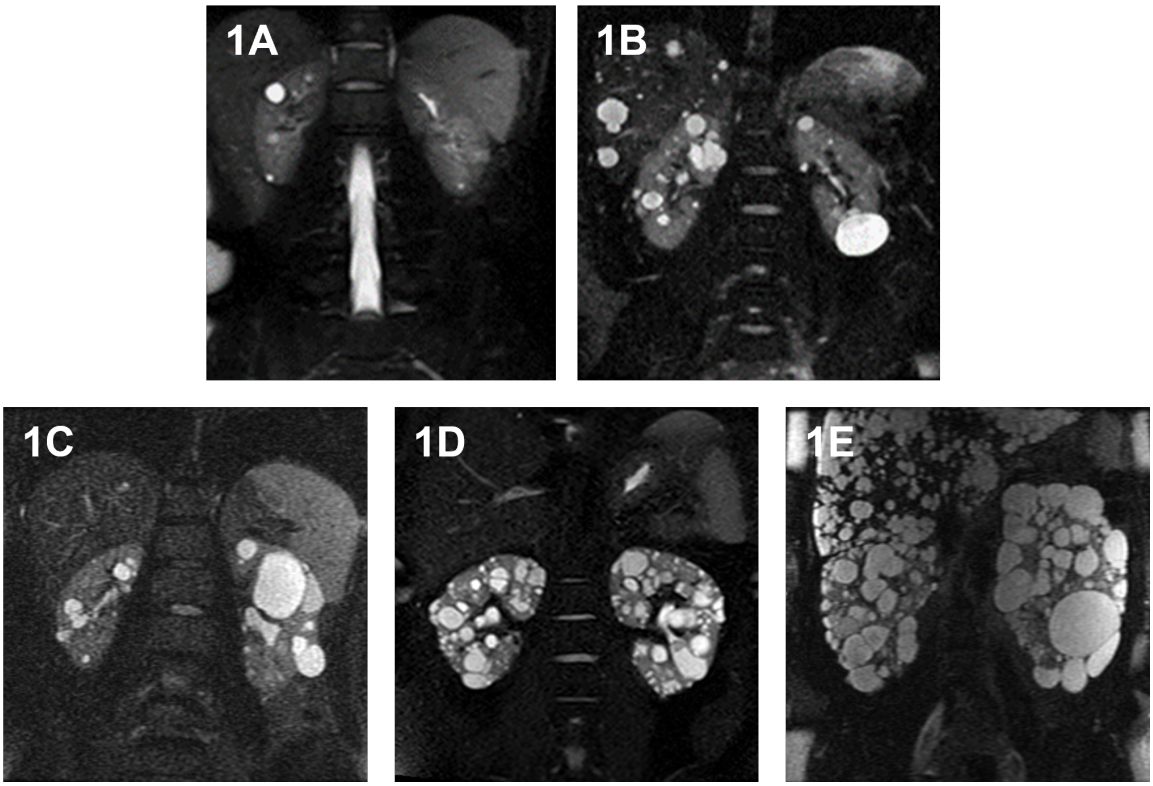

Figure 3.1: Coronal MR images of the kidneys of an 18-year-old Mayo Imaging Classification (MIC) class 1A patient (top left), 34-year-old MIC class 1B patient (top right), 17-year-old MIC class 1C patient (bottom left), 23-year-old MIC class 1D patient (bottom center), and 43-year-old MIC class 1E patient (bottom right). MIC classes (based on age and height-corrected total kidney volume (htTKV)) range from low risk (1A) to high risk (1E) for kidney function decline.

3.2 Methods

3.2.1 Database

T2-weighted fat saturated (T2W-FS) MRI images in this study had been acquired at baseline from a longitudinal, multi-site randomized clinical trial, HALT Progression of Polycystic Kidney Disease (HALTA-PKD) (NCT00283686). The cohort included 138 MIC 1A/1B patients and 324 MIC 1C/1D/1E patients for a total of 462 patients. Table 3.1 includes patient demographic characteristics with respect to gender, age, and average htTKV of the low/intermediate- and high-risk groups, and is differentiated by genetic mutation: *PKD1*, *PKD2*, and no mutation detected (*NMD*).

Table 3.1: HALT subject characteristics for the differentiation among Mayo Imaging Classification (MIC) 1A-1E.

Clinical	*PKD1*	*PKD2*	*NMD*
No. Patients	329	72	61
Mean age ±SD			
Male (n=230)	35 ±7.34 (n=167)	40 ±7.84 (n=34)	35 ±8.76 (n=29)
Female (n=232)	36 ±8.51 (n=162)	40 ±7.91 (n=38)	37 ±8.94 (n=32)
Mean htTKV ±SD			
MIC 1A (n=27)	235.07 ±40.29 (n=9)	220.02 ±34.58 (n=9)	237.18 ±44.38 (n=9)
MIC 1B (n=111)	400.28 ±97.28 (n=69)	380.98 ±91.28 (n=30)	378.20 ±138.72 (n=12)
MIC 1C (n=160)	587.57 ±167.79 (n=119)	659.54 ±229.07 (n=23)	586.59 ±166.01 (n=18)
MIC 1D (n=107)	909.41 ±316.91 (n=89)	1146.24 ±351.69 (n=8)	869.40 ±335.40 (n=10)
MIC 1E (n=57)	1218.37 ±468.70 (n=43)	1627.36 ±349.44 (n=2)	1174.01 ±509.60 (n=12)

T2W-FS MR images were acquired on 1.5T scanners using single-shot fast spin echo/half-Fourier acquisition (SSFSE/HASTE) with fat saturation and a 3-mm fixed slice thickness. All HALT MR image matrix sizes ranged from 256x256-560x560 pixels, and pixel sizes ranged from 0.59-1.8 mm. Representative 2D MR images of the left and right kidney were chosen

based on the coronal MR image that maximized the longitudinal length for each kidney individually.

3.2.2 U-Net for segmentation

U-Net is a deep convolutional neural network (CNN) that is used for the task of segmenting medical images with limited training images [74]. The U-Net architecture developed by Ronneberger et al. [74] gets its name due to the u-shaped contracting path (capturing context) and symmetric expanding path (precise localization) (Figure 3.2). The contracting path captures image context using high-resolution feature maps and consists of repeated convolutions, rectified linear unit (ReLU), and max pooling operations. The expanding path consists of up-sampling, up-convolutions, concatenations (with the contracting path's corresponding cropped feature map), and ReLU operations that all work to combine image context from the feature maps and spatial information for a localization of the predicted pixel class label as either object or background.

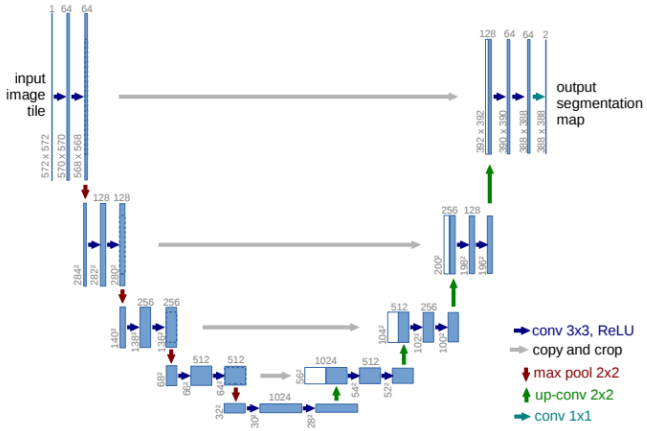

Figure 3.2: U-net architecture (example for 32x32 pixels in the lowest resolution). Each blue box corresponds to a multi-channel feature map. The number of channels is denoted on top of the box. The x-y size is provided at the lower left edge of the box. White boxes represent copied feature maps. The arrows denote the different operations. Reprinted from Ronnebeger, et al. [74].

The 2D U-Net implemented a 75% train, 10% validation, and 15% test scheme, by patient, on a total of 227 left kidney and 226 right kidney manual segmentations of both kidney and non-cystic kidney parenchyma (using one representative, largest coronal image slice per kidney, per patient) from the HALTA-PKD dataset using T2W-FS MR images. All images were resized to a 256x256 matrix size. In total, there were four U-Net models used for segmentation: left kidney, left non-cystic kidney parenchyma, right kidney, and right non-cystic kidney parenchyma (Figure 3.3). In removing the cysts from the kidney, the U-Net model for segmenting the kidney was applied to the original image for a resultant masked kidney image. Masked kidney T2W-FS images were used to train, validate, and test the U-Net model instead of the original image (including anatomy other than the kidney). Removing anatomy other than the kidney improved the segmentation performance. The output of the U-Net models is a probability of a pixel belonging to the desired segmentation task (object or background). The images were then subsequently binarized using Otsu's method of threshold selection from gray-level histograms [75]. The U-Net was used as a semi-automated tool, and the output of the U-Net was reviewed and corrected if needed using an in-house MATLAB segmentation tool [67]. The final segmentations were used for feature extraction after pre-processing (MRI normalization, pixel resampling, gray-level discretization).

The performance of the U-Net model was evaluated using the Dice similarity coefficient (DSC) and Hausdorff distance (HD) (Table 3.2) [76, 77]; DSC evaluates the agreement or overlap (0 being no overlap and 1 being perfect overlap) between manual segmentations and the U-Net output, while HD measures how well the margins of the two segmentations agree. There was a total of 35 patients from the HALT dataset for testing. Segmentations from T2W-FS MR images of 15 patients from the CRISP dataset were used as an external test set. Of the total 50 patients for testing of the U-Net models, 49 had measured htTKV; the breakdown of MIC across the 49 patients were: 1A (3), 1B (21), 1C (13), 1D (5), 1E (7).

Median DSC ranged from 0.922-0.957 across the left and right kidney segmentations using CRISP and HALT datasets; the maximum and minimum values across DSC were 0.986 and 0.153, respectively. The median DSC ranged from 0.905-0.948 for the non-cystic kidney parenchyma across the left and right kidney using the CRISP and HALT datasets; the maximum and minimum values across DSC were 0.991 and 0.586, respectively.

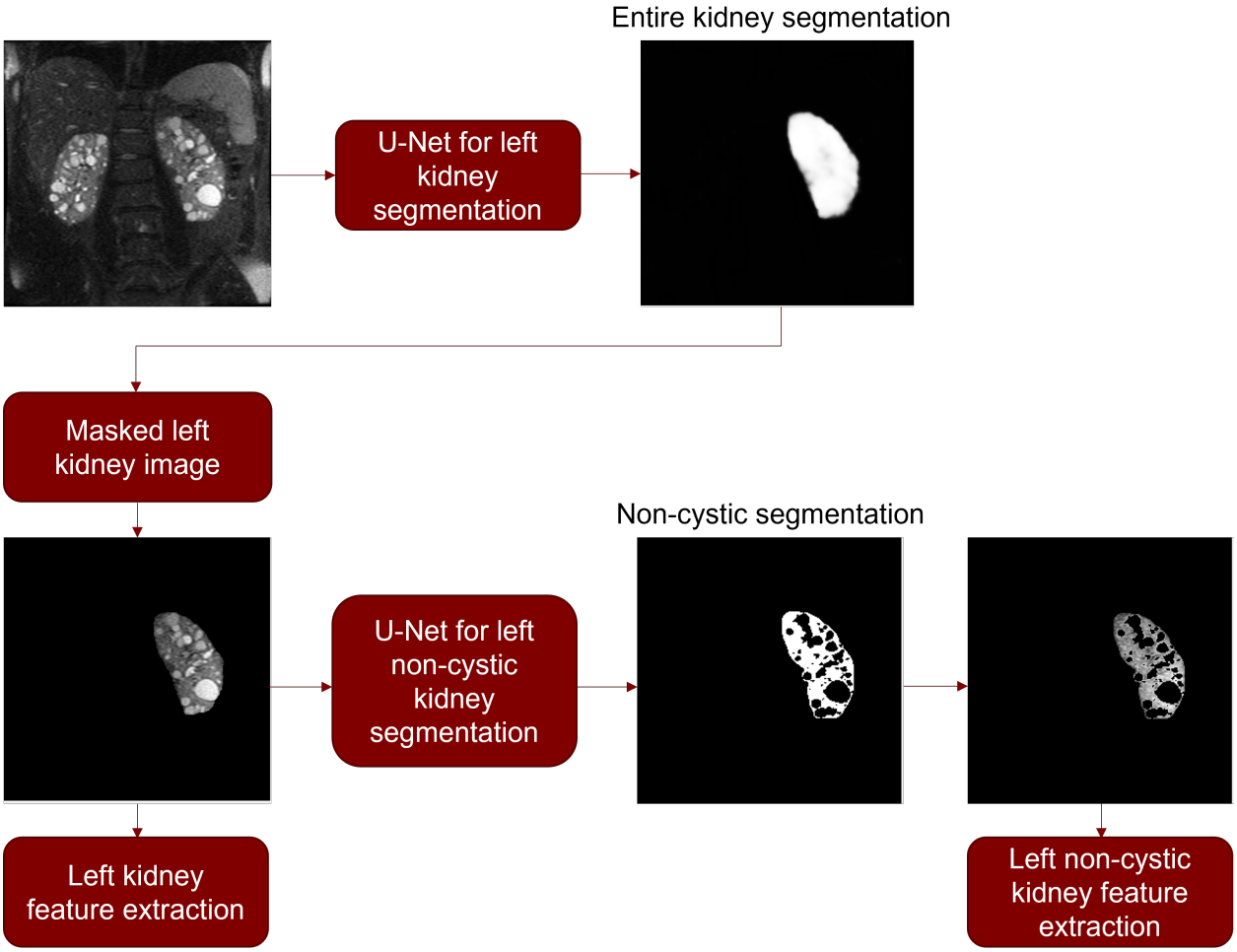

Figure 3.3: Pipeline using the U-Net models (4 in total) to segment left kidney and left non-cystic kidney parenchyma (the same process was done for the right kidney and right non-cystic kidney parenchyma), first starting with the original T2-weighted fat saturated (T2W-FS) MR image.

Table 3.2: U-Net model performance. The images of the Consortium for Radiologic Imaging Studies of Polycystic Kidney Disease (CRISP) dataset (n=14 patients) were used as an external test set in addition to the test set from the HALT dataset (n=35 patients).

Kidney	U-Net Segmentation	Dataset	Median DSC [95% CI]	Median Average HD [95% CI] (mm)
Left	Kidney	HALT	0.948 [0.926, 0.960]	1.28 [0.872, 2.24]
Left	Kidney	CRISP	0.922 [0.886, 0.952]	2.17 [1.52, 3.37]
Left	Non-cystic	HALT	0.913 [0.898, 0.921]	1.82 [0.753, 3.99]
Left	Non-cystic	CRISP	0.948 [0.919, 0.973]	0.725 [0.427, 1.85]
Right	Kidney	HALT	0.957 [0.947, 0.963]	1.26 [0.875, 1.87]
Right	Kidney	CRISP	0.934 [0.902, 0.953]	1.92 [1.35, 2.54]
Right	Non-cystic	HALT	0.905 [0.889, 0.933]	3.36 [1.83, 7.27]
Right	Non-cystic	CRISP	0.933 [0.908, 0.946]	1.41 [0.528, 2.97]

3.2.3 MR image pre-processing and feature extraction

The established pre-processing pipeline in Chapter 2 (Figure 2.2) is continued for MIC 1A/1B vs. MIC 1C/1D/1E classification. This includes MR normalization of signal intensity (original image, z-score normalization, reference-tissue normalization using the psoas muscle), pixel resampling using either upsampling (1.0x1.0 mm) or downsampling (2.0x2.0 mm) schemes, gray-level discretization method (fixed bin size (FBS), fixed bin number (FBN)), and number of gray levels for discretization (8, 16, 32, 64, 128, 256).

3.2.4 Correlation of radiomic features with kidney size

Previous research has shown a dependency of potential imaging biomarkers for different disease cohorts on voxel size, gray levels, and other imaging parameters [59, 69, 78]. The dependency of feature values on kidney size (i.e., number of voxels inside the ROI) and the impact of MRI normalization, pixel resampling, and gray-level discretization on this

dependency has not been established. This dependency is of importance because high-risk MIC (1C-1E) classes usually have significantly larger kidneys than low- and interemediate-risk MIC (1A-1B) classes. Radiomic features incorporated in classification models should capture actual texture-based differences rather than reflecting differences in the number of voxels. Therefore, this work investigated the impact of radiomic features correlated with kidney size on the radiomics-based assessment of MIC risk class. The Spearman rank correlation coefficient, ρ, was calculated between feature values and the number of voxels in the kidney; features with $\rho < 0.5$, $0.5 < \rho < 0.9$, and $\rho > 0.9$ were considered to have low, moderate, and high correlation with the number of voxels, respectively. This was calculated across pre-processing parameters from the established pipeline in Chapter 2 (Figure 2.2). The number of voxels after resampling voxel sizes, either up-sampling or down-sampling, was used (calculated and provided by Pyradiomics [61]).

3.2.5 Feature selection and classification

Features were selected using least absolute shrinkage and selection operator (LASSO) as implemented in Statistics and Machine Learning Toolbox using Matlab [67], and were merged in a logistic regression (LR) classifier using 5-fold cross-validation [79]. LASSO trains and fits least-squares regression coefficients that were used to train a linear regression model [67]. Features were standardized to have a zero mean and a variance of one, and a 10-fold cross-validation was used to estimate the mean square error (MSE). Features from the training data with a non-zero coefficient after the regularization process and those that minimized the MSE were chosen. LASSO feature selection was implemented in each fold of the cross-validation scheme, and a Pearson correlation threshold of 0.7 removed correlated features. This process used a repeated cross-validation (rCV) of 10 to account for variance in 5-fold cross-validation.

3.2.6 MIC 1B (intermediate-risk) vs. MIC 1C (high-risk) classification

ADPKD has a heterogeneous phenotype, and although MIC classification remains stable in most patients, there are some patients who can move to an immediate lower MIC class or progress to a higher MIC class during follow-up imaging (more often patients progressing to a higher-risk MIC class) [11, 13, 80]. Patients with an intermediate-risk (MIC 1B) are suggested to be re-evaluated at yearly intervals to more accurately classify their risk for progression [11]. Investigating radiomic differences between intermediate-risk MIC 1B patients and high-risk MIC 1C patients may provide additional insights on patients who are at higher risk for kidney function decline.

In addition to using LASSO and the LR machine learning classifier, a supervised learning algorithm with known truth labels, fuzzy c-means (FCM) clustering was utilized. FCM is an unsupervised technique in which data was clustered into N clusters with a datapoint having a certain probability of belonging to a particular cluster [67]. FCM clustering minimizes an objective function over iterations to locate the cluster centers of each cluster:

$$J_m = \sum_{i=1}^{C} \sum_{j=1}^{N} \mu_{ij}^{m} (D_{ij})^2, \tag{3.1}$$

where C is the number of clusters, N is the number of data points, m is the fuzzy partition matrix that controls the degree of fuzzy overlap between the clusters, D_{ij} is the distance from the jth data point to the ith cluster, and μ_{ij}^{m} is the degree of membership of the jth data point in the ith cluster (the sum of membership values for all clusters is one). The cluster centers are first initiated randomly, the distance for each data point is calculated, membership values are updated for each data point, the objective function J_m is calculated, and this process is repeated until the minimum threshold is met. Default values for the exponent for fuzzy partition matrix, maximum number of iterations, minimum improvement in objective function, and distance metric (euclidean) in Matlab were implemented [67]. FCM clustering

was applied to the MIC 1B and 1C patients using clinical features, radiomic features, and combined clinical and radiomic features; ROC analysis was subsequently applied to the output of FCM, which are class probabilities for each patient.

3.3 Results

3.3.1 Low- and intermediate-risk MIC (1A-1B) vs. high-risk MIC (1C-1E)

Figure 3.4 show shows the MRI signal intensity range for all patient data after normalization and downsampling of pixel size; similar to that of Chapter 2, the non-cystic kidney parenchyma has a lower range of signal intensity than the entire kidney parenchyma due to the exclusion of cysts.

Tables 3.3-3.4 show the AUC values across pre-processing parameters using the original image, z-score normalization, and psoas muscle normalization using features extracted from the non-cystic kidney parenchyma. The range of AUC values was between 0.69-0.85. Across normalizations, the range of AUC values for the original image, z-score normalization, and psoas muscle normalization were 0.69-0.83, 0.75-0.85, and 0.74-0.84, respectively. Across normalizations, FBS discretization, and up-sampling only, the range of AUC values for the original image, z-score normalization, and psoas muscle normalization were 0.69-0.77, 0.77-0.84, and 0.75-0.84, respectively (Figure 3.5); the range of AUC values for the down-sampling method for the original image, z-score normalization, and psoas muscle normalization were 0.70-0.79, 0.75-0.85, and 0.74-0.82, respectively (Figure 3.5). Across normalizations, FBN discretization, and up-sampling only, the range of AUC values for the original image, z-score normalization, and psoas muscle normalization were 0.71-0.82, 0.78-0.83, and 0.77-0.82, respectively (Figure 3.6); the range of AUC values for the down-sampling method for the original image, z-score normalization, and psoas muscle normalization were 0.79-0.83, 0.75-0.84, and 0.76-0.84, respectively (Figure 3.6).

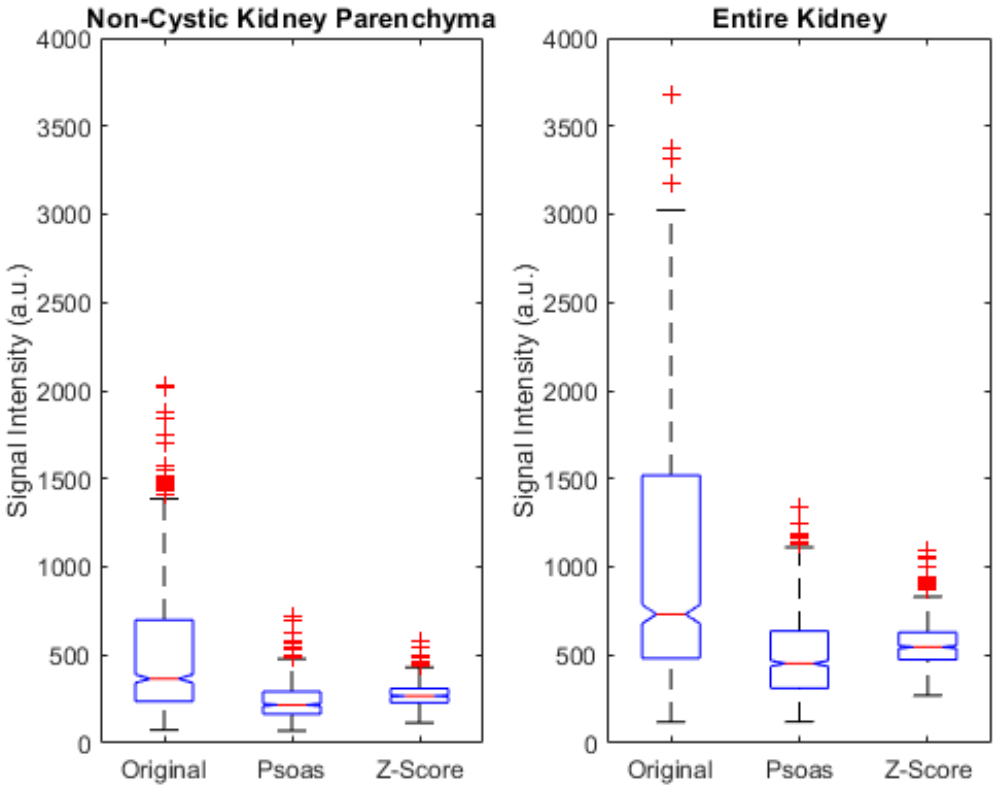

Figure 3.4: Box and whisker plot of the MRI signal intensity range in arbitrary units (a.u.) across HALTA-PKD MIC patient (1A-1E) images and normalization methods after down-sampling. The non-cystic kidney parenchyma (left) had extreme outliers (n=4) with signal intensities ranging between 4901-13540 using the original image; similarly using the original image, the entire kidney range (right) had extreme outliers (n=5) with ranges between 4596-21805. The box represents the interquartile range, with the central line indicating the median value; the whiskers extend to the minimum and maximum values, while outliers are represented as individual data points. The extreme outliers mentioned are not included in the figure.

Table 3.5-3.6 show the AUC values across pre-processing parameters using the original image, z-score normalization, and psoas muscle normalization using features extracted from the entire kidney parenchyma. The range of AUC values was between 0.86-0.90. Across normalizations, the range of AUC values for the original image, z-score normalization, and psoas muscle normalization were 0.86-0.90, 0.88-0.90, and 0.87-0.90, respectively. Across

normalizations, FBS discretization, and up-sampling only, the range of AUC values for the original image, z-score normalization, and psoas muscle normalization were 0.87-0.88, 0.88-0.90, and 0.87-0.90, respectively (Figure 3.7); the range of AUC values for the down-sampling method for the original image, z-score normalization, and psoas muscle normalization were 0.86-0.89, 0.88-0.89, and 0.89-0.90, respectively (Figure 3.7). Across normalizations, FBN discretization, and up-sampling only, the range of AUC values for the original image, z-score normalization, and psoas muscle normalization were 0.88-0.89, 0.88-0.89, and 0.88-0.89, respectively (Figure 3.8); the range of AUC values for the down-sampling method for the original image, z-score normalization, and psoas muscle normalization were 0.88-0.90, 0.88-0.90, and 0.89-0.90, respectively (Figure 3.8).

Table 3.3: Area under the receiver operating characteristic curve (AUC) along with the 95% confidence interval (CI) of AUC for features extracted from the non-cystic kidney parenchyma using fixed bin size (FBS) discretization for Mayo Imaging Classification (MIC) 1A-1B vs. MIC 1C-1E classification.

	Number of Gray Levels					
Normalization	8	16	32	64	128	256
Original						
1.0x1.0	0.77	0.75	0.69	0.70	0.71	0.74
	[0.76,0.79]	[0.74,0.77]	[0.67,0.71]	[0.68,0.71]	[0.70,0.73]	[0.72,0.75]
2.0x2.0	0.79	0.77	0.72	0.71	0.71	0.70
	[0.78,0.80]	[0.75,0.78]	[0.70,0.73]	[0.69,0.72]	[0.69,0.72]	[0.69,0.72]
Psoas						
1.0x1.0	0.84	0.80	0.75	0.75	0.76	0.77
	[0.83,0.85]	[0.78,0.81]	[0.73,0.76]	[0.74,0.77]	[0.74,0.77]	[0.75,0.78]
2.0x2.0	0.82	0.77	0.74	0.76	0.79	0.78
	[0.81,0.83]	[0.75,0.78]	[0.72,0.76]	[0.75,0.78]	[0.78,0.80]	[0.77,0.80]
Z-score						
1.0x1.0	0.84	0.83	0.79	0.77	0.80	0.78
	[0.83,0.85]	[0.82,0.84]	[0.77,0.80]	[0.76,0.79]	[0.78,0.81]	[0.76,0.79]
2.0x2.0	0.85	0.81	0.82	0.80	0.75	0.77
	[0.83,0.86]	[0.79,0.82]	[0.80,0.83]	[0.78,0.81]	[0.73,0.77]	[0.76,0.79]

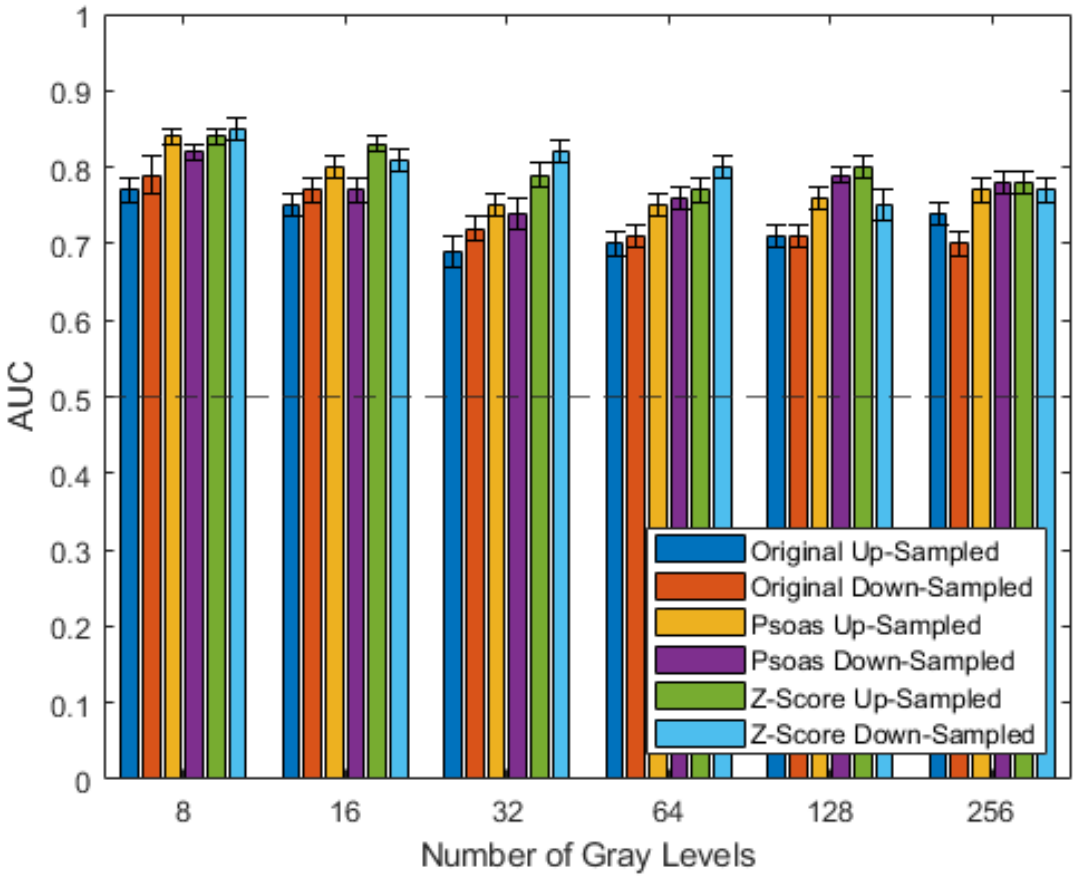

Figure 3.5: Area under the receiver operating characteristic curve (AUC) values in classifying Mayo Imaging Classification (MIC) using radiomic features extracted from the non-cystic kidney parenchyma using fixed bin size (FBS) discretization. The dotted line at an AUC of 0.5 is random guessing.

Table 3.4: Area under the receiver operating characteristic curve (AUC) along with the 95% confidence interval (CI) of AUC for features extracted from the non-cystic kidney parenchyma using fixed bin number (FBN) discretization for Mayo Imaging Classification (MIC) 1A-1B vs. MIC 1C-1E classification.

	Number of Gray Levels					
Normalization	8	16	32	64	128	256
Original						
1.0x1.0	0.82	0.75	0.74	0.71	0.74	0.78
	[0.81,0.84]	[0.73,0.76]	[0.73,0.76]	[0.69,0.72]	[0.72,0.75]	[0.77,0.80]
2.0x2.0	0.83	0.80	0.79	0.79	0.82	0.75
	[0.81,0.84]	[0.79,0.81]	[0.78,0.80]	[0.78,0.81]	[0.81,0.83]	[0.73,0.76]
Psoas						
1.0x1.0	0.82	0.79	0.77	0.77	0.77	0.80
	[0.81,0.83]	[0.77,0.80]	[0.76,0.79]	[0.75,0.78]	[0.75,0.78]	[0.79,0.81]
2.0x2.0	0.84	0.81	0.81	0.80	0.80	0.76
	[0.82,0.85]	[0.79,0.82]	[0.79,0.82]	[0.79,0.81]	[0.79,0.82]	[0.74,0.77]
Z-score						
1.0x1.0	0.83	0.80	0.78	0.78	0.78	0.81
	[0.82,0.84]	[0.78,0.81]	[0.77,0.79]	[0.77,0.80]	[0.77,0.80]	[0.80,0.82]
2.0x2.0	0.84	0.82	0.81	0.80	0.82	0.75
	[0.83,0.85]	[0.81,0.83]	[0.79,0.82]	[0.79,0.82]	[0.81,0.84]	[0.74,0.76]

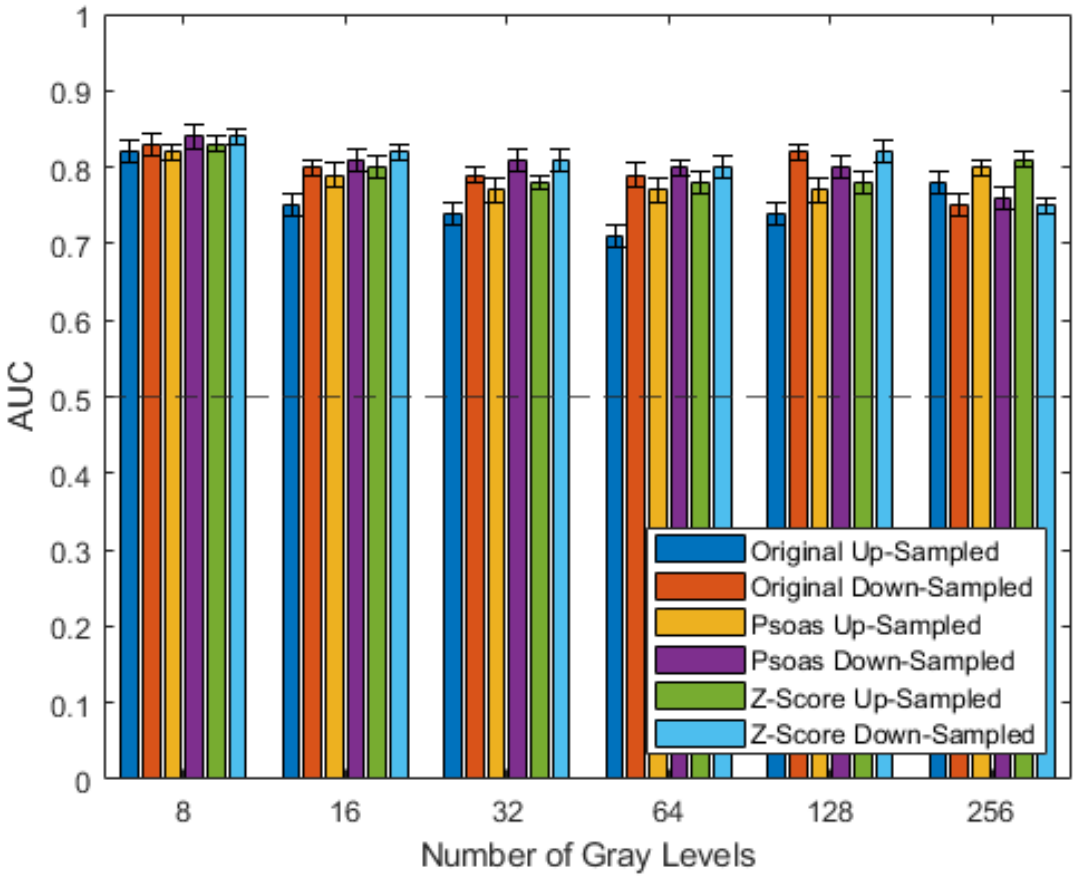

Figure 3.6: Area under the receiver operating characteristic curve (AUC) values in classifying Mayo Imaging Classification (MIC) using radiomic features extracted from the non-cystic kidney parenchyma using fixed bin number (FBN) discretization. The dotted line at an AUC of 0.5 is random guessing.

Table 3.5: Area under the receiver operating characteristic curve (AUC) along with the 95% confidence interval (CI) of AUC for features extracted from the entire kidney parenchyma using fixed bin size (FBS) discretization for Mayo Imaging Classification (MIC) 1A-1B vs. MIC 1C-1E classification.

	Number of Gray Levels					
Normalization	8	16	32	64	128	256
Original						
1.0x1.0	0.87	0.88	0.87	0.88	0.88	0.88
	[0.86,0.88]	[0.87,0.89]	[0.85,0.88]	[0.87,0.89]	[0.86,0.89]	[0.87,0.89]
2.0x2.0	0.87	0.87	0.86	0.88	0.89	0.88
	[0.86,0.88]	[0.86,0.88]	[0.85,0.88]	[0.87,0.89]	[0.88,0.90]	[0.87,0.89]
Psoas						
1.0x1.0	0.90	0.87	0.89	0.89	0.88	0.88
	[0.89,0.90]	[0.86,0.88]	[0.89,0.90]	[0.88,0.90]	[0.87,0.89]	[0.87,0.89]
2.0x2.0	0.89	0.89	0.90	0.90	0.89	0.89
	[0.88,0.90]	[0.88,0.90]	[0.89,0.90]	[0.89,0.91]	[0.89,0.90]	[0.88,0.90]
Z-score						
1.0x1.0	0.89	0.89	0.90	0.89	0.89	0.88
	[0.88,0.90]	[0.88,0.89]	[0.89,0.90]	[0.88,0.89]	[0.88,0.89]	[0.87,0.89]
2.0x2.0	0.88	0.89	0.89	0.89	0.89	0.89
	[0.87,0.89]	[0.88,0.90]	[0.88,0.90]	[0.88,0.90]	[0.88,0.90]	[0.88,0.90]

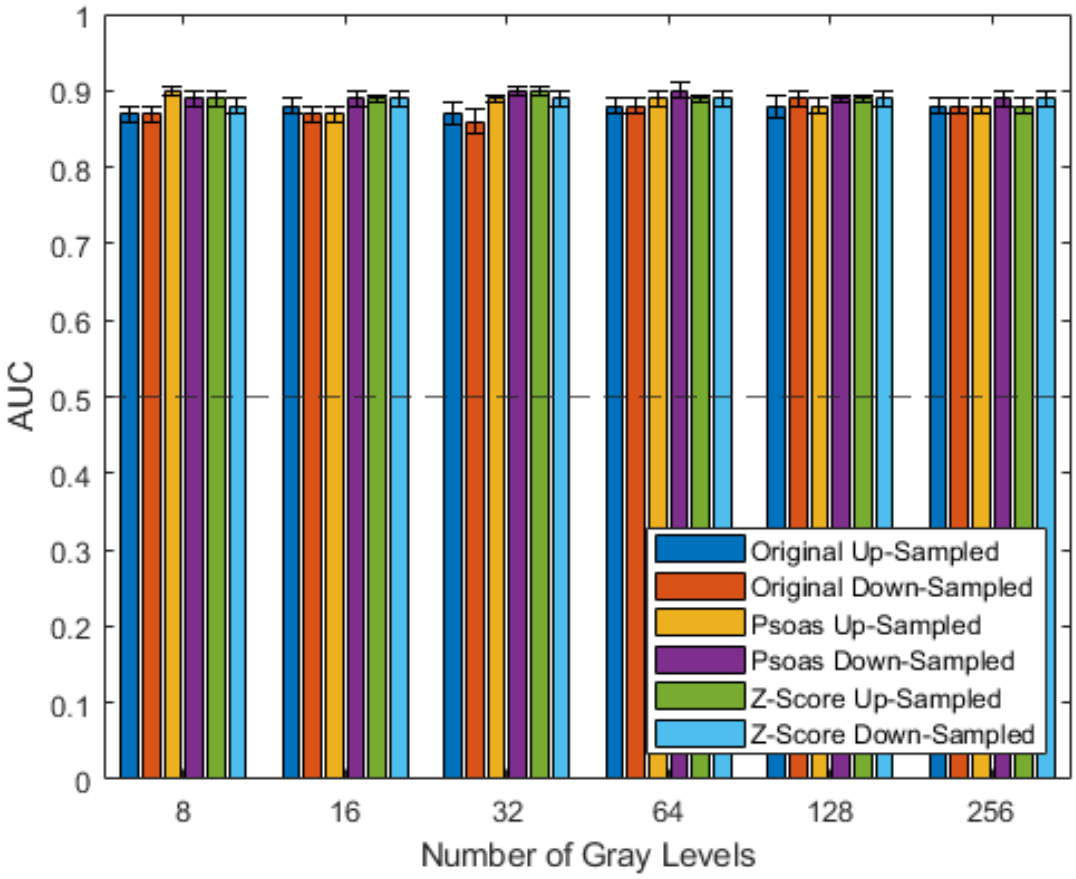

Figure 3.7: Area under the receiver operating characteristic curve (AUC) values in classifying Mayo Imaging Classification (MIC) using radiomic features extracted from the entire kidney using fixed bin size (FBS) discretization. The dotted line at an AUC of 0.5 is random guessing.

Table 3.6: Area under the receiver operating characteristic curve (AUC) along with the 95% confidence interval (CI) of AUC for features extracted from the entire kidney parenchyma using fixed bin number (FBN) discretization for Mayo Imaging Classification (MIC) 1A-1B vs. MIC 1C-1E classification.

	Number of Gray Levels					
Normalization	8	16	32	64	128	256
Original						
1.0x1.0	0.88	0.88	0.89	0.89	0.88	0.88
	[0.87,0.89]	[0.88,0.89]	[0.88,0.90]	[0.88,0.90]	[0.87,0.89]	[0.87,0.89]
2.0x2.0	0.89	0.90	0.89	0.89	0.88	0.89
	[0.88,0.90]	[0.89,0.91]	[0.88,0.90]	[0.88,0.90]	[0.87,0.89]	[0.88,0.90]
Psoas						
1.0x1.0	0.88	0.89	0.89	0.89	0.89	0.89
	[0.87,0.89]	[0.88,0.90]	[0.88,0.90]	[0.88,0.90]	[0.88,0.90]	[0.88,0.90]
2.0x2.0	0.89	0.90	0.90	0.89	0.89	0.89
	[0.88,0.90]	[0.89,0.91]	[0.89,0.90]	[0.88,0.90]	[0.88,0.90]	[0.88,0.90]
Z-score						
1.0x1.0	0.88	0.88	0.89	0.89	0.89	0.89
	[0.87,0.89]	[0.87,0.89]	[0.88,0.90]	[0.88,0.90]	[0.88,0.89]	[0.88,0.90]
2.0x2.0	0.88	0.90	0.90	0.89	0.89	0.89
	[0.87,0.89]	[0.89,0.91]	[0.89,0.91]	[0.88,0.90]	[0.88,0.90]	[0.88,0.90]

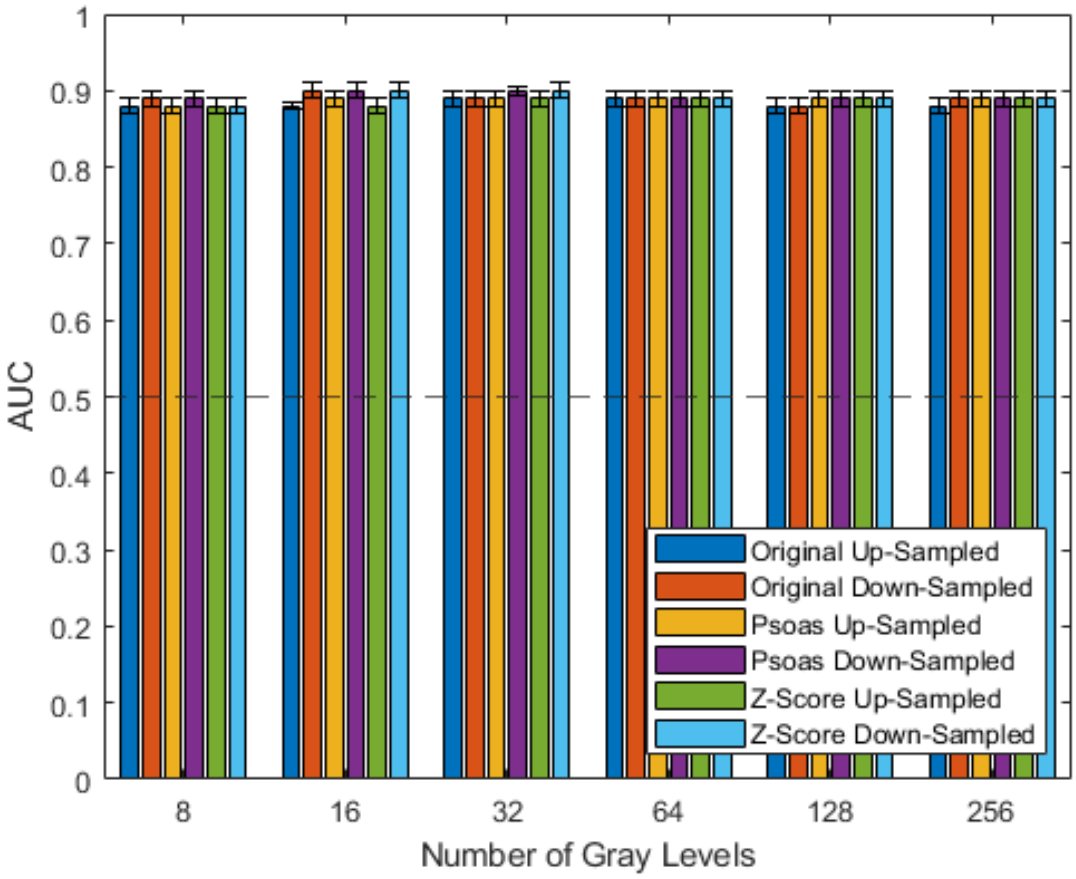

Figure 3.8: Area under the receiver operating characteristic curve (AUC) values in classifying Mayo Imaging Classification (MIC) using radiomic features extracted from the entire kidney using fixed bin number (FBN) discretization. The dotted line at an AUC of 0.5 is random guessing.

To identify which radiomic features were being used for classification, radiomic features chosen using LASSO feature selection in the training phase of the cross-validation was explored with a representative pre-processing method (psoas muscle normalization, pixel resampling to 2.0x2.0 mm, and FBN discretization using 32 gray levels). Table 3.7 contains the radiomic features used when features were extracted from the entire kidney, non-cystic kidney parenchyma, and the radiomic features that were used in classification regardless of

Table 3.7: The radiomic features chosen during the training phase of LASSO feature selection across the 5-fold repeated cross-validation (rCV) in classifying low/intermediate- and high-risk MIC class using a representative pre-processing method (psoas muscle normalization, pixel resampling to 2.0x2.0 mm, and fixed bin number (FBN) discretization using 32 gray levels).

Radiomic features	Entire kidney (n=16)	Union (n=26)	Non-cystic kidney (n=16)
First-order	Interquartile Range, Maximum, Median, Range	10 Percentile, Energy, Kurtosis, Minimum, Skewness	Entropy, Total Energy
GLCM	Inverse Difference Moment Normalized, Inverse Variance, Joint Entropy, Sum Entropy, Sum Squares	Cluster Prominence, Cluster Shade, Contrast, Correlation, Difference Variance, Informational Measure of Correlation 2, Maximum Probability	Autocorrelation, Cluster Tendency, Difference Average, Inverse Difference, Inverse Difference Moment, Informational Measure of Correlation 1, Maximal Correlation Coefficient
GLSZM	Gray-Level Non-Uniformity Normalized, High-Gray Level Zone-Emphasis, Zone Variance	Gray-Level Non-Uniformity, Gray-Level Variance, Large Area High-Gray Level Emphasis, Large Area Low-Gray Level Emphasis, Size-Zone Non-Uniformity, Small Area Emphasis, Small Area High Gray-Level Emphasis, Small Area Low Gray-Level Emphasis, Zone Entropy	Low-Gray Level Zone Emphasis
GLRLM	Gray-Level Non-Uniformity Normalized, Run-Length Non-Uniformity, Run Variance	Gray-Level Non-Uniformity	
NGTDM	Complexity	Busyness, Coarseness	
GLDM		Dependence Entropy, Dependence Non-Uniformity	Gray-Level Non-Uniformity, Gray-Level Variance, Large Dependence High-Gray Level Emphasis, Large Dependence Low-Gray Level Emphasis, Low Gray-Level Emphasis, Small Dependence Low-Gray Level Emphasis

what part of the kidney was used for feature extraction. There were 26 radiomic features that were used in both classification schemes (union column of Table 3.7).

3.3.2 Impact of feature correlation with kidney size across pre-processing

Figure 3.9 shows the box and whisker plot of the distribution of the number of voxels in the kidney segmentations of MIC 1A-1B classes and MIC 1C-1E classes. The number of voxels were statistically significantly different between MIC 1A-1B and MIC 1C-1E classes.

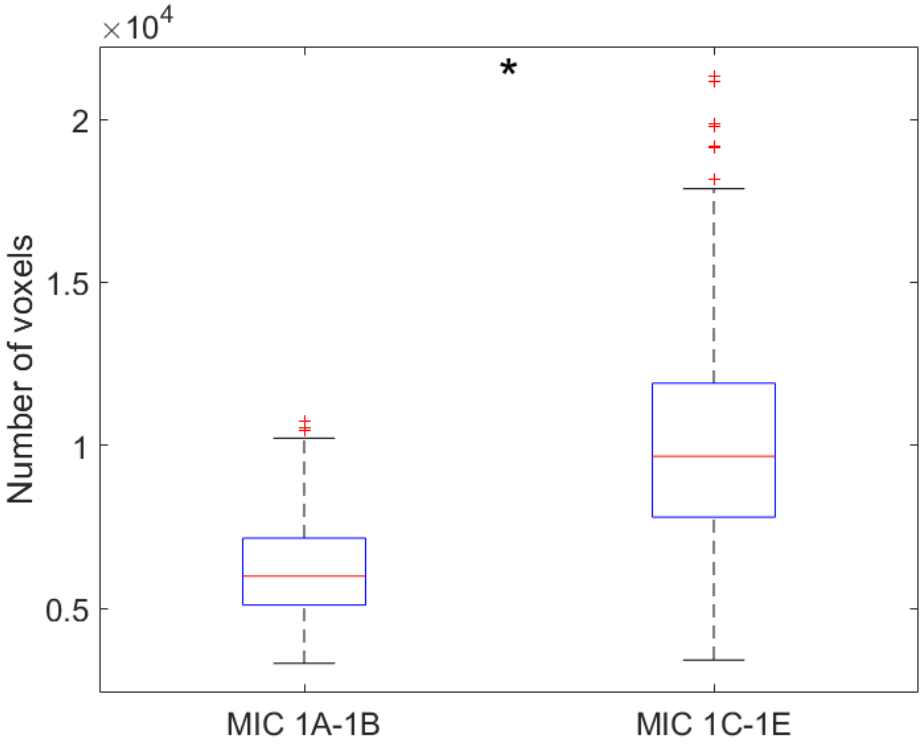

Figure 3.9: Box and whisker plot of the distribution of kidney size from the representative coronal section (i.e., number of voxels) of low/intermediate- and high-risk Mayo Imaging Classification (MIC) patients, showing a statistically significant difference using the Wilcoxon rank sum test ($p < 0.05$). The box represents the interquartile range, with the central line indicating the median value; the whiskers extend to the minimum and maximum values, while outliers are represented as individual data points.

Four of the 93 features extracted exhibited a high correlation ($\rho > 0.9$) with number of voxels across normalization, pixel resampling, and gray-level discretization: GLCM joint en-

tropy, GLDM dependence non-uniformity, GLRLM run-length non-uniformity, and GLSZM size-zone non-uniformity. Figures 3.10-3.13 show the Spearman rank correlation coefficient between feature values and number of voxels across pre-processing for the four features using both FBS and FBN discretization methods. As the gray levels available for discretization increase, the correlation between the feature value and the voxel size increases, with the exception of GLDM dependence non-uniformity that stayed relatively stable across gray levels except for the original image using the downsampling method. Downsampling across normalizations tended to have higher Spearman rank correlation values than upsampling. For all features using FBN the Spearman rank correlation values were consistent across pre-processing, with the downsampling method consistently higher in Spearman rank correlation values than upsampling. There were 33 of the 93 features extracted that exhibited moderate ($0.5 < \rho < 0.9$) correlation with number of voxels across pre-processing (Table 3.8).

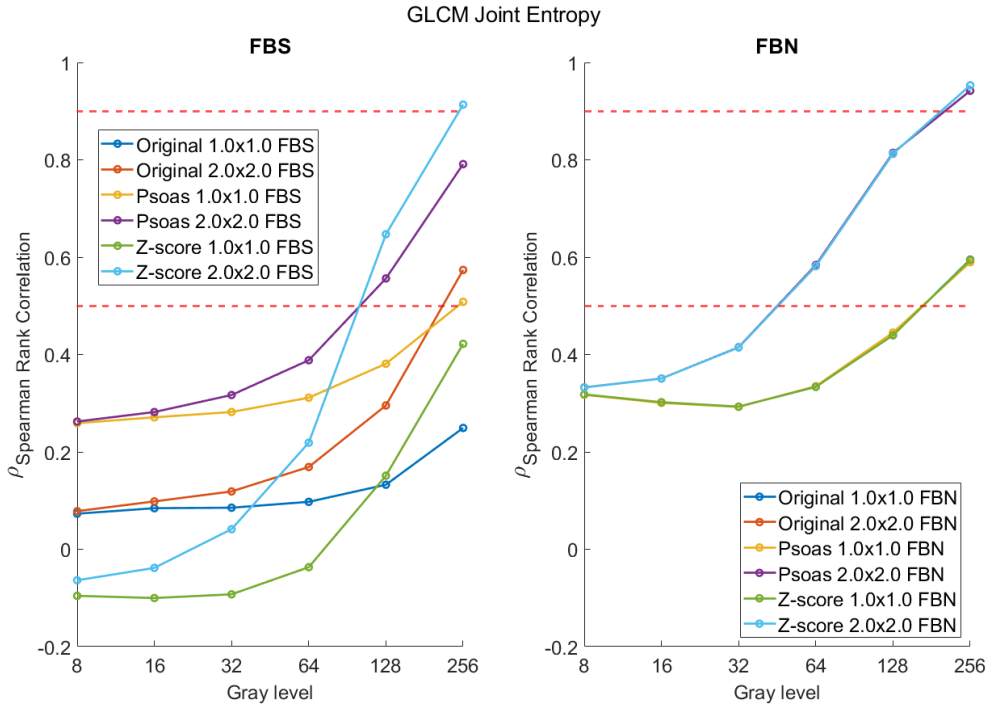

Figure 3.10: GLCM joint entropy Spearman rank correlation values ranging from -0.10-0.91 using FBS discretization and 0.29-0.95 using FBN discretization. The red dashed lines represent the moderate ($0.5 < \rho < 0.9$) and high correlation ($\rho > 0.9$) thresholds.

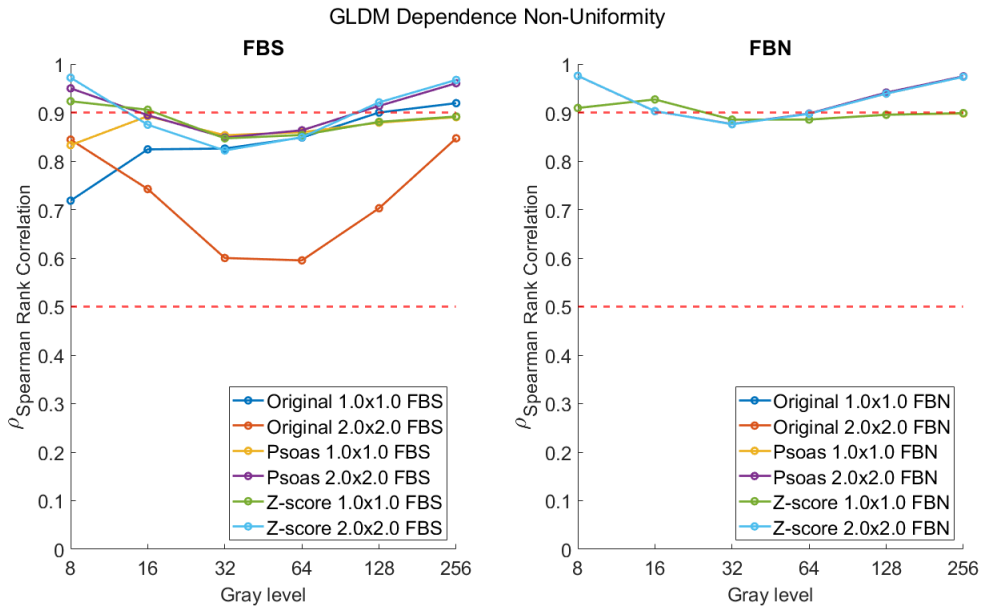

Figure 3.11: GLDM dependence non-uniformity Spearman rank correlation values ranging from 0.60-0.97 using FBS discretization and 0.88-0.98 using FBN discretization. The red dashed lines represent the moderate ($0.5 < \rho < 0.9$) and high correlation ($\rho > 0.9$) thresholds.

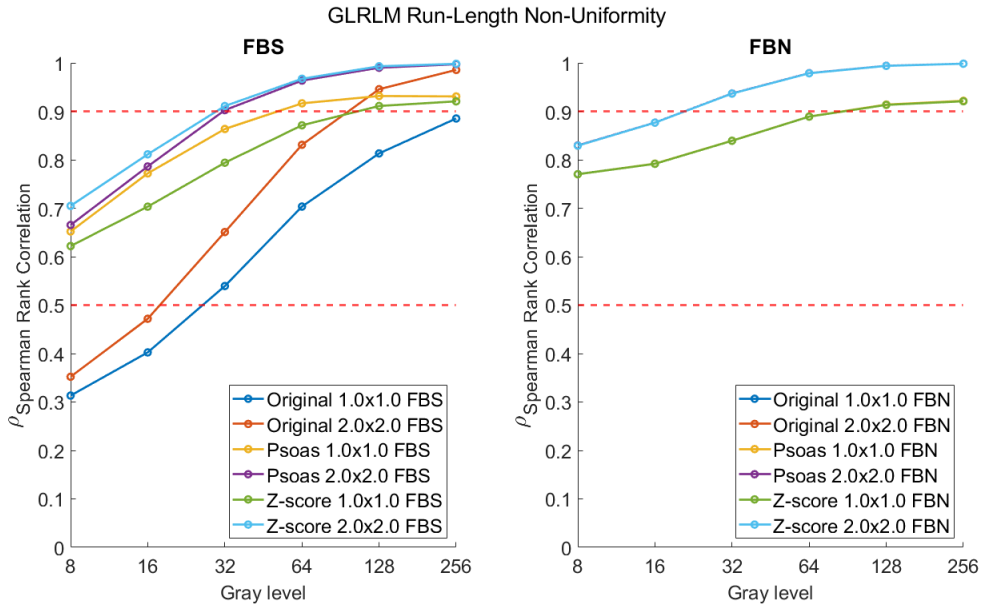

Figure 3.12: GLRLM run-length non-uniformity Spearman rank correlation values ranging from 0.31-1.0 using FBS discretization and 0.77-1.0 using FBN discretization. The red dashed lines represent the moderate ($0.5 < \rho < 0.9$) and high correlation ($\rho > 0.9$) thresholds.

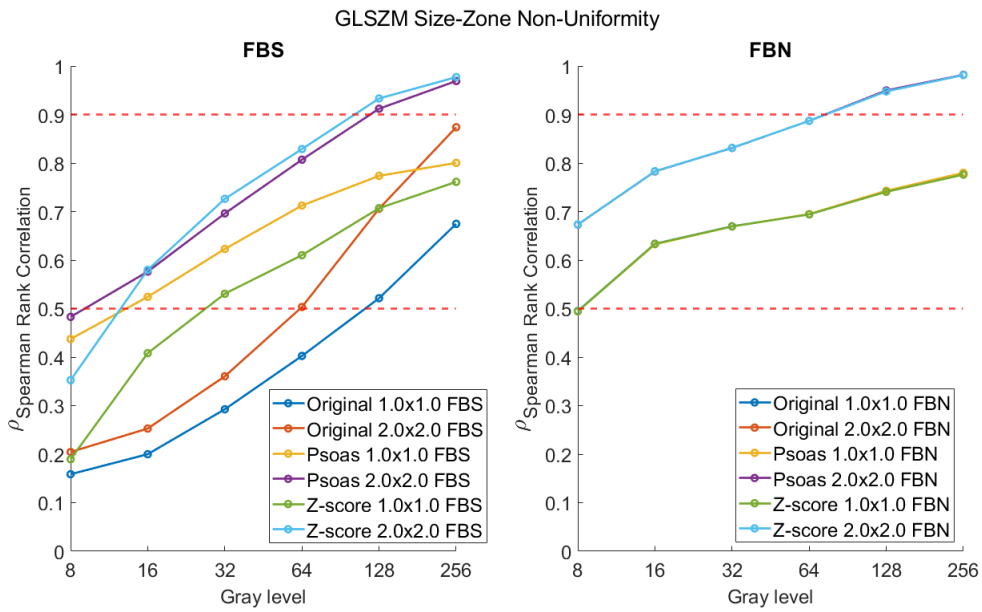

Figure 3.13: GLSZM size-zone non-uniformity Spearman rank correlation values ranging from 0.16-0.98 using FBS discretization and 0.49-0.98 using FBN discretization. The red dashed lines represent the moderate ($0.5 < \rho < 0.9$) and high correlation ($\rho > 0.9$) thresholds.

Table 3.8: Radiomic features that exhibited moderate ($0.5 < \rho < 0.9$) Spearman rank correlation values with number of voxels across pre-processing parameters.

Radiomic features	Moderate ($0.5 < \rho < 0.9$) correlation (n=33)
First-order (n=5)	Energy, Entropy, Skewness, Total Energy, Uniformity
GLCM (n=5)	Cluster Shade, IMC 1, Joint Energy, Maximum Probability, Sum Entropy
GLSZM (n=7)	Gray-Level Non-Uniformity, Gray-Level Non-Uniformity Normalized, Large Area High Gray-Level Emphasis, Large Area Low Gray-Level Emphasis, Low Gray-Level Zone Emphasis, Small Area Low Gray-Level Emphasis, Zone Entropy
GLRLM (n=7)	Gray-Level Non-Uniformity, Gray-Level Non-Uniformity Normalized, Long Run High Gray-Level Emphasis, Long Run Low Gray-Level Emphasis, Low Gray-Level Run Emphasis, Run Entropy, Short Run Low Gray-Level Emphasis
NGTDM (n=3)	Busyness, Coarseness, Strength
GLDM (n=6)	Dependence Entropy, Gray-Level Non-Uniformity, Large Dependence High Gray-Level Emphasis, Large Dependence Low Gray-Level Emphasis, Low Gray-Level Emphasis, Small Dependence Low Gray-Level Emphasis

The impact on differentiating low/intermediate- vs. high-risk MIC classes using the radiomic features from the entire kidney was investigated and the AUC values are listed in Table 3.9, with significant differences assessed after Bonferroni-Holm correction [81]. The representative pre-processing method for classification was psoas muscle normalization, pixel resampling to 2.0x2.0 mm, and FBS discretization using 64 gray levels. A radiomic feature that exhibited moderate ($0.5 < \rho < 0.9$) or high ($\rho > 0.9$) correlation across any of the pre-processing parameters was classified as such for inclusion or exclusion in feature selection. Classification of low-/intermediate- vs. high-risk patients yielded an AUC value of 0.90 [0.89, 0.91] using feature selection on all available 93 features. When feature selection was limited to features with (1) low or moderate correlation and (2) only low correlation, the resulting AUC values were 0.89 [0.88, 0.90] and 0.78 [0.76, 0.79], respectively.

Table 3.9: Area under the receiver operating characteristic curve (AUC) along with the 95% confidence interval (CI) of AUC in the task of classifying patients between low/intermediate- vs. high-risk Mayo Imaging Classification (MIC) classes using radiomic features with respect to their correlation with number of voxels (i.e., kidney size). Asterisks denote significance after accounting for multiple comparisons using Bonferroni-Holm corrections.

Features available for LASSO	AUC [95% CI]	Statistical significance
All Features (n=93)	0.90 [0.89,0.91]	$\rho < 0.9$: $p = 0.061$ $\rho < 0.5$: $p =< 0.001$*
Features with $\rho < 0.9$ (n=89)	0.89 [0.88,0.90]	$\rho < 0.5$: $p =< 0.001$*
Features with $\rho < 0.5$ (n=56)	0.78 [0.76,0.79]	

3.3.3 MIC 1B (intermediate-risk) vs. MIC 1C (high-risk) classification

There were a total of 271 MIC 1B (n=111) and MIC 1C (n=160) patients available for classification (Table 3.10); there was a statistically significant difference in mean age and mean htTKV, but not a statistically significant difference in eGFR. The AUC performances across clinical, radiomics, and combined clinical-radiomics models can be found in Table

3.11. The pre-processing method chosen was the z-score normalization and 32 gray levels for discretization using FBS for both the non-cystic kidney and entire kidney classification.

Table 3.10: HALT subject characteristics for Mayo Imaging Classification (MIC) 1B vs. MIC 1C classification. Asterisks indicate statistically significant differences ($p < 0.05$).

Clinical	1B	1C	p-value
No. Patients	111	160	
Mean age ±SD	40 ±7.15	37 ±7.67	<0.001*
Male (n=124)	38 ±7.43 (n=37)	37 ±7.08 (n=87)	0.516
Female (n=147)	41 ±6.79 (n=74)	36 ±8.35 (n=73)	<0.001*
Mean htTKV ±SD	392.68 ±100.34	597.81 ±178.17	<0.001*
Male	356.044 ±81.55	616.42 ±183.13	<0.001*
Female	410.99 ±104.26	575.63 ±170.67	<0.001*
Mean eGFR ±SD	85.98 ±16.74	86.78 ±20.00	0.783
Male	89.12 ±15.60	86.74 ±21.14	0.181
Female	84.42 ±17.17	86.82 ±18.70	0.480

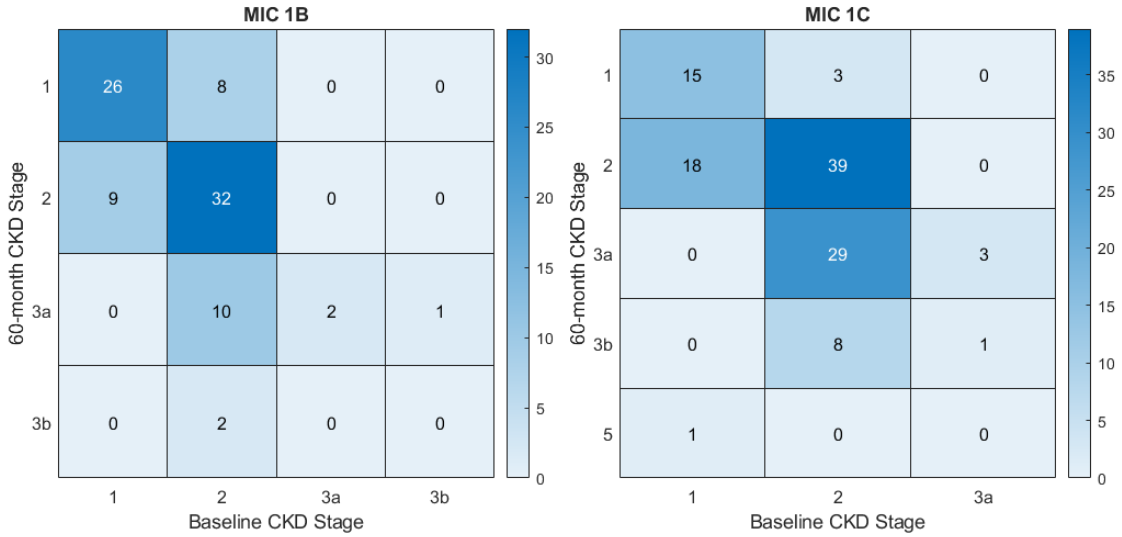

Figure 3.14: Heatmaps of the number of patients for MIC 1B (left) and MIC 1C (right) with their baseline chronic kidney disease (CKD) stage (x-axis) and CKD stage at 60-months follow-up (y-axis). The color intensity represents the number of patients, with darker shades indicating a higher patient count.

Table 3.11: Area under the receiver operating characteristic curve (AUC) along with the 95% confidence interval (CI) of AUC for 1B vs. 1C classification using age, sex, estimated glomerular filtration rate (eGFR), radiomic features, and combined clinical-radiomics model. The combined clinical features incorporate age, sex, and eGFR, but do not include height-corrected total kidney volume (htTKV). Although Mayo Imaging Classification (MIC) classification considers htTKV and age, htTKV was investigated as a standalone feature.

Features	AUC [95% CI]
htTKV	0.84 [0.83, 0.86]
Clinical	
Age	0.62 [0.59, 0.64]
Sex	0.56 [0.54, 0.58]
eGFR	0.59 [0.57, 0.61]
Combined clinical	0.66 [0.64, 0.68]
Radiomics model	
Entire kidney	0.79 [0.77, 0.80]
Non-cystic kidney	0.67 [0.65, 0.69]
Clinical and radiomics model	
Entire kidney	0.89 [0.88, 0.90]
Non-cystic kidney	0.77 [0.76, 0.79]

FCM clustering using clinical features (age, sex, and eGFR), radiomic features from the entire kidney, radiomic features from the non-cystic kidney, and combined clinical-radiomics features were used to cluster MIC 1B and MIC 1C patients. Radiomic features identified with LASSO feature selection in the classification of MIC 1B and MIC 1C were subsequently input for FCM clustering: 20 features extracted from the entire kidney, and 42 features extracted from the non-cystic kidney. ROC analysis was performed on the output of fuzzy class membership, listed in Table 3.12. The AUC values of FCM clustering of MIC 1B and MIC 1C using clinical features was 0.52, 0.62-0.63 using radiomic features, and 0.62-0.63

using combined clinical and radiomic features. An example of FCM clustering in graphical form using age and eGFR for simplicity can be found on Figure 3.15.

Table 3.12: Area under the receiver operating characteristic curve (AUC) along with the 95% confidence interval (CI) of AUC for fuzzy c-means (FCM) clustering memberships of Mayo Imaging Classification (MIC) 1B and MIC 1C patients using age, sex, estimated glomerular filtration rate (eGFR), radiomic features, and combined clinical-radiomics model.

Features	AUC [95% CI]
Clinical	0.52 [0.46, 0.59]
Radiomics features	
Entire kidney	0.63 [0.56, 0.70]
Non-cystic kidney	0.62 [0.55, 0.69]
Clinical and radiomic features	
Entire kidney	0.63 [0.57, 0.69]
Non-cystic kidney	0.62 [0.55, 0.69]

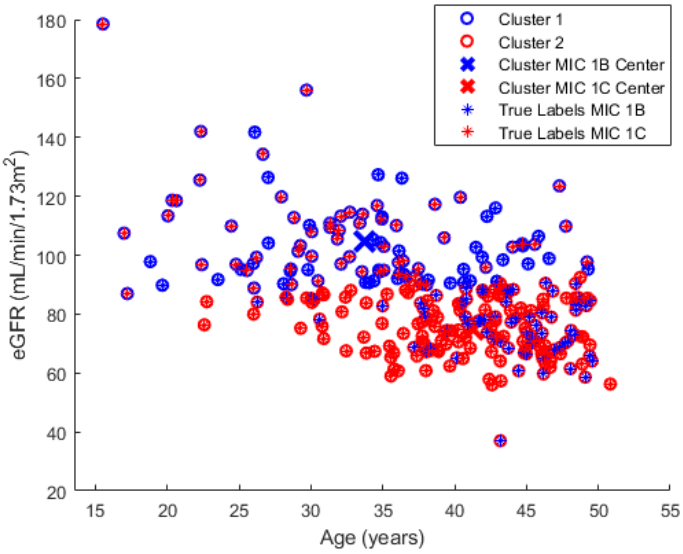

Figure 3.15: Fuzzy c-means (FCM) clustering using clinical features (age, sex, and eGFR). For simplicity, the figure shows the age on the x-axis and eGFR on the y-axis. The cluster center for 1B is (33.74, 104.71) and for 1C is (41.24, 74.75). The scatter plot shows the cluster membership as well as the true class label of each patient (n=271).

3.4 Discussion

The aim of this chapter was to determine whether radiomic features can differentiate patients into low/intermediate- and high-risk MIC classes and to observe the effect of pre-processing on subsequent classification. There is a need to investigate texture-based differences in the non-cystic kidney parenchyma among risk-stratified groups because features may quantify potential prognostic information. The results indicate that radiomic features extracted from the non-cystic kidney parenchyma and the entire kidney have the potential to correctly classify patients in low/intermediate- and high-risk MIC classes. The non-cystic parenchyma AUC values ranged from 0.69-0.85 across pre-processing parameters, and the entire kidney AUC values ranged from 0.86-0.90 across pre-processing parameters.

The range of AUC values using features extracted from the entire kidney across pre-processing parameters was narrower than the range of features extracted from the non-cystic parenchyma: 0.04 compared to 0.16. The impact of pre-processing on subsequent classification reveals that features extracted from the non-cystic kidney parenchyma were more sensitive to pre-processing than those from the entire kidney. The classification of patients with low/intermediate-risk (low cyst burden) from those with high-risk (high cyst burden with higher signal intensity) was unaffected by pre-processing parameters such as MR normalization, gray-level discretization, and pixel resampling. This underscores the influence of cysts on texture features, indicating that these imaging characteristics remain robust indicators of disease severity and risk classification regardless of pre-processing variations. Unlike those of the entire kidney, features extracted from the non-cystic kidney parenchyma show an increase in sensitivity to pre-processing parameters. Figure 3.4 displays the box and whisker plot of signal ranges from the non-cystic kidney parenchyma and entire kidney across normalization methods, revealing that the median signal range of the non-cystic kidney parenchyma is lower than that of the entire kidney (similar to results in Chapter 2). The difference in sensitivity to pre-processing in subsequent classification between the non-cystic

kidney parenchyma and the entire kidney may be attributed to the variance in signal range of gray levels for discretization and the removal of cyst texture compared to that of the entire kidney.

The classification task of distinguishing intermediate-risk (1B) MIC class and high-risk (1C) MIC class patients proved to be a more difficult task than distinguishing low/intermediate-risk (1A-1B) MIC classes from high-risk (1C-1E) MIC classes. Although MIC 1C patients had statistically significant differences in age and htTKV compared to MIC 1B patients, eGFR was not found to be statistically significantly different. Classification models using clinical features, radiomics features, and combined clinical-radiomics models were investigated. The AUC value for the clinical model (age, sex, and eGFR) was 0.66 [0.64, 0.68], and for the radiomics model using the entire kidney and non-cystic kidney parenchyma was 0.79 [0.77, 0.80] and 0.67 [0.65, 0.69], respectively; in combining clinical and radiomics features, there was an increase in both the entire kidney and non-cystic kidney parenchyma of 0.89 [0.88, 0.90] and 0.77 [0.76, 0.79], respectively. The performance of combining clinical features with radiomic features versus clinical features alone demonstrates the potential benefit of incorporating kidney texture, which is not captured by clinical features alone.

This study investigated the correlation between radiomic feature values and number of voxels (i.e., kidney size) for feature extraction. High-risk MIC class patients have by definition larger kidney sizes than patients in low/intermediate-risk MIC classes, and radiomic features that exhibited low ($\rho < 0.5$), moderate ($0.5 < \rho < 0.9$), or high ($\rho > 0.9$) correlation with number of voxels (i.e., kidney size) across pre-processing parameters were identified. There were four features that exhibited a high correlation with number of voxels across MR normalization, pixel resampling, and gray-level discretization: GLCM joint entropy, GLDM dependence non-uniformity, GLRLM run-length non-uniformity, and GLSZM size-zone non-uniformity; GLDM dependence non-uniformity and GLRLM run-length non-uniformity have been found to be important biomarkers in previous radiomics research [82, 83]. The cor-

relation coefficient between the radiomic feature value and the number of voxels was found to be dependent on the pre-processing parameters used. The inclusion of features based on their correlation to kidney size in classification models impacted the performance in distinguishing between low/intermediate-risk and high-risk MIC classes. Pyradiomics [61] has identified features that are volume-confounded, and Shafiq-ul-Hassan et al. [78] provided normalization factors based on the number of voxels in an ROI/volume of interest (VOI), pixel size, and slice thickness for radiomic features. Future research will explore the impact of MRI pre-processing on radiomic feature values following correction and their dependence on the number of voxels.

A limitation of this study is the use of a single representative coronal MRI image, and future work could assess texture of the total kidney and total non-cystic kidney parenchyma. Additionally, classification could extend to specific classes of MIC, beyond just intermediate-risk (1B) and high-risk (1C) patients; this analysis is limited due to the small number of patients, and future work could incorporate larger imaging datasets. Radiomic features that associate across multi-class MIC classification could be compared to those of 1A/1B vs. 1C-1E MIC classes. Although the work in this chapter is based on differentiating MIC classes, this work could be extended to future kidney function decline using the radiomic features associated with low/intermediate-risk and high-risk MIC classes.

3.5 Conclusion

Based on ROC performance, the data indicate the potential to distinguish between patients based on low/intermediate- and high-risk MIC classification. Furthermore, the effect of MRI pre-processing (MR normalization, gray-level discretization, and pixel resampling) was investigated. The potential to correctly classify between low/intermediate- and high-risk MIC classes was demonstrated across pre-processing parameters. Results reveal that features extracted from the non-cystic kidney parenchyma were more sensitive to pre-processing than

those from the entire kidney and that extracted features are associated with MIC risk-stratified classes and may be used for patient prognosis. Ultimately, this work will pave the way to extract radiomic features of the non-cystic kidney parenchyma to better identify low- and high-risk patients for onset of ESKD and establish optimal pre-processing parameters for feature extraction.

CHAPTER 4

TEMPORAL ASSESSMENT OF MRI RADIOMIC FEATURES TO PREDICT KIDNEY FUNCTION DECLINE IN PATIENTS WITH ADPKD

4.1 Introduction

Autosomal dominant polycystic kidney disease (ADPKD) is a hereditary kidney disease with gradual kidney cyst growth and expansion causing deterioration in kidney function and total kidney volume (TKV) growth, ultimately leading to end-stage kidney disease (ESKD). Although TKV increases continuously during a patient's life, kidney function may remain intact for decades. ADPKD biomarkers are needed that may both detect current damage to the kidneys while also predicting disease progression during the period of maintained kidney function [14]. Radiomics analysis of the kidney has been shown to predict estimated glomerular filtration rate (eGFR) decline and classify chronic kidney disease (CKD) stage in ADPKD and has been shown to provide additive power with clinical features such as age, baseline height-corrected TKV (htTKV), and baseline eGFR [52, 53, 54, 55]. Incorporating radiomic features in predictive models would provide quantitative information of kidney texture that is not currently captured by standard clinical features. TKV is monitored throughout a patient's life, and magnetic resonance imaging (MRI) is the predominant imaging modality to measure TKV because it does not use ionizing radiation, has excellent soft-tissue contrast, and provides good delineation of kidney parenchyma and cysts. Previous ADPKD radiomics studies used features from one imaging timepoint, although temporal imaging is clinically indicated to monitor disease progression [53, 54, 55, 84, 85].

Delta radiomics captures feature variation at different acquisition timepoints and utilizes longitudinal imaging data to assess intra-patient changes [86, 87]. Delta radiomics is used to assess response to therapy and prediction of patient outcomes in breast, lung, and cervical

cancers using imaging modalities such as computed tomography (CT), positron emission tomography (PET)-CT systems, and MRI [88, 89, 90, 91]. The most common method to calculate delta radiomics is the difference in the radiomic feature value between two imaging timepoints divided by the radiomic feature value at the first timepoint [86]. Incorporating radiomic features from different acquisition timepoints takes advantage of feature variation in ADPKD patients and may provide insight in changes of kidney texture with disease progression. Utilizing delta radiomics may provide additive power in predicting future kidney function decline for patients with ADPKD rather than one timepoint alone.

Therefore, the purpose of this work was to determine whether radiomic features from 24-month and 48-month timepoints provide additive power versus baseline texture alone in predicting kidney function decline to (1) CKD stage 3A or greater ("≥CKD stage 3A") or (2) a greater than 30% reduction in eGFR ("≥30% reduction in eGFR") of 281 patients with ADPKD at 60-months follow-up. Additionally, radiomic features extracted from the non-cystic kidney parenchyma (excluding cysts) from CT images have shown predictive power in future kidney function decline, suggesting that radiomic features from the non-cystic kidney parenchyma may be a sensitive biomarker of kidney function decline [55]. The work in this chapter investigated the ability of radiomic features extracted from the non-cystic parenchyma from MR images to predict future kidney function decline. This is the first work to investigate the utility of incorporating multiple imaging timepoints in ADPKD radiomics prediction models using delta radiomics.

4.2 Methods

4.2.1 Database

T2-weighted fat saturated (T2W-FS) MRI images in this study had been acquired from a longitudinal, multi-site randomized clinical trial, HALT Progression of Polycystic Kidney

Disease (HALTA-PKD) (NCT00283686). T2W-FS MR images were acquired with the same scanning protocol on 1.5T scanners using single-shot fast spin echo/half-Fourier acquisition single-shot turbo spin echo imaging (SSFSE/HASTE) with fat saturation and a 3-mm fixed slice thickness. Of this cohort, a subset of 281 patients was chosen who had baseline, 24-month, 48-month, and 60-month follow-up imaging, measured htTKV, and measured eGFR.

CKD is defined by a presence of albuminuria (i.e., albumin in the urine) of at least 30 mg per 24 hours or markers of kidney damage persisting more than 3 months [92]. The staging of CKD is classified as CKD stage 1 (GFR $> 90\,\mathrm{mL/min/1.73\ m^2}$), CKD stage 2 (GFR $60 - 89\,\mathrm{mL/min/1.73\ m^2}$), CKD stage 3A ($45 - 59\,\mathrm{mL/min/1.73\ m^2}$), CKD stage 3B ($30 - 44\,\mathrm{mL/min/1.73\ m^2}$), CKD stage 4 ($15 - 29\,\mathrm{mL/min/1.73\ m^2}$), and CKD stage 5 ($< 15\,\mathrm{mL/min/1.73\ m^2}$) [92, 93]. All 281 patients in this study were classified in CKD stage 1 or 2 at baseline, and the clinical stage at 60 months spanned all CKD stages (Table 4.1). The subject characteristics were compared between those who did or did not progress to $\geq$CKD stage 3A and between those who did or did not experience a $\geq$30% reduction in eGFR (Table 4.2).

Table 4.1: Subject characteristics based on baseline age, baseline sex, baseline estimated glomerular filtration rate (eGFR), and baseline height-corrected total kidney volume (ht-TKV) for 281 ADPKD patients. The chronic kidney disease (CKD) stage is the patient stage at 60-months follow-up.

Clinical stage at 60 months	Mean age ±SD	Sex (M:F)	Mean eGFR ±SD	Mean htTKV ±SD
CKD stage 1 (n=60)	34 ±8.07	32:28	103.87 ±18.57	434.55 ±186.23
CKD stage 2 (n=141)	38 ±7.71	68:73	84.86 ±11.93	676.38 ±379.60
CKD stage 3A (n=57)	39 ±5.84	34:23	71.91 ±6.58	848.00 ±369.43
CKD stage 3B (n=20)	40 ±8.44	12:8	68.50 ±7.21	972.76 ±549.47
CKD stage 4 (n=1)	25	1:0	62.26	1751.16
CKD stage 5 (n=2)	37 ±8.73	1:1	76.50 ±19.55	1287.01 ±1038.54

Table 4.2: A Wilcoxon rank sum test between the subject characteristics at baseline for those who progressed or did not progress to ≥ CKD stage 3A and between those who experienced or did not experience a ≥30% reduction in eGFR ($p < 0.05$ were significant). Asterisks indicate statistically significant differences. The "≥30% reduction in eGFR" cohort contains patients in CKD stages 1-5.

Clinical features at baseline	<CKD stage 3A (n=201)	≥CKD stage 3A (n=80)	p-value	<30% reduction in eGFR (n=236)	≥30% reduction in eGFR (n=45)	p-value
Mean age ±SD	37 ±8.09	39 ±6.73	0.035*	38 ±7.58	36 ±8.79	0.450
Mean eGFR ±SD	90.53 ±16.66	71.05 ±7.21	≤ 0.001*	86.09 ±15.98	79.21 ±21.06	≤ 0.001*
Mean htTKV ±SD	604.19 ±351.29	901.46 ±446.39	≤ 0.001*	647.57 ±373.91	905.18 ±479.51	≤ 0.001*

4.2.2 MR image selection, segmentation, and pre-processing

For each patient, the largest coronal slice of both the left and right kidney was selected from the baseline scan as the representative baseline image for each kidney. Subsequent images at the 24-month, 48-month, and 60-month follow-up scan timepoints were matched to the baseline left and right kidney representative images. A convolutional neural network (CNN) U-Net architecture introduced in Chapter 3 was employed for the purpose of segmenting the entire kidney parenchyma (including cysts) and non-cystic kidney parenchyma (excluding cysts). The U-Net segmentations were visually examined to ensure quality of segmentation and manually adjusted if necessary. All MR image matrix sizes ranged from 224x224-640x640 pixels, and pixel sizes ranged from 0.59-1.8 mm. All images matrix sizes were resized to 256x256 pixels for feature extraction using nearest-neighbor interpolation. All pixel sizes were harmonized to 2.0x2.0 mm using nearest-neighbor interpolation, and all images had the same slice thickness of 3 mm.

4.2.3 Feature extraction and predictive model

Image features from baseline and 24-month, 48-month, and 60-month timepoints were extracted using Pyradiomics [61]. Z-score normalization was used to standardize pixel in-

tensities and a fixed bin number discretization method was implemented using 64 bins for gray-level discretization. Ninety-three features were extracted per kidney: first-order (18), gray-level co-occurrence matrix (GLCM) (24), gray-level run length matrix (GLRLM) (16), gray-level size zone matrix (GLSZM) (16), neighboring gray-tone difference matrix (NGTDM) (5), and gray-level dependence matrix (GLDM) (14) features. Three images anterior and three images posterior to the selected baseline and follow-up images were selected for a total of seven images per kidney per timepoint for feature extraction. Radiomic features extracted from the seven images were combined using a weighted average for the left and right kidney individually:

$$RF_{\text{weighted average}} = \frac{\sum_i RF_{slice,i} \cdot \text{weights}_i}{\sum_i \text{weights}_i}. \tag{4.1}$$

The weights were calculated by dividing the number of voxels in a kidney segmentation of a given image by the total number of voxels across all of the kidney segmentations of the MRI sections, i:

$$\text{weights}_i = \frac{\text{Number of voxels}_{seg,i}}{\sum_i \text{Number of voxels}_{seg,i}}. \tag{4.2}$$

Finally, the left and right kidney features were averaged for a representative feature value for each timepoint. Radiomic features were calculated from images at baseline, 24-month, 48-month, and 60-month timepoints of both the non-cystic kidney parenchyma and entire kidney parenchyma for each patient. Delta radiomics to assess feature changes over time was calculated using:

$$\Delta RF_{\text{Delta}} = \frac{RF_{\text{Follow-up}} - RF_{\text{Baseline}}}{RF_{\text{Baseline}}}. \tag{4.3}$$

Predictive models using clinical features, radiomic features, and clinical-radiomic features combined were investigated to assess whether a patient (1) progressed to $\geq$CKD stage 3A or (2) experienced a $\geq$30% reduction in eGFR at 60-month follow-up. The radiomics

models used features extracted from the non-cystic kidney parenchyma and entire kidney: (1) baseline, (2) 24-month delta radiomics, (3) 24-month timepoint, (4) 48-month delta radiomics, (5) 48-month timepoint, (6) combined baseline and 24-month delta radiomics, (7) combined baseline and 24-month radiomics, (8) combined baseline and 48-month delta radiomics, (9) combined baseline and 48-month timepoint, (10) combined baseline, 24-month delta radiomics, and 48-month delta radiomics, and (11) combined baseline, 24-month timepoint, and 48-month timepoint. Figure 4.1 shows the summary of workflow and analysis methods.

A logistic-regression (LR) model using 5-fold cross validation was utilized for a total of 20 different models: single timepoint radiomics models (3), delta radiomics models (2), combined timepoint (including delta radiomics) radiomics models (6), clinical models (4), and combined clinical-radiomics models (5). For each training partition of any radiomics model, the top-10 uncorrelated (Pearson correlation threshold of 0.7 with any other selected feature) performing features from the 93 total extracted features were determined using the area under the receiver operating characteristic curve (AUC); for example, the top-10 performing, uncorrelated features from the baseline timepoint for the training partition were merged with the top-10 performing, uncorrelated features from the 24-month timepoint of the same training partition for a total of 20 features. The clinical model included baseline age, baseline htTKV, and baseline eGFR. This process used a repeated cross-validation (rCV) of 10 to account for variance in 5-fold cross-validation (50 iterations of feature selection), and AUC was used to evaluate the predictive performance across these 50 iterations. This process was performed separately for radiomic features extracted from the non-cystic kidney parenchyma and the entire kidney parenchyma for prediction of progression to $\geq$CKD stage 3A and a $\geq$30% reduction in eGFR at 60-month follow-up (64 models in total).

The predictive performance (AUC) of single timepoint models, delta radiomics models, and combined radiomics models were compared with the predictive performance of the

baseline model alone using the DeLong test [94]. Predictive performances of the combined clinical-radiomics models were compared with the predictive performance of the combined clinical model of baseline age, baseline htTKV, and baseline eGFR also using the DeLong test [94]. Bonferroni-Holm corrections were used to account for multiple comparisons [81].

A single representative radiomic feature that was frequently used in all single timepoint models (baseline, 24-month timepoint, and 48-month timepoint) was investigated (4 features in total): a feature from both the non-cystic kidney and entire kidney for prediction of progression to $\geq$CKD stage 3A, and a feature from the non-cystic kidney and entire kidney for prediction to a $\geq$30% reduction in eGFR. The Spearman rank correlation coefficient was used to assess the relationship of the chosen radiomic feature between (1) baseline and the 24-month timepoint, (2) baseline and the 48-month timepoint, and (3) baseline and the 60-month timepoint.

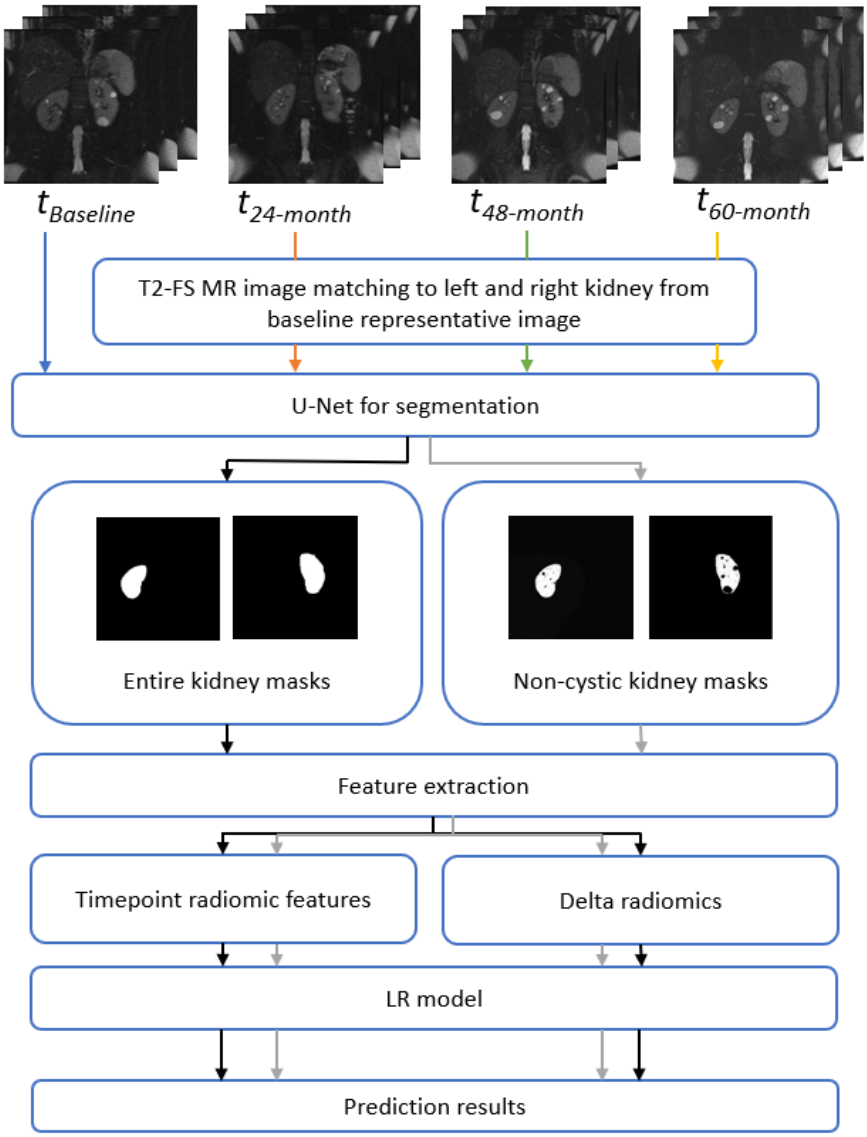

Figure 4.1: Image selection and feature extraction pipeline used to predict kidney function decline to ≥CKD stage 3A or ≥30% reduction in eGFR at 60 months using radiomic features extracted from MR images at baseline (t_{baseline}), 24 months ($t_{\text{24-month}}$), and 48 months ($t_{\text{48-month}}$). Features were extracted from the 60-month timepoint ($t_{\text{60-month}}$) but were not used in the logistic regression (LR) prediction models.

4.3 Results

4.3.1 Prediction to ≥CKD Stage 3A and ≥30% reduction in eGFR

The model performances using radiomic features to predict kidney function decline to (1) ≥CKD stage 3A or (2) ≥30% reduction in eGFR are presented in Table 4.3 and Table 4.4, respectively. Delta-radiomic features alone yielded AUC values close to random guessing (AUC=0.5) for both prediction models regardless of whether the features were extracted from the non-cystic kidney parenchyma or the entire kidney; the AUC values using delta radiomics ranged from 0.50-0.62 for prediction of progression to ≥CKD stage 3A and 0.50-0.56 for prediction of a ≥30% reduction in eGFR. The highest AUC value using delta-radiomics features was 0.62 from the entire kidney at the 48-month timepoint for prediction of progression to ≥CKD stage 3A.

For prediction of progression to ≥CKD stage 3A, AUC values using radiomic features at single timepoints (baseline, 24-month, and 48-month) from the non-cystic kidney parenchyma ranged from 0.72-0.75, and radiomic features from the entire kidney ranged from 0.75-0.78 (Figure 4.2). For prediction of a ≥30% reduction in eGFR, AUC values using radiomic features at single timepoints from the non-cystic kidney parenchyma ranged from 0.68-0.70, and radiomic features from the entire kidney ranged from 0.68-0.69 (Figure 4.3). In combining features, the combined-model of baseline and delta features did not result in an increase in AUC values for features extracted from the non-cystic kidney parenchyma or entire kidney. In combining features from timepoints, the largest increase in AUC to baseline was in combining baseline, 24-month, and 48-month timepoint features from the entire kidney in the prediction of progression to ≥CKD stage 3A, with an increase in AUC from 0.75 at baseline to 0.81; this increase was statistically significant after applying Bonferroni-Holm for multiple comparisons.

Table 4.3: Area under the receiver operating characteristic curve (AUC) along with the 95% confidence interval (CI) of AUC for radiomics-based models to predict progression to ≥CKD stage 3A at 60-month timepoint using single timepoints, delta-radiomics features, and combined timepoints using radiomic features extracted from the non-cystic kidney parenchyma and entire kidney. Asterisks indicate statistically significant differences from baseline model performance (after Bonferroni-Holm correction for 10 multiple comparisons).

Radiomics model	Non-cystic kidney	Entire kidney
Baseline	0.75 [0.72, 0.76]	0.75 [0.72, 0.77]
Δ 24-Months	0.50 [0.48, 0.53] ($p \leq 0.001$)*	0.50 [0.48, 0.53] ($p \leq 0.001$)*
24-Months	0.73 [0.71, 0.75] ($p = 0.260$)	0.78 [0.76, 0.79] ($p = 0.047$)
Δ 48-Months	0.56 [0.53, 0.58] ($p \leq 0.001$)*	0.62 [0.60, 0.64] ($p \leq 0.001$)*
48-Months	0.72 [0.70, 0.74] ($p = 0.132$)	0.77 [0.75, 0.79] ($p = 0.471$)
Baseline + Δ 24-Months	0.73 [0.71, 0.75] ($p = 0.265$)	0.72 [0.70, 0.75] ($p = 0.002$)*
Baseline + 24-Months	0.76 [0.74, 0.78] ($p = 0.695$)	0.78 [0.76, 0.79] ($p = 0.187$)
Baseline + Δ 48-Months	0.75 [0.73, 0.76] ($p = 0.438$)	0.73 [0.71, 0.75] ($p = 0.012$)
Baseline + 48-Months	0.76 [0.74, 0.78] ($p = 0.893$)	0.78 [0.76, 0.80] ($p = 0.352$)
Baseline + Δ 24-Months + Δ 48-Months	0.72 [0.70, 0.74] ($p = 0.088$)	0.71 [0.69, 0.73] ($p \leq 0.001$)*
Baseline + 24-Months + 48-Months	0.75 [0.73, 0.77] ($p = 0.613$)	0.81 [0.79, 0.82] ($p \leq 0.001$)*

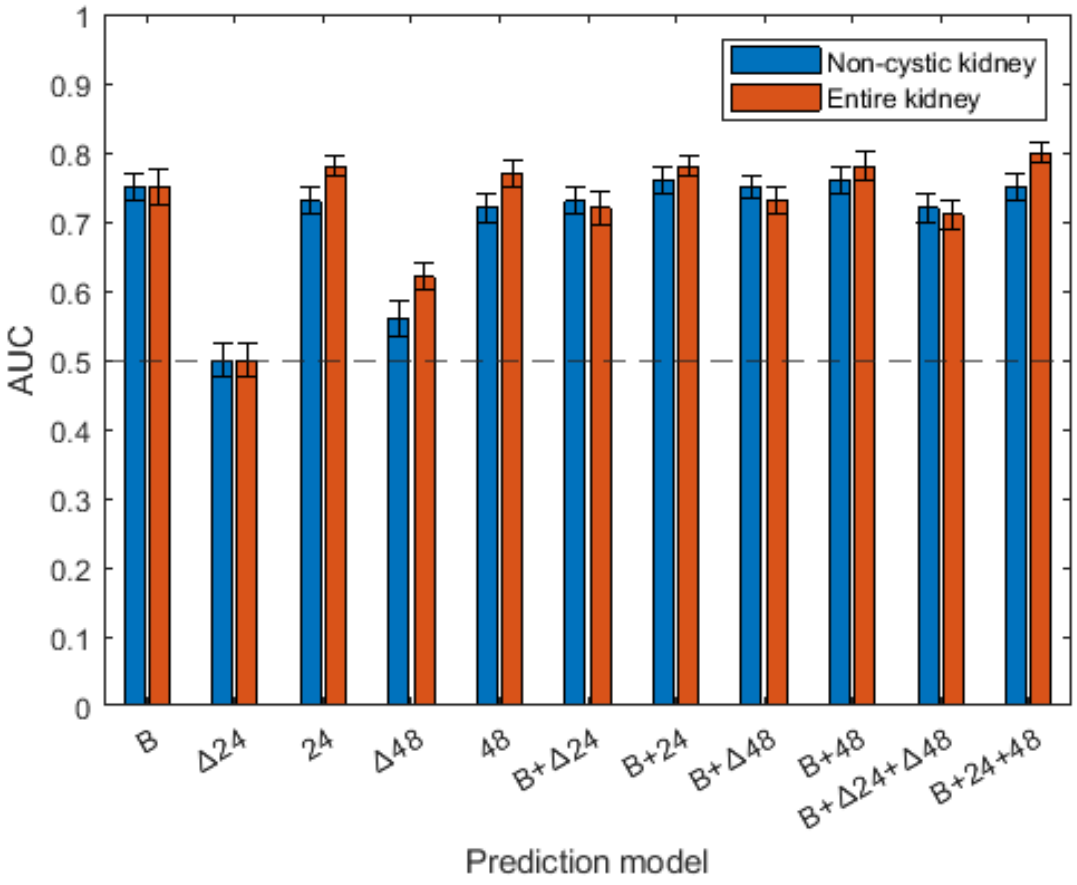

Figure 4.2: Area under the receiver operating characteristic curve (AUC) values for the prediction of progression to ≥CKD stage 3A using radiomic features from the non-cystic kidney parenchyma (blue) and the entire kidney (orange). The dotted line at an AUC of 0.5 is random guessing. There were 11 radiomics models investigated using baseline and follow-up timepoints: (1) baseline (B); (2) 24-month delta radiomics (Δ24); (3) 24-month timepoint (24); (4) 48-month delta radiomics (Δ48); (5) 48-month timepoint (48); (6) combined baseline and 24-month delta radiomics (B+Δ24); (7) combined baseline and 24-month radiomics (B+24); (8) combined baseline and 48-month delta radiomics (B+Δ48); (9) combined baseline and 48-month timepoint (B+48); (10) combined baseline, 24-month delta radiomics, and 48-month delta radiomics (B+Δ24+Δ48); and (11) combined baseline 24-month timepoint, and 48-month timepoint(B+24+48).

Table 4.4: Area under the receiver operating characteristic curve (AUC) along with the 95% confidence interval (CI) of AUC for radiomics-based models to predict progression to ≥30% reduction in eGFR at 60-month timepoint using single timepoints, delta-radiomics features, and combined timepoints using radiomic features extracted from the non-cystic kidney parenchyma and entire kidney. Asterisks indicate statistically significant differences from baseline model performance (after Bonferroni-Holm correction for 10 multiple comparisons).

Radiomics model	Non-cystic kidney	Entire kidney
Baseline	0.68 [0.65, 0.71]	0.68 [0.66, 0.71]
Δ 24-Months	0.52 [0.49, 0.55] ($p \leq 0.001$)*	0.55 [0.52, 0.58] ($p \leq 0.001$)*
24-Months	0.70 [0.68, 0.72] ($p = 0.675$)	0.68 [0.65, 0.71] ($p = 0.932$)
Δ 48-Months	0.50 [0.47, 0.53] ($p \leq 0.001$)*	0.56 [0.53, 0.59] ($p \leq 0.001$)*
48-Months	0.70 [0.67, 0.72] ($p = 0.763$)	0.69 [0.66, 0.71] ($p = 0.606$)
Baseline + Δ 24-Months	0.66 [0.63, 0.69] ($p = 0.133$)	0.67 [0.64, 0.69] ($p = 0.834$)
Baseline + 24-Months	0.67 [0.64, 0.70] ($p = 0.084$)	0.67 [0.64, 0.70] ($p = 0.355$)
Baseline + Δ 48-Months	0.66 [0.63, 0.69] ($p = 0.038$)	0.67 [0.64, 0.69] ($p = 0.372$)
Baseline + 48-Months	0.67 [0.65, 0.70] ($p = 0.399$)	0.67 [0.64, 0.69] ($p = 0.918$)
Baseline + Δ 24-Months + Δ 48-Months	0.66 [0.63, 0.68] ($p = 0.052$)	0.65 [0.62, 0.67] ($p = 0.025$)
Baseline + 24-Months + 48-Months	0.64 [0.61, 0.67] ($p = 0.006$)*	0.65 [0.62, 0.68] ($p = 0.023$)

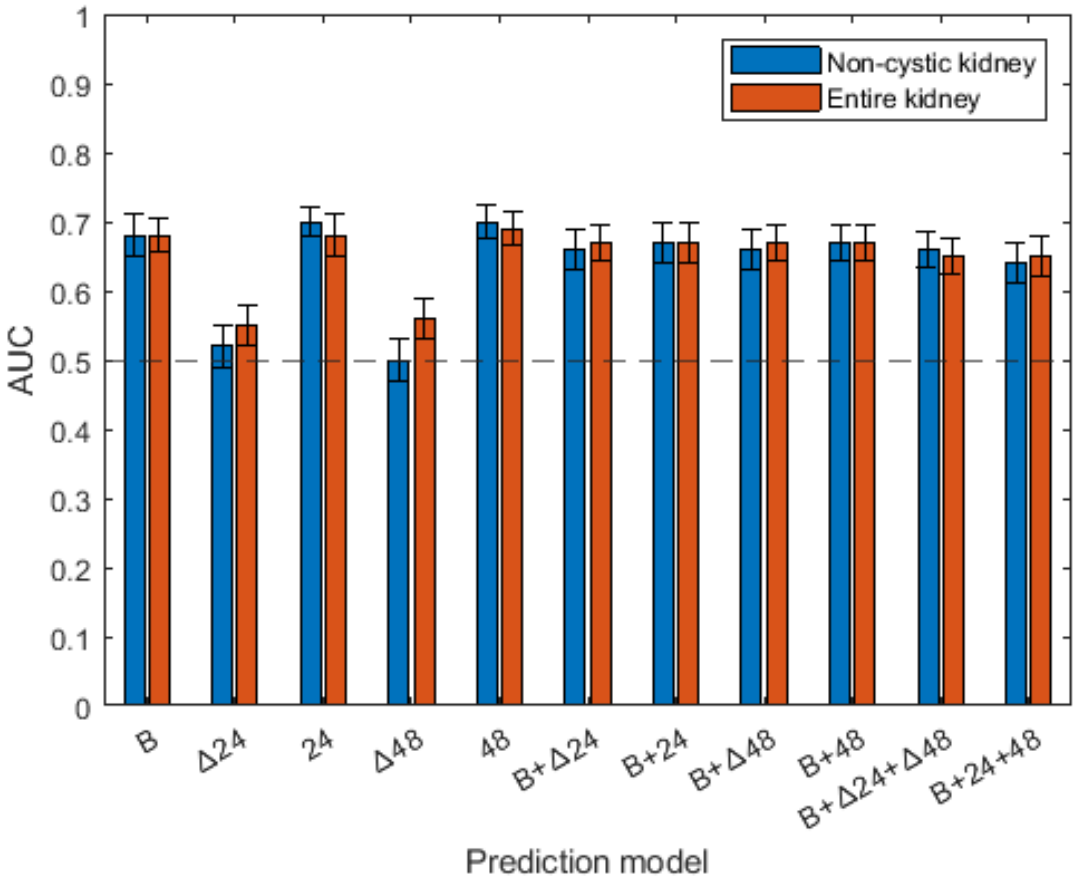

Figure 4.3: Area under the receiver operating characteristic curve (AUC) values for prediction to a ≥30% reduction in eGFR using radiomic features from the non-cystic kidney parenchyma (blue) and the entire kidney (orange). The dotted line at an AUC of 0.5 is random guessing. There were 11 radiomics models investigated using baseline and follow-up timepoints: (1) baseline (B); (2) 24-month delta radiomics (Δ24); (3) 24-month timepoint (24); (4) 48-month delta radiomics (Δ48); (5) 48-month timepoint (48); (6) combined baseline and 24-month delta radiomics (B+Δ24); (7) combined baseline and 24-month radiomics (B+24); (8) combined baseline and 48-month delta radiomics (B+Δ48); (9) combined baseline and 48-month timepoint (B+48); (10) combined baseline, 24-month delta radiomics, and 48-month delta radiomics (B+Δ24+Δ48); and (11) combined baseline 24-month timepoint, and 48-month timepoint(B+24+48).

The model performances for the clinical model and clinical-radiomics combined models are presented in Table 4.5. Baseline age had an AUC value of 0.57 and 0.51 for prediction of progression to ≥CKD stage 3A and a ≥30% reduction in eGFR, respectively. Baseline htTKV had modest AUC values of 0.72 and 0.68 for prediction to ≥CKD stage 3A and ≥30% reduction in eGFR, respectively. Finally, baseline eGFR had the highest AUC value for prediction of progression to ≥CKD stage 3A of 0.88 and an AUC value of 0.65 for a ≥30% reduction in eGFR. The clinical model incorporating all clinical features had an AUC value of 0.88 for prediction of progression to ≥CKD stage 3A and an AUC value of 0.68 for a ≥30% reduction in eGFR. The largest increase in AUC for prediction of progression to ≥CKD stage 3A for the combined clinical-radiomics model was from 0.88 to 0.90, using features extracted from the entire kidney at the 48-month timepoint. The largest increase in AUC for prediction of a ≥30% reduction in eGFR was from 0.68 to 0.71 using features extracted from the kidney at the 24-month timpeoint. In summary, the performance of combined clinical-radiomics models did not show statistically significant differences, as determined by the DeLong test, compared to the performance of combined clinical models.

Table 4.5: Area under the receiver operating characteristic curve (AUC) along with the 95% confidence interval (CI) of AUC for the clinical and combined clinical-radiomics models to predict kidney function decline at the 60-month timepoint. Asterisks on AUC results indicate performance statistically significant to combined clinical model performance. Asterisks indicate statistically significant differences from baseline model performance (after Bonferroni-Holm correction for 5 multiple comparisons).

Model	≥CKD stage 3A	≥30% reduction in eGFR
Clinical		
Age	0.57 [0.55, 0.60]	0.51 [0.48, 0.54]
htTKV	0.72 [0.70, 0.74]	0.68 [0.65, 0.70]
eGFR	0.88 [0.87, 0.89]	0.65 [0.62, 0.68]
Combined clinical	0.88 [0.87, 0.89]	0.68 [0.65, 0.70]
Clinical + Baseline		
Non-cystic kidney	0.89 [0.88, 0.90] ($p = 0.930$)	0.69 [0.67, 0.72] ($p = 0.650$)
Entire kidney	0.88 [0.86, 0.89] ($p = 0.287$)	0.70 [0.67, 0.72] ($p = 0.485$)
Clinical + 24-Months		
Non-cystic kidney	0.89 [0.88, 0.90] ($p = 0.425$)	0.69 [0.67, 0.72] ($p = 0.654$)
Entire kidney	0.89 [0.88, 0.90] ($p = 0.511$)	0.71 [0.69, 0.74] ($p = 0.238$)
Clinical + 48-Months		
Non-cystic kidney	0.87 [0.86, 0.89] ($p = 0.200$)	0.68 [0.65, 0.71] ($p = 0.649$)
Entire kidney	0.90 [0.88, 0.91] ($p = 0.276$)	0.70 [0.67, 0.73] ($p = 0.209$)
Clinical + Baseline + 24-Months		
Non-cystic kidney	0.88 [0.87, 0.90] ($p = 0.623$)	0.66 [0.63, 0.68] ($p = 0.060$)
Entire kidney	0.87 [0.86, 0.89] ($p = 0.220$)	0.67 [0.64, 0.70] ($p = 0.505$)
Clinical + Baseline + 24-Months + 48-Months		
Non-cystic kidney	0.87 [0.86, 0.89] ($p = 0.167$)	0.61 [0.58, 0.64] ($p = 0.027$)
Entire kidney	0.88 [0.87, 0.89] ($p = 0.956$)	0.66 [0.64, 0.69] ($p = 0.142$)

4.3.2 Radiomic features at timepoints

The Spearman rank correlation coefficient was calculated to assess the relationship of radiomic features at baseline, 24-month, 48-month, and 60-month timepoints. The Spearman correlation coefficient was calculated on a single representative radiomic feature chosen frequently during feature selection and subsequently used in the four predictive tasks: non-cystic kidney parenchyma and entire kidney for the prediction of progression to ≥CKD stage 3A and from the non-cystic kidney parenchyma and entire kidney parenchyma for prediction to a ≥30% reduction in eGFR; the same features were analyzed using their calculated delta feature to assess the relationship with the baseline feature.

For the prediction of kidney function decline to ≥CKD stage 3A, top features include GLSZM size-zone non-uniformity normalized (SZNN) from the non-cystic kidney parenchyma and NGTDM strength from the entire kidney (Figure 4.4-Figure 4.5). GLSZM quantifies gray level zones, or number of connected voxels that share the same gray-level intensity, in an image; GLSZM SZNN measures the variability of size zones, with a lower value indicating more homogeneity among zone size volumes in an image [61]. The AUC values of GLSZM SZNN in predicting kidney function decline with all of the data at baseline, 24-month timepoint, 48-month timepoint, and 60-month timepoint was 0.75 [0.69, 0.81], 0.73 [0.66, 0.79], 0.70 [0.63, 0.77], and 0.68 [0.60, 0.75], respectively. NGTDM quantifies the sum of absolute differences of the difference between a gray level and the average gray levels of its neighbors; NGTDM strength measures primitives in an image, with high values indicating a slow change in intensity with more large coarse differences in gray-level intensities [61]. The AUC values of NGTDM strength in predicting kidney function decline with all of the data at baseline, 24-month timepoint, 48-month timepoint, and 60-month timepoint was 0.75 [0.70, 0.82], 0.74 [0.69, 0.80], 0.74 [0.68, 0.81], and 0.75 [0.68, 0.81], respectively. Table 4.6 and Table 4.8 contain the Spearman correlation coefficients between baseline and timepoint features for the non-cystic kidney and entire kidney features, respectively. Table 4.7 and Table 4.9 are

the comparison of feature values between baseline and timepoints and between progressors and non-progressors at timepoints using the Wilcoxon rank sum test for the non-cystic kidney and entire kidney features, respectively. Figure 4.7 and Figure 4.9 are the histograms of feature values for progressors and non-progressors to ≥CKD stage 3A for the non-cystic kidney and entire kidney features, respectively.

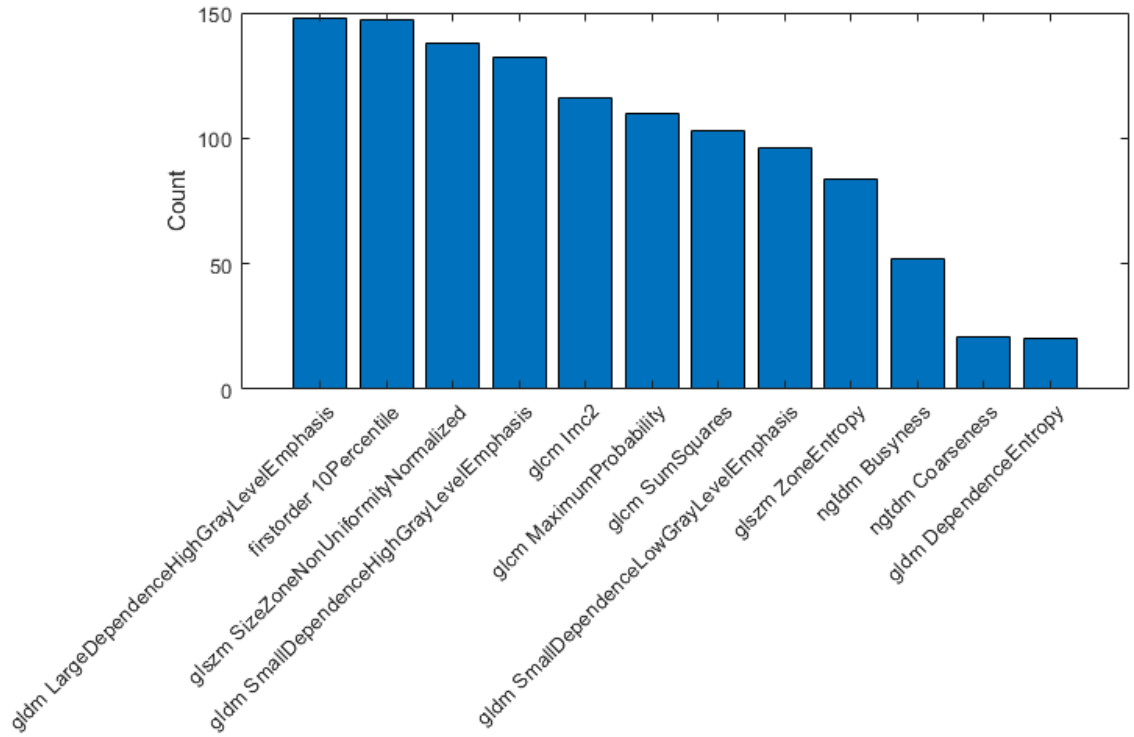

Figure 4.4: Histogram counts of radiomic features extracted from the non-cystic kidney for prediction of progression to ≥CKD stage 3A. The radiomic features were selected in all single timepoint models of baseline, 24-month, and 48-month timepoints.

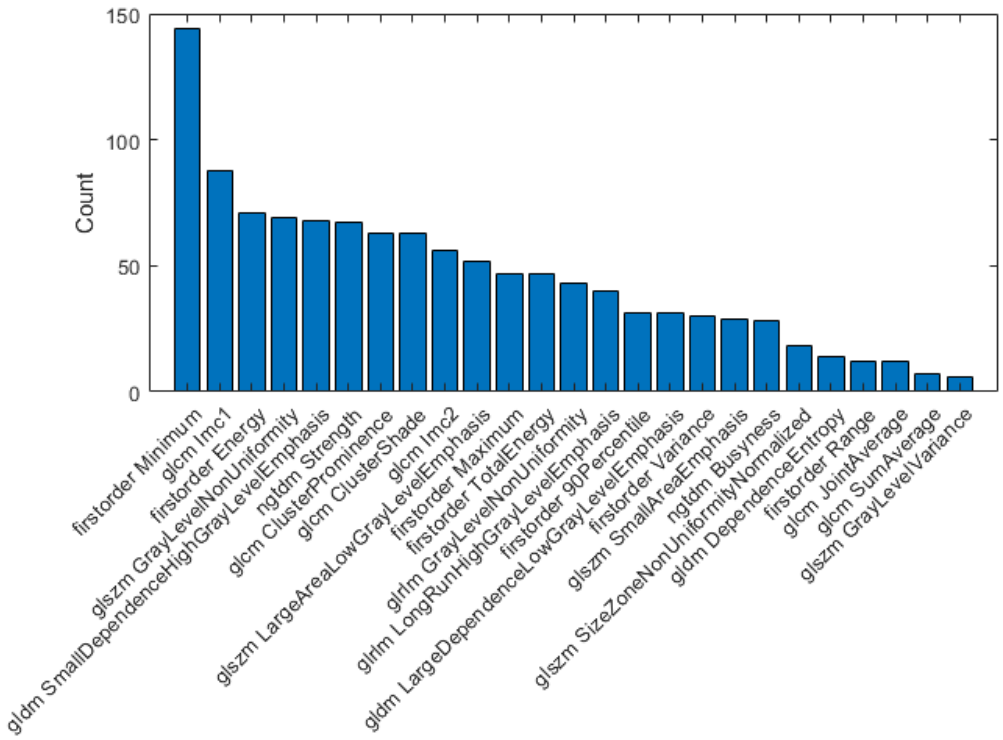

Figure 4.5: Histogram counts of radiomic features extracted from the entire kidney for prediction of progression to ≥CKD stage 3A. The radiomic features were selected in all single timepoint models of baseline, 24-month, and 48-month timepoints.

Table 4.6: The Spearman rank correlation coefficients between the baseline and timepoint feature and between the baseline and delta feature of GLSZM SZNN, used in baseline, 24-month timepoint, and 48-month timepoint models for prediction of progression to ≥CKD stage 3A using the non-cystic kidney for feature extraction.

Timepoint	Timepoint Feature vs. Baseline	Delta Feature vs. Baseline
24-month	0.636	-0.336
48-month	0.625	-0.349
60-month	0.601	-0.368

Table 4.7: The p-values obtained from the Wilcoxon rank sum test using radiomic feature GLSZM SZNN from the non-cystic kidney for prediction of progression to ≥CKD stage 3A. A corrected $p < 0.05$ was significant for comparisons with baseline feature values, and asterisks denote statistical significance after multiple comparison corrections.

Timepoint	<CKD stage 3A vs. Baseline	≥CKD stage 3A vs. Baseline	≥CKD stage 3A vs. <CKD stage 3A
Baseline	-	-	<0.001*
24-month	0.837	0.771	<0.001*
48-month	0.041	0.711	<0.001*
60-month	0.015*	0.699	<0.001*

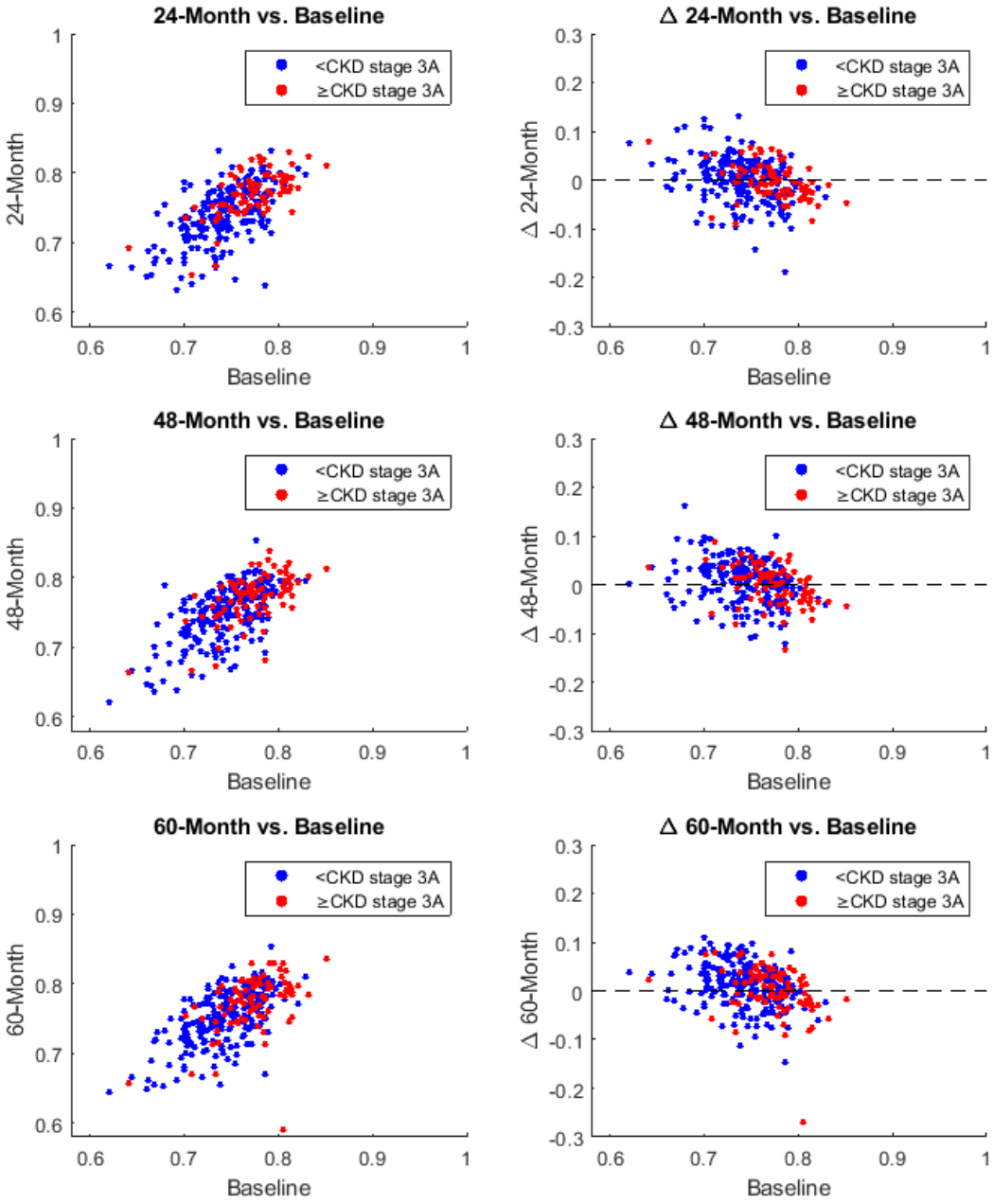

Figure 4.6: Scatter plots of radiomic feature GLSZM SZNN from the non-cystic kidney used for the prediction of progression to ≥CKD stage 3A at 24-month, 48-month, and 60-month timepoints, and its corresponding delta-feature from baseline at these timepoints.

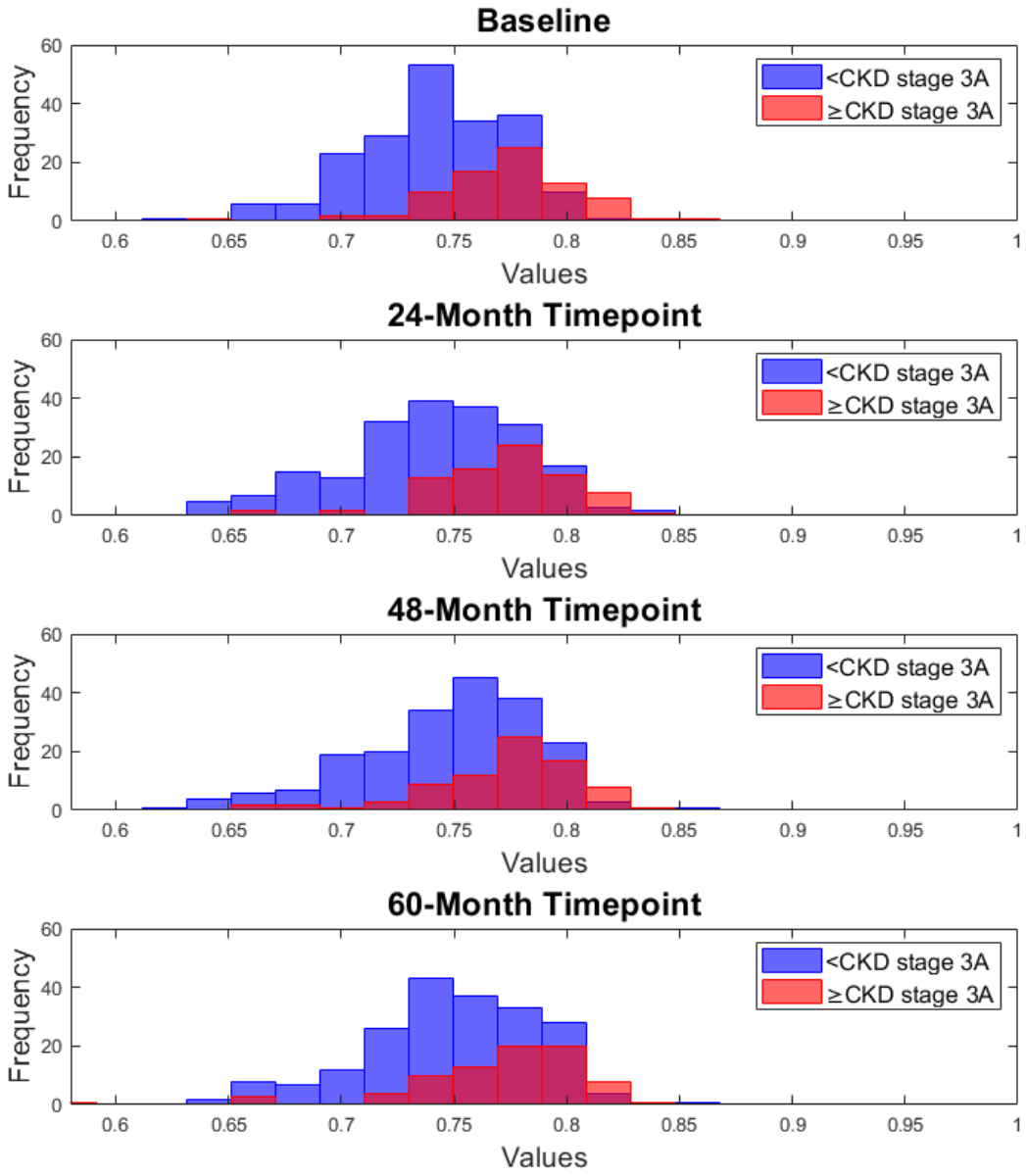

Figure 4.7: Histograms of radiomic feature values of GLSZM SZNN from the non-cystic kidney for the prediction of progression to ≥CKD stage 3A histogram at baseline, 24-month, 48-month, and 60-month timepoints. GLSZM SZNN measures the variability of size zones with a lower value indicating more homogeneity among zone size volumes (i.e., number of connected voxels that share the same gray-level intensity) in an image.

Table 4.8: The Spearman rank correlation coefficient was computed between the baseline and timepoint feature and between the baseline and delta feature of NGTDM strength, used in baseline, 24-month timepoint, and 48-month timepoint models for prediction of progression to ≥CKD stage 3A using the entire kidney for feature extraction.

Timepoint	Timepoint Feature vs. Baseline	Delta Feature vs. Baseline
24-month	0.869	-0.304
48-month	0.865	-0.291
60-month	0.868	-0.276

Table 4.9: The p-values obtained from the Wilcoxon rank sum test using radiomic feature NGTDM strength from the entire kidney for prediction of progression to ≥CKD stage 3A. A corrected $p < 0.05$ was significant for comparisons with baseline feature values, and asterisks denote statistical significance after multiple comparison corrections.

Timepoint	<CKD stage 3A vs. Baseline	≥CKD stage 3A vs. Baseline	≥CKD stage 3A vs. <CKD stage 3A
Baseline	-	-	<0.001*
24-month	0.963	0.678	<0.001*
48-month	0.184	0.494	<0.001*
60-month	0.040	0.163	<0.001*

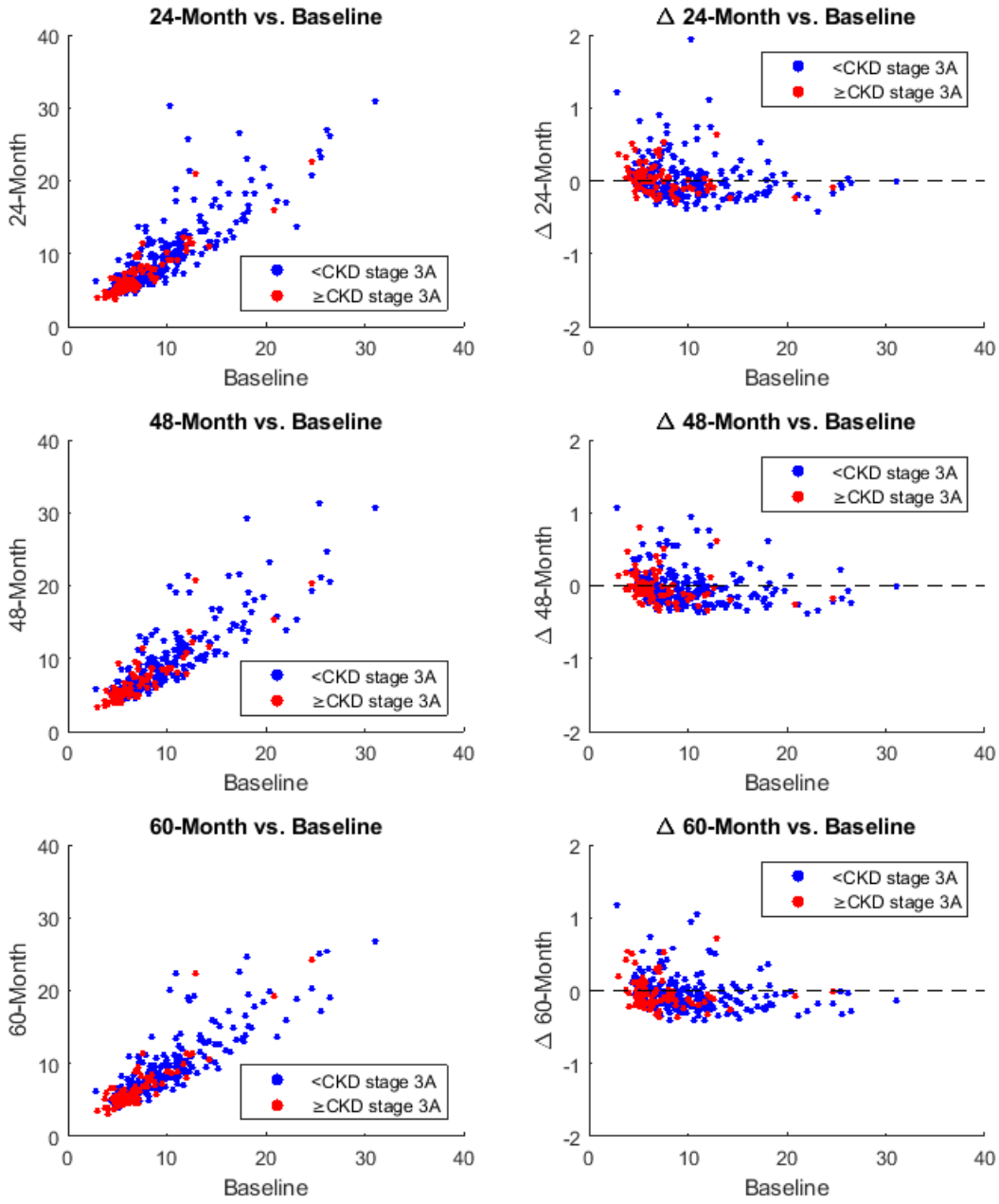

Figure 4.8: Scatter plots of radiomic feature NGTDM strength from the entire kidney used for the prediction of progression to ≥CKD stage 3A at 24-month, 48-month, and 60-month timepoints, and its corresponding delta-feature from baseline at these timepoints.

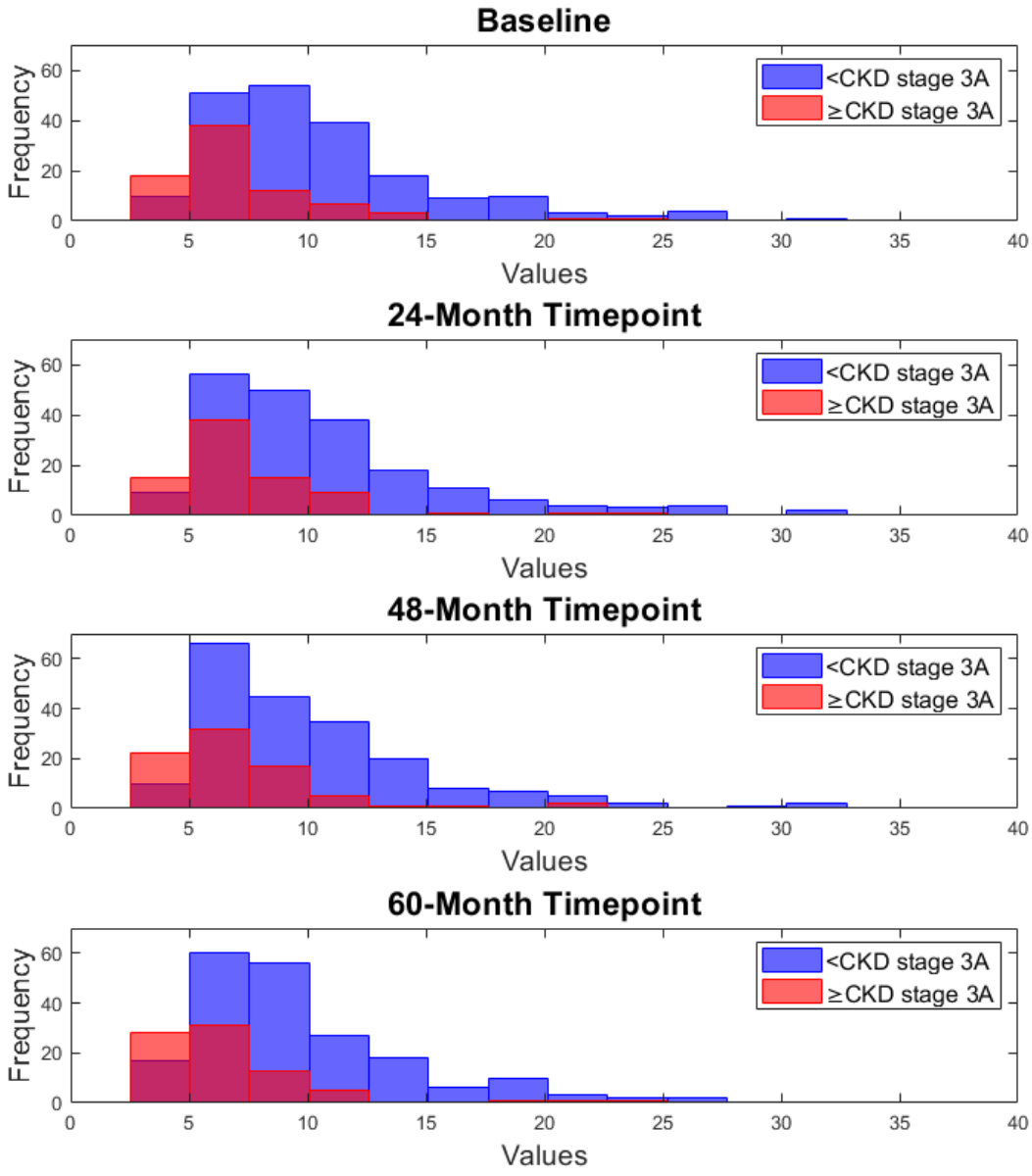

Figure 4.9: Histograms of radiomic feature values of NGTDM strength from the entire kidney for the prediction of progression to ≥CKD stage 3A histogram at baseline, 24-month, 48-month, and 60-month timepoints. NGTDM strength measures primitives in an image with high values indicating a slow change in intensity with more large coarse differences in gray-level intensities.

For prediction to a ≥30% reduction in eGFR, top features include GLCM informational measure of correlation (IMC) 2 from the non-cystic kidney parenchyma and GLCM joint energy from the entire kidney (Figure 4.10-Figure 4.11). GLCM is the second-order joint probability function of the number of times the combinations of two gray levels occur in two pixels in an image, and this work computed the GLCM at a distance of 1 pixel from the center voxel. GLCM IMC2 assesses the correlation between the probability distribution of pixel pairs and describes the complexity of texture, with a maximal value representing two fully dependent and uniform distributions that share maximal mutual information with a value approaching one [61]. The AUC values of GLCM IMC2 in predicting kidney function decline with all of the data at baseline, 24-month timepoint, 48-month timepoint, and 60-month timepoint was 0.64 [0.55, 0.71], 0.68 [0.60, 0.75], 0.67 [0.58, 0.75], and 0.62 [0.52, 0.71], respectively. GLCM joint energy is a measure of homogeneous patterns in the image, with a greater energy having more instances of intensity value pairs that neighbor each other at higher frequencies in the image [61]. The AUC values of GLCM joint energy in predicting kidney function decline with all of the data at baseline, 24-month timepoint, 48-month timepoint, and 60-month timepoint was 0.73 [0.65, 0.81], 0.73 [0.65, 0.80], 0.70 [0.60, 0.77], and 0.68 [0.58, 0.75], respectively. Table 4.10 and Table 4.12 contain the Spearman correlation coefficients between baseline and timepoint features for the non-cystic kidney and entire kidney features, respectively. Table 4.11 and Table 4.13 are the comparison of feature values between baseline and timepoints and between progessors and non-progressors at timepoints using the Wilcoxon rank sum test for the non-cystic kidney and entire kidney features, respectively. Figure 4.13 and Figure 4.15 are the histograms of feature values for progressors and non-progressors to ≥30% reduction in eGFR for the non-cystic kidney and entire kidney features, respectively.

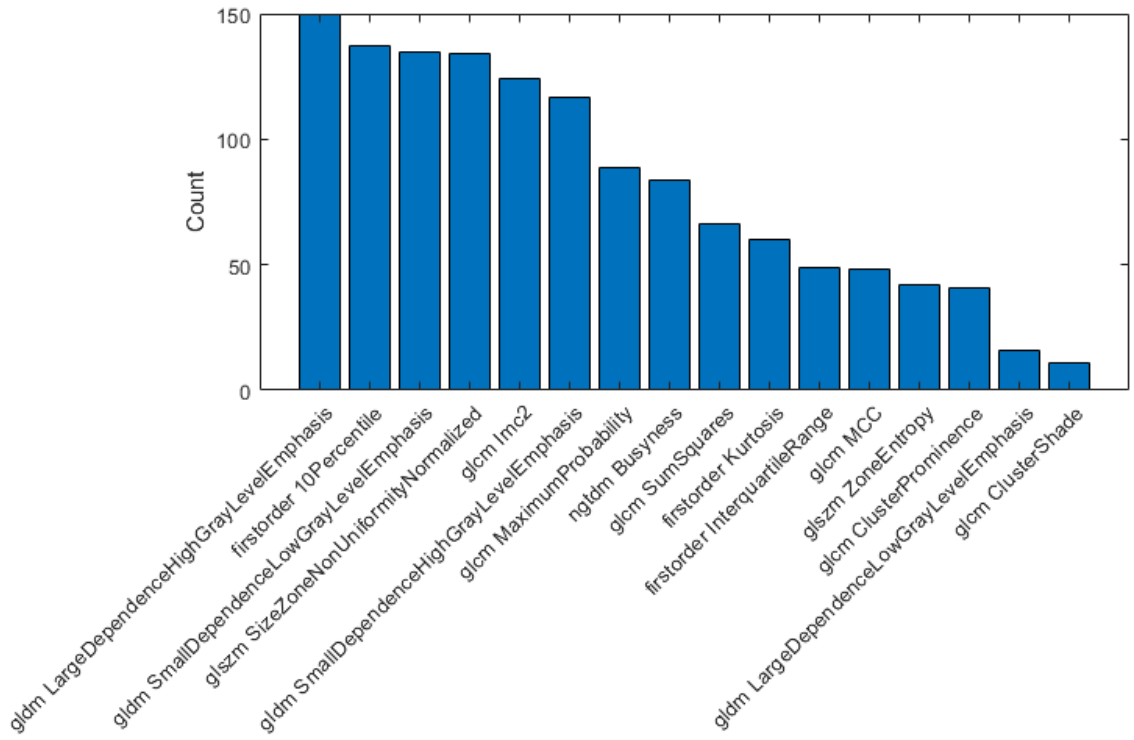

Figure 4.10: Histogram counts of radiomic features extracted from the non-cystic kidney for prediction to a $\geq$30% reduction in eGFR. The radiomic features were selected in all single timepoint models of baseline, 24-month, and 48-month timepoints.

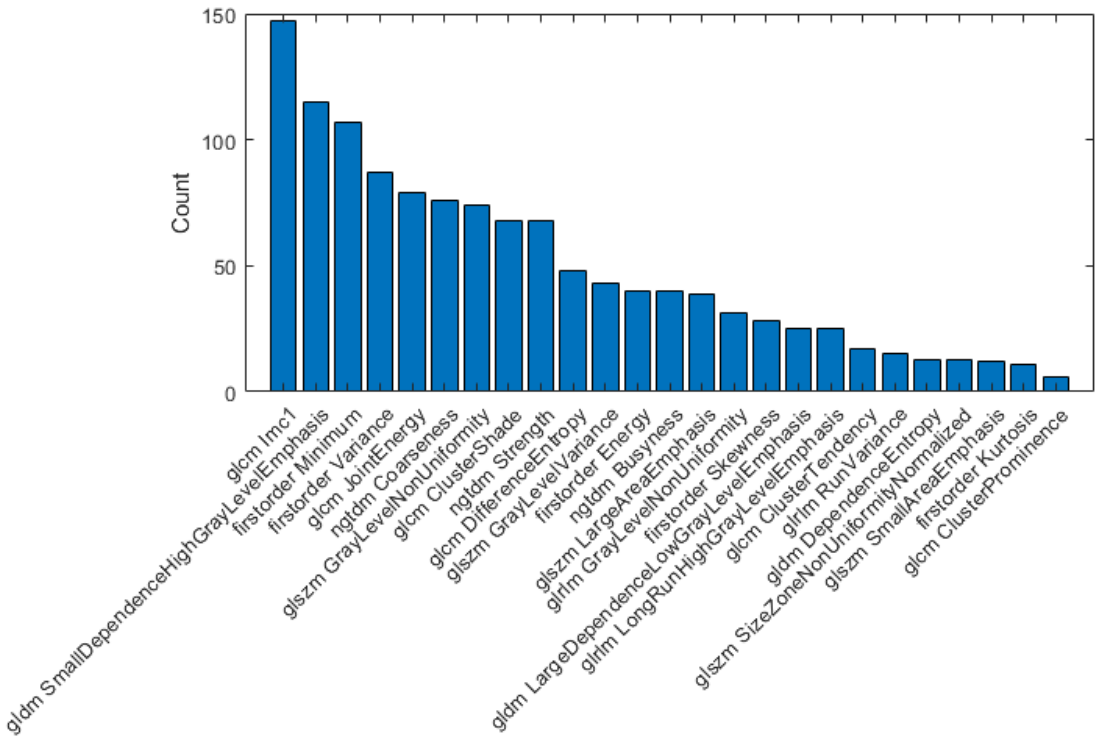

Figure 4.11: Histogram counts of radiomic features extracted from the entire kidney for prediction to a $\geq$30% reduction in eGFR. The radiomic features were selected in all single timepoint models of baseline, 24-month, and 48-month timepoints.

Table 4.10: The Spearman rank correlation coefficient was computed between the baseline and timepoint feature and between the baseline and delta feature of GLCM IMC2, used in baseline, 24-month timepoint, and 48-month timepoint models for prediction to a ≥30% reduction in eGFR using the non-cystic kidney for feature extraction.

Timepoint	Timepoint Feature vs. Baseline	Delta Feature vs. Baseline
24-month	0.634	-0.504
48-month	0.574	-0.491
60-month	0.524	-0.545

Table 4.11: The p-values obtained from the Wilcoxon rank sum test using radiomic feature GLCM IMC2 from the non-cystic kidney for prediction to a ≥30% reduction in eGFR. A corrected $p < 0.05$ was significant for comparisons with baseline feature values, and asterisks denote statistical significance after multiple comparison corrections.

Timepoint	<30% reduction in eGFR vs. Baseline	≥30% reduction in eGFR vs. Baseline	<30% reduction in eGFR vs. ≥30% reduction in eGFR
Baseline	-	-	0.004*
24-month	<0.001*	0.012*	<0.001*
48-month	<0.001*	<0.001*	<0.001*
60-month	<0.001*	0.002*	0.011*

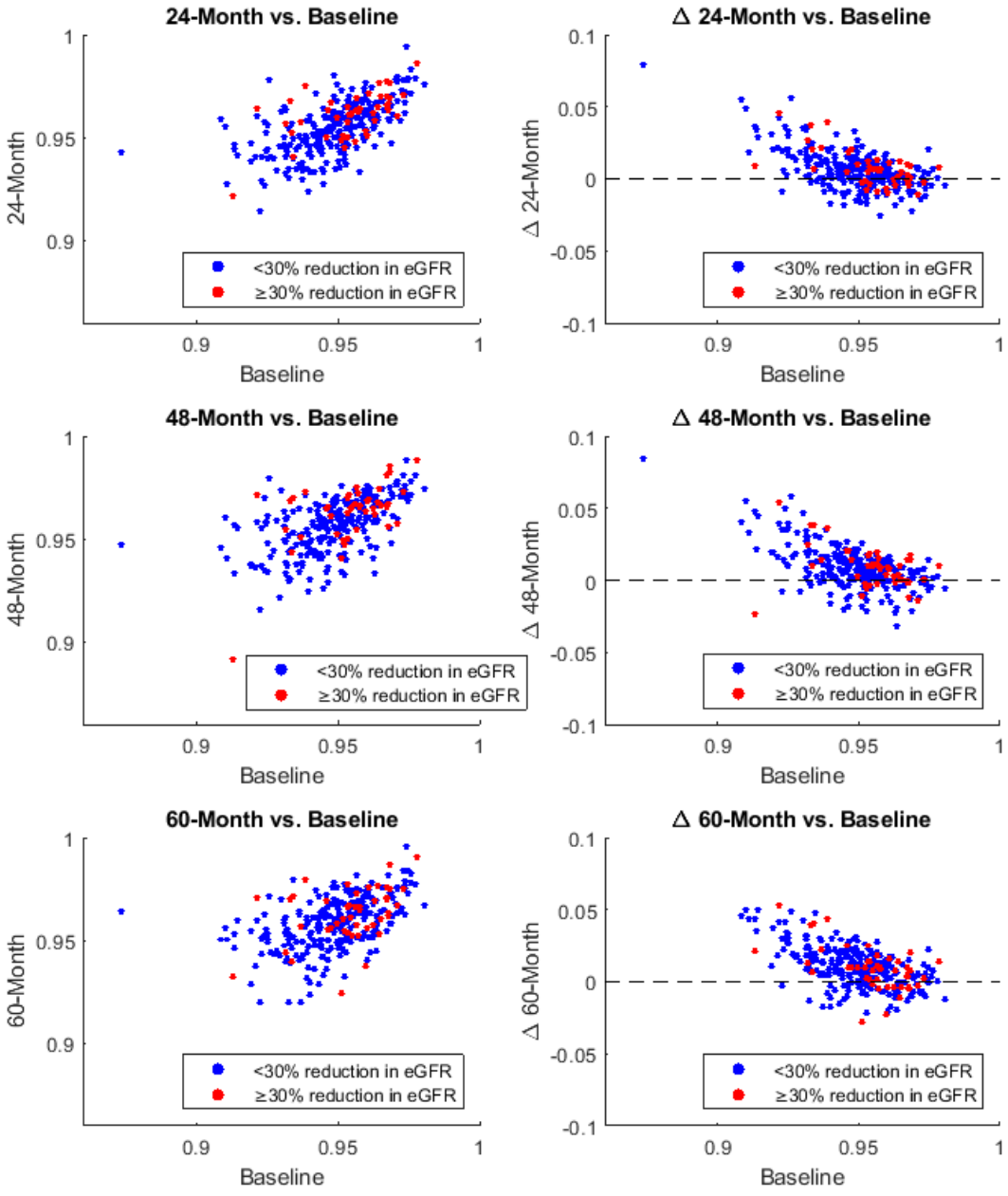

Figure 4.12: Scatter plots of radiomic feature GLCM IMC2 from the non-cystic kidney used for the prediction to a ≥30% reduction in eGFR at 24-month, 48-month, and 60-month timepoints, and its corresponding delta-feature from baseline at these timepoints.

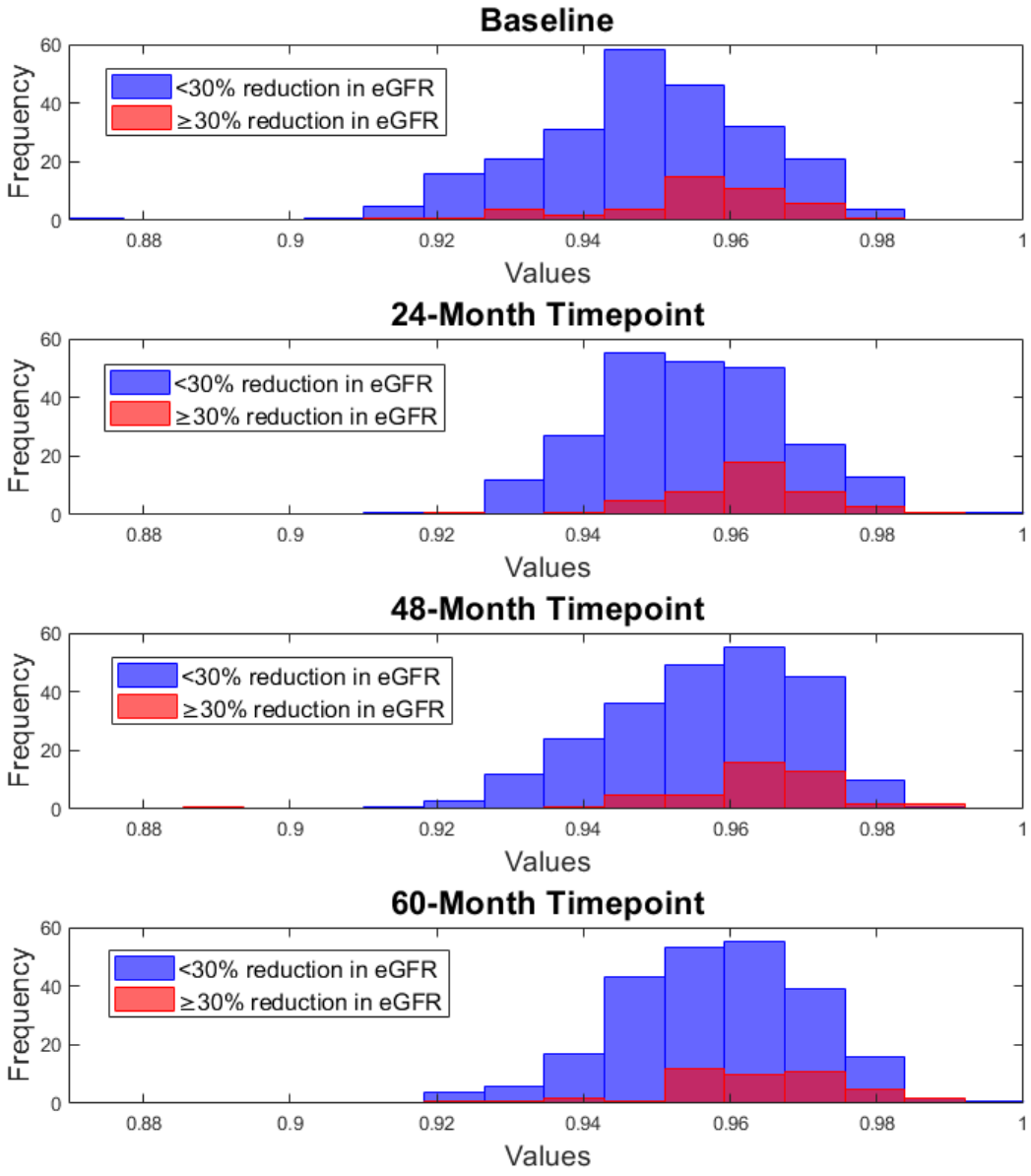

Figure 4.13: Histograms of radiomic feature values of GLCM IMC2 from the non-cystic kidney for the prediction to a ≥30% reduction in eGFR at baseline, 24-month, 48-month, and 60-month timepoints. GLCM IMC2 assesses the correlation between the probability distribution of pixel pairs and describes the complexity of texture, with a value approaching one representing two fully dependent and uniform distributions.

Table 4.12: The Spearman rank correlation coefficient was computed between the baseline and timepoint feature and between the baseline and delta feature of GLCM joint energy, used in baseline, 24-month timepoint, and 48-month timepoint models for prediction to a ≥30% reduction in eGFR using the entire kidney for feature extraction.

Timepoint	Timepoint Feature vs. Baseline	Delta Feature vs. Baseline
24-month	0.807	-0.292
48-month	0.755	-0.388
60-month	0.738	-0.407

Table 4.13: The p-values obtained from the Wilcoxon rank sum test using radiomic feature GLCM joint energy from the entire kidney for prediction to a ≥30% reduction in eGFR. A corrected $p < 0.05$ was significant for comparisons with baseline feature values, and asterisks denote statistical significance after multiple comparison corrections.

Timepoint	<30% reduction in eGFR vs. Baseline	≥30% reduction in eGFR vs. Baseline	<30% reduction in eGFR vs. ≥30% reduction in eGFR
Baseline	-	-	<0.001*
24-month	0.359	0.478	<0.001*
48-month	0.719	0.287	<0.001*
60-month	0.965	0.163	<0.001*

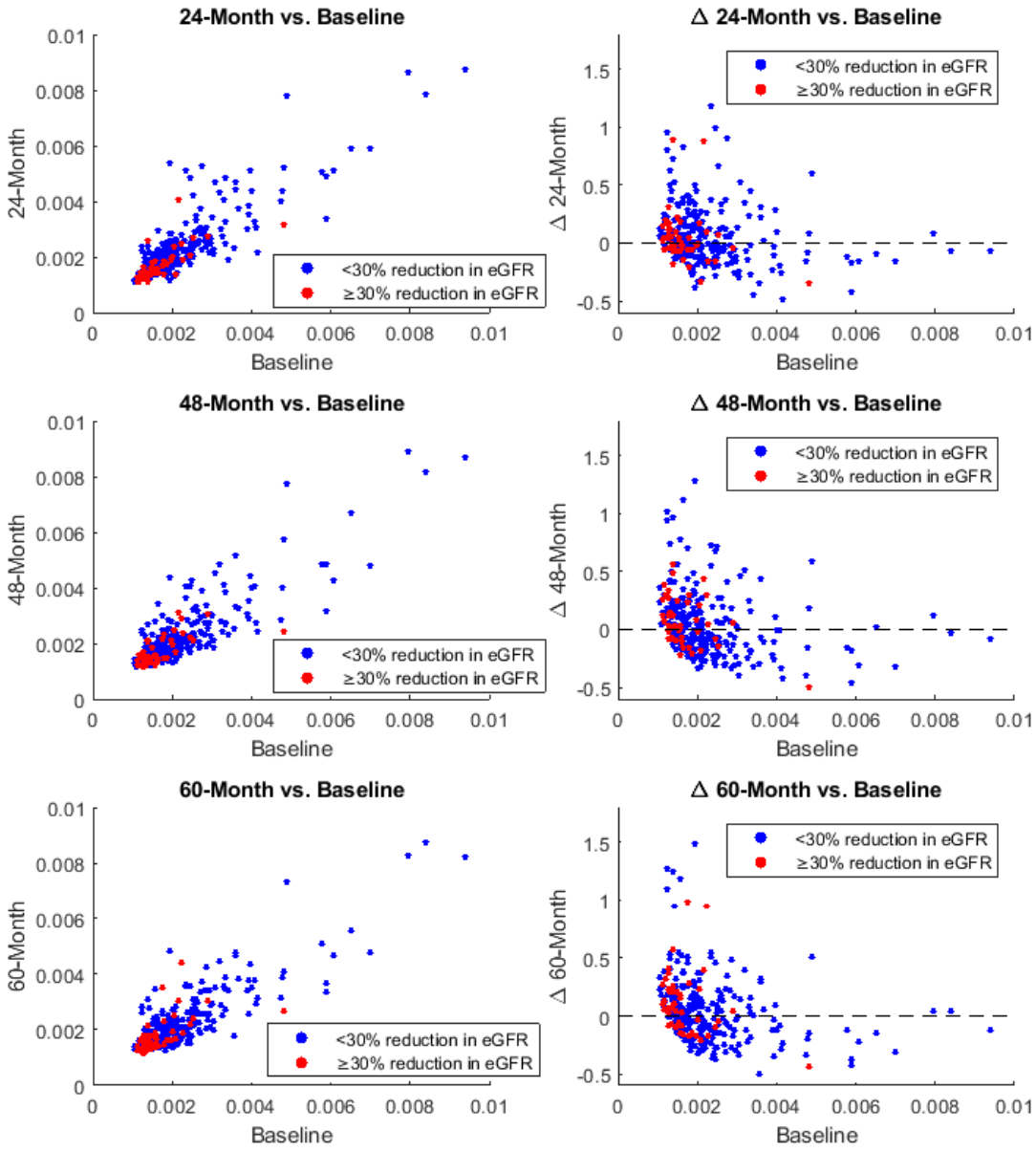

Figure 4.14: Scatter plots of radiomic feature GLCM joint energy from the entire kidney used for the prediction to a ≥30% reduction in eGFR at 24-month, 48-month, and 60-month timepoints, and its corresponding delta-feature from baseline at these timepoints.

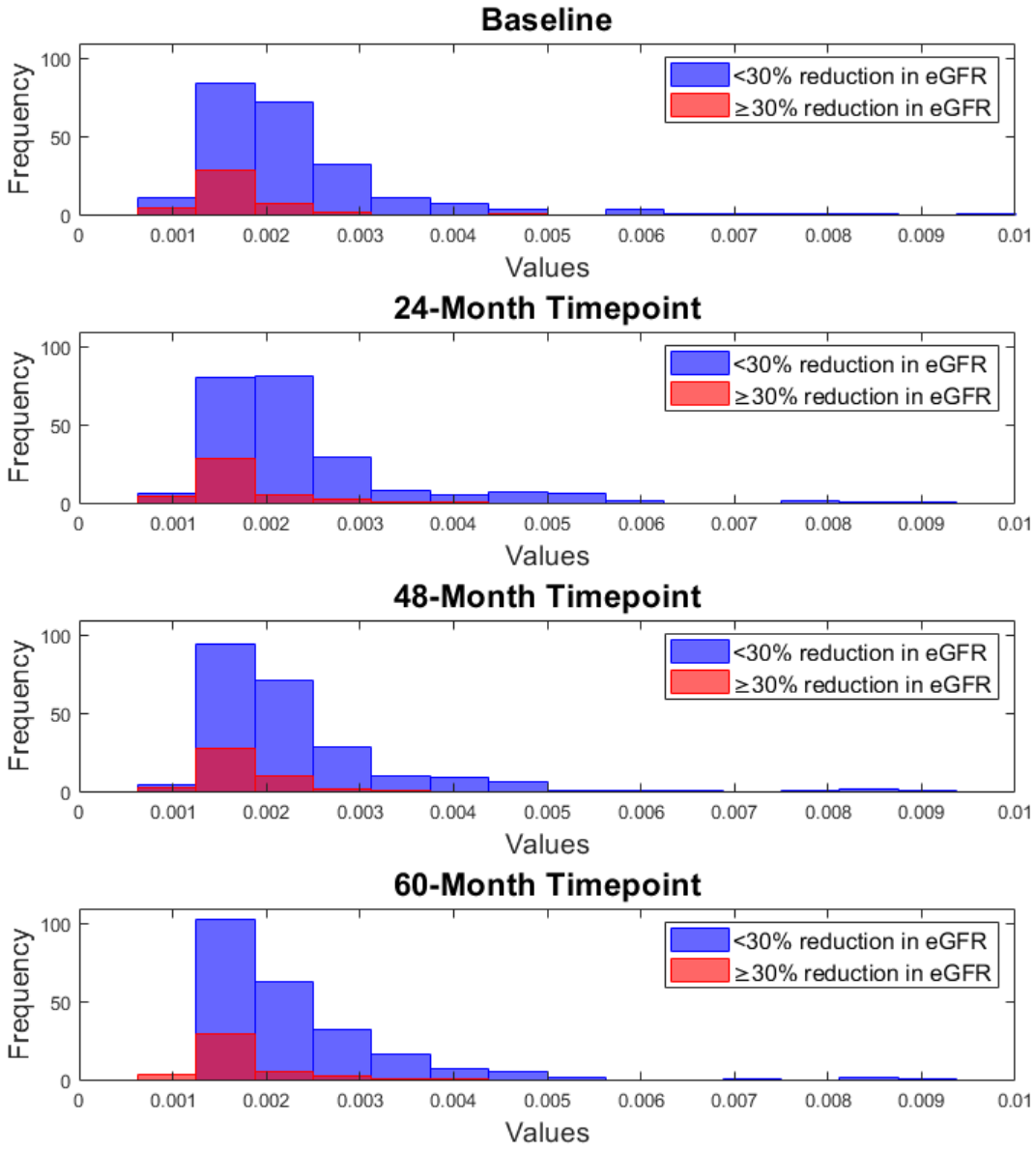

Figure 4.15: Histograms of radiomic feature values of GLCM joint energy from the entire kidney for the prediction to a ≥30% reduction in eGFR at baseline, 24-month, 48-month, and 60-month timepoints. GLCM joint energy is a measure of homogeneous patterns in the image, with a greater energy having more instances of intensity value pairs that neighbor each other at higher frequencies in the image.

4.4 Discussion

The purpose of this work was to determine whether radiomic features from 24-month and 48-month timepoints provide additive power versus baseline texture alone in predicting kidney function decline of patients with ADPKD. Utilizing delta radiomic features did not improve the predictive power for progression to ≥CKD stage 3A or a ≥30% reduction in eGFR, and delta radiomic features alone had close to random guessing performance (AUC=0.5). Furthermore, it is interesting to note that the performance of single timepoint models were not statistically significantly different from baseline performance in prediction for progression to ≥CKD stage 3A and a >30% reduction in eGFR. In combining features from timepoints versus baseline texture alone, the largest increase in AUC value was from 0.75 [0.72, 0.77] to 0.81 [0.79, 0.82] using baseline, 24-month, and 48-month texture from the entire kidney in predicting progression to ≥CKD stage 3A, which was statistically significantly different from baseline performance. Radiomic features extracted from the non-cystic kidney parenchyma had moderate performance for the prediction of kidney function decline, and, unlike the results from the entire kidney, there was no improvement in performance with the non-cystic kidney parenchyma in either predictive task by combining timepoints.

The Spearman rank correlation coefficient was used to assess the relationship between a single representative radiomic feature from each prediction model at baseline, 24-month, and 48-month timepoints and also the relationship between the same feature at baseline and its corresponding delta feature. In general, a strong, positive correlation coefficient with baseline and timepoint features was observed and this relationship was lost using delta radiomics. The relationship between the features at baseline and delta features had low to moderate negative correlation coefficients. The positive correlation coefficients between baseline and follow-up timepoints reflect the similar AUC values for prediction tasks at single timepoints using the representative radiomic features. The radiomic features from baseline to 60 months were highly correlated and highlights their consistency over time. This consistency

likely contributes to the lack of statistically significant differences in performance between the 24-month and 48-month timepoints compared to baseline.

The feature values at 24-month, 48-month, and 60-month timepoints of patients who did or did not progress to ≥CKD stage 3A were not statistically significantly different from the corresponding baseline features, except for the GLSZM SZNN feature extracted from non-cystic kidney parenchyma at the 60-month timepoint for patients who did not progress to ≥CKD stage 3A. Patients who experienced a ≥30% reduction and patients who did not experience a ≥30% reduction in eGFR had similar trends with the GLCM joint energy feature extracted from the entire kidney, but the GLCM IMC2 feature extracted from the the non-cystic kidney at 24-month, 48-month, and 60-month timepoints were statistically significantly different from baseline values. At baseline, 24-month, 48-month, and 60-month timepoints, the feature values were statistically significantly different between those who did or did not progress to ≥CKD stage 3A or experience a ≥30% reduction in eGFR. Exploring single features and how they change over time within and between groups provided insights as to why there was a poor predictive performance using delta radiomics. Regardless if patients progressed or did not progress to either predictive tasks, features were not statistically significantly different from baseline and the change in radiomic features at follow-up timepoints was similar between patients who progressed or did not progress.

The performance of the combined clinical-radiomics models did not exhibit statistically significant differences compared to that of the combined clinical model, which included baseline age, baseline htTKV, and baseline eGFR. These results contrast with findings from previous ADPKD radiomics studies that utilized one imaging timepoint. The work presented in this chapter differed from previous literature in a few ways. The first study utilizing radiomic features for kidney function decline in patients with ADPKD was published by Kline et al. [52], and this work extracted radiomic features from T2W-FS MR images from 122 patients who had eGFR values > 70 mL/min/1.73m^2 at baseline and assessed the ability to pre-

dict kidney function decline at eight-year follow up; incorporating texture into a traditional model with baseline age, htTKV, and eGFR, the AUC increased for CKD stage 3A patients from 0.86 to 0.94, for CKD stage 3B patients from 0.90 to 0.96, and for a ≥30% reduction in eGFR from 0.75 to 0.85. The previous work predicted kidney function decline at eight year follow-up which allows for three more years of disease progression relative to the five years presented in this work; additionally, the mean eGFR of the patients that progressed was 80.8-87.4 mL/min/1.73m^2, and for the patients who did not progress the mean eGFR range was 94.2-100 mL/min/1.73m^2. The cohort of patients presented in this chapter that progressed to either prediction task had a lower baseline mean eGFR of 71.05 mL/min/1.73m^2 (≥CKD stage 3A) and 79.21 mL/min/1.73m^2 (≥30% reduction in eGFR), and for the patients who did not progress the baseline mean eGFR was 90.5 mL/min/1.73m^2 (≥CKD stage 3A) and 86.09 mL/min/1.73m^2 (≥30% reduction in eGFR); the ability of eGFR alone to predict kidney function in this study was higher than the study by Kline et al. [52]: 0.75-0.77 (≥CKD stage 3A) and 0.55 (≥30% reduction in eGFR) versus the eGFR prediction in this work of 0.88 (≥CKD stage 3A) and 0.65 (≥30% reduction in eGFR). Xie et al. [55] examined the non-contrast CT images of 340 ADPKD patients and assessed the predictive value of radiomic features extracted from the renal (kidney) parenchyma volume for ADPKD patients that had a reduction of 30% of baseline eGFR at five-year follow up. When radiomics and renal parenchymal volume were combined, the predictive power improved versus radiomic features alone from 0.85 to 0.90 and outperformed the performance using TKV alone [55]; however, this study did not report the performance of baseline eGFR in the prediction of kidney function decline. The patients from the previous work who experienced a ≥30% reduction in eGFR also had a higher baseline mean eGFR than the patient cohort presented in this chapter: 86.9-87.1 mL/min/1.73m^2 for patients who progressed and 97.3-98.3 mL/min/1.73m^2 for the patients who did not progress. In summary, the baseline subject characteristics (age, htTKV, and eGFR) and time from baseline to follow-up differed

from previous published work in predicting kidney function decline of ADPKD patients using radiomic features.

Although radiomic features showed predictive power, future work will investigate change in texture of patients with either a stricter baseline eGFR >90 mL/min/1.73m^2 or match patients for baseline eGFR. In chapter 2, *PKD1* and *PKD2* patients were matched for age, gender, and MIC to account for confounding clinical factors that would impact the ability to capture texture differences that were associated with genotype in the non-cystic kidney parenchyma. Additionally, as patient age there are age-related factors that contribute to disease progression that may impact texture features, versus capturing cyst growth in younger patients who progress earlier and faster to ESKD [56]. This reiterates the complexity and heterogeneity of ADPKD phenotype, and accounting for clinical characteristics in radiomics studies is of utmost importance.

A limitation of this study was that only one MR pulse sequence was used for feature extraction. Li et al. [54] combined radiomic features from T1-weighted fat suppression (T1W-FS) MR images and T2W-FS MR images with clinical features, and the ability to classify patients in CKD stage 1 and 2 versus ≥CKD stage 3A improved from using radiomic features alone from 0.75 to 0.84. T2W-FS has high contrast between cyst and parenchyma and T1W-FS shows hemorrhagic cyst and residual kidney parenchyma, and radiomic features extracted from different MR pulse sequences provide complementary quantitative information that may associate with risk to kidney function decline [54]. Future work could investigate multiparametric MRI texture across timepoints such as T1W-FS, T1 and T2 quantitative maps, and diffusion-weighted imaging, all of which have shown importance in kidney imaging [95, 96].

Since medical imaging is clinically indicated for ADPKD patients to monitor disease progression to measure TKV, investigating change in texture takes advantage of the wealth of imaging information across timepoints for ADPKD patients. Delta radiomics is one method

of utilizing change in texture features over time, but there are other methods that could be of interest in ADPKD radiomics studies. For example, future work could investigate long short-term memory (LSTM) recurrent neural networks that exploit long-term dependencies between time steps of sequential data by incorporating distinctive features that are responsible for retaining information [47, 97]. In using LSTM networks of temporal imaging of ADPKD there may not be a need to calculate delta features. LSTM networks have been used in breast cancer to incorporate temporal information to improve overall classification performance to discriminate future lesions as malignant or benign, but its application in ADPKD has not been investigated [47, 98].

4.5 Conclusion

Based on ROC performance, the data indicate that delta-radiomic features did not improve the predictive power for progression to ≥CKD stage 3A or a ≥30% reduction in eGFR at 60-months follow-up. The performance using features from single timepoint models (24-month and 48-month timepoints) were not statistically significantly different from baseline performance. There was a statistically significant difference from baseline performance in combining baseline, 24-month, and 48-month texture from the entire kidney in predicting progression to ≥CKD stage 3A. MRI radiomic features extracted from the non-cystic kidney parenchyma show prognostic power across baseline, 24-month, and 48-month timepoints, revealing prognostic power in predicting kidney function decline without the presence of cysts.

CHAPTER 5

MAGNETIC RESONANCE FINGERPRINTING RADIOMICS ANALYSIS IN HEALTHY CONTROLS AND ADPKD PATIENTS: A PILOT STUDY

5.1 Introduction

Quantitative magnetic resonance imaging (MRI) of autosomal dominant polycystic kidney disease (ADPKD) patients has the ability to characterize the non-cystic kidney tissue microstructure, stratify based on risk-class, and differentiate ADPKD cohorts from healthy controls [20, 23, 24]. Quantitative MRI allows for direct quantitation of an image, thus removing the variability of interpretation, and helps further personalized tracking of pathological changes in tissue. Siedek et al. [24] proposed mean T2 of kidney parenchyma as an imaging biomarker for ADPKD, since it showed strong correlation with cyst fraction and served as a surrogate to differentiate between kidney cyst fraction groups. However, limitations of conventional quantitative MRI techniques include speed of acquisition and accurate quantitation. Magnetic resonance fingerprinting (MRF) is the simultaneous non-invasive quantification of tissues using pseudorandom acquisition parameters (e.g., repetition time and flip angle), that cause unique signal evolutions known as "fingerprints" [26]. MRF has dramatically decreased the acquisition time of quantitative MRI maps by deliberately varying acquisition parameters in a pseudo-random fashion, which differs from conventional methods of repeated, serial acquisition of data. MRF allows for simultaneous calculation of various MR contrasts, such as T1 and T2 relaxation times, in a single acquisition [25]. The MRF acquisition schemes were initially developed for brain imaging. However, a rapid kidney MRF method has recently been established [27]. The high efficiency and accuracy of MRF multiparametric mapping may provide personalized medicine opportunities and clinical promise for patients with ADPKD.

Mayo Imaging Classification (MIC) stratifies ADPKD patients from low and intermediate risk (MIC 1A-1B) to high risk (1C-1E) for kidney function decline and serves as a prognostic biomarker by incorporating a patient's height-corrected total kidney volume (htTKV) and age. Currently, there is no established risk-stratification method for children with ADPKD because MIC classification has not been validated in ADPKD patients less than 15 years of age. There is a need to better identify ADPKD children at higher risk of disease progression to both improve clinical care and allow for their inclusion in clinical trials.

Radiomic features have been shown to predict kidney function decline in ADPKD patients, classify ADPKD genotype and MIC using texture analysis of the non-cystic kidney parenchyma, and classify chronic kidney disease (CKD) stage [52, 54, 84, 85]. A recent published study used radiomic features to predict risk of progression in young ADPKD patients from T2-weighted (T2W) MR images [99]; in predicting fast or slow progression in htTKV growth (median annual change of $\geq$7.4%) at a 3-year follow-up, the AUC value increased from 0.56 when baseline htTKV alone was used to 0.70 when baseline htTKV was combined with kidney texture features [99]. In extracting radiomic features from T2W MR and T1-weighted (T1W) MR images, a normalization of signal intensity is required for feature extraction; although normalization has shown to improve classification and prediction performance, this step alters the original texture in an image, an effect that varies among normalization methods [40, 84]. Quantitative MRI measures tissue-specific properties (e.g., T1 relaxation, T2 relaxation), thus removing the pre-processing step of image normalization for feature extraction. Radiomic features from MRF-acquired T1 and T2 maps of ADPKD children and young adults may provide the ability to characterize the non-cystic kidney tissue and cystic tissue to better identify those at a high-risk for kidney function decline, without the need for image normalization with the ability to acquire images in a clinically relevant time-frame.

This exploratory aim will (1) compare kidney segmentation between two segmenters from different imaging sites and the effect on downstream non-cystic kidney parenchyma segmentation using three cyst removal methods and (2) investigate radiomic features from MRF-acquired T1 and T2 maps from a cohort of ADPKD children and young adults in comparison to a cohort of healthy controls. The work in this aim furthers the use of radiomic features extracted from the non-cystic kidney parenchyma of ADPKD children and young adults to stratify risk and capture early cystogenesis.

5.2 Methods

5.2.1 Database

In this work there were a total of 25 individuals: 15 ADPKD children and young adults and 10 healthy adult controls (Table 5.1). The healthy controls cohort were not age matched to the ADPKD cohort. A total of 15 ADPKD children and young adults were imaged at two sites: Northshore Hospital and the University of Chicago (n=8) and Cleveland Clinic Hospital and Case Western Reserve University (n=7). Completed measurements of the ADPKD cohort include body surface area-total kidney volume (BSA-TKV) and htTKV; total kidney volume (TKV) values are adjusted for body surface area (BSA) to account for the impact of somatic kidney growth on both TKV and total cyst volume (TCV) [100].

Table 5.1: ADPKD subject characteristics (n=14) based on mean age, sex, mean BSA-TKV, and mean htTKV. There was one patient whose clinical information was not completed at the time of this study.

Age group	Mean age [min, max]	Sex (M:F)	Mean BSA-TKV (ml/m) [min, max]	Mean htTKV (ml/m) [min, max]
<15 years (n=7)	11.25	1:6	165.86	144.91
	[8.50, 12.50]		[137.37, 203.21]	[117.20, 178.50]
>15 years (n=7)				
MIC 1A (n=1)	18.17	0:1	150.58	141.40
MIC 1B (n=3)	20.19	3:0	227.42	252.50
	[18.25, 23.83]		[218.16, 239.95]	[243.60, 270.20]
MIC 1C (n=3)	20.14	2:1	277.37	280.70
	[17.25, 25.75]		[245.28, 296.36]	[249.50, 305.10]

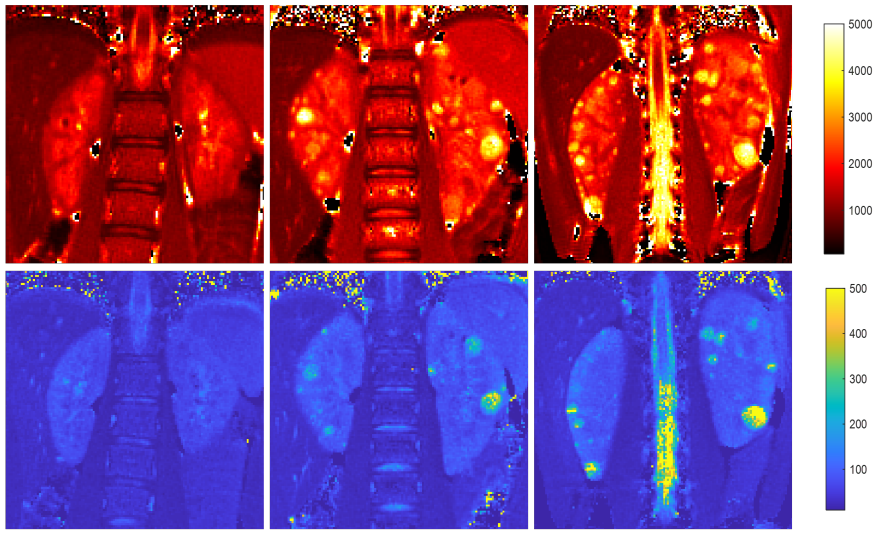

Figure 5.1: T1 (top) and T2 (bottom) MRF maps of three ADPKD patients ranging from the lowest cyst burden (left) to the highest cyst burden (right). The images shown are cropped to the kidneys from the original image and are in units of milliseconds (ms).

5.2.2 MRF acquisition and imaging parameters

The MRF images acquired are based on acquisition parameters published by MacAskill et. al [27], a rapid kidney MRF technique at 3.0 T for rapid and simultaneous quantification of multiple tissue parameters in a single scan. This 15-second kidney MRF acquisition was designed with 12 acquisition segments, a range of low flip angles (5°–12°) to limit the impact of B_1 heterogeneities, multiple magnetization preparation schema (T1, T2, and fat suppression), and an undersampled spiral trajectory (Figure 5.2). Each of the 12 segments consisted of different magnetization preparations, and a total of 144 consecutive MRF readouts were acquired after each preparation module. The two-dimensional kidney MRF technique implements parameters from previous brain and heart MRF acquisition schemes [101, 102]. The kidney MRF dictionary was generated using Bloch simulations with the MRF acquisition parameters and incorporated T1 and T2 values; each image voxel of the T1 and T2 maps was generated using the best-matching entry in the dictionary with the acquired MRF signal evolution profile of the acquired MRF data.

Coronal T1 and T2 maps were acquired on 3.0 T MRI scanners. MRF imaging parameters were field of views between 400x400 mm-450x450 mm, matrix sizes between 256x256-288x288, image slice thickness of 5.0 mm, and a pixel size of 1.5625x1.5625 mm. The MRF data for the healthy controls were acquired on the central sections of the kidneys and did not include full kidney volumes.

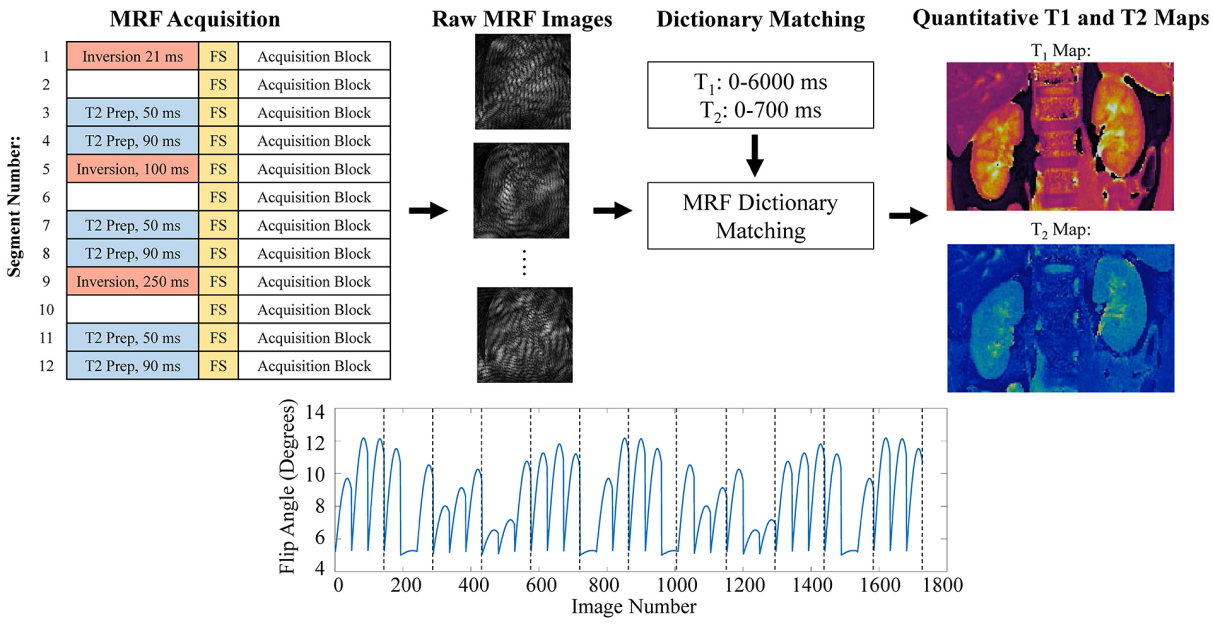

Figure 5.2: Schematic of the MRF acquisition and postprocessing methodology. The kidney MRF acquisition combines 12 acquisition segments with each having a magnetization preparation (T1, T2, or no preparation), fat suppression (FS), and 144 fast imaging with steady-state free precession (FISP) MRF imaging readouts. The flip angle pattern over the 1728 total MRF data points is shown at the bottom. Vertical dashed lines within the flip angle profile indicate the 12 different segments. Reprinted with permission from MacAskill, et al. [27].

5.2.3 Segmentation

Manual kidney segmentations were completed for both the ADPKD and healthy control cohorts. For the ADPKD cohort, kidney segmentations were completed by two segmenters from different imaging sites; the two segmenters were investigators familiar with kidney anatomy. All kidneys were manually segmented from T1 maps; the same segmentations were applied to T2 for downstream feature extraction because T1 and T2 maps are inherently co-registered. To remove cysts from the ADPKD kidney, three methods were used: (1) a fixed threshold of 2160 ms from T1 maps, (2) an adaptive manual threshold on a per-image basis for each patient from their T1 maps, and (3) an open-source fuzzy c-means (FCM) clustering

algorithm [103]. The fixed threshold of 2160 ms was based on previous work at Case Western Reserve University. The FCM algorithm clustering process analyzes the histogram of image intensities rather than on the raw image data [103].

This work used kidney and non-cystic kidney segmentations from (1) all available images and (2) ±1 from the largest coronal image of the left and right kidney (total of three central kidney sections per kidney). The ability to differentiate ADPKD from a healthy cohort using the central sections of the kidney will be investigated in comparison to using all available images; this approach aims to reduce the time required for segmentation and feature extraction. The Dice similarity coefficient (DSC) was used to (1) assess similarity between the manual kidney segmentations from two segmenters from different imaging sites (the University of Chicago and Case Western Reserve University) and (2) assess similarity of the subsequent non-cystic kidney parenchyma masks from these manual kidney segmentations using the three cyst removal methods: fixed threshold, adaptive manual threshold, and FCM. DSC evaluates the agreement or overlap (0 being no overlap and 1 being perfect overlap).

5.2.4 Feature extraction and classification

Pyradiomics [61] was used for feature extraction on the images, and a fixed bin number discretization method was implemented using 64 bins for gray-level discretization. Features were extracted from the non-cystic kidney parenchyma and the entire kidney of the ADPKD cohort and from the entire kidney from the healthy controls cohort. Fourty-two features were extracted per kidney: first-order (18) and gray-level co-occurrence matrix (GLCM) (24). Radiomic features extracted from images, i, were combined using a weighted-average for the left and right kidney individually (introduced in Chapter 4):

$$RF_{\text{weighted average}} = \frac{\sum_i RF_{slice,i} \cdot \text{weights}_i}{\sum_i \text{weights}_i}. \tag{5.1}$$

The weights were calculated by dividing the number of voxels in a kidney segmentation of a given image by the total number of voxels across all of the kidney segmentations of the MRI sections, i:

$$\text{weights}_i = \frac{\text{Number of voxels}_{seg,i}}{\sum_i \text{Number of voxels}_{seg,i}}. \tag{5.2}$$

Finally, the left and right kidney features were averaged for a representative feature value. Radiomic features were calculated from both the non-cystic kidney parenchyma and entire kidney for each patient.

The first-order means of the T1 and T2 maps were first investigated as previous research has proposed mean T2 as a potential imaging biomarker in ADPKD [20, 24]. The mean T1 and T2 of the ADPKD and healthy cohort was compared using the Wilcoxon rank sum test for significance for both the non-cystic parenchyma and entire kidney (including cysts) [68]. The relationship of mean T1 and mean T2 values between the entire kidney and non-cystic kidney parenchyma and the mean T1 and mean T2 with cyst fraction was assessed using the Pearson correlation coefficient; additionally, the relationship between cyst fraction and imaging biomarkers, BSA-TKV and htTKV, was investigated. The cyst fraction for each ADPKD patient was determined as follows: (1) dividing the total number of voxels in the non-cystic kidney segmentations by the total number of voxels in the kidney segmentations, yielding the non-cystic kidney parenchyma fraction, and (2) subtracting this fraction from one to derive the cyst fraction.

Two radiomic feature schemes were used to investigate the ability of features to classify the ADPKD cohort from the healthy controls cohort: (1) using just the mean T1 and T2 and (2) merging radiomic features (first order and GLCM texture features) for classification. A linear discriminant analysis (LDA) classifier using a leave-one-patient-out training/testing approach was implemented. A leave-one-patient-out training/testing approach performs feature selection each iteration when one patient's data is left out of the training partition (15 iterations total). For each training partition the top-two uncorrelated (Pearson correlation

threshold of 0.7 with any other selected feature) performing features from the 42 total extracted features were determined using the area under the receiver operating characteristic curve (AUC). Performance of the LDA was assessed using ROC analysis with the AUC used as the figure of merit.

In comparing mean T1 and mean T2 values and using radiomic features for classification, the adaptive manual threshold cyst removal method for the non-cystic kidney parenchyma was used.

5.3 Results

5.3.1 Segmentation comparisons

The DSC calculated in comparing manual kidney segmentations and cyst removal methods are listed in Table 5.2. The mean DSC increased when restricting the segmentations to the central kidney sections. The mean DSC [95% CI] of the manual kidney segmentations for all images was 0.857 [0.835, 0.880] versus 0.911 [0.901, 0.922] for the central kidney sections (Figures 5.3-5.4).

Table 5.3 is the DSC of cyst removal methods of one imaging site (the University of Chicago). The fixed and adaptive manual threshold methods had higher DSC than FCM: 0.969-0.975 and 0.836-0.855. Figure 5.5 show an example of the FCM cyst removal segmentation for ADPKD patients with both a high and low cyst burden, as well as healthy control kidneys; this example shows that when the cyst burden was low, the FCM method clustered the signal intensities with respect to the kidney medulla and cortex.

In implementing the adaptive manual threshold for cyst removal, the mean T1 threshold was 2218 ms with a minimum value of 1920 ms and a maximum value of 2820 ms (Figure 5.6). The percent difference between the mean T1 threshold value using the adaptive manual threshold and the fixed threshold of 2160 ms was 2.65%.

Table 5.2: Mean Dice similarity coefficient (DSC) [95% CI] of manual kidney segmentations and subsequent non-cystic kidney parenchyma segmentations between two different segmenters.

Segmentation	Method	Mean DSC All images	Mean DSC Central images
Kidney	Manual	0.857 [0.835, 0.880]	0.911 [0.901, 0.922]
Non-cystic	Fixed threshold	0.856 [0.833, 0.879]	0.907 [0.896, 0.918]
Non-cystic	Adaptive manual threshold	0.856 [0.833, 0.879]	0.909 [0.897, 0.920]
Non-cystic	FCM	0.781 [0.753, 0.809]	0.853 [0.834, 0.872]

Table 5.3: Mean Dice similarity coefficient (DSC) [95% CI] of cyst removal methods to obtain the non-cystic kidney parenchyma masks using kidney segmentations from one segmenter (the University of Chicago).

Comparison	Mean DSC All images	Mean DSC Central images
FCM vs. Fixed	0.842 [0.820, 0.863]	0.855 [0.834, 0.876]
FCM vs. Adaptive manual	0.836 [0.814, 0.858]	0.846 [0.823, 0.870]
Adaptive manual vs. Fixed	0.975 [0.971, 0.978]	0.969 [0.963, 0.976]

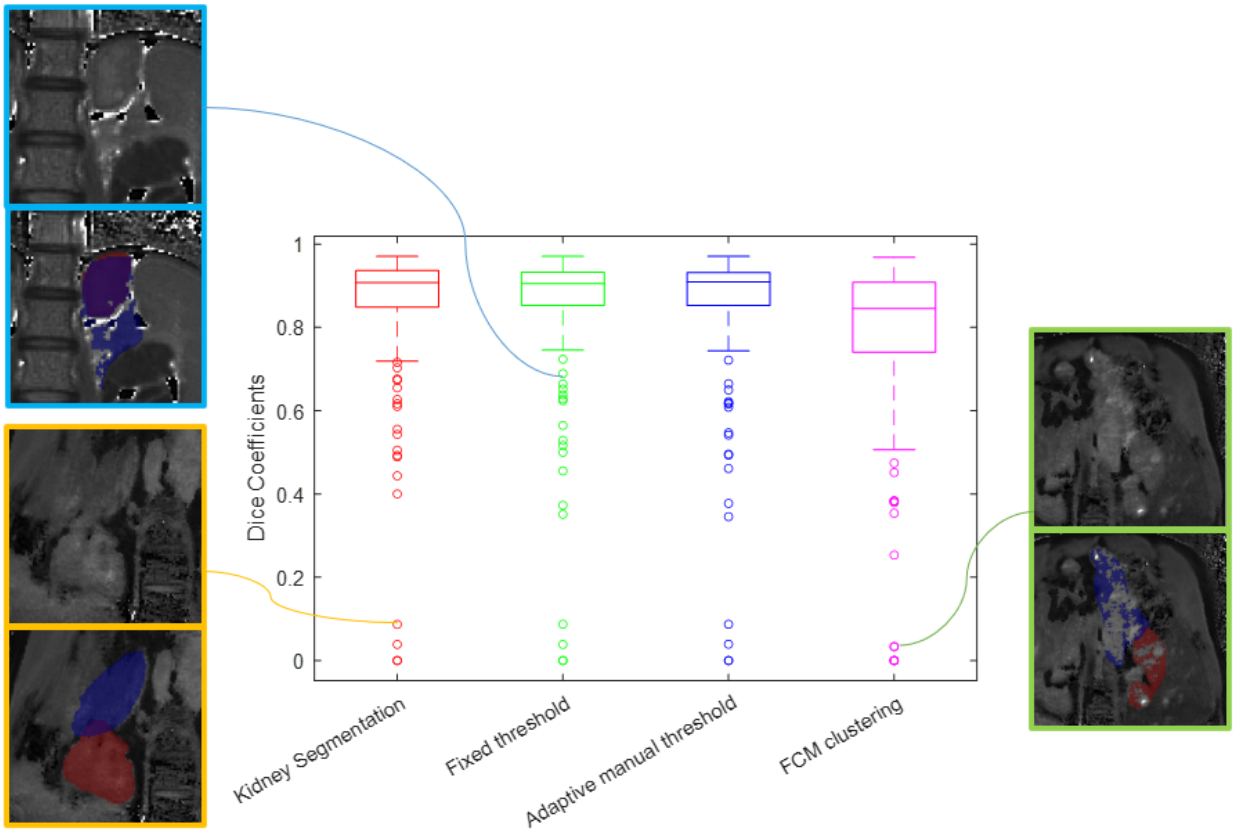

Figure 5.3: Box and whisker plots of Dice similarity coefficient (DSC) between two segmenters using segmentations of all available images of the left and right kidney. The images show the red and blue corresponding to the segmentations, and purple is the overlap between the two segmentations. The box represents the interquartile range, with the central line indicating the median value; the whiskers extend to the minimum and maximum values, while outliers are represented as individual data points.

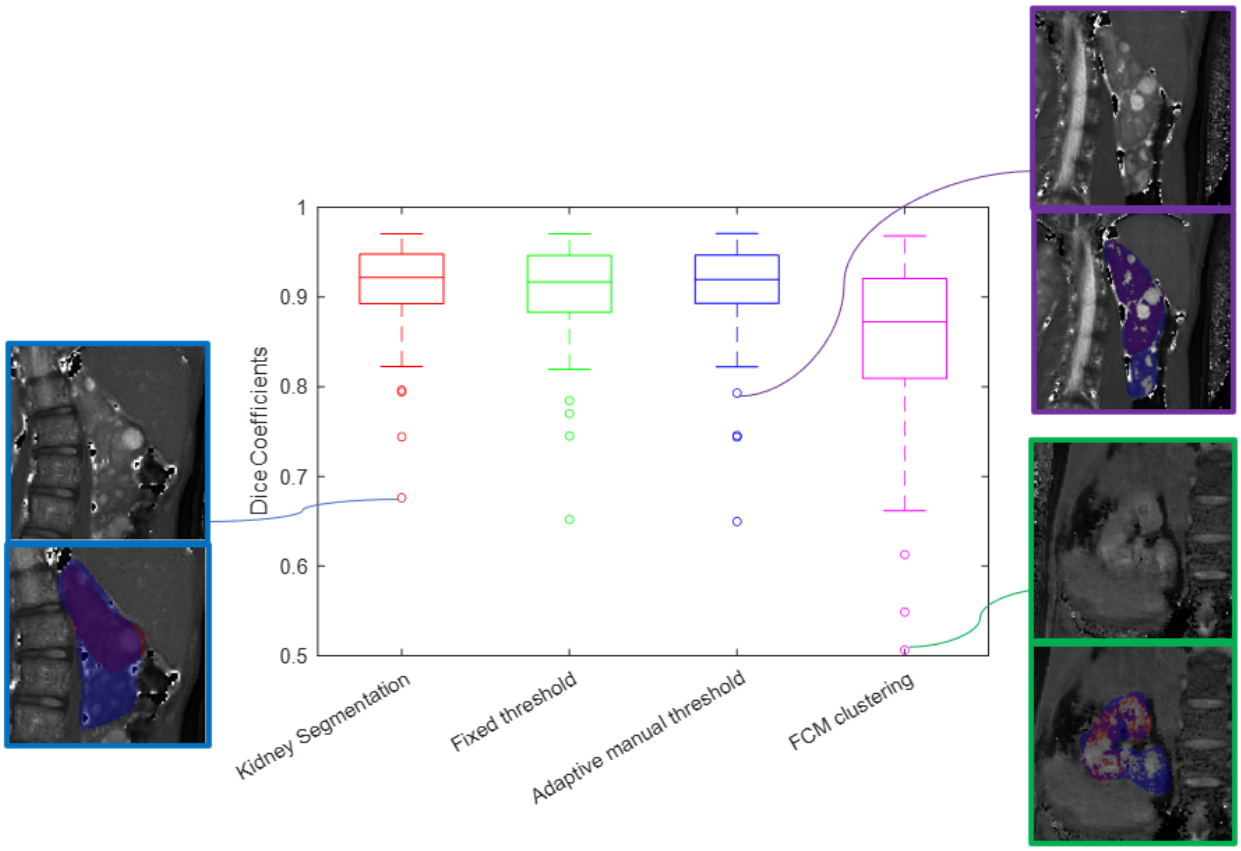

Figure 5.4: Box and whisker plots of Dice similarity coefficient (DSC) between two segmenters using three central sections of the left and right kidney. The images show the red and blue corresponding to the segmentation, and purple is the overlap. The box represents the interquartile range, with the central line indicating the median value; the whiskers extend to the minimum and maximum values, while outliers are represented as individual data points.

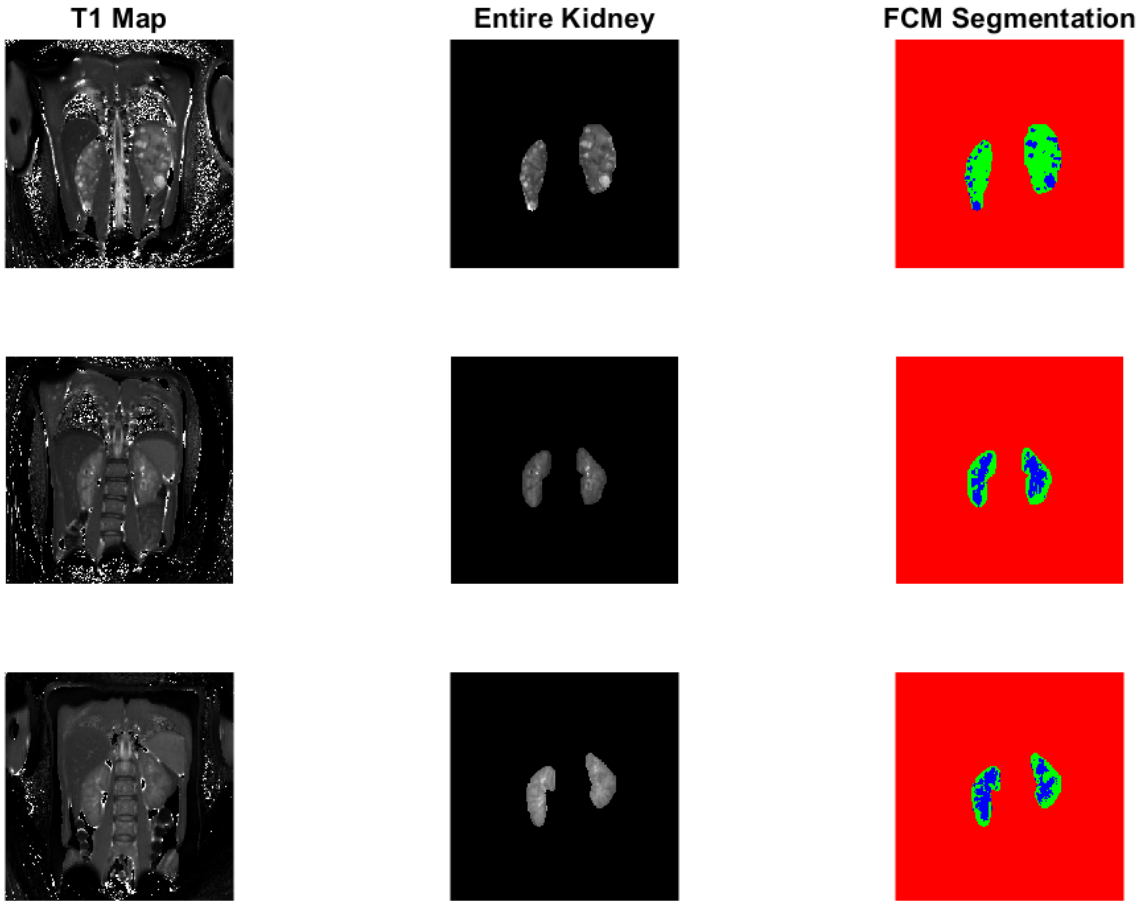

Figure 5.5: The fuzzy c-means (FCM) cyst removal from the entire kidney from an ADPKD patient with a high total cyst burden (37.99%) (top), an ADPKD patient with a low total cyst burden (1.98%) (middle), and a healthy control (bottom). Note that the FCM cyst removal method segmented the kidney medulla and cortex of the ADPKD patient with a low cyst burden and the healthy control. The colors represent the number of clusters (n=3) in the FCM algorithm: background (red), kidney (green), and cyst (blue).

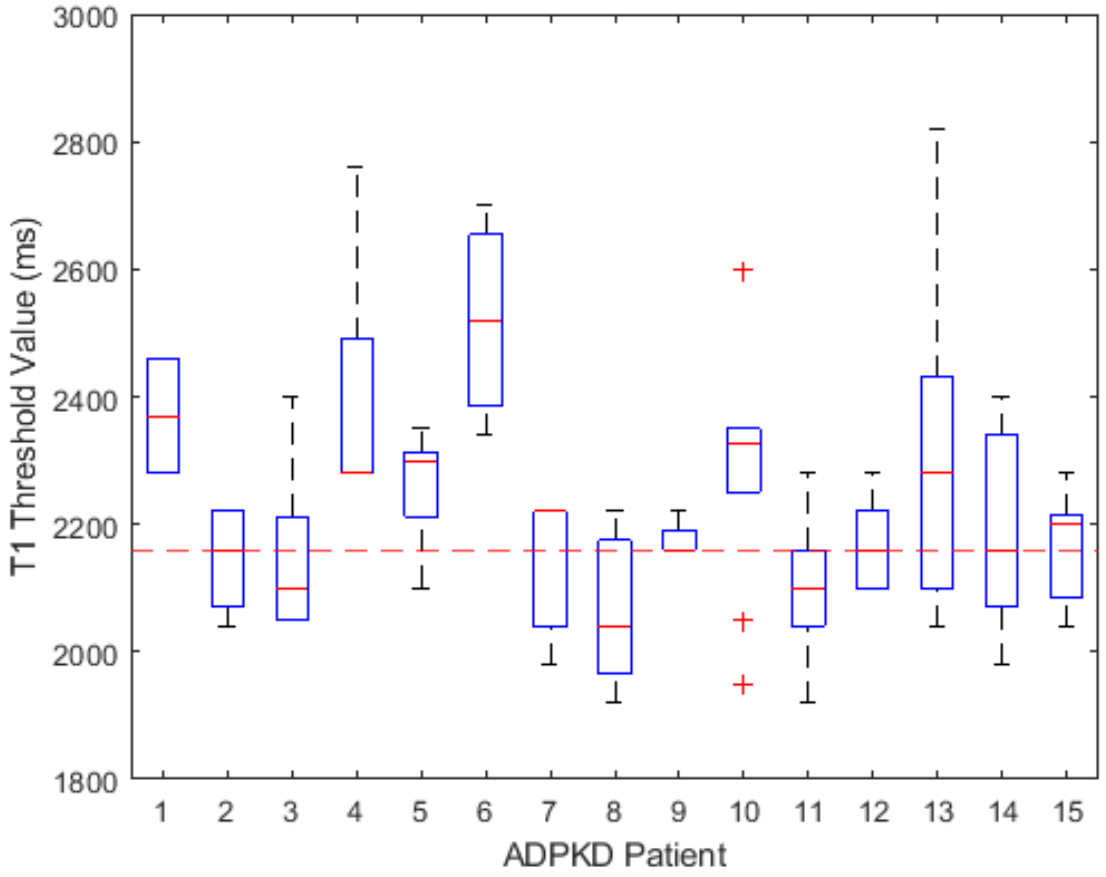

Figure 5.6: Box and whisker plots of the adaptive manual T1 threshold value chosen on a per image basis for each patient. The box and whisker plots of each patient are images that presented macroscopic cysts and subsequently used for cyst removal. The red dashed line represents the fixed threshold value of 2160 ms for cyst removal. The box represents the interquartile range, with the central line indicating the median value; the whiskers extend to the minimum and maximum values, while outliers are represented as individual data points.

5.3.2 Mean T1 and T2

Table 5.4 contains the mean T1 and T2 values of the segmented kidneys from the healthy control cohort and of the kidney and non-cystic kidney parenchyma of the ADPKD cohort. The mean T1 and mean T2 values of the kidney and non-cystic kidney segmentation methods were significantly different from healthy controls, except for the FCM cyst removal method

from the T1 maps. Figure 5.7 and Figure 5.8 contain the box and whisker plots of T1 and T2 mean values, respectively, from using all of the images and the central kidney sections for the ADPKD patients. Figure 5.9 and Figure 5.10 are the scatter plots of T2 values versus T1 values for the healthy and ADPKD cohorts using all images and central kidney sections, respectively; the scatter plots contain paired numbers that match the non-cystic values to the entire kidney values from the same ADPKD patient.

Table 5.4: Mean T1 and T2 [95% CI] of manual kidney segmentations and non-cystic kidney segmentations using the cyst removal methods. A $p < 0.05$ was significant from the healthy control cohort using Wilcoxon rank sum test. Asterisks indicate statistically significant differences (after Bonferroni-Holm correction for 8 multiple comparisons).

MRF Map	Healthy control kidney	ADPKD kidney	Fixed threshold non-cystic	Adaptive manual threshold non-cystic	FCM non-cystic
T1 (All)	-	1869.06	1663.48	1677.36	1633.12
		[1770.26, 1967.87]	[1619.07, 1707.89]	[1629.82, 1724.91]	[1555.99, 1710.26]
		($p \leq 0.001$)*	($p = 0.002$)*	($p = 0.001$)*	($p = 0.157$)
T1 (Central)	1544.50	1909.41	1680.47	1699.74	1656.07
	[1501.62, 1587.37]	[1803.35, 2015.48]	[1634.82, 1726.13]	[1647.23, 1752.25]	[1576.57, 1735.57]
		($p \leq 0.001$)*	($p = 0.002$)*	($p \leq 0.001$)*	($p = 0.063$)
T2 (All)	-	89.72	77.83	78.07	78.64
		[81.99, 97.44]	[75.04, 80.62]	[75.28, 80.87]	[75.38, 81.90]
		($p \leq 0.001$)*	($p = 0.007$)*	($p = 0.004$)*	($p = 0.007$)*
T2 (Central)	72.15	90.82	78.14	78.44	79.07
	[68.93, 75.36]	[83.37, 98.27]	[75.34, 80.95]	[75.61, 81.27]	[75.77, 82.36]
		($p \leq 0.001$)*	($p = 0.005$)*	($p = 0.004$)*	($p = 0.005$)*

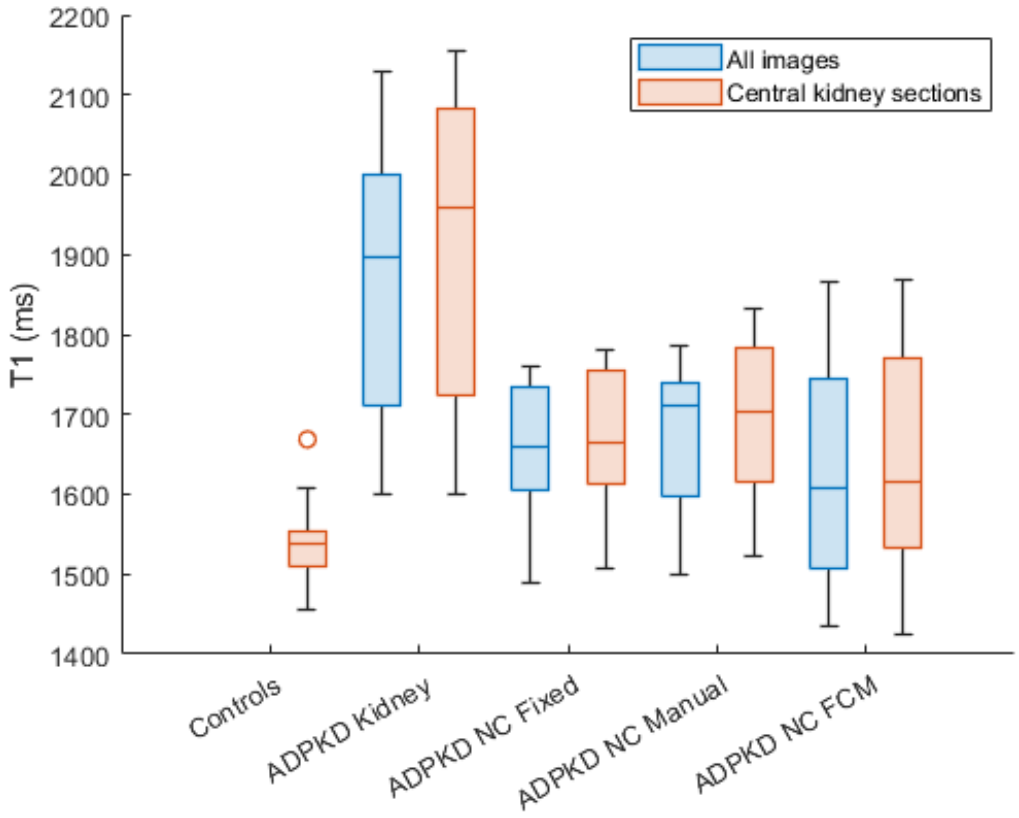

Figure 5.7: Box and whisker plots comparing the mean T1 values obtained from the entire kidney and non-cystic (NC) kidney parenchyma of the ADPKD cohort, using all available images and central kidney sections, alongside the healthy controls' kidneys. The cyst removal methods from the ADPKD kidney were (1) a fixed threshold of 2160 ms from T1 maps, (2) an adaptive manual threshold on a per-image basis for each patient from their T1 maps, and (3) a fuzzy c-means (FCM) clustering algorithm. The box represents the interquartile range, with the central line indicating the median value; the whiskers extend to the minimum and maximum values, while outliers are represented as individual data points.

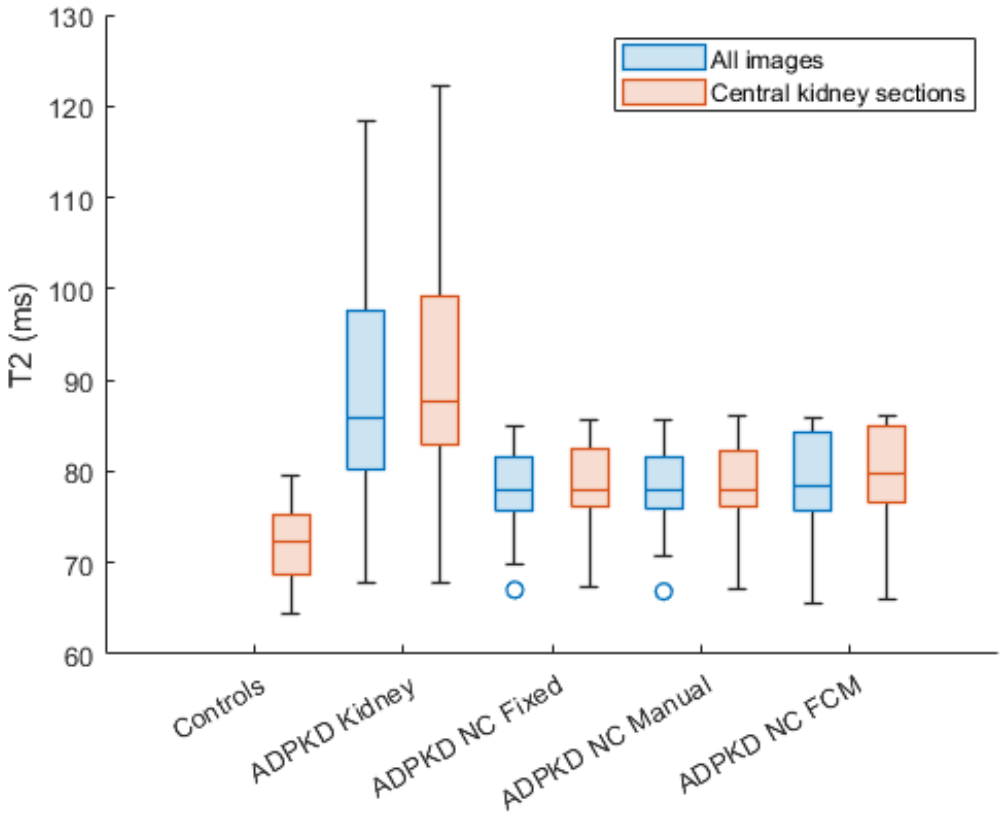

Figure 5.8: Box and whisker plots comparing the mean T2 values obtained from the entire kidney and non-cystic (NC) kidney parenchyma of the ADPKD cohort, using all available images and central kidney sections, alongside the healthy controls' kidneys. The cyst removal methods from the ADPKD kidney were (1) a fixed threshold of 2160 ms from T1 maps, (2) an adaptive manual threshold on a per-image basis for each patient from their T1 maps, and (3) a fuzzy c-means (FCM) clustering algorithm. The box represents the interquartile range, with the central line indicating the median value; the whiskers extend to the minimum and maximum values, while outliers are represented as individual data points.

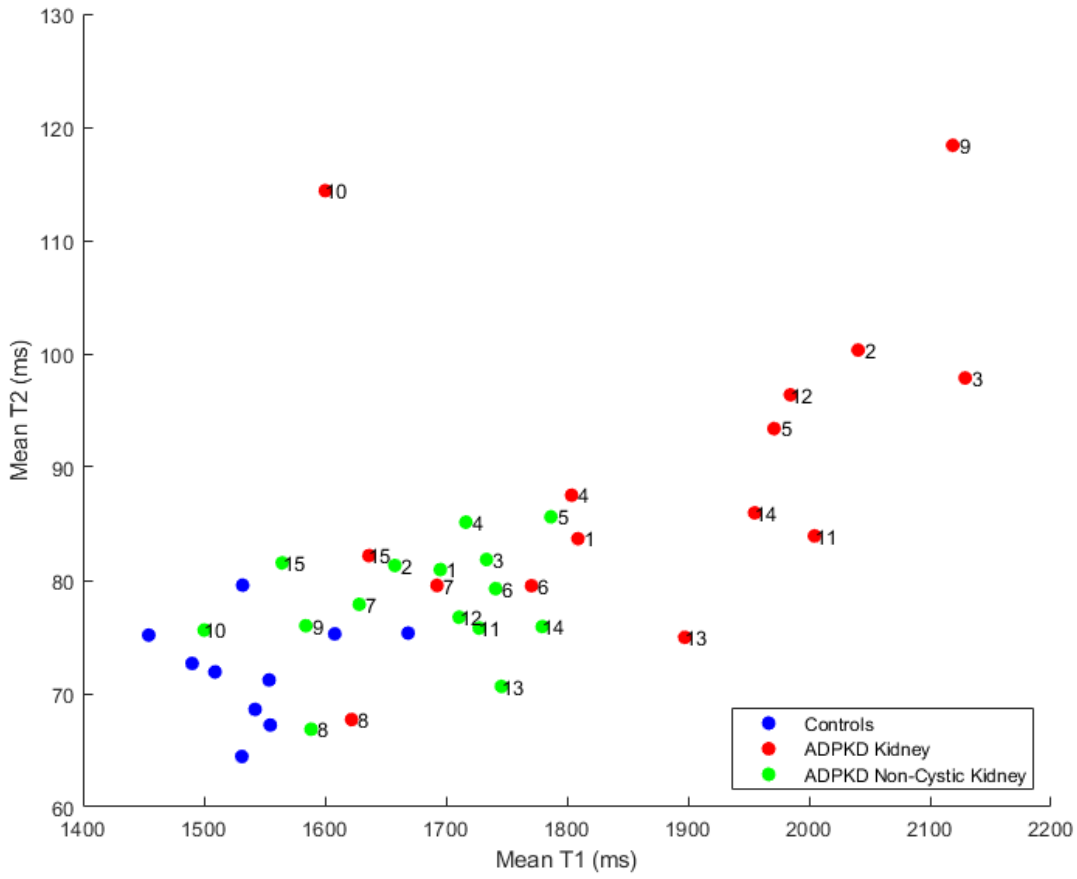

Figure 5.9: Mean T2 versus T1 of the entire kidney and non-cystic kidney parenchyma compared with healthy controls using all the images from the ADPKD cohort. Each red data point (kidney) corresponds with a green data point (non-cystic kidney parenchyma) of the same patient.

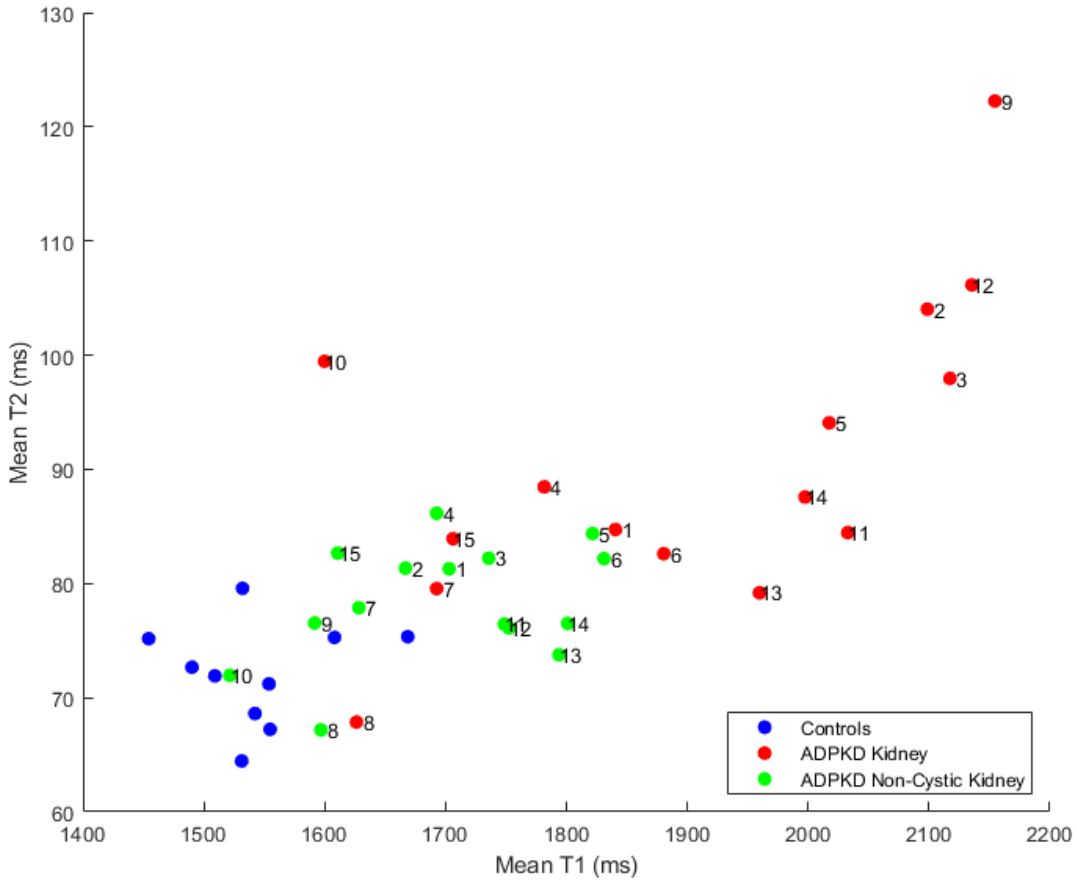

Figure 5.10: Mean T2 versus T1 of the entire kidney and non-cystic kidney parenchyma compared with healthy controls using the central kidney sections from the ADPKD cohort. Each red data point (kidney) corresponds with a green data point (non-cystic kidney parenchyma) of the same patient.

Pearson correlation coefficient between the mean T1 values of the kidney and non-cystic kidney parenchyma using all images and central kidney sections were 0.54 and 0.53, respectively, and the the mean T2 of the kidney and non-cystic kidney parenchyma were 0.28 and 0.18, respectively (Figure 5.11-Figure 5.12). Pearson correlation coefficient between cyst fraction and mean T1 values from the entire kidney using all images and central kidney sections were 0.91 and 0.84, respectively (Figure 5.13); Pearson correlation coefficient between

cyst fraction and mean T1 values from the non-cystic kidney parenchyma using all images and central kidney sections were 0.28 and 0.18, respectively (Figure 5.14). Pearson correlation coefficient between cyst fraction and mean T2 values from the entire kidney using all images and central kidney sections were 0.49 and 0.60, respectively (Figure 5.13); Pearson correlation coefficient between cyst fraction and mean T2 values from the non-cystic kidney parenchyma using all images and central kidney sections were 0.26 and 0.13, respectively (Figure 5.14).

The Pearson correlation coefficient between cyst fraction and htTKV was 0.51, and the Pearson correlation coefficient between cyst fraction and BSA-TKV was 0.61 (Figure 5.17-Figure 5.18).

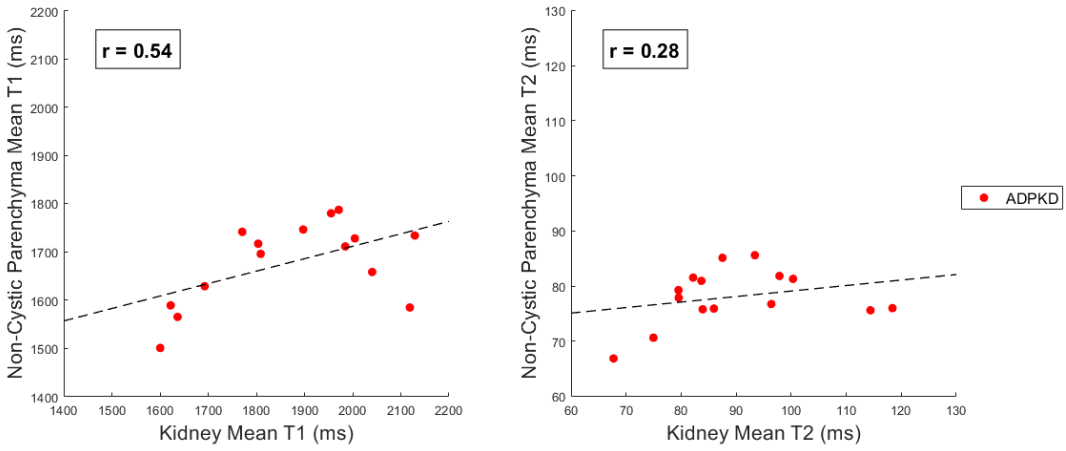

Figure 5.11: Mean T1 (left) and mean T2 (right) of the non-cystic kidney parenchyma versus the entire kidney using all the images from the ADPKD cohort, shown with the fitted Pearson line and Pearson correlation coefficient.

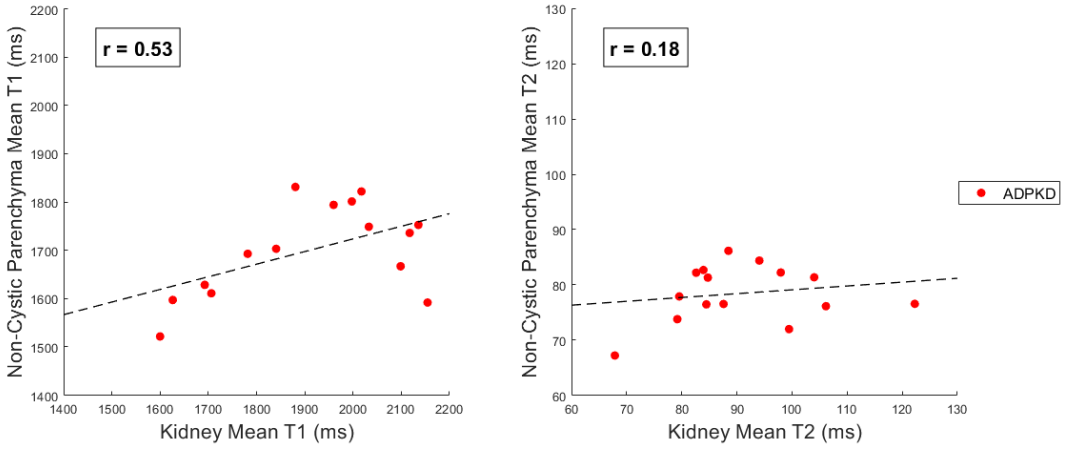

Figure 5.12: Mean T1 (left) and mean T2 (right) of the non-cystic kidney parenchyma versus the entire kidney using the central kidney sections from the ADPKD cohort, shown with the fitted Pearson line and Pearson correlation coefficient.

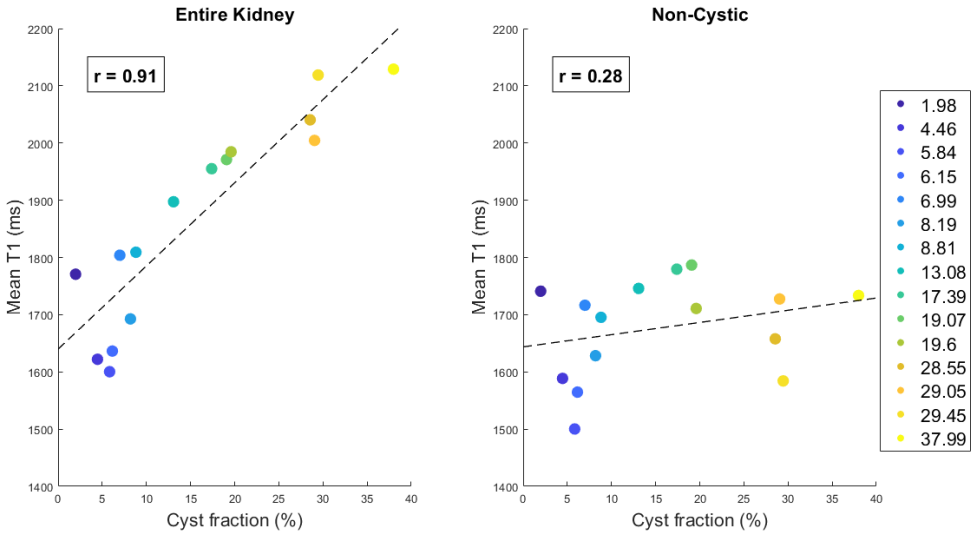

Figure 5.13: Mean T1 of the entire kidney (left) and non-cystic kidney parenchyma (right) with respect to cyst volumes using all images from the ADPKD cohort, shown with the fitted Pearson line and Pearson correlation coefficient. The legend contains the lowest to highest cyst fractions values.

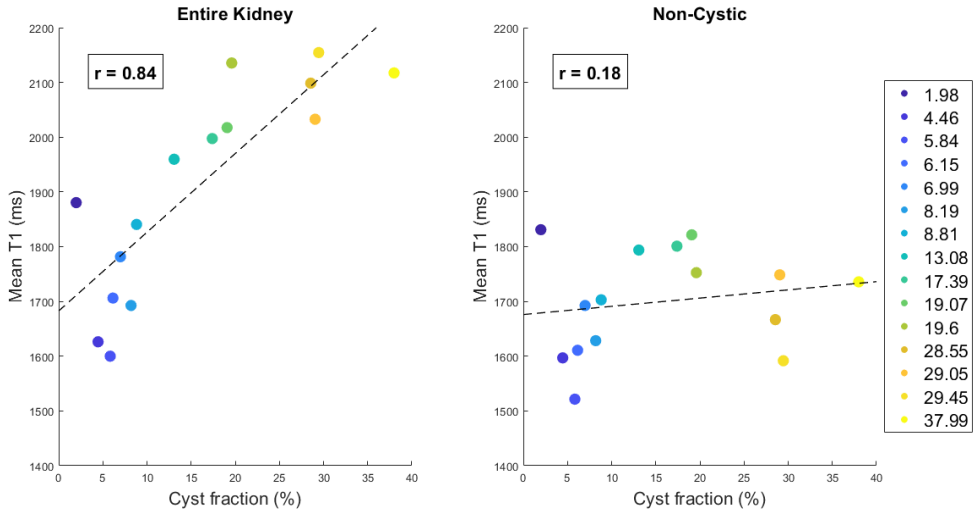

Figure 5.14: Mean T1 of the entire kidney (left) and non-cystic kidney parenchyma (right) with respect to cyst volumes using the central kidney sections from the ADPKD cohort, shown with the fitted Pearson line and Pearson correlation coefficient. The legend contains the lowest to highest cyst fraction values.

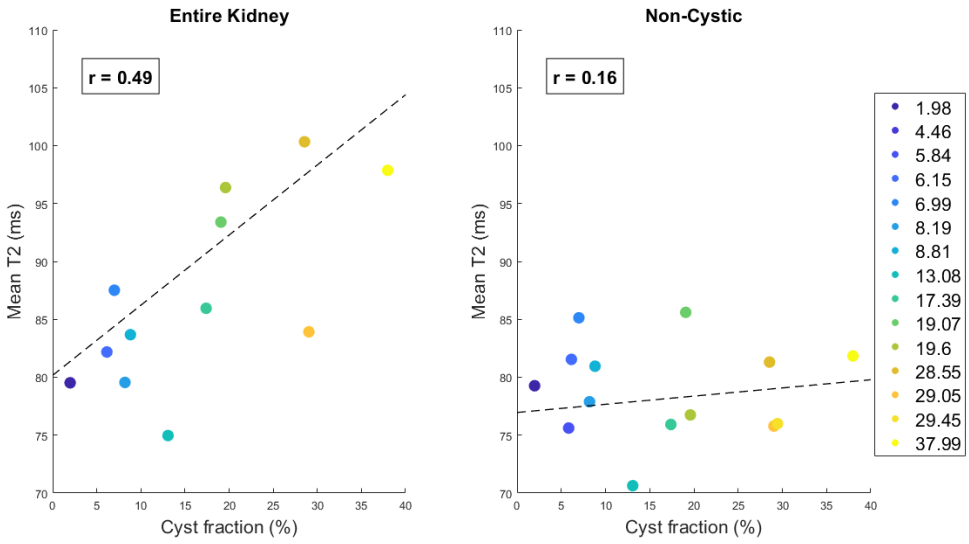

Figure 5.15: Mean T2 of the entire kidney (left) and non-cystic kidney parenchyma (right) with respect to cyst volumes using all images from the ADPKD cohort, shown with the fitted Pearson line and Pearson correlation coefficient. The legend contains the lowest to highest cyst fraction values.

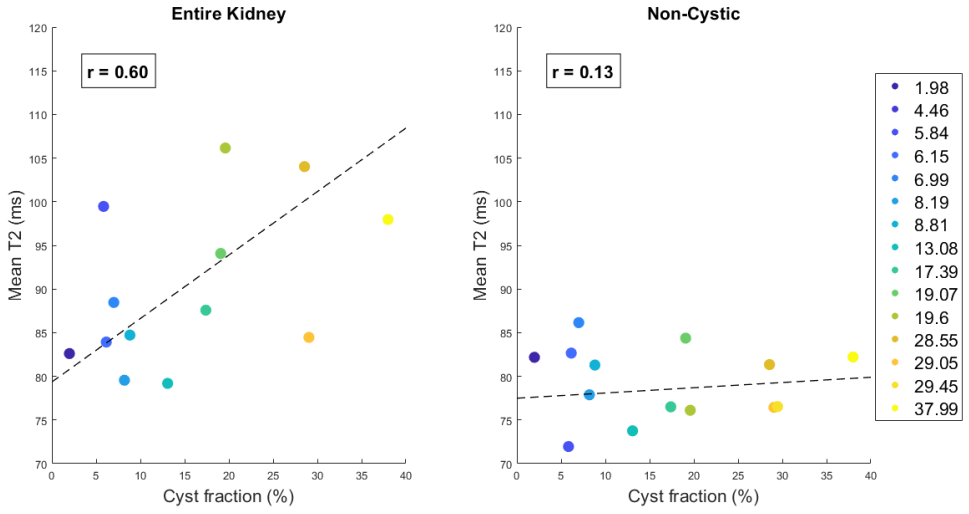

Figure 5.16: Mean T2 of the entire kidney (left) and non-cystic kidney parenchyma (right) with respect to cyst volumes using the central kidney sections from the ADPKD cohort, shown with the fitted Pearson line and Pearson correlation coefficient. The legend contains the lowest to highest cyst fraction values.

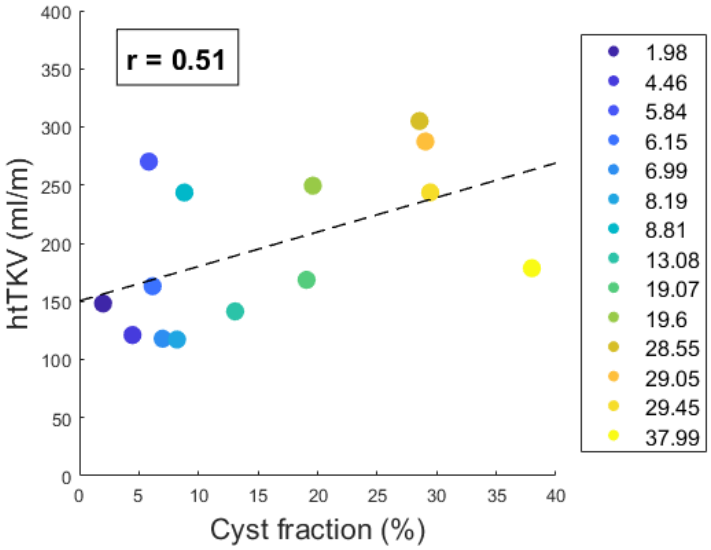

Figure 5.17: Height-corrected total kidney volume (htTKV) with respect to cyst volumes, shown with the fitted Pearson line and Pearson correlation coefficient. 14 of the 15 patients had htTKV data available. The legend contains the lowest to highest cyst fraction values.

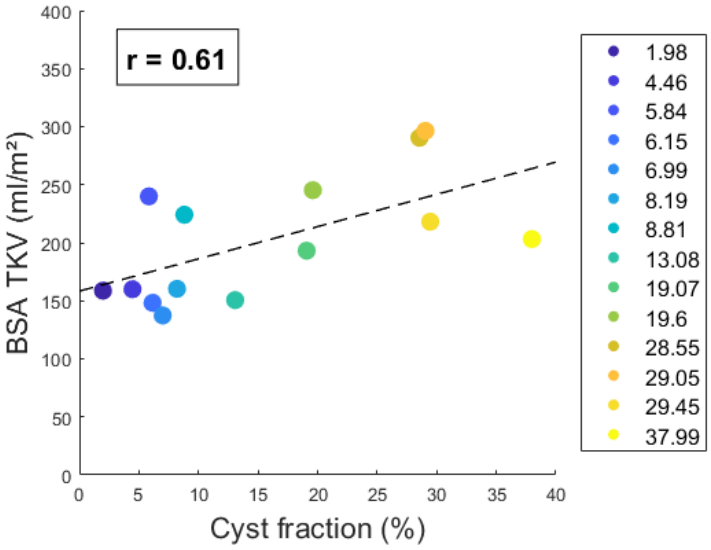

Figure 5.18: Body surface area-total kidney volume (BSA-TKV) with respect to cyst volumes, shown with the fitted Pearson line and Pearson correlation coefficient. 14 of the 15 patients had BSA-TKV data available. The legend contains the lowest to highest cyst fraction values.

5.3.3 Classification of ADPKD patients vs. a healthy cohort

Classification schemes employed an LDA classifier and a leave-one-patient-out training/testing approach, incorporating two feature selection techniques for distinguishing ADPKD from the healthy controls cohort: (1) using just mean T1 and T2 for classification (2) using the top-two performing radiomics features for classification. The AUC values are listed in Table 5.5.

The AUC values for mean T1 for all images was 0.94 and for the central kidney sections was 0.95; the AUC values for mean T2 for all images was 0.89 and for the central kidney sections was 0.92. The AUC values across all segmentation methods of non-cystic kidney parenchyma segmentation using mean T1 was 0.63-0.88 and for mean T2 was 0.73-0.80. The FCM method for cyst removal from T1 maps resulted in lower AUC values (AUC=0.63-0.69) than the fixed threshold and adaptive manual threshold methods (AUC=0.84-0.88).

There was an increase in AUC values in merging the top-two performing, uncorrelated radiomic features extracted from the non-cystic kidney parenchyma of T1 maps versus mean T1 alone; the AUC values ranged from 0.99 to 1.0 across segmentation methods. The 95% confidence interval (CI) ranged from 0.84-1.0 using features extracted from the non-cystic kidney parenchyma in separating ADPKD and healthy controls, which was smaller than the 95% CI using the mean T1 value that ranged from 0.58-0.98. The AUC values using the top-two performing, uncorrelated radiomic features extracted from the non-cystic kidney parenchyma of T2 maps ranged from 0.69-0.86 across segmentation methods.

The radiomic features selected during feature selection of the training partition, which were merged to form the top-two performing radiomic features for classification, were identified. The features selected from the entire kidney of T1 maps are listed in Figure 5.19, and the features selected from the non-cystic kidney parenchyma of T1 maps are listed in Figure 5.20; Figure 5.21 depicts two representative radiomic features selected from the list of features for distinguishing the non-cystic kidney parenchyma of ADPKD patients from

healthy controls using T1 maps. The features selected from the entire kidney of T2 maps are listed in Figure 5.22, and the features extracted from the non-cystic kidney parenchyma of T1 maps are listed in Figure 5.23.

Table 5.5: Area under the receiver operating characteristic curve (AUC) values for differentiating ADPKD children and young adults from a healthy adult cohort using radiomic features from the non-cystic kidney parenchyma and entire kidney from T1 and T2 quantitative MRF maps. The mean T1 and mean T2 values and the top-two performing radiomic features were used for classification.

MRF Map	Entire kidney	Fixed threshold non-cystic	Adaptive manual threshold non-cystic	FCM non-cystic
T1				
Mean: All images	0.94	0.84	0.86	0.63
	[0.78, 0.99]	[0.58, 0.97]	[0.61, 0.97]	[0.35, 0.84]
Mean: Central images	0.95	0.87	0.88	0.69
	[0.76, 1.0]	[0.59, 0.98]	[0.65, 0.97]	[0.45, 0.89]
Top-2: All images	0.81	0.99	1.0	1.0
	[0.57, 0.94]	[0.84, 1.0]		
Top-2: Central images	0.83	0.99	1.0	1.0
	[0.61, 0.97]	[0.87, 1.0]		
T2				
Mean: All images	0.89	0.73	0.73	0.77
	[0.63, 0.99]	[0.49, 0.90]	[0.45, 0.90]	[0.49, 0.92]
Mean: Central images	0.92	0.76	0.80	0.80
	[0.63, 1.0]	[0.51, 0.91]	[0.54, 0.95]	[0.55, 0.95]
Top-2: All images	0.84	0.72	0.73	0.86
	[0.61, 0.96]	[0.49, 0.91]	[0.46, 0.92]	[0.57, 1.0]
Top-2: Central images	0.81	0.69	0.77	0.85
	[0.57, 0.98]	[0.44, 0.89]	[0.53, 0.93]	[0.58, 1.0]

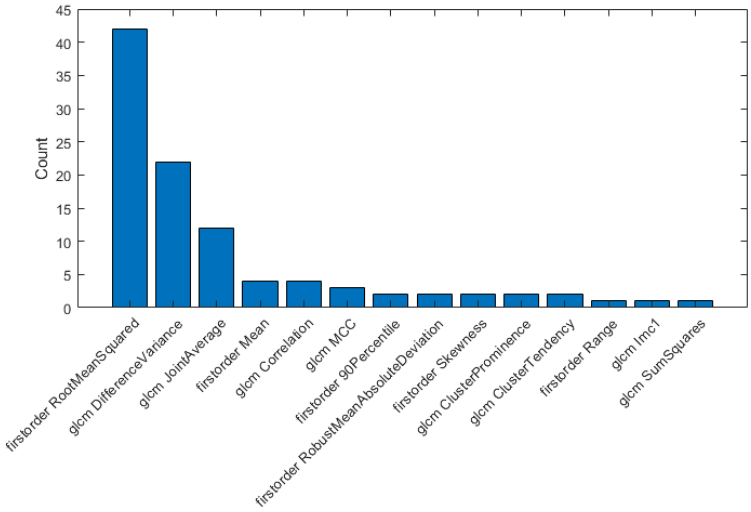

Figure 5.19: The radiomic features extracted from the entire kidney of T1 maps that were chosen during feature selection in merging radiomic features for classification. The counts refer to the number of times the radiomic feature was selected across training partitions of the leave-one-out approach.

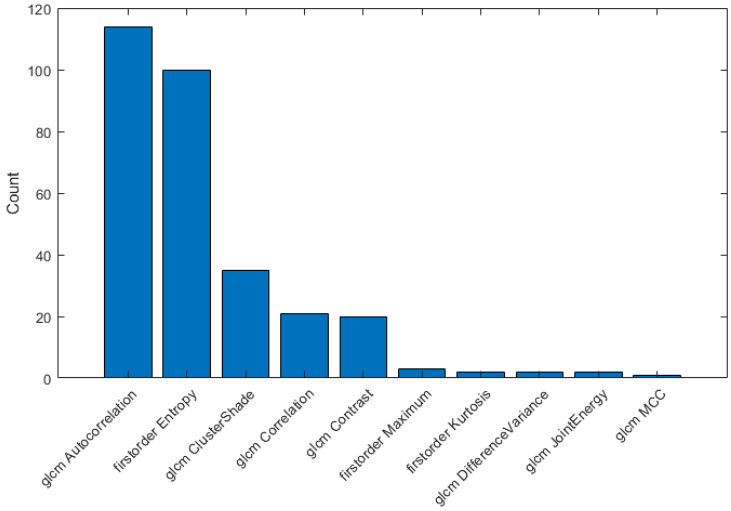

Figure 5.20: The radiomic features extracted from the non-cystic kidney parenchyma of T1 maps that were chosen during feature selection in merging radiomic features for classification across all segmentation methods. The counts refer to the number of times the radiomic feature was selected across training partitions of the leave-one-out approach.

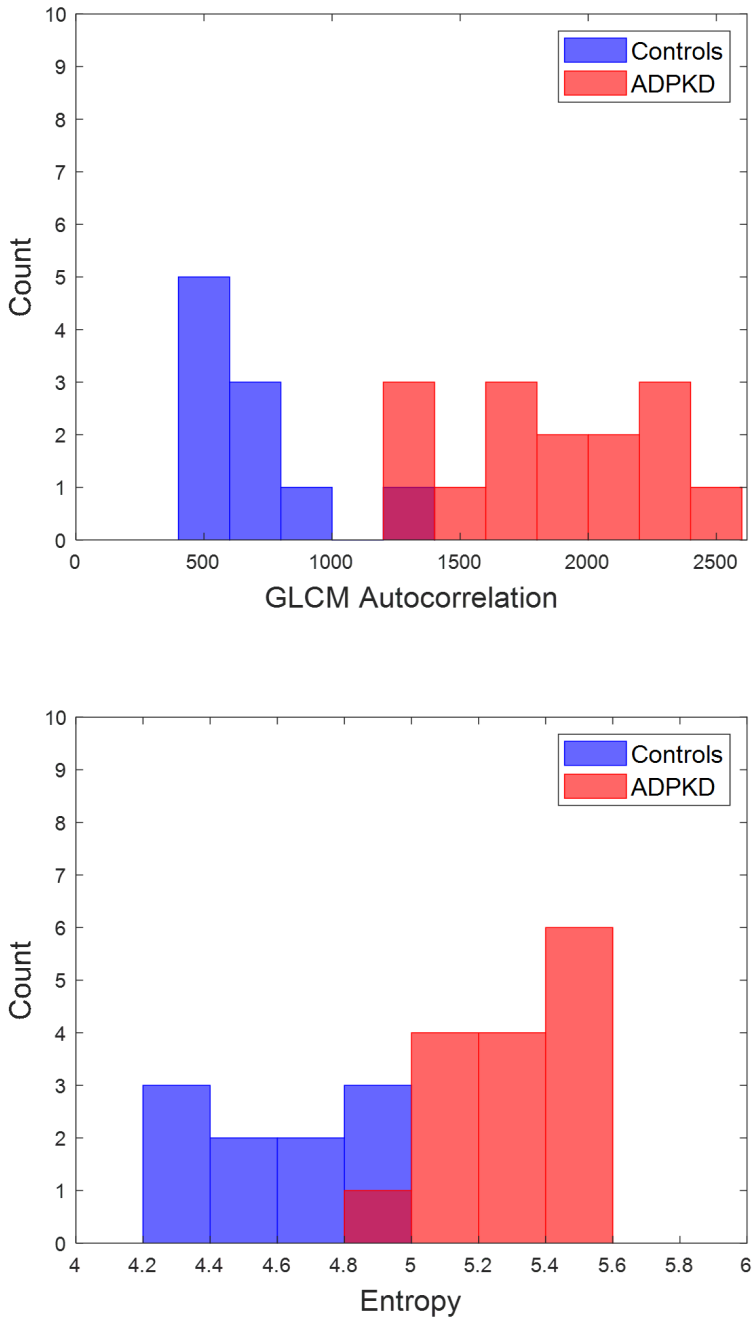

Figure 5.21: Histogram of radiomic features extracted from the non-cystic kidney parenchyma of the ADPKD cohort versus healthy controls using T1 maps (adaptive manual threshold cyst removal). GLCM autocorrelation measures the coarseness of texture (top) and entropy captures the randomness of signal intensities (bottom).

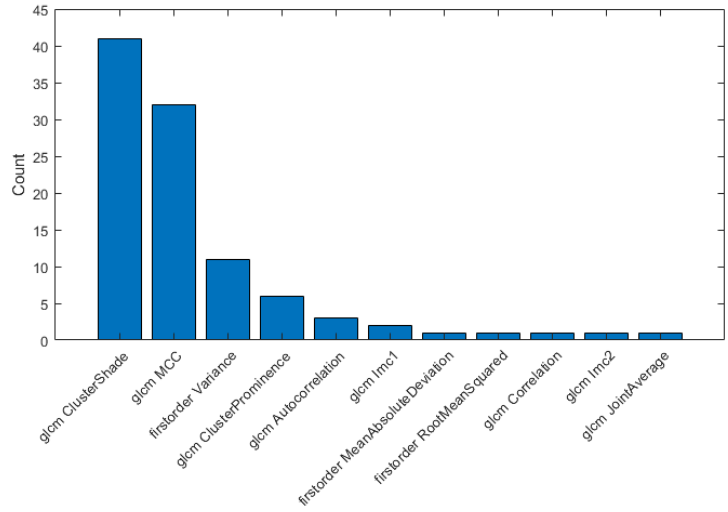

Figure 5.22: The radiomic features extracted from the entire kidney of T2 maps that were chosen during feature selection in merging radiomic features for classification. The counts refer to the number of times the radiomic feature was selected across training partitions of the leave-one-out approach.

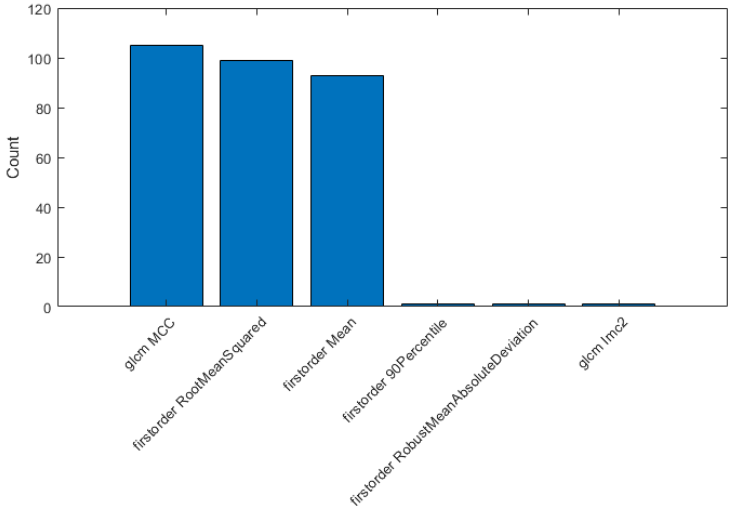

Figure 5.23: The radiomic features extracted from the non-cystic kidney parenchyma of T2 maps that were chosen during feature selection in merging radiomic features for classification across all segmentation methods. The counts refer to the number of times the radiomic feature was selected across training partitions of the leave-one-out approach.

5.4 Discussion

This exploratory aim investigated radiomic features extracted from MRF-acquired T1 and T2 maps from a cohort of ADPKD children and young adults and healthy controls. This is the first work to utilize radiomic features from MRF T1 and T2 maps to potentially capture early cystogenesis in ADPKD children and young adults using the non-cystic kidney parenchyma. Additionally, kidney segmentations were compared between two segmenters from different imaging sites, and their impact on subsequent non-cystic kidney parenchyma segmentation using three cyst removal methods was assessed.

There was a trend of higher DSC between two segmenters when restricting to the central kidney sections for segmentation, increasing the DSC from 0.857 [0.835, 0.880] to 0.911 [0.901, 0.922]. The decrease in DSC observed when using all images was attributed to the partial volume effects present in the anterior and posterior regions of the kidney in the T1 maps, which present a more challenging segmentation task. DSC of subsequent non-cystic kidney parenchyma segmentation using three cyst removal methods demonstrates the propagation of discrepancies observed in the manual kidney segmentations, particularly evident in the removal of cysts for the fixed and adaptive threshold methods. For the FCM method, the ability to segment cysts from the parenchyma depended on the cyst burden present in the image of the patient; the mean T1 from the non-cystic kidney parenchyma segmentation using FCM failed to show a statistically significant difference with healthy controls compared to the fixed and adaptive threshold methods. Future research could explore the correlation between cyst fraction and the efficacy of FCM in accurately removing cysts from the kidney.

The mean T1 and T2 of the ADPKD cohort were statistically significantly different from healthy controls in the entire kidney (including cysts) and non-cystic parenchyma using the fixed and adaptive threshold cyst removal methods for all images and central kidney sections. The correlation of mean T1, mean T2, htTKV, and BSA-TKV to cyst fraction is of interest because patients at higher risk for end-stage kidney disease (ESKD) are associated with

higher cyst fractions, and an inverse correlation between cyst volume and glomerular filtration rate (GFR) has been observed in adults with ADPKD [36, 100]. A strong correlation was shown between the cyst fraction and mean T1 (0.91), which was higher than that of the imaging biomarkers of htTKV (0.51) and BSA-TKV (0.61). Siedek et al. [24] analyzed T2 maps of ADPKD and healthy controls, and their findings indicate that the mean T2 value of the non-cystic kidney parenchyma could potentially serve as a novel imaging biomarker for ADPKD. The results presented in the study showed a strong correlation of mean T2 of the non-cystic kidney parenchyma to cyst fraction (0.77), which was similar to that of the correlation of the mean T2 of the entire kidney to cyst fraction (0.76) [24], and both the mean T2 of the non-cystic kidney parenchyma and entire kidney had stronger correlation coefficients to cyst fraction than htTKV (0.48) [24]. The findings in this chapter contrast with those of the previous study, as stronger correlation coefficients to cyst fraction were observed using the mean T1 of the entire kidney and non-cystic kidney parenchyma, compared to the mean T2 of the entire kidney and non-cystic kidney parenchyma.

The ability of radiomic features to classify the ADPKD cohort from the healthy controls cohort was investigated using two feature selection schemes: (1) using just the mean T1 and T2 and (2) merging the top-two performing radiomic features for classification. The AUC values in classifying the ADPKD cohort from the healthy controls using mean T1 and mean T2 of the entire kidney ranged between 0.94-0.95 and 0.89-0.92, respectively. There was an increase in AUC values when merging the top-two performing radiomic features from the non-cystic kidney parenchyma versus mean T1 alone. Two features extracted from the non-cystic kidney parenchyma of T1 maps that were frequently selected across training partitions were investigated; the ADPKD cohort had higher randomness in image values than that of the healthy cohort (entropy), and also had higher coarseness of texture than that of the healthy cohort (GLCM autocorrelation). Overall, the radiomic features from the non-cystic kidney parenchyma from T1 maps show a stronger ability to classify the ADPKD cohort from

healthy controls than radiomic features from T2 maps. Radiomic features extracted from T1 maps may capture early signs of cyst growth in children and young adults with ADPKD, which might not be captured solely by quantifying the mean T1 from the non-cystic kidney parenchyma. There is potential benefit in characterizing the non-cystic kidney tissue using texture features. The ability to quantify cyst growth using texture features would allow for better risk-stratification of patients with ADPKD, as cyst growth is the predominant mechanism in younger patients who progress earlier and faster to ESKD [56].

A limitation of this work is the small cohort of ADPKD children and young adults (N=15) and healthy controls (N=10). The MRF maps of the healthy controls did not include full volumes of the kidney. Future work could acquire full volumes of healthy kidneys to accurately determine the mean T1, mean T2, and other radiomic features values to compare with ADPKD kidneys. In addition, future work could incorporate additional clinical features such as eGFR and the association of disease progression with change in T1 and T2 feature values over time.

In using T1 and T2 maps to quantify mean T1 and T2 values as well as for feature extraction, there may be change in values if hemorrhages in kidney cysts are present. Therefore, hemorrhagic cysts represent a potential confounder in quantifying mean T1, mean T2, and radiomic features. Hemorrhage in kidney cysts is typically observed during advanced cystic degeneration rather than during initial cystogenesis stages [20]. Therefore, its impact on feature extraction from T1 and T2 maps in pediatric ADPKD patients might be less significant compared to adults with ADPKD.

Radiomic features extracted from MRF-acquired T1 maps and T2 maps have been used to differentiate common adult brain tumors using 2D and 3D MRF data [104, 105]. Multiparametric PET-MRI and MRF research shows potential to serve as an alternative to invasive tissue characterization, furthering imaging-based tumor decoding and phenotyping [106]. Although the potential of using MRF is promising, recent research studied the effect

on first- and second-order radiomic features extracted from MRF data reconstructed using various dictionaries with different step sizes and evaluating intra-dictionary repeatability and inter-dictionary reproducibility [107]. Radiomic features exhibited high repeatability with intraclass correlation coefficients (ICCs) ranging from 0.86 to 0.95 across various dictionary step sizes [107]. However, lower ICCs (ranging from 0.62 to 0.99) were observed when comparing features extracted from MRF data reconstructed from different dictionaries [107]. Future ADPKD studies should account for the impact of data reconstruction from different dictionaries on the extracted radiomic features from MRF-acquired T1 and T2 maps.

5.5 Conclusion

MRF-derived mean T1 and T2 values of ADPKD patients were significantly different than those of healthy controls when extracted from the entire kidney and non-cystic kidney parenchyma. There was a higher correlation of cyst fraction with mean T1 values from the kidney (including cysts) than that of cyst fraction and the biomarkers htTKV and BSA-TKV, showing potential for mean T1 of the kidney to capture cyst growth. Trends of increasing AUC values in classifying ADPKD from healthy controls were observed when merging radiomic features extracted from the non-cystic kidney parenchyma compared to quantifying the mean T1 alone from the non-cystic kidney parenchyma, indicating the potential benefit of incorporating texture features in capturing cystogenesis. Radiomic features extracted from the non-cystic kidney parenchyma revealed higher randomness of image values and coarser texture than those of healthy controls. The characterization of non-cystic kidney parenchyma using radiomic features warrants further investigation in a larger cohort of children and young adults in ADPKD.

CHAPTER 6

SUMMARY AND FUTURE DIRECTIONS

In summary, this work investigated texture-based magnetic resonance imaging (MRI) biomarkers of autosomal dominant polycystic kidney disease (ADPKD). Radiomics-based quantitative models were developed for classification of genotype and Mayo Imaging Classification (MIC) class using features extracted from the non-cystic kidney parenchyma, the additive power of temporal imaging for the prediction of kidney function decline, and the differentiation of magnetic resonance fingerprinting (MRF)-acquired T1 and T2 maps of ADPKD patients from those of a healthy cohort.

The field of radiomics lacks a standardized approach to extracting features, and previous disease-specific radiomic studies investigated the effect of pre-processing on radiomic features. In this work, the effect of pre-processing on radiomic features extracted from the kidney and on subsequent classification of genotype (Chapter 2) and differentiation of risk-stratified MIC class (Chapter 3) was established.

In Chapter 2, the aim of the presented work was to investigate radiomic features from T2-weighted fat saturated (T2W-FS) MR images for the differentiation between *PKD1* and *PKD2* patients in a cohort matched for age, gender, and MIC class. In addition, the impact of MRI pre-processing (MR normalization, gray-level discretization, and pixel resampling) on radiomic feature reproducibility was evaluated with the intra-class correlation coefficient, and the effect of pre-processing parameters on subsequent genotype classification was assessed. The results from this chapter identified texture features indicative of genotype expression in ADPKD in both the non-cystic and entire kidney parenchyma regions. Reproducibility of radiomic features extracted from the noncystic kidney parenchyma was dependent on the pre-processing parameters used. Additionally, there was an increase in feature reproducibility across MR normalizations with the number of gray levels available for discretization. The ability to classify *PKD1* and *PKD2* patients based on radiomic features extracted from

either the non-cystic kidney parenchyma or entire kidney depended on the pre-processing parameters used, and there was not one combination of pre-processing parameters that optimized classification when using features extracted from the non-cystic kidney parenchyma or entire kidney. Future work building upon the ability of texture-based image features to classify genotype include the investigation of *PKD1* patients and textural differences of truncating and non-truncating mutations, as the median age at onset of end-stage kidney disease (ESKD) of patients with truncating mutations is less than that of patients with non-truncating mutations [108]. Texture-based image features could identify patients with non-truncating mutations that are associated with more severe kidney disease.

The effect of pre-processing on subsequent classification was continued into Chapter 3 for differentiating among low/intermediate-risk (1A-1B) and high-risk (1C-1E) MIC classes using radiomic features; furthermore, the classification task was focused on just intermediate-risk (1B) and high-risk (1C) MIC class. The results suggest the potential to distinguish between patients based on low/intermediate- and high-risk MIC class; the impact of pre-processing on subsequent classification reveals that features extracted from the non-cystic kidney parenchyma were more sensitive to pre-processing than those from the entire kidney. The classification of patients with low/intermediate-risk (low cyst burden) from those with high-risk (high cyst burden with higher signal intensity) was unaffected by pre-processing parameters such as MR normalization, gray-level discretization, and pixel resampling. This underscores the influence of cysts on texture features, indicating that these imaging characteristics remain robust indicators of disease severity and risk classification regardless of pre-processing variations. High-risk MIC class patients have by definition larger kidney sizes than patients in low/intermediate-risk MIC classes, and radiomic features that exhibited low, moderate, or high correlation with number of voxels (i.e., kidney size) across pre-processing parameters were identified; the incorporation of features based on their correlation to kidney size into classification models affected the performance in distinguishing

between low/intermediate-risk and high-risk MIC classes, and warrants further investigation. Radiomic features were able to distinguish intermediate-risk (1B) and high-risk (1C) MIC class, with higher AUC values when combining clinical features (age, sex, and estimated glomerular filtration rate (eGFR)) with radiomic features versus clinical features alone, showing the potential benefit of kidney texture not captured by clinical features.

Chapter 4 determined whether radiomic features from T2W-FS MR images from 24-month and 48-month timepoints provide additive power versus baseline texture alone in predicting kidney function decline to (1) chronic kidney disease (CKD) stage 3A or greater or (2) a greater than 30% reduction in eGFR of ADPKD patients at 60-months follow-up. Delta-radiomic features were calculated at 24-month and 48-month timepoints and exhibited close to random guessing performance (AUC=0.5) for either prediction task. The 24-month and 48-month timepoints were not statistically different from baseline performance for either prediction task, but there was a statistically significant difference from baseline performance using combined baseline, 24-month, and 48-month texture from the entire kidney for the prediction of progression to CKD stage 3A or greater. Unlike for the entire kidney, there was no improvement in performance using radiomic features extracted from the non-cystic kidney parenchyma; however, features from the non-cystic kidney parenchyma did exhibit moderate performance for either prediction task. Radiomic features from the 24-month, 48-month, and 60-month timepoints exhibited strong correlation coefficients with baseline feature values, and radiomic features were generally not statistically significantly different from baseline, regardless if patients progressed or did not progress in either prediction tasks. The relationship between radiomic features at baseline and follow-up timepoints provided insight for why delta radiomics failed to offer additive power in predicting kidney function decline. There was not a statistically significant difference between performance using combined clinical-radiomics models and performance from a combined-clinical model. Future studies could involve patient matching based on baseline characteristics such as height-corrected total kidney volume

(htTKV), eGFR, and age. Additionally, implementing a more stringent minimum baseline eGFR cutoff value may be beneficial, given that age, htTKV, and eGFR are recognized as confounding factors in predicting kidney function decline.

Chapter 5 was an exploratory aim that analyzed the first MRF-derived T1 and T2 quantitative coronal maps of an ADPKD cohort (children and young adults) and the texture-based differences from that of a healthy cohort. Additionally, this aim compared kidney segmentations performed by two image segmenters familiar with kidney anatomy from different sites, and assessed their impact on subsequent segmentation of the non-cystic kidney parenchyma using three cyst removal methods. The results from this study show good agreement between manual kidney segmentations, and there was a trend of lower Dice similarity coefficients (DSC) between two segmenters of all image sections compared to the DSC when restricting to the central kidney sections for segmentation. The decrease in DSC was attributed to the partial volume effect observed in the anterior and posterior regions in the kidney of the T1 maps, which pose a more challenging segmentation task. MRF-derived mean T1 and T2 values of ADPKD patients were significantly different than those of healthy controls when extracted from the entire kidney and non-cystic kidney parenchyma. Trends of increasing AUC values in classifying ADPKD from healthy controls were observed when merging the top-two performing radiomic features extracted from the non-cystic kidney parenchyma relative to the performance of mean T1 alone. Radiomic features extracted from T1 maps may capture early signs of cyst growth, suggesting potential benefits in utilizing texture features to characterize cystogenesis. This may provide insights not captured solely by quantifying mean T1. Mean T1 values showed a strong correlation with cyst fraction, exhibiting a higher correlation coefficient compared with the biomarkers htTKV (computed from image-based measurements of kidney size) and body surface area total kidney volume (BSA-TKV). Future work can extend the work to a larger cohort and investigate change in texture over time with disease progression.

Future directions to build upon the work presented in this thesis may include the effect of pre-processing on other MR pulse sequences such as T1-weighted MRI and the effect on subsequent ADPKD classification tasks. A limitation of Chapters 2 and 3 was that representative 2D coronal MR images were used, and future work could investigate radiomics analyses of total kidney volume (TKV) texture-based differences in genotype and risk-stratified MIC class.

The effect of normalization and gray levels for discretization impacts the original texture of an image, but, in utilizing quantitative MR, one can remove the normalization step before feature extraction to allow for better comparison of radiomic features for the longitudinal assessment of ADPKD patients; future work could associate change in features extracted from MRF-derived T1 and T2 maps with change in htTKV and eGFR, along with quantitative measurements such as total parenchymal volume, total cyst number, cyst surface area, and cyst parenchyma surface area using automatic segmentation AI algorithms (e.g., U-Net, semantic segmentation).

In utilizing radiomic features as imaging biomarkers for ADPKD, ensuring radiomic features are repeatable (e.g., features remain temporally stable with patients and image protocols) and reproducible (e.g., features remain unchanged using different acquisition or processing) is of the utmost importance; therefore, future work could quantify feature repeatability and reproducibility under varying conditions such as day-to-day repeatability and the effect of MRF dictionary matching on features extracted from the kidney.

Furthermore, deep-learning-based features and classification models have the ability to complement or even replace traditional radiomic features ("human-engineered") and machine learning models [109]; deep learning is a subarea of machine learning and uses neural networks that remove traditional feature extraction and feature selection processes and "learns" directly from medical images, bypassing explicit feature definitions [109]. In general, AI has and will continue to impact decision-making processes of clinicians; there is an interesting

crossroads between "big data" (e.g., the mining of thousands of quantitative image features or deep-learning-based features) and human interpretability and explainability. Nephrologists use explainable features such as change in htTKV and eGFR, but there is a "black box" of AI in medicine that challenges the traditional methods of care. In fostering conversation between clinicians, medical physicists, data scientists, and other stakeholders in human health, computer-aided diagnosis systems can be created to mine informative "big data" while maintaining the human element of interpretable and explainable care in ADPKD.

In conclusion, the presented work will ultimately improve identification of low- and high-risk individuals for progression to ESKD and advance personalized medicine using radiomics. Radiomic feature extraction from MR images provides a non-invasive method to evaluate ADPKD disease severity classification and disease progression to ultimately assist clinical decision-making.

REFERENCES

[1] A. B. Chapman et al. "Autosomal-dominant polycystic kidney disease (ADPKD): executive summary from a Kidney Disease: Improving Global Outcomes (KDIGO) Controversies Conference". In: *Kidney International* 88 (2015), pp. 17–27. DOI: 10.1038/ki.2015.59.

[2] A. B. Chapman. "Approaches to testing new treatments in autosomal dominant polycystic kidney disease: insights from the CRISP and HALT-PKD studies". In: *Clinical Journal of the American Society of Nephrology* 3 (2008), pp. 1197–1204. DOI: 10.2215/CJN.00060108.

[3] J. J. Grantham et al. "Volume Progression in Polycystic Kidney Disease". In: *The New England Journal of Medicine* 354 (2006), pp. 2122–2130. DOI: 10.1056/NEJMoa054341.

[4] E. Cornec-Le Gall, A. Alam, and R. D. Perrone. "Autosomal dominant polycystic kidney disease". In: *Lancet* 393 (2019), pp. 919–935. DOI: 10.1016/S0140-6736(18)32782-X.

[5] L. Grau et al. "Interactions between FGF23 and Genotype in Autosomal Dominant Polycystic Kidney Disease". In: *Kidney360* 1 (2020), pp. 648–656. DOI: 10.34067/KID.0001692020.

[6] F. T. Chebib et al. "A Practical Guide for Treatment of Rapidly Progressive ADPKD with Tolvaptan". In: *Journal of the American Society of Nephrology* 29 (2018), pp. 2458–2470. DOI: 10.1681/ASN.2018060590.

[7] J. J. Grantham and V. E. Torres. "The importance of total kidney volume in evaluating progression of polycystic kidney disease". In: *Nature Reviews Nephrology* 12 (2016), pp. 667–677. DOI: 10.1038/nrneph.2016.135.

[8] M. V. Irazabal et al. "Prognostic Enrichment Design in Clinical Trials for Autosomal Dominant Polycystic Kidney Disease: The TEMPO 3:4 Clinical Trial". In: *Kidney International Reports* 1 (2016), pp. 213–220. DOI: 10.1016/j.ekir.2016.08.001.

[9] M. V. Irazabal et al. "Prognostic Enrichment Design in Clinical Trials for Autosomal Dominant Polycystic Kidney Disease: The HALT-PKD Clinical Trial". In: *Nephrology, Dialysis, Transplantation* 32 (2017), pp. 1857–1865. DOI: 10.1093/ndt/gfw294.

[10] G. M. Brosnahan et al. "Patterns of Kidney Function Decline in Autosomal Dominant Polycystic Kidney Disease: A Post Hoc Analysis From the HALT-PKD Trials". In: *American Journal of Kidney Diseases* 71 (2018), pp. 666–676. DOI: 10.1053/j.ajkd.2017.10.023.

[11] M. V. Irazabal et al. "Imaging classification of autosomal dominant polycystic kidney disease: a simple model for selecting patients for clinical trials". In: *Journal of the American Society of Nephrology* 26 (2015), pp. 160–172. DOI: 10.1681/ASN.2013101138.

[12] P. C. Harris et al. "Cyst Number but Not the Rate of Cystic Growth Is Associated with the Mutated Gene in Autosomal Dominant Polycystic Kidney Disease". In: *Journal of the American Society of Nephrology* 17 (2006), pp. 3013–3019. DOI: 10.1681/ASN.2006080835.

[13] S. Lavu et al. "The value of genotypic and imaging information to predict functional and structural outcomes in ADPKD". In: *JCI Insight* 5 (2020), e138724. DOI: 10.1172/jci.insight.138724.

[14] A. Caroli and T. L. Kline. "Abdominal Imaging in ADPKD: Beyond Total Kidney Volume". In: *Journal of Clinical Medicine* 12 (2023), p. 5133. DOI: 10.3390/jcm12155133.

[15] E. Higashihara et al. "Kidney Volume Estimations with Ellipsoid Equations by Magnetic Resonance Imaging in Autosomal Dominant Polycystic Kidney Disease". In: *Nephron* 129 (2015), pp. 253–262. DOI: 10.1159/000381476.

[16] N. Demoulin et al. "Limited Performance of Estimated Total Kidney Volume for Follow-up of ADPKD". In: *Kidney International Reports* 6 (2021), pp. 2821–2829. DOI: 10.1016/j.ekir.2021.08.013.

[17] P. Gaur, W. Gedroyc, and P. Hill. "ADPKD-what the radiologist should know". In: *The British Journal of Radiology* 92 (2019), p. 20190078. DOI: 10.1259/bjr.20190078.

[18] T. W. Redpath. "MRI Developments in Perspective". In: *The British Journal of Radiology* 70 (1997), S70–S80. DOI: 10.1259/bjr.1997.0010.

[19] D. B. Plewes and W. Kucharczyk. "Physics of MRI: A Primer". In: *Journal of Magnetic Resonance Imaging* 35 (2012), pp. 1038–1054. DOI: 10.1002/jmri.23642.

[20] M. Franke et al. "Magnetic Resonance T2 Mapping and Diffusion-weighted Imaging for Early Detection of Cystogenesis and Response to Therapy in a Mouse Model of Polycystic Kidney Disease". In: *Kidney International* 92 (2017), pp. 1544–1554. DOI: 10.1016/j.kint.2017.05.024.

[21] R. Lupica et al. "3 Tesla-Diffusion Tensor Imaging in Autosomal Dominant Polycystic Kidney Disease: The Nephrologist's Point of View". In: *Nephron* 134 (2016), pp. 73–80. DOI: 10.1159/000442296.

[22] S. T. Francis, N. M. Selby, and M. W. Taal. "Magnetic Resonance Imaging to Evaluate Kidney Structure, Function, and Pathology: Moving Toward Clinical Application". In: *American Journal of Kidney Diseases* 82 (2023), pp. 491–504. DOI: 10.1053/j.ajkd.2023.02.007.

[23] A. Caroli et al. "Diffusion Magnetic Resonance Imaging for Kidney Cyst Volume Quantification and Non-cystic Tissue Characterisation in ADPKD". In: *European Radiology* 33 (2023), pp. 6009–6019. DOI: 10.1007/s00330-023-09601-4.

[24] F. Siedek et al. "Magnetic Resonance Kidney Parenchyma-T2 as a Novel Imaging Biomarker for Autosomal Dominant Polycystic Kidney Disease". In: *Investigative Radiology* 55 (2020), pp. 217–225. DOI: 10.1097/RLI.0000000000000633.

[25] D. Ma et al. "Magnetic Resonance Fingerprinting". In: *Nature* 495 (2013), pp. 187–192. DOI: 10.1038/nature11971.

[26] C. Tippareddy et al. "Magnetic Resonance Fingerprinting: An Overview". In: *European Journal of Nuclear Medicine and Molecular Imaging* 48 (2021), pp. 4189–4200. DOI: 10.1007/s00259-021-05384-2.

[27] C. J. MacAskill et al. "Rapid B1-Insensitive MR Fingerprinting for Quantitative Kidney Imaging". In: *Radiology* 300 (2021), pp. 380–387. DOI: 10.1148/radiol.2021202302.

[28] J. C. Gore. "Artificial Intelligence in Medical Imaging". In: *Magnetic Resonance Imaging* 68 (2020), A1–A4. DOI: 10.1016/j.mri.2019.12.006.

[29] M. L. Giger, H. P. Chan, and J. Boone. "Anniversary Paper: History and Status of CAD and Quantitative Image Analysis: The Role of Medical Physics and AAPM". In: *Medical Physics* 35 (2008), pp. 5799–5820. DOI: 10.1118/1.3013555.

[30] M. L. Giger, N. Karssemeijer, and 3rd Armato S. G. "Computer-Aided Diagnosis in Medical Imaging". In: *IEEE Transactions on Medical Imaging* 20 (2001), pp. 1205–1208. DOI: 10.1109/TMI.2001.974915.

[31] T. Y. Shin et al. "Expert-level segmentation using deep learning for volumetry of polycystic kidney and liver". In: *Investigative and Clinical Urology* 61 (2020), pp. 555–564. DOI: 10.4111/icu.20200086.

[32] A. Sharbatdaran et al. "Deep Learning Automation of Kidney, Liver, and Spleen Segmentation for Organ Volume Measurements in Autosomal Dominant Polycystic Kidney Disease". In: *Tomography* 8 (2022), pp. 1804–1819. DOI: 10.3390/tomography8040152.

[33] A. Goel et al. "Deployed Deep Learning Kidney Segmentation for Polycystic Kidney Disease MRI". In: *Radiology Artificial Intelligence* 4 (2022), e210205. DOI: 10.1148/ryai.210205.

[34] T. A. Potretzke et al. "Clinical Implementation of an Artificial Intelligence Algorithm for Magnetic Resonance-Derived Measurement of Total Kidney Volume". In: *Mayo Clinic Proceedings* 98 (2023), pp. 689–700. DOI: 10.1016/j.mayocp.2022.12.019.

[35] J. Taylor et al. "An Artificial Intelligence Generated Automated Algorithm to Measure Total Kidney Volume in ADPKD". In: *Kidney International Reports* 9 (2023), pp. 249–256. DOI: 10.1016/j.ekir.2023.10.029.

[36] S. Wigerinck et al. "Evaluation of Advanced Imaging Biomarkers at Kidney Failure in Patients with ADPKD: A Pilot Study". In: *Clinical Kidney Journal* 16 (2023), pp. 1691–1700. DOI: 10.1093/ckj/sfad114.

[37] P. Lambin et al. "Radiomics: Extracting More Information from Medical Images Using Advanced Feature Analysis". In: *European Journal of Cancer* 48 (2012), pp. 441–446. DOI: 10.1016/j.ejca.2011.11.036.

[38] P. Lambin et al. "Radiomics: the bridge between medical imaging and personalized medicine". In: *Nature Reviews Clinical Oncology* 14 (2017), pp. 749–762. DOI: 10.1038/nrclinonc.2017.141.

[39] A. Zwanenburg et al. "The Image Biomarker Standardization Initiative: Standardized Quantitative Radiomics for High-Throughput Image-based Phenotyping". In: *Radiology* 295 (2020), pp. 328–338. DOI: 10.1148/radiol.2020191145.

[40] A. Carré et al. "Standardization of brain MR images across machines and protocols: bridging the gap for MRI-based radiomics". In: *Scientific Reports* 10 (2020), p. 12340. DOI: 10.1038/s41598-020-69298-z.

[41] M. Schwier et al. "Repeatability of Multiparametric Prostate MRI Radiomics Features". In: *Scientific Reports* 9 (2019), p. 9441. DOI: 10.1038/s41598-019-45766-z.

[42] D. C. Sullivan et al. "Metrology Standards for Quantitative Imaging Biomarkers". In: *Radiology* 277 (2015), pp. 813–825. DOI: 10.1148/radiol.2015142202.

[43] D. Sheth and M. L. Giger. "Artificial Intelligence in the Interpretation of Breast Cancer on MRI". In: *Journal of Magnetic Resonance Imaging* 51 (2020), pp. 1310–1324. DOI: 10.1002/jmri.26878.

[44] J. D. Budzikowski et al. "Radiomics-Based Assessment of Idiopathic Pulmonary Fibrosis Is Associated with Genetic Mutations and Patient Survival". In: *Journal of Medical Imaging* 8 (2021), p. 031903. DOI: 10.1117/1.JMI.8.3.031903.

[45] S. Hindocha et al. "Author Correction: Gross Tumour Volume Radiomics for Prognostication of Recurrence & Death Following Radical Radiotherapy for NSCLC". In: *NPJ Precision Oncology* 6 (2022), p. 87. DOI: 10.1038/s41698-022-00332-1.

[46] D. Liu et al. "Imaging-Genomics in Glioblastoma: Combining Molecular and Imaging Signatures". In: *Frontiers in Oncology* 11 (2021), p. 699265. DOI: 10.3389/fonc.2021.699265.

[47] N. Antropova et al. "Breast lesion classification based on dynamic contrast-enhanced magnetic resonance images sequences with long short-term memory networks". In: *Journal of Medical Imaging* 6 (2019), p. 011002. DOI: 10.1117/1.JMI.6.1.011002.

[48] S. Starke et al. "Longitudinal and Multimodal Radiomics Models for Head and Neck Cancer Outcome Prediction". In: *Cancers* 15 (2023), p. 673. DOI: 10.3390/cancers15030673.

[49] R. R. Wu et al. "Delta Radiomics Analysis for Prediction of Intermediary- and High-Risk Factors for Patients with Locally Advanced Cervical Cancer Receiving Neoadjuvant Therapy". In: *Scientific Reports* 13 (2023), p. 19409. DOI: 10.1038/s41598-023-46621-y.

[50] H. Nasief et al. "A Machine Learning Based Delta-Radiomics Process for Early Prediction of Treatment Response of Pancreatic Cancer". In: *NPJ Precision Oncology* 3 (2019), p. 25. DOI: 10.1038/s41698-019-0096-z.

[51] S. Ahrari et al. "Application of PET Imaging Delta Radiomics for Predicting Progression-Free Survival in Rare High-Grade Glioma". In: *Scientific Reports* 14 (2024), p. 3256. DOI: 10.1038/s41598-024-53693-x.

[52] T. L. Kline et al. "Image texture features predict renal function decline in patients with autosomal dominant polycystic kidney disease". In: *Kidney International* 92 (2017), pp. 1206–1216. DOI: 10.1016/j.kint.2017.03.026.

[53] L. Cong et al. "A radiomics method based on MR FS-T2WI sequence for diagnosing of autosomal dominant polycystic kidney disease progression". In: *European Review for Medical and Pharmacological Sciences* 25 (2021), pp. 5769–5780. DOI: 10.26355/eurrev_202109_26795.

[54] X. Li et al. "A MRI-based radiomics nomogram for evaluation of renal function in ADPKD". In: *Abdominal Radiology* 47 (2022), pp. 1385–1395. DOI: 10.1007/s00261-022-03433-4.

[55] Y. Xie et al. "The predictive value of renal parenchymal information for renal function impairment in patients with ADPKD: a multicenter prospective study". In: *Abdominal Radiology* 47 (2022), pp. 2845–2857. DOI: 10.1007/s00261-022-03554-w.

[56] S. S. Shukoor et al. "Characteristics of Patients with End-Stage Kidney Disease in ADPKD". In: *Kidney International Reports* 6 (2020), pp. 755–767. DOI: 10.1016/j.ekir.2020.12.016.

[57] J. E. van Timmeren et al. "Radiomics in medical imaging-"how-to" guide and critical reflection". In: *Insights into Imaging* 11 (2020), p. 91. DOI: 10.1186/s13244-020-00887-2.

[58] M. R. S. Sunoqrot et al. "Automated reference tissue normalization of T2-weighted MR images of the prostate using object recognition". In: *Magma* 34 (2021), pp. 309–321. DOI: 10.1007/s10334-020-00871-3.

[59] E. Scalco et al. "T2w-MRI signal normalization affects radiomics features reproducibility". In: *Medical Physics* 47 (2020), pp. 1680–1691. DOI: 10.1002/mp.14038.

[60] A. Fedorov et al. "3D Slicer as an image computing platform for the Quantitative Imaging Network". In: *Magnetic Resonance Imaging* 30 (2012), pp. 1323–1341. DOI: 10.1016/j.mri.2012.05.001.

[61] J. J. M. van Griethuysen et al. "Computational Radiomics System to Decode the Radiographic Phenotype". In: *Cancer Research* 77 (2017), pp. 104–107. DOI: 10.1158/0008-5472.CAN-17-0339.

[62] F. Tixier et al. "Intratumor heterogeneity characterized by textural features on baseline ^{18}F-FDG PET images predicts response to concomitant radiochemotherapy in esophageal cancer". In: *Journal of Nuclear Medicine* 52 (2011), pp. 369–378. DOI: 10.2967/jnumed.110.082404.

[63] L. Duron et al. "Gray-level discretization impacts reproducible MRI radiomics texture features". In: *PloS One* 14 (2019), e0213459. DOI: 10.1371/journal.pone.0213459.

[64] A. K. Jha et al. "Repeatability and reproducibility study of radiomic features on a phantom and human cohort". In: *Scientific Reports* 11 (2021), p. 2055. DOI: 10.1038/s41598-021-81526-8.

[65] M. Bologna et al. "Repeatability and reproducibility of MRI-radiomic features: A phantom experiment on a 1.5 T scanner". In: *Medical Physics* 50 (2023), pp. 750–762. DOI: 10.1002/mp.16054.

[66] T. K. Koo and M. Y. Li. "A Guideline of Selecting and Reporting Intraclass Correlation Coefficients for Reliability Research". In: *Journal of Chiropractic Medicine* 15 (2016), pp. 155–163. DOI: 10.1016/j.jcm.2016.02.012.

[67] MATLAB. *MATLAB version: 9.13.0 (R2022b)*. Natick, Massachusetts, United States: The MathWorks Inc., 2022. URL: https://www.mathworks.com.

[68] F. Wilcoxon. "Probability tables for individual comparisons by ranking methods". In: *Biometrics* 3 (1947), pp. 119–22.

[69] Muhammad Shafiq-Ul-Hassan et al. "Voxel size and gray level normalization of CT radiomic features in lung cancer". In: *Scientific Reports* 8 (2018), p. 10545. DOI: 10.1038/s41598-018-28895-9.

[70] A. Crombé et al. "Assessment of Repeatability, Reproducibility, and Performances of T2 Mapping-Based Radiomics Features: A Comparative Study". In: *Journal of Magnetic Resonance Imaging* 54 (2021), pp. 537–548. DOI: 10.1002/jmri.27558.

[71] L. J. Isaksson et al. "Effects of MRI image normalization techniques in prostate cancer radiomics". In: *Physica Medica* 71 (2020), pp. 7–13. DOI: 10.1016/j.ejmp.2020.02.007.

[72] A. V. Gregory et al. "Semantic Instance Segmentation of Kidney Cysts in MR Images: A Fully Automated 3D Approach Developed Through Active Learning". In: *Journal of Digital Imaging* 34 (2021), pp. 773–787. DOI: 10.1007/s10278-021-00452-3.

[73] E. Higashihara et al. "Estimation of Changes in Kidney Volume Growth Rate in ADPKD". In: *Kidney International Reports* 5 (2020), pp. 1459–1471. DOI: 10.1016/j.ekir.2020.06.011.

[74] O. Ronneberger, P. Fischer, and T. Brox. "U-Net: Convolutional Networks for Biomedical Image Segmentation". In: *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015*. Ed. by N. Navab et al. Vol. 9351. Lecture Notes in Computer Science. Springer, 2015, pp. 234–241. DOI: 10.1007/978-3-319-24574-4_28.

[75] N. Otsu. "A Threshold Selection Method from Gray-Level Histograms". In: *IEEE Transactions on Systems, Man, and Cybernetics* 9 (1979), pp. 62–66. DOI: 10.1109/TSMC.1979.4310076.

[76] L. R. Dice. "Measures of the Amount of Ecologic Association Between Species". In: *Ecology* 26 (1945), pp. 297–302. DOI: 10.2307/1932409.

[77] F. Hausdorff. *Grundzüge der Mengenlehre (Basics of Set Theory)*. Leipzig Viet, 1914.

[78] M. Shafiq-Ul-Hassan et al. "Intrinsic dependencies of CT radiomic features on voxel size and number of gray levels". In: *Medical Physics* 44 (2017), pp. 1050–1062. DOI: 10.1002/mp.12123.

[79] R. Tibshirani. "Regression Shrinkage and Selection via the Lasso". In: *Journal of the Royal Statistical Society Series B* 58 (1996), pp. 267–88.

[80] H. C. Park et al. “Mayo imaging classification is a good predictor of rapid progress among Korean patients with autosomal dominant polycystic kidney disease: results from the KNOW-CKD study”. In: *Kidney Research and Clinical Practice* 41 (2022), pp. 432–441. DOI: 10.23876/j.krcp.21.261.

[81] S. Holm. “A Simple Sequentially Rejective Multiple Test Procedure”. In: *Scandinavian Journal of Statistics* 6 (1979), pp. 65–70.

[82] S. Fiset et al. “Repeatability and reproducibility of MRI-based radiomic features in cervical cancer”. In: *Radiotherapy and Oncology* 135 (2019), pp. 107–114. DOI: 10.1016/j.radonc.2019.03.001.

[83] Z. Hou et al. “Radiomic analysis in T2W and SPAIR T2W MRI: predict treatment response to chemoradiotherapy in esophageal squamous cell carcinoma”. In: *Journal of Thoracic Disease* 10 (2018), pp. 2256–2267. DOI: 10.21037/jtd.2018.03.123.

[84] L. E. Kremer, A. B. Chapman, and 3rd Armato S. G. “Magnetic Resonance Imaging Preprocessing and Radiomic Features for Classification of Autosomal Dominant Polycystic Kidney Disease Genotype”. In: *Journal of Medical Imaging* 10 (2023), p. 064503. DOI: 10.1117/1.JMI.10.6.064503.

[85] L. E. Kremer et al. “Radiomics-based classification of autosomal dominant polycystic kidney disease (ADPKD) Mayo imaging classification (MIC) and the effect of gray-level discretization”. In: *Medical Imaging 2023: Computer-Aided Diagnosis*. Vol. 12465. International Society for Optics and Photonics. 2023, 124652Y. DOI: 10.1117/12.2654476.

[86] J. E. Park et al. “Reproducibility and Generalizability in Radiomics Modeling: Possible Strategies in Radiologic and Statistical Perspectives”. In: *Korean Journal of Radiology* 20 (2019), pp. 1124–1137. DOI: 10.3348/kjr.2018.0070.

[87] V. Nardone et al. “Delta radiomics: a systematic review”. In: *La Radiologia Medica* 126 (2021), pp. 1571–1583. DOI: 10.1007/s11547-021-01436-7.

[88] R. R. Wu et al. “Delta radiomics analysis for prediction of intermediary- and high-risk factors for patients with locally advanced cervical cancer receiving neoadjuvant therapy”. In: *Scientific Reports* 13 (2023), p. 19409. DOI: 10.1038/s41598-023-46621-y.

[89] L. Guo et al. “Delta-Radiomics Based on Dynamic Contrast-Enhanced MRI Predicts Pathologic Complete Response in Breast Cancer Patients Treated with Neoadjuvant Chemotherapy”. In: *Cancers* 14 (2022), p. 3515. DOI: 10.3390/cancers14143515.

[90] E. Abbas et al. “Delta-radiomics in cancer immunotherapy response prediction: A systematic review”. In: *European Journal of Radiology Open* 11 (2023), p. 100511. DOI: 10.1016/j.ejro.2023.100511.

[91] X. Han et al. “Delta-radiomics features for predicting the major pathological response to neoadjuvant chemoimmunotherapy in non-small cell lung cancer”. In: *European Radiology* 34 (2024), pp. 2716–2726. DOI: 10.1007/s00330-023-10241-x.

[92] T. K. Chen, D. H. Knicely, and M. E. Grams. "Chronic Kidney Disease Diagnosis and Management: A Review". In: *JAMA* 322 (2019), pp. 1294–1304. DOI: 10.1001/jama.2019.14745.

[93] A. S. Levey et al. "Definition and classification of chronic kidney disease: a position statement from Kidney Disease: Improving Global Outcomes (KDIGO)". In: *Kidney International* 67 (2005), pp. 2089–2100. DOI: 10.1111/j.1523-1755.2005.00365.x.

[94] E. R. DeLong, D. M. DeLong, and D. L. Clarke-Pearson. "Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach". In: *Biometrics* 44 (1988), pp. 837–845.

[95] T. L. Kline et al. "Quantitative MRI of kidneys in renal disease". In: *Abdominal Radiology* 43 (2018), pp. 629–638. DOI: 10.1007/s00261-017-1236-y.

[96] I. A. Dekkers et al. "Consensus-based technical recommendations for clinical translation of renal T1 and T2 mapping MRI". In: *Magma* 33 (2020), pp. 163–176. DOI: 10.1007/s10334-019-00797-5.

[97] S. Hochreiter and J. Schmidhuber. "Long short-term memory". In: *Neural computation* 9 (1997), pp. 1735–1780. DOI: 10.1162/neco.1997.9.8.1735.

[98] H. Li et al. "Temporal Machine Learning Analysis of Prior Mammograms for Breast Cancer Risk Prediction". In: *Cancers* 15 (2023), p. 2141. DOI: 10.3390/cancers15072141.

[99] X. Yang et al. "Imaging Biomarkers in Young Patients With ADPKD". In: *Kidney International Reports* 8 (2023), pp. 2153–2155. DOI: 10.1016/j.ekir.2023.07.004.

[100] M. A. Cadnapaphornchai et al. "Magnetic Resonance Imaging of Kidney and Cyst Volume in Children With ADPKD". In: *Clinical Journal of the American Society of Nephrology* 6 (2011), pp. 369–376. DOI: 10.2215/CJN.03780410.

[101] J. I. Hamilton et al. "Investigating and reducing the effects of confounding factors for robust T1 and T2 mapping with cardiac MR fingerprinting". In: *Magnetic Resonance Imaging* 53 (2018), pp. 40–51. DOI: 10.1016/j.mri.2018.07.018.

[102] Y. Chen et al. "Three-dimensional MR Fingerprinting for Quantitative Breast Imaging". In: *Radiology* 290 (2019), pp. 33–40. DOI: 10.1148/radiol.2018180921.

[103] A. Semechko. *Fast fuzzy c-means image segmentation*. GitHub. 2024. URL: https://github.com/AntonSemechko/Fast-Fuzzy-C-Means-Segmentation.

[104] S. Dastmalchian et al. "Radiomic Analysis of Magnetic Resonance Fingerprinting in Adult Brain Tumors". In: *European Journal of Nuclear Medicine and Molecular Imaging* 48 (2021), pp. 683–693. DOI: 10.1007/s00259-020-05037-w.

[105] C. Tippareddy et al. "Novel 3D Magnetic Resonance Fingerprinting Radiomics in Adult Brain Tumors: A Feasibility Study". In: *European Radiology* 33 (2023), pp. 836–844. DOI: 10.1007/s00330-022-09067-w.

[106] J. Haubold et al. “Non-Invasive Tumor Decoding and Phenotyping of Cerebral Gliomas Utilizing Multiparametric ^{18}F-FET PET-MRI and MR Fingerprinting”. In: *European Journal of Nuclear Medicine and Molecular Imaging* 47 (2020), pp. 1435–1445. DOI: 10.1007/s00259-019-04602-2.

[107] S. Fujita et al. “Radiomics with 3-dimensional Magnetic Resonance Fingerprinting: Influence of Dictionary Design on Repeatability and Reproducibility of Radiomic Features”. In: *European Radiology* 32 (2022), pp. 4791–4800. DOI: 10.1007/s00330-022-08555-3.

[108] E. Cornec-Le Gall et al. “Type of PKD1 Mutation Influences Renal Outcome in ADPKD”. In: *Journal of the American Society of Nephrology* 24 (2013), pp. 1006–1013. DOI: 10.1681/ASN.2012070650.

[109] A. Demircioğlu. “Are Deep Models in Radiomics Performing Better Than Generic Models? A Systematic Review”. In: *European Radiology Experimental* 7 (2023), p. 11. DOI: 10.1186/s41747-023-00325-0.