Article

Using MRI Texture Analysis Machine Learning Models to Assess Graft Interstitial Fibrosis and Tubular Atrophy in Patients with Transplanted Kidneys

Valeria Trojani [1], Filippo Monelli [2,*], Giulia Besutti [2,3], Marco Bertolini [1], Laura Verzellesi [1], Roberto Sghedoni [1], Mauro Iori [1,*], Guido Ligabue [3], Pierpaolo Pattacini [2], Paolo Giorgi Rossi [4], Marta Ottone [4], Alessia Piccinini [5], Gaetano Alfano [5], Gabriele Donati [5,6] and Francesco Fontana [5]

[1] Medical Physics Unit, Azienda USL-IRCCS di Reggio Emilia, 42123 Reggio Emilia, Italy; valeria.trojani@ausl.re.it (V.T.); marco.bertolini@ausl.re.it (M.B.); laura.verzellesi@ausl.re.it (L.V.); roberto.sghedoni@ausl.re.it (R.S.)

[2] Radiology Unit, Azienda USL-IRCCS di Reggio Emilia, 42123 Reggio Emilia, Italy; giulia.besutti@ausl.re.it (G.B.); pierpaolo.pattacini@ausl.re.it (P.P.)

[3] Division of Radiology, Department of Medical and Surgical Sciences of Children and Adults, University of Modena and Reggio Emilia, 41224 Modena, Italy; guido.ligabue@unimore.it

[4] Epidemiology Unit, Azienda USL-IRCCS di Reggio Emilia, 42123 Reggio Emilia, Italy; paolo.giorgirossi@ausl.re.it (P.G.R.); marta.ottone@ausl.re.it (M.O.)

[5] Nephrology Dialysis and Transplant Unit, University Hospital of Modena, 41125 Modena, Italy; 163520@studenti.unimore.it (A.P.); gaetano.alfano@unimore.it (G.A.); gabriele.donati@unimore.it (G.D.); francesco.fontana@unimore.it (F.F.)

[6] Surgical, Medical and Dental Department of Morphological Sciences (CHIMOMO), Section of Nephrology, University Hospital of Modena, 41124 Modena, Italy

* Correspondence: filippo.monelli@ausl.re.it (F.M.); mauro.iori@ausl.re.it (M.I.)

Abstract: Objective: Interstitial fibrosis/tubular atrophy (IFTA) is a common, irreversible, and progressive form of chronic kidney allograft injury, and it is considered a critical predictor of kidney allograft outcomes. The extent of IFTA is estimated through a graft biopsy, while a non-invasive test is lacking. The aim of this study was to evaluate the feasibility and accuracy of an MRI radiomic-based machine learning (ML) algorithm to estimate the degree of IFTA in a cohort of transplanted patients. Approach: Patients who underwent MRI and renal biopsy within a 6-month interval from 1 January 2012 to 1 March 2021 were included. Stable MRI sequences were selected, and renal parenchyma, renal cortex and medulla were segmented. After image filtering and pre-processing, we computed radiomic features that were subsequently selected through a LASSO algorithm for their highest correlation with the outcome and lowest intercorrelation. Selected features and relevant patients' clinical data were used to produce ML algorithms using 70% of the study cases for feature selection, model training and validation with a 10-fold cross-validation, and 30% for model testing. Performances were evaluated using AUC with 95% confidence interval. Main results: A total of 70 coupled tests (63 patients, 35.4% females, mean age 52.2 years) were included and subdivided into a wider cohort of 50 for training and a smaller cohort of 20 for testing. For IFTA $\geq 25\%$, the AUCs in test cohort were 0.60, 0.59, and 0.54 for radiomic features only, clinical variables only, and a combined radiomic–clinical model, respectively. For IFTA $\geq 50\%$, the AUCs in training cohort were 0.89, 0.84, and 0.96, and in the test cohort, they were 0.82, 0.83, and 0.86, for radiomic features only, clinical variables only, and the combined radiomic–clinical model, respectively. Significance: An ML-based MRI radiomic algorithm showed promising discrimination capacity for IFTA > 50%, especially when combined with clinical variables. These results need to be confirmed in larger cohorts.

Keywords: kidney; magnetic resonance; transplantation; radiomics

1. Introduction

End-stage kidney disease (ESKD) is a major cause of morbidity and mortality worldwide, with a high impact on health systems due to the high costs of renal replacement therapies (RRT) [1]. ESKD has a growing incidence due to increasing rates of chronic diseases such as diabetes and hypertension, with a projection of 5 million people receiving RRT in 2030 [2]. Kidney transplantation represents the treatment of choice for ESKD regardless of the cause [3], improving both patients' survival and quality of life with respect to dialysis [4–6]. With constant advances in surgical techniques and medical management substantially reducing the risk of early post-transplant complications, the current focus is set on improving the long-term survival of kidney grafts. IFTA is a common, irreversible, and progressive form of chronic allograft injury, with a multi-factorial aetiology (ischemia-reperfusion damage, episodes of rejection, and immunosuppressive medication nephrotoxicity), characterized by the deposition of extracellular collagen and development of tubular atrophy with rarefaction of peritubular capillaries [7,8]. Interstitial fibrosis/tubular atrophy is unanimously considered a critical predictor of kidney allograft outcomes [9]. Since treatments for rejection and other causes of graft dysfunction bear substantial toxicity and could have limited efficacy, the extent of irreversible graft scarring is crucial information for the clinician in order to evaluate the risks and benefits of specific therapies. The assessment of IFTA is based on the pathological evaluation of renal graft biopsies and its extent is estimated according to the updated Banff classification [10], which differentiates three grades depending on the percentage of the affected cortical tissue: <25% grade one, ≥25% and <50% grade two, and ≥50% grade three. Kidney graft biopsy is considered a relatively safe procedure when performed by a skilled operator; nevertheless, it is an invasive manoeuvre, and its risks are not negligible, with significant bleeding occurring in up to 4% of patients and minor complications in almost one-fifth [11–13].

Radiomics is a quantitative imaging approach that extracts information from diagnostic images produced by routine examinations in the form of numeric continuous variables. Those variables describe intrinsic characteristics of the imaged region of interest that may be related to specific aspects of tissue or its cellular, metabolic, and genetic characteristics. Although radiomics-based works have been developed and validated mainly in oncologic settings [14], there is a growing interest in adopting these techniques to study non-cancerous tissues [15].

In this study, we aimed to evaluate whether radiomics applied to the analysis of kidney-graft MRI may potentially assess a binary outcome of IFTA grading according to Banff classification using different models based on clinical data, radiomic data, and a combination of both.

2. Materials and Methods

2.1. Study Design, Ethics, and Population

We conducted a single-centre retrospective study to explore the feasibility and potential of MRI radiomics in reproducing the histologic IFTA grading in kidney graft biopsies. This study was approved by the "Area Vasta Emilia Nord" Ethical Committee (protocol number 0007669/21, 11 March 2021). Whenever it was impossible to contact the patients, informed consent was waived according to the study's retrospective nature.

We included all consecutive kidney-transplanted patients in a clinical follow-up at the Nephrology and Dialysis Unit of University Hospital of Modena who received a graft biopsy and a contrast-enhanced MRI within six months on clinical indication from 1 January 2012 to 1 March 2021.

Among acquired sequences, SShT2 weighted and post-contrast fat-saturated T1 3D GRE sequences were adopted because of the possible presence of tissue information related to organ water content and vascularization. Two radiologists made a preliminary analysis of MRI images to select sequences with more stable acquisition parameters among different exams. Axial SSh T2 weighted sequence and post-contrast fat-saturated T1 3D

GRE sequence in the venous phase (Supplementary Table S1) were selected due to lower variability and better reproducibility.

Exclusion criteria were unsuitable biopsies (inadequate material) or MRI examinations (incomplete transplanted kidney coverage, artifacts, absence of contrast administration, or adoption of a different acquisition protocol with highly divergent acquisition parameters) (Figure 1).
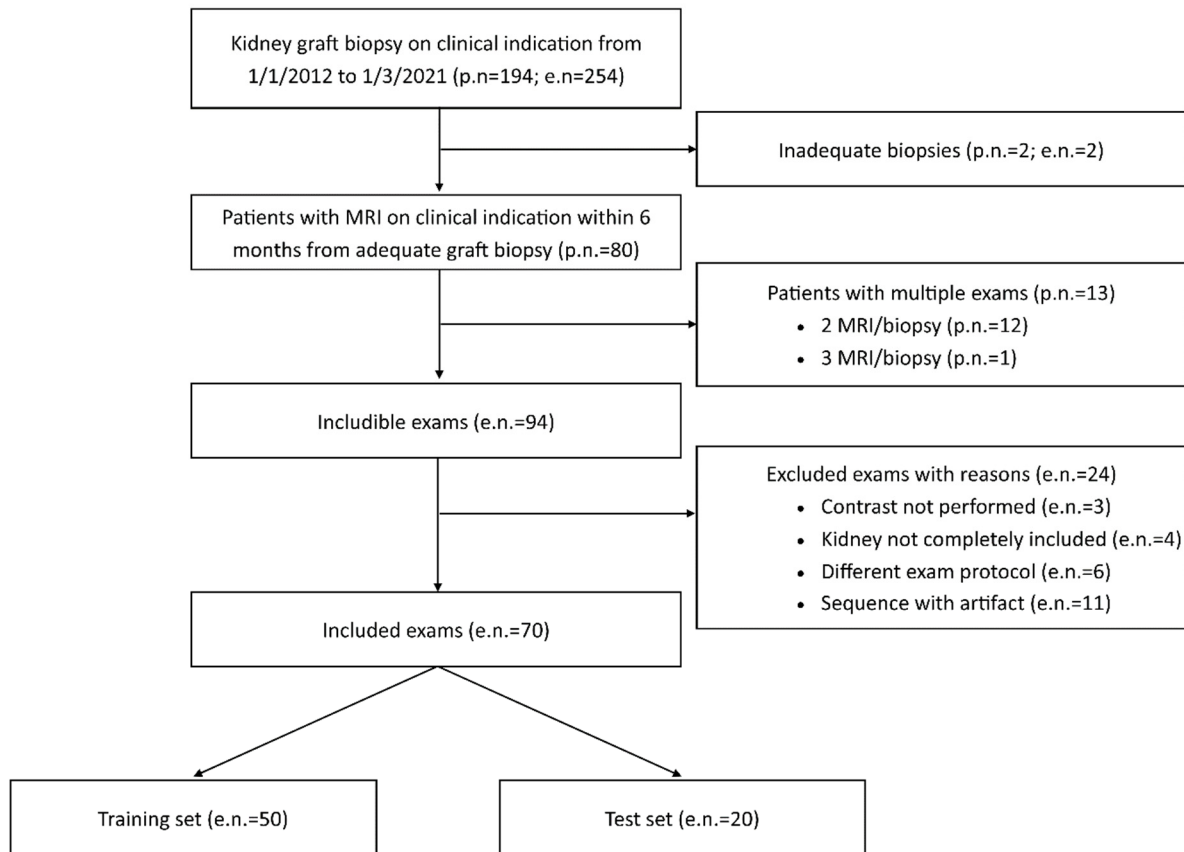
Figure 1. Patient and exam inclusion flow chart. p.n.: patient number; e.n.: exam number, referring to the number of couples of MR/biopsy tests.

Each suitable pair of tests (MRI and biopsy within six months) was considered separately for patients who underwent multiple MRI exams and biopsies.

Our work is a pilot study primarily aimed at exploring the potential of kidney graft MRI radiomics and how radiomic features were related to IFTA. We also attempted to evaluate the accuracy of clinical and radiomic signatures in the assessment of a binary IFTA-based outcome. For the data preparation, we followed the methodology suggested by the Image Biomarker Standardisation Initiative (IBSI) [16]. Hence, we chose a hold-out validation method among the whole cohort to assess our model performance metric. The initial cohort was thus divided into two cohorts, one composed of 70% of samples, used firstly for feature selection and then for model training and validation (10-fold cross-validation), and the other with the remaining 30% of samples used for testing the final model. Samples in the training and validation sets were chosen randomly, but a similar distribution of the primary binary outcome in both sets was maintained.

2.2. Clinical Data

Relevant clinical and laboratory data were collected from patient records after the acquisition of informed consent and personal data treatment agreement from all subjects involved in the study. These included age, sex, ethnicity, BMI, transplant type, transplant

age, estimated glomerular filtration rate (eGFR) according to CKD-EPI equation at biopsy, and proteinuria/creatininuria ratio on single void urine sample immediately before biopsy.

2.3. Kidney Graft Biopsy

All biopsies were performed on clinical indication by one of three expert nephrologists in hospitalized patients, who were adequately informed and provided written consent for the procedure.

Adequate blood pressure control was ensured before the procedure; coagulation parameters and platelet levels were assessed before biopsy; a minimum level of patient collaboration was required. In brief, graft kidney biopsy was carried out with the patient in a supine position, with adequate local anaesthesia and under real-time ultrasound assistance, using a 16- or 18-gauge spring-loaded automatic needle (Bard Magnum, C.R. Bard Inc., Murray Hill, NJ, USA) and targeting the lower pole of the transplanted kidney through a percutaneous anterior abdominal approach. In general, two cores of renal tissue were obtained and allocated to light microscopy, immunofluorescence, and, whenever indicated, electron microscopy technique for diagnosis. Biopsies were revised by one expert nephropathologist (FF) according to the updated Banff classification; the amount of IFTA in renal cortical tissue was assessed using light microscopy through analysis of multiple slides with different stains and reported as deciles and terciles classes (<25%, >25%, >50%). Due to the reduced dimensions of the cohort, we decided to build a model for predicting a binary IFTA outcome. The deciles and terciles IFTA values were thus used as thresholds for our two examined binary IFTA outcomes: IFTA > 25% and IFTA > 50%.

2.4. Image Data Acquisition and Analysis

All the included MRI scans were acquired using a 1.5 T scanner (Philips Achieva, The Best, The Netherlands). SSh T2 sequences were performed without the injection of contrast medium in the axial and coronal planes, and those providing an axial view of the transplanted kidney (short axis) were selected. Axial post-contrast fat-saturated T1 3D GRE sequences were acquired with a dynamic protocol in the arterial phase with bolus tracking technique and after 30, 60, and 90 s, respectively, from the threshold. The last phase was selected due to higher reproducibility among different examinations in terms of contrast distribution. Acquisition parameters of selected sequences are reported in Supplementary Table S1.

Image pre-processing was conducted following IBSI guidelines [16], and the whole radiomic pipeline is summarized in Supplementary Figure S1. Among the included exams, we decided to split the cohort into a training cohort (70%), with which our predictive model was trained and the feature selection process was carried out, and a separate test cohort (30%), with which the model's final performance was tested. The test cohort was never used in the training step. The splitting process was randomized with a balanced separation of the IFTA > 25% binary outcome. The two included exam sequences (T1 3D and T2 2D) were spatially resampled with the B-Spline interpolation method. T1 3D sequences were resampled to an isotropic voxel of $2 \times 2 \times 2$ mm^3 to obtain rotationally invariant features. T2 2D sequences were resampled to a voxel isotropic in two dimensions of $2 \times 2 \times 3$ mm^3, and consequently, features from that image set were computed only in 2 dimensions (2D features). Then, we performed an image intensity normalization using a chosen range, set to -3 standard deviations (σ) and $+3\sigma$ from the mean value of the analysed ROI. Finally, the images were binned by intensity, grouping the original intensity values into five bins (fixed bin number) with the aim of reducing image noise and computational burden, according to the IBSI guidelines.

After the image pre-processing steps, we computed radiomic features from the image sets. Features (intensity-based, shape-based, and second-order) were extracted both from the original images and filtered images, namely wavelets (high–low, low–high, high–high, low–low filters), Laplacian of Gaussian (LoG) with sigma of 1 and 3 mm, and gamma modifier filters [17]. Radiomic features were calculated using an

in-house software pipeline employing the widely used pyRadiomics package (v. 3.0.1) running on Python (v. 3.7.9). The list of the extracted radiomic features is compliant with the image biomarker standardization initiative guidelines [16] and can be found at https://pyradiomics.readthedocs.io/en/latest/features.html (accessed on 29 August 2024).

Since the study is focused on transplanted non-cancerous organs, we chose to avoid the extraction of shape features that were not expected to have a clear correlation with the outcome of interest (IFTA), which is different from tumour shape in oncologic studies. As detailed before, from T2 2D images, we extracted only features derived from 2D matrices, while for T1 3D images, the whole set of features derived from 2D and 3D matrices was used.

A radiologist (FM) blinded to the clinical outcomes segmented the transplanted kidneys using 3D Slicer software version 4.10.1 [18]. For T2 2D images, the manual contouring tool was used to select the whole parenchyma (Figure 2A), the medulla (Figure 2B), and the cortex (Figure 2C) obtained through previous segment subtraction. For T1 3D postcontrast images, the semiautomatic contouring tool based on intensity levels was used for the delineation of the whole parenchyma only since it was not possible to differentiate cortex and medulla reliably (Figure 2D). A different tool was adopted because, on T2-weighted sequences, the intensities of parenchyma and surrounding tissue were too similar to be reliably differentiated by the semiautomatic contouring tool. Moreover, the higher number of images of T1 3D sequences, caused by a lower thickness, made the manual contouring tool too slow to be adopted.
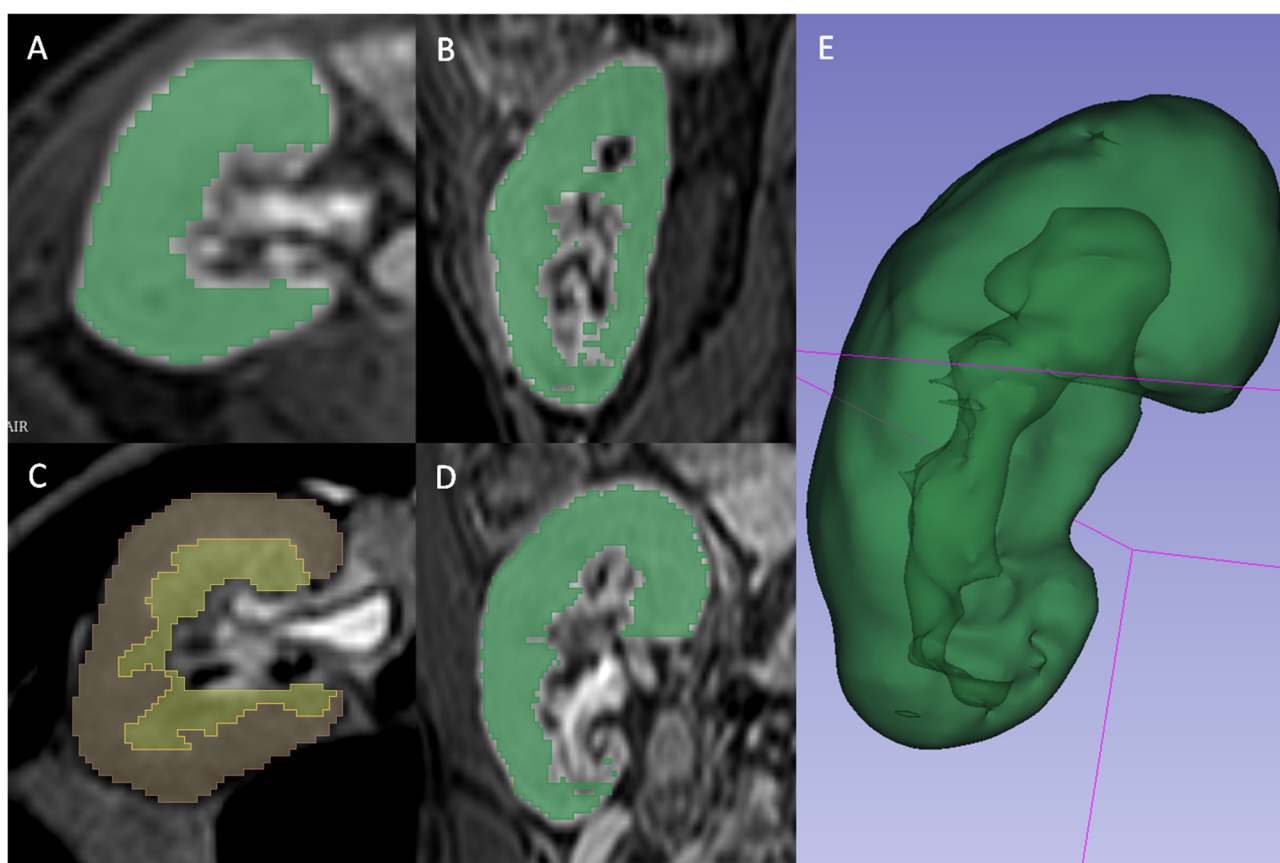
Figure 2. (**A**,**B**,**D**) Gradient Echo T1 weighted 3D venous-phase post-contrast sequence segmented in whole parenchyma by using the intensity-levels based semiautomatic contouring tool of 3D Slicer software, and depicted in axial view (**A**), sagittal view (**B**) and coronal view (**D**). (**C**) Single-shot T2 weighted sequences segmented with 3D Slicer software manual contouring tool, respectively subdivided in medulla (yellow) and cortex (red). (**E**) Three-dimensional rendering of whole parenchyma using T1 3D semiautomatic segmentation.

2.5. Statistical Analyses and Model Building

Associations of ordinal clinical features and IFTA were evaluated using the Kendall rank correlation coefficient "τ" to investigate a possible correlation without assuming a linearity between the two variables. The association between transplant type (categorical variable) and IFTA was assessed using the Spearman rank correlation coefficient. To reduce the dimensionality of our extracted features set, we employed a selection process based on the association between radiomic features and the two aforementioned binary IFTA outcomes (IFTA > 25% and IFTA > 50%), assessed through a least absolute shrinkage and selection operator (LASSO). LASSO is a regression analysis method that performs feature selection and regularization to enhance the prediction accuracy and interpretability of the statistical model it produces. It incorporates a shrinkage penalty term in its optimization function that is proportional to the absolute value of the regression coefficients. Therefore, LASSO drives the coefficients of less relevant variables to zero. This results in a model that retains only a subset of the predictors, deeming them most essential for predicting the outcome variable. The parameter λ controls the strength of the regularization. The choice of λ determines the level of penalty applied: larger values of λ create simpler models by forcing more coefficients to zero, thus preventing overfitting. In comparison, smaller values allow more complex models by retaining more features in the model. In our application, λ was chosen using 10-fold cross-validation (CV). In addition, LASSO handles sets of collinear features by increasing the weight of one of them while setting the others to zero. Because IFTA was considered binary, we used a binomial function to perform our LASSO regression. Relevant features were thus selected by setting a threshold on the LASSO regression coefficients associated with the involved radiomic features.

We built two Ensemble Subspace Discriminant (ESD) models to predict IFTA > 25% and IFTA > 50% using the selected features. ESD models are based on an ensemble technique that combines multiple discriminant analysis classifiers. These classifiers are trained on distinct random subsamples of the feature space, aiming to improve model robustness and accuracy by aggregating predictions from multiple models. Three models were built for both outcomes using radiomic features (radiomic model), clinical variables (clinical model), and a combination of both (combined model). Therefore, we trained a total of 6 ML models through a 10-fold cross-validation method, with which we computed model performance on the training cohort using areas under the curve (AUCs) with 95% CI. Similarly, we calculated the models' performances on our test cohort, which was composed of new cases unseen by the previously trained models.

Feature selection and model-building were conducted using MATLAB (The MathWorks Inc., version R2012b).

3. Results

3.1. Study Population

During the study period (from 1 January 2012 to 1 March 2021), 254 graft biopsies were performed for clinical indication in 194 transplanted patients. For 94 of these 254 procedures, an MRI performed within six months from the kidney biopsy was available for 80 patients (twelve patients had two pairs of tests, and one patient had three pairs of tests). A total of 24 MRI examinations were judged unsuitable and excluded from the study (Figure 1), including 70 paired tests performed on 63 patients.

Those 70 paired tests were randomly subdivided into a training set of 50 tests and a test set of 20 tests, with a fixed proportion of IFTA over 25% and 50% in the two sets. Clinical variables from the whole study population and the two cohorts (training and test sets) are reported in Table 1.

Table 1. Clinical characteristics of the included patients/paired tests. DBD donation after brain death; DCD donation after cardiac death; LD living donor.

Clinical Variables		Total (*n* = 70)	Training Set (*n* = 50)	Test Set (*n* = 20)
Sex (M:F)		45:25	33:17	12:8
Ethnicity (Caucasian/Sub-Saharan)		62:8	45:5	17:3
Age (years) (mean $\pm$ SD)		52.19 $\pm$ 12.76	54.10 $\pm$ 12.36	47.41 $\pm$ 12.78
RM/Biopsy interval (days) (median, IQR)		16, 4–48.75	16, 4–48.75	15, 4.75–49
RM/Biopsy interval > 90 days (*n*, %)		13 (18.57%)	10 (20%)	3 (15%)
BMI (median, IQR)		24.59, 22.47–27.30	25.39, 22.68–27.90	23.50, 22.46–25.39
eGRF at biopsy (median, IQR)		25.68, 11.88–35.51	26.90, 13.08–34.95	20.17, 11.10–38.08
Proteinuria/creatininuria, g/g (median, IQR)		0.79, 0.30–2.10	0.74, 0.21–2.09	0.79, 0.35–2.00
Transplant type (n, %)	DBD	59 (84.29%)	42 (84.00%)	17 (85.00%)
	DCD	2 (2.86%)	2 (4.00%)	0 (0.00%)
	LD	9 (12.86%)	6 (12.00%)	3 (15.00%)
Transplant age (years) (median, IQR)		0.78, 0.31–6.36	1.03, 0.36–0.77	0.62, 0.24–1.78
IFTA % (median, IQR)		20, 10–30	20, 10–37.5	20, 10–30
IFTA $\geq$ 25% (*n*, %)		29 (41.42%)	21 (42.00%)	8 (40.00%)
IFTA $\geq$ 50% (*n*, %)		14 (19.72%)	11 (22.00%)	3 (15.00%)

3.2. MRI Radiomic Feature Selection and Radiomic Signatures

Feature extraction produced 2595 features (1120 2D-spaced from T2 sequences and 1475 3D-spaced from T1 sequences). To select radiomic features associated with the outcome of interest, we performed a binomial LASSO feature selection for both MRI sequences. Selected features for IFTA $\geq$ 25% and IFTA $\geq$ 50% were merged in two radiomic signatures associated with the outcomes. These signatures are reported in Table 2 with every selected radiomic feature and its respective LASSO regression coefficient.

Table 2. T1 and T2 radiomic features used for the prediction of IFTA $\geq$ 25% and IFTA $\geq$ 50% with their respective LASSO regression coefficients at the chosen penalization parameter.

IFTA $\geq$ 25%	LASSO RC	IFTA $\geq$ 50%	LASSO RC
T1 Logsigma30mm3D glrlm LongRunLowGrayLevelEmphasis	1.8	T1 logsigma30mm3D glcm ClusterShade	−0.071
T1 waveletHHL glcm Idmn	-2.6×10^3	T1 waveletHLH glcm Idmn	0.0071
T1 waveletHHH firstorder Skewness	2.4	T1 squareroot firstorder Kurtosis	1.6
T1 logarithm glszm SizeZoneNonUniformity	1.5	T1 exponential glcm Imc2	−8.5
T1 exponential glcm Imc2	−84	T1 exponential gldm SmallDependenceLowGrayLevelEmphasis	610
T1 exponential gldm SmallDependenceLowGrayLevelEmphasis	2.6×10^3	T1 gradient glcm Imc2	57
T2 waveletLH firstorder Mean	8.4	T2 logsigma30mm3D firstorder Median	0.0012
T2 waveletLH firstorder Median	2.5	T2 waveletLH glszm ZoneEntropy	−1.9
T2 waveletLH glszm ZoneEntropy	−88	T2 waveletHH glcm Idmn	2.3
T2 waveletHH glcm Imc1	3.7	T2 waveletHH glcm Imc1	−0.67
T2 waveletHH ngtdm Busyness	2.4	T2 waveletHH ngtdm Busyness	0.10
T2 waveletLL glcm MaximumProbability	-2.9×10^3	T2 logarithm gldm SmallDependenceLowGrayLevelEmphasis	−0.34
T2 exponential glrlm ShortRunLowGrayLevelEmphasis	8.9	T2 exponential glrlm LongRunHighGrayLevelEmphasis	0.56

3.3. Clinical Variables and Signature

The association between meaningful clinical variables and IFTA in our cohort was evaluated using the Kendall test (Supplementary Table S2). Among analysed variables, transplant age and proteinuria/creatininuria ratio at biopsy were associated with the outcome. Other relevant clinical variables were combined with radiomic features in a clinical model since they are considered prognostic factors and are thus commonly used in clinical practice.

3.4. Machine Learning Model Performances

Model performances are reported in Table 3, providing AUCs with respective 95% CI for IFTA $\geq$ 50% and $\geq$25% both in training and in test sets. The models showed a better performance for IFTA $\geq$ 50%. Radiomics performed slightly better than clinical-only models, and models using both radiomic and clinical variables showed a marked increase in AUC. Models showed higher AUCs, especially for IFTA $\geq$ 50%, in the training set than in the test set. The decrease in AUC from training to test set was particularly significant for the model, including both radiomic and clinic characteristics (for IFTA $\geq$ 50%, AUC 0.96, 95%CI 0.84–0.98, and 0.86, 95% CI 0.41–0.99, for the training and test set, respectively) (Figure 3).

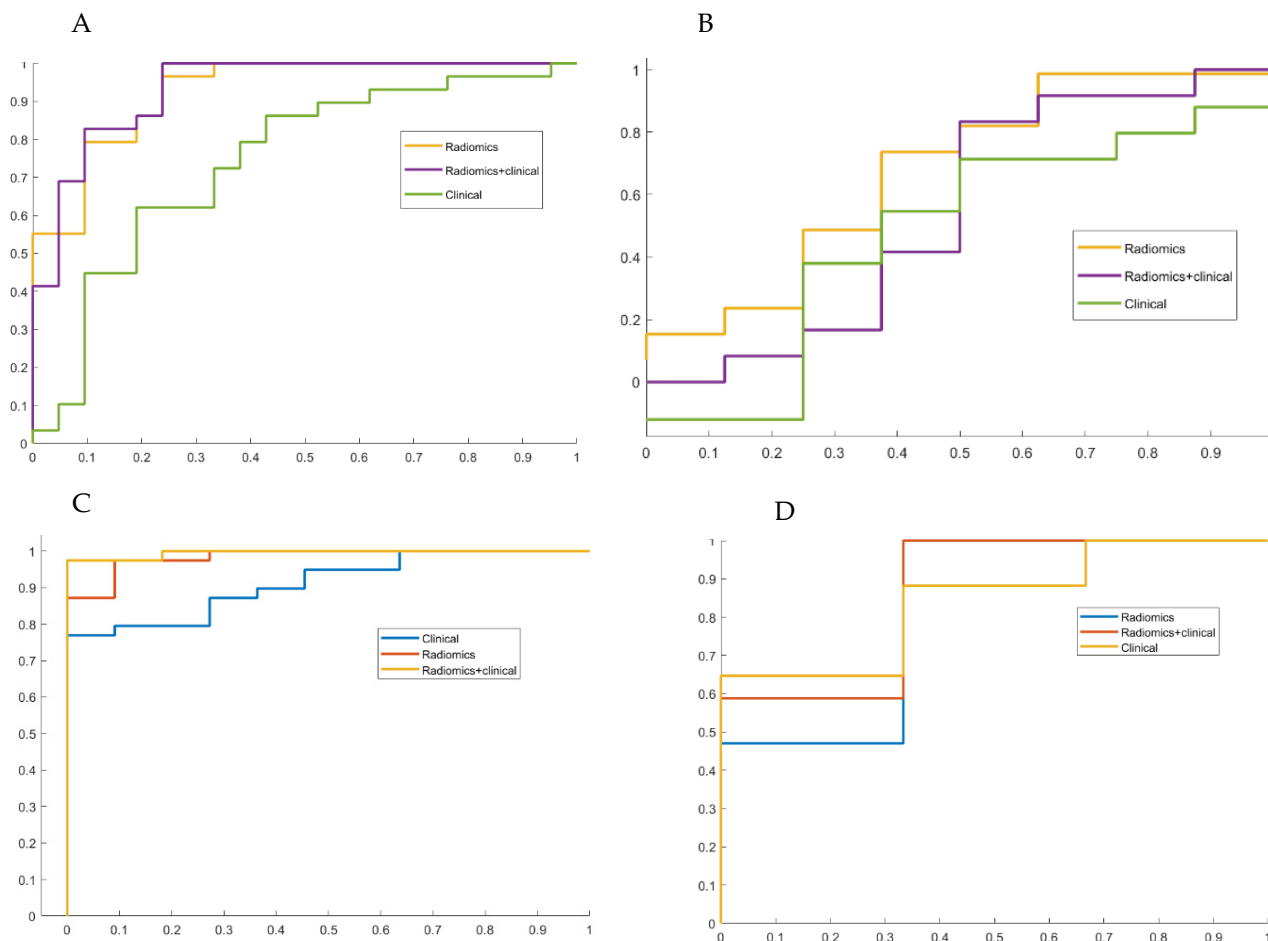**Figure 3.** AUC of the two models based on clinical, radiomic and mixed model. (**A**,**C**) AUC obtained using an internal 10-fold cross validation method on the training cohort ((**A**) IFTA $\geq$ 25%, (**C**) IFTA $\geq$ 50%). (**B**,**D**) AUC obtained applying models deployed on training cohort to a validation cohort composed by new tests ((**B**) IFTA $\geq$ 25%, (**D**) IFTA $\geq$ 50%).

Table 3. Areas under the curve (AUCs) with respective 95% confidence intervals of different models (radiomic-based, clinical-based, mixed) for IFTA ≥ 50% and ≥25% in the training and test sets.

Model Performance			AUC	95% Confidence Interval
IFTA ≥ 25%	Training	Radiomic Model	0.80	(0.64/0.90)
		Clinical Model	0.64	(0.45/0.79)
		Mixed Model	0.83	(0.66/0.93)
	Test	Radiomic Model	0.60	
		Clinical Model	0.59	
		Mixed Model	0.54	
IFTA ≥ 50%	Training	Radiomic Model	0.89	(0.84/0.94)
		Clinical Model	0.83	(0.75/0.91)
		Mixed Model	0.94	(0.90/0.98)
	Test	Radiomic Model	0.82	
		Clinical Model	0.83	
		Mixed Model	0.86	

4. Discussion

In this cohort of 63 transplanted patients who underwent graft biopsy and contrast-enhanced MRI within six months (for a total of 70 coupled procedures), radiomic features were extracted from T2-weighed and post-contrast T1-weighted MR images and selected according to their association with histology outcome (IFTA), and relevant clinical variables of all patients were collected. Using those data, different ML-based MRI radiomic algorithms were deployed. They demonstrated a good discrimination capacity for IFTA > 50%, especially when radiomic features were merged with clinical variables (AUCs for clinical–radiomic combined model were 0.96 and 0.86 in the training set and test set, respectively).

Diagnostic kidney graft biopsy currently remains an essential procedure to provide a precise diagnosis and guide the management of transplanted patients. However, the decision to proceed to biopsy could not always be as straightforward, especially with the increased complexity of patients' co-morbidities and treatments potentially interfering with adequate haemostasis (antiplatelets or anticoagulants). Moreover, the prediction of severe chronic parenchymal damage (i.e., IFTA > 50%) in the transplanted kidney may justify avoidance of graft biopsy to diagnose rejection since the patient would likely not be a candidate for further immunosuppressive treatment. In addition, a higher degree of parenchymal scarring has been reported as a risk factor for bleeding, at least in native kidneys [19]. Although biological parameters, such as eGFR and proteinuria, are commonly used in clinical practice to estimate the degree of chronic injury, their reliability is questionable [20]. Consequently, the development of robust and non-invasive tools to detect the presence of extensive IFTA in kidney grafts would represent a step forward in personalizing patients' management.

In recent years, increased attention has been devoted to using MRI for the study of transplanted kidneys, with a specific focus on multiparametric/functional techniques [7,21,22] and elastography [23]. Indeed, even if results are promising, especially for the prediction of kidney fibrosis, the availability of these techniques has not been widespread yet.

Radiomics is an advanced technique that allows the extraction and analysis of a large number of quantitative features from medical images, acquired during the course of the routine clinical pathway of the patient. Available data in non-oncological kidney diseases are scarce, with only a few available studies investigating T2-weighted MR-based radiomics for the evaluation of autosomal dominant polycystic kidney disease progression [24,25] and decreased glomerular filtration rate during chronic lithium treatment [26], or US-derived radiomics in the classification of IgA nephropathy [27]. Few studies have

evaluated the potential role of US-based radiomics in the assessment of post-transplant renal function [28–30] or the identification and discrimination of different stages of chronic kidney disease [31–34]. Zhu et al. [35] studied radiomics signatures on elastography ultrasound images to predict kidney injury progression. Recently, Zhi et al. [36] integrated clinical variables and MRI radiomic features into a convolutional neural network, reporting stable performance in discriminating the cause of kidney allograft dysfunction in a cohort of 252 patients from China. Several studies [21,37,38] have confirmed the correlation between renal function indicators (serum creatinine, albuminuria, and eGFR) and conventional quantitative fMRI parameters. Furthermore, radiomic studies based on MRI images carry the intrinsic issue of dealing with the variability of the image sets in the context of multicentric studies.

We selected T2 and post-contrast T1 MR-derived radiomic features with the highest association with histopathological IFTA in transplanted patients who underwent graft biopsy and MRI. In our cohort of patients, models produced to assess IFTA $\geq$ 50% showed a good performance. In particular, the combined model (with clinical and radiomics variables) lowered its performance, in terms of AUC, from 0.96 in the training set to 0.86 in the test set. With respect to models produced to assess IFTA > 25%, their performances can be considered poor, with a much higher decrease in terms of AUC from training to test sets. As expected, the decrease in AUC from training to test set was stronger for the combined model, probably due to the higher number of variables used to build the model. Notwithstanding the small size of the test set, the consistency of AUC in training and test sets in the prediction of IFTA > 50% seems to suggest that our methodology allowed for a reduction in overfitting. These results highlight the presence of valuable information in MRI images, which may be predictive of the chosen threshold of the histological IFTA grading. Indeed, MRI T2 relaxation time is related to kidney tissue water content, which is a sensitive indicator of IFTA [36], and T1 postcontrast images are influenced by organ vascularization and penetration of contrast medium through the whole parenchyma, which could be reduced by advanced parenchymal scarring. The information that more than half of kidney parenchyma is affected could be clinically meaningful, preventing further invasive procedures since they would not alter patients' management.

The main limitation of our study is the small sample size; moreover, its limited number of events (IFTA $\geq$ 25% and $\geq$50%) in the test set does not allow us to derive definitive conclusions, emphasizing the feasibility of our proposed study. In addition, the number of selected features, based on the relevance of their associated LASSO coefficient, could be high compared to the sample size. Thus, a more extensive training cohort combined with an external validation test set to exclude the risk of overfitting needs must be accounted for in future studies to confirm our results. In addition, the single-centre design implied a lack of external validation, which could also produce a bias due to the different clinical protocols employed in the image acquisition step. We tried to overcome this limitation by building a sub-cohort of patients for feature selection and model training and a separate sub-cohort for testing. Another limitation of the study was its retrospective nature, which prevented us from designing specific MR sequences that were more fit for radiomic feature extraction. Hence, some sequences were affected by high parameter variability; in fact, only SSh T2 and post-Gd fat-saturated T1 3D sequences were selected. However, the selection of both T1 and T2 weighting and post-contrast images probably allowed us to collect most MRI-based information for transplanted kidney tissue characterization. Moreover, including transplanted patients referred to MRI for clinical reasons could have led to a selection bias. Finally, we acknowledge that the allowed temporal interval between biopsy and MRI (within six months) may be considered to be long enough for important changes in kidney tissue to develop. Nevertheless, less than 20% of the enrolled patients had an interval longer than three months. Furthermore, since IFTA is a chronic condition following kidney transplantation, we did not expect a relevant IFTA modification in the considered time interval.

Another limitation is that, in our research, we only used one segmentation, which was manually performed by an expert radiologist. This approach may represent a limitation of our study, as it does not allow for an evaluation of how alternative segmentation methods might have impacted the outcomes. The impact of the segmentation method in MRI images could be addressed in future work to further validate and potentially enhance the robustness of our findings.

The major strength of this study is our effort to pursue the standardization guidelines, especially critical in the pre-processing step for MRI images due to their intrinsic nature, following IBSI indications [14], despite the retrospective nature of the study and the small sample size. Shape features were not extracted. While these may be of interest when collecting radiomics in oncology, where size, shape, and borders are expected to correlate with the nature of the tissue and its histology, diagnosis, and prognosis, in the case of a transplanted organ, they are not likely to introduce meaningful, informative value. Still, they could add unwanted complexity to the analysis and increase the probability of observing false associations and overfitting. Another strength of our study is that the ground truth of our prediction models is represented by kidney graft biopsy, which is currently considered the gold standard for histopathological diagnosis of kidney allograft (and IFTA quantification), in opposition to the only available study on MRI radiomics in kidney transplant patients [36], where a relevant proportion of predicted diagnosis was assigned based on clinical judgment.

5. Conclusions

In conclusion, this preliminary study shows that MR-derived radiomics of kidney grafts is feasible and may add informative value to clinical factors in predicting an IFTA binary outcome (>50%). Prospective studies, possibly with predefined MR acquisition parameters, are necessary to confirm this hypothesis. Further studies are needed to evaluate the potential role of MR radiomics as a predictor of long-term graft dysfunction. Our data establish a new radiological/histopathological correlation, which could support future clinical applications; should the good performance of MR radiomics be confirmed, it could be used as a non-invasive tool to assess and monitor chronic parenchymal injury in kidney-transplanted patients.

Supplementary Materials: The following supporting information can be downloaded at: https://www.mdpi.com/article/10.3390/info15090537/s1, Table S1: Acquisition parameters of every included study; Table S2: Associations between clinical variables and IFTA as a continuous variable computed using Kendall's Tau correlation coefficient/Spearman for transplant type variable; higher absolute value of τ corresponds to higher correlation between variables and outcome; Figure S1: Radiomic pipeline followed in our study.

Author Contributions: Conceptualization, F.M., G.B., M.B., R.S. and F.F.; Data curation, F.M., G.B. and M.O.; Formal analysis, V.T.; Funding acquisition, F.F.; Investigation, V.T., F.M., L.V. and P.G.R.; Methodology, V.T., F.M., G.B., P.G.R. and M.O.; Project administration, G.B., M.I., P.P. and F.F.; Resources, G.B., M.B., R.S., M.I., G.L., P.P., A.P., G.A. and G.D.; Software, V.T. and L.V.; Supervision, G.B., M.B., M.I. and F.F.; Validation, V.T., F.M., M.B., P.G.R. and M.O.; Visualization, V.T., F.M., L.V., R.S., G.L., P.P., P.G.R., M.O., A.P., G.A. and G.D.; Writing—original draft, V.T. and F.M.; Writing—review and editing, G.B., M.B., L.V., P.G.R., M.O. and F.F. All authors have read and agreed to the published version of the manuscript.

Institutional Review Board Statement: The study was conducted in accordance with the Declaration of Helsinki, and approved by the Ethics Committee of "Area Vasta Emilia Nord" (protocol code 0007669/21 and date of approval 11 March 2021).

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study. Written informed consent has been obtained from the patient(s) to publish this paper.

Data Availability Statement: Informed consent and personal data treatment were obtained from all subjects involved in the study. Written informed consent was obtained from patients to publish this paper. Participant data that underlie the results reported in this manuscript will be shared after de-identification, beginning 6 months and ending at least 7 years after article publication, to researchers who provide a methodologically sound proposal consistent with the original study's objectives. Proposals and data access requests should be directed to the Area Vasta Emilia Nord (AVEN) Ethics Committee at CEReggioemilia@ausl.re.it as well as to the Authors at the Epidemiology Unit of AUSL–IRCCS di Reggio Emilia at info.epi@ausl.re.it, who are the data guardians. Data requestors must sign a data access agreement to gain access.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Jha, V.; Garcia-Garcia, G.; Iseki, K.; Li, Z.; Naicker, S.; Plattner, B.; Saran, R.; Wang, A.Y.-M.; Yang, C.-W. Chronic Kidney Disease: Global Dimension and Perspectives. *Lancet* **2013**, *382*, 260–272. [CrossRef] [PubMed]
2. Liyanage, T.; Ninomiya, T.; Jha, V.; Neal, B.; Patrice, H.M.; Okpechi, I.; Zhao, M.; Lv, J.; Garg, A.X.; Knight, J.; et al. Worldwide Access to Treatment for End-Stage Kidney Disease: A Systematic Review. *Lancet* **2015**, *385*, 1975–1982. [CrossRef]
3. Suthanthiran, M.; Strom, T.B. Renal Transplantation. *N. Engl. J. Med.* **1994**, *331*, 365–376. [CrossRef]
4. Schnuelle, P.; Lorenz, D.; Trede, M.; Van Der Woude, F.J. Impact of Renal Cadaveric Transplantation on Survival in End-Stage Renal Failure: Evidence for Reduced Mortality Risk Compared with Hemodialysis during Long-Term Follow-Up. *J. Am. Soc. Nephrol.* **1998**, *9*, 2135–2141. [CrossRef] [PubMed]
5. Port, F.K.; Wolfe, R.A.; Mauger, E.A.; Berling, D.P.; Jiang, K. Comparison of Survival Probabilities for Dialysis Patients vs. Cadaveric Renal Transplant Recipients. *JAMA* **1993**, *270*, 1339–1343. [CrossRef] [PubMed]
6. Ojo, A.O.; Port, F.K.; Wolfe, R.A.; Mauger, E.A.; Williams, L.; Berling, D.P. Comparative Mortality Risks of Chronic Dialysis and Cadaveric Transplantation in Black End-Stage Renal Disease Patients. *Am. J. Kidney Dis.* **1994**, *24*, 59–64. [CrossRef]
7. Wang, W.; Yu, Y.; Wen, J.; Zhang, M.; Chen, J.; Cheng, D.; Zhang, L.; Liu, Z. Combination of Functional Magnetic Resonance Imaging and Histopathologic Analysis to Evaluate Interstitial Fibrosis in Kidney Allografts. *Clin. J. Am. Soc. Nephrol.* **2019**, *14*, 1372–1380. [CrossRef]
8. Poggio, E.D. Imaging as a Noninvasive Tool for Evaluating Interstitial Fibrosis in Kidney Allografts. *Clin. J. Am. Soc. Nephrol.* **2019**, *14*, 1286–1287. [CrossRef]
9. Chand, S.; Atkinson, D.; Collins, C.; Briggs, D.; Ball, S.; Sharif, A.; Skordilis, K.; Vydianath, B.; Neil, D.; Borrows, R. The Spectrum of Renal Allograft Failure. *PLoS ONE* **2016**, *11*. [CrossRef]
10. Roufosse, C.; Simmonds, N.; Clahsen-van Groningen, M.; Haas, M.; Henriksen, K.J.; Horsfield, C.; Loupy, A.; Mengel, M.; Perkowska-Ptasińska, A.; Rabant, M.; et al. Reference Guide to the Banff Classification of Renal Allograft Pathology. *Transplantation* **2018**, *102*, e497. [CrossRef]
11. Morgan, T.A.; Chandran, S.; Burger, I.M.; Zhang, C.A.; Goldstein, R.B. Complications of Ultrasound-Guided Renal Transplant Biopsies. *Am. J. Transplant.* **2016**, *16*, 1298–1305. [CrossRef] [PubMed]
12. Furness, P.N.; Philpott, C.M.; Chorbadjian, M.T.; Nicholson, M.L.; Bosmans, J.L.; Corthouts, B.L.; Bogers, J.J.; Schwarz, A.; Gwinner, W.; Haller, H.; et al. Protocol Biopsy of the Stable Renal Transplant: A Multicenter Study of Methods and Complication Rates. *Transplantation* **2003**, *76*, 969–973. [CrossRef] [PubMed]
13. Schwarz, A.; Gwinner, W.; Hiss, M.; Radermacher, J.; Mengel, M.; Haller, H. Safety and Adequacy of Renal Transplant Protocol Biopsies. *Am. J. Transplant.* **2005**, *5*, 1992–1996. [CrossRef]
14. Reuzé, S. Prediction of Cervical Cancer Recurrence Using Textural Features Extracted from 18F-FDG PET Images Acquired with Different Scanners. *Oncotarget* **2017**, *8*, 43169–43179. [CrossRef]
15. Zhao, D.; Wang, W.; Tang, T.; Zhang, Y.-Y.; Yu, C. Current Progress in Artificial Intelligence-Assisted Medical Image Analysis for Chronic Kidney Disease: A Literature Review. *Comput. Struct. Biotechnol. J.* **2023**, *21*, 3315–3326. [CrossRef] [PubMed]
16. Zwanenburg, A. The Image Biomarker Standardization Initiative: Standardized Quantitative Radiomics for High-Throughput Image-based Phenotyping. *Radiology* **2020**, *295*, 328–338. [CrossRef]
17. Griethuysen, J.J.M.; Fedorov, A.; Parmar, C.; Hosny, A.; Aucoin, N.; Narayan, V.; Beets-Tan, R.G.H.; Fillon-Robin, J.C.; Pieper, S.; Aerts, H.J.W.L. Computational Radiomics System to Decode the Radiographic Phenotype. *Cancer Res.* **2017**, *77*, 104–107. [CrossRef]
18. Fedorov, A.; Beichel, R.; Kalpathy-Cramer, J.; Finet, J.; Fillion-Robin, J.C.; Pujol, S.; Bauer, C.; Jennings, D.; Fennessy, F.; Sonka, M.; et al. 3D Slicer as an Image Computing Platform for the Quantitative Imaging Network. *Magn. Reson. Imaging* **2012**, *30*, 1323–1341. [CrossRef]
19. Chantaduly, C.; Troutt, H.R.; Perez Reyes, K.A.; Zuckerman, J.E.; Chang, P.D.; Lau, W.L. Artificial Intelligence Assessment of Renal Scarring (AIRS Study). *Kidney360* **2021**, *3*, 83–90. [CrossRef]
20. Berchtold, L.; Crowe, L.A.; Friedli, I.; Legouis, D.; Moll, S.; De Perrot, T.; Martin, P.Y.; Vallee, J.P.; De Seigneux, S. Diffusion Magnetic Resonance Imaging Detects an Increase in Interstitial Fibrosis Earlier than the Decline of Renal Function. *Nephrol. Dial. Transplant.* **2020**, *35*, 1274–1276. [CrossRef]

21. Berchtold, L.; Crowe, L.A.; Combescure, C.; Kassaï, M.; Aslam, I.; Legouis, D.; Moll, S.; Martin, P.Y.; de Seigneux, S.; Vallée, J.P. Diffusion-Magnetic Resonance Imaging Predicts Decline of Kidney Function in Chronic Kidney Disease and in Patients with a Kidney Allograft. *Kidney Int.* **2022**, *101*, 804–813. [CrossRef] [PubMed]

22. Bane, O.; Hectors, S.J.; Gordic, S.; Kennedy, P.; Wagner, M.; Weiss, A.; Khaim, R.; Yi, Z.; Zhang, W.; Delaney, V.; et al. Multiparametric Magnetic Resonance Imaging Shows Promising Results to Assess Renal Transplant Dysfunction with Fibrosis. *Kidney Int.* **2020**, *97*, 414–420. [CrossRef] [PubMed]

23. Kirpalani, A.; Hashim, E.; Leung, G.; Kim, J.K.; Krizova, A.; Jothy, S.; Deeb, M.; Jiang, N.N.; Glick, L.; Mnatzakanian, G.; et al. Magnetic Resonance Elastography to Assess Fibrosis in Kidney Allografts. *Clin. J. Am. Soc. Nephrol.* **2017**, *12*, 1671–1679. [CrossRef] [PubMed]

24. Cong, L.; Hua, Q.Q.; Huang, Z.Q.; Ma, Q.L.; Wang, X.M.; Huang, C.C.; Xu, J.X.; Ma, T. A Radiomics Method Based on MR FS-T2WI Sequence for Diagnosing of Autosomal Dominant Polycystic Kidney Disease Progression. *Eur. Rev. Med. Pharmacol. Sci.* **2021**, *25*, 5769–5780. [CrossRef]

25. Kremer, L.E.; Chapman, A.B.; Armato, S.G. Magnetic Resonance Imaging Preprocessing and Radiomic Features for Classification of Autosomal Dominant Polycystic Kidney Disease Genotype. *J. Med. Imaging (Bellingham)* **2023**, *10*, 064503. [CrossRef]

26. Beunon, P.; Barat, M.; Dohan, A.; Cheddani, L.; Males, L.; Fernandez, P.; Etain, B.; Bellivier, F.; Marlinge, E.; Vrtovsnik, F.; et al. MRI-Based Kidney Radiomic Analysis during Chronic Lithium Treatment. *Eur. J. Clin. Investig.* **2022**, *52*. [CrossRef]

27. Zhang, L.; Chen, Z.; Feng, L.; Guo, L.; Liu, D.; Hai, J.; Qiao, K.; Chen, J.; Yan, B.; Cheng, G. Preliminary Study on the Application of Renal Ultrasonography Radiomics in the Classification of Glomerulopathy. *BMC Med. Imaging* **2021**, *21*, 115. [CrossRef]

28. Zhu, L.; Huang, R.; Li, M.; Fan, Q.; Zhao, X.; Wu, X.; Dong, F. Machine Learning-Based Ultrasound Radiomics for Evaluating the Function of Transplanted Kidneys. *Ultrasound Med. Biol.* **2022**, *48*, 1441–1452. [CrossRef]

29. Zhu, L.; Huang, R.; Zhou, Z.; Fan, Q.; Yan, J.; Wan, X.; Zhao, X.; He, Y.; Dong, F. Prediction of Renal Function 1 Year After Transplantation Using Machine Learning Methods Based on Ultrasound Radiomics Combined With Clinical and Imaging Features. *Ultrason. Imaging* **2023**, *45*, 85–96. [CrossRef]

30. Abbasian Ardakani, A.; Mohammadi, A.; Khalili Najafabad, B.; Abolghasemi, J. Assessment of Kidney Function After Allograft Transplantation by Texture Analysis. *Iran. J. Kidney Dis.* **2017**, *11*, 157–164.

31. Chen, C.-J.; Pai, T.-W.; Hsu, H.-H.; Lee, C.-H.; Chen, K.-S.; Chen, Y.-C. Prediction of Chronic Kidney Disease Stages by Renal Ultrasound Imaging. *Enterp. Inf. Syst.* **2020**, *14*, 178–195. [CrossRef]

32. Bandara, M.S.; Gurunayaka, B.; Lakraj, G.; Pallewatte, A.; Siribaddana, S.; Wansapura, J. Ultrasound Based Radiomics Features of Chronic Kidney Disease. *Acad. Radiol.* **2022**, *29*, 229–235. [CrossRef]

33. Iqbal, F.; Pallewatte, A.S.; Wansapura, J.P. Texture Analysis of Ultrasound Images of Chronic Kidney Disease. In Proceedings of the 17th International Conference on Advances in ICT for Emerging Regions, ICTer 2017—Proceedings, Colombo, Sri Lanka, 6–9 September 2017; pp. 299–303. [CrossRef]

34. Lee, S.; Kang, M.; Byeon, K.; Lee, S.E.; Lee, I.H.; Kim, Y.A.; Kang, S.-W.; Park, J.T. Machine Learning-Aided Chronic Kidney Disease Diagnosis Based on Ultrasound Imaging Integrated with Computer-Extracted Measurable Features. *J. Digit. Imaging* **2022**, *35*, 1091–1100. [CrossRef]

35. Zhu, M.; Tang, L.; Yang, W.; Xu, Y.; Che, X.; Zhou, Y.; Shao, X.; Zhou, W.; Zhang, M.; Li, G.; et al. Predicting Progression of Kidney Injury Based on Elastography Ultrasound and Radiomics Signatures. *Diagnostics* **2022**, *12*, 2678. [CrossRef] [PubMed]

36. Zhi, R.; Zhang, X.-D.; Hou, Y.; Jiang, K.-W.; Li, Q.; Zhang, J.; Zhang, Y.-D. RtNet: A Deep Hybrid Neural Networks for the Identification of Acute Rejection and Chronic Allograft Nephropathy after Renal Transplantation Using Multiparametric MRI. *Nephrol. Dial. Transplant.* **2022**, *37*, 2581–2590. [CrossRef] [PubMed]

37. Li, X.-S.; Zhang, Q.-J.; Zhu, J.; Zhou, Q.-Q.; Yu, Y.-S.; Hu, Z.-C.; Xia, Z.-Y.; Wei, L.; Yin, X.-D.; Zhang, H. Assessment of Kidney Function in Chronic Kidney Disease by Combining Diffusion Tensor Imaging and Total Kidney Volume. *Int. Urol. Nephrol.* **2022**, *54*, 385–393. [CrossRef]

38. Dillman, J.R.; Benoit, S.W.; Gandhi, D.B.; Trout, A.T.; Tkach, J.A.; VandenHeuvel, K.; Devarajan, P. Multiparametric Quantitative Renal MRI in Children and Young Adults: Comparison between Healthy Individuals and Patients with Chronic Kidney Disease. *Abdom. Radiol.* **2022**, *47*, 1840–1852. [CrossRef]