Full length article

A deep learning image analysis method for renal perfusion estimation in pseudo-continuous arterial spin labelling MRI

Anne Oyarzun-Domeño [a,b,*], Izaskun Cia [a], Rebeca Echeverria-Chasco [b,d],
María A. Fernández-Seara [b,d], Paloma L. Martin-Moreno [b,e], Nuria Garcia-Fernandez [b,e],
Gorka Bastarrika [b,d], Javier Navallas [a,b], Arantxa Villanueva [a,b,c]

[a] *Electrical Electronics and Communications Engineering, Public University of Navarre, 31006 Pamplona, Spain*
[b] *IdiSNA, Health Research Institute of Navarra, 31008, Spain*
[c] *Institute of Smart Cities (ISC), Health Research Institute of Navarra, 31006, Pamplona, Spain*
[d] *Department of Radiology, Clínica Universidad de Navarra, 31008 Pamplona, Spain*
[e] *Department of Nephrology, Clínica Universidad de Navarra, 31008 Pamplona, Spain*

ABSTRACT

Accurate segmentation of renal tissues is an essential step for renal perfusion estimation and postoperative assessment of the allograft. Images are usually manually labeled, which is tedious and prone to human error. We present an image analysis method for the automatic estimation of renal perfusion based on perfusion magnetic resonance imaging. Specifically, non-contrasted pseudo-continuous arterial spin labeling (PCASL) images are used for kidney transplant evaluation and perfusion estimation, as a biomarker of the status of the allograft. The proposed method uses machine/deep learning tools for the segmentation and classification of renal cortical and medullary tissues and automates the estimation of perfusion values. Data from 16 transplant patients has been used for the experiments. The automatic analysis of differentiated tissues within the kidney, such as cortex and medulla, is performed by employing the time-intensity-curves of non-contrasted T_1-weighted MRI series. Specifically, using the Dice similarity coefficient as a figure of merit, results above 93%, 92% and 82% are obtained for whole kidney, cortex, and medulla, respectively. Besides, estimated cortical and medullary perfusion values are considered to be within the acceptable ranges within clinical practice.

1. Introduction

Chronic kidney disease (CKD) is characterized by a progressive and irreversible loss of kidney function and currently considered one of the major public health challenges by the World Health Organization [1]. CKD remains a widespread and crucial public health problem afflicting >12% of the population worldwide [1]. It can lead to kidney function loss, cardiovascular disease, and premature death. In case of renal failure, dialysis or transplant is needed. While both treatments have advantages and disadvantages, studies show that patients who have a successful kidney transplant live longer than patients treated with dialysis, and also have better quality of life [2]. In recent years, the risk of renal transplant loss has been reduced in the short term [3], but it still remains a concern in the mid- and long- term. The renal blood flow

(RBF) or perfusion has an enormous potential value for clinical nephrologists as it enables the identification of perfusion impairment as an emerging biomarker of renal dysfunction in kidney transplant patients [4]. The major portion of blood flow is delivered to the cortex and a small quantity is delivered to the medulla [5]. However, the blood flow in the medulla is also considerably sensitive to reductions in RBF and oxygen delivery [5]. The vasoconstriction of renal arterial tree and changes in the intrarenal vascular resistance are determinant of the overall RBF [5]. Thus, precise perfusion measurement is desirable as it may assess in the research, diagnosis, prognosis and the evaluation of CKD condition.

The quantitative measurement of RBF has been achieved by means of several imaging techniques, as positron emission tomography (PET) [6,7], computed tomography (CT) [8,9] and magnetic resonance

imaging (MRI) techniques [10,11]. MRI is able to assess renal micro-anatomy, perfusion, and diffusion, among others [12]. More specifically, arterial spin labeling (ASL) is a non-invasive MRI technique that allows the characterization of RBF using magnetically labeled arterial blood water spins as endogenous tracer, offering an interesting alternative for renal perfusion estimation, especially in patients with renal dysfunction for whom the administration of gadolinium-based contrast agents could be contraindicated [10]. Pseudo-continuous ASL (PCASL) is a variant of ASL that uses a combination of a train of radiofrequency pulses with slice-selective gradients to alter the longitudinal magnetization of the blood that irrigates the organ of interest. PCASL has been accepted as consensus solution for brain perfusion measurement [11]. Furthermore, it shows up as a potential alternative to dynamic contrast enhanced - MRI (DCE-MRI). The quality of DCE-MRI imaging modality has expanded its number of applications and thus the number of technological developments involved. However, DCE-MRI is limited by the use of contrast agent that can affect CKD patient safety in renal impairment cases. For that reason, the non-invasive nature of the ASL method provides highly beneficial added value for clinical practice and has already become a valid technique for renal function measurements. On the other hand, for transplant patients, which are the core of this study, it emerges as the only viable solution given the high limitations when using contrast agents. The RBF estimation obtained from ASL images provides renal functional measurements; however, it entails extra segmentation work. The segmentation of the kidney in the image is a necessary and crucial stage for perfusion estimation and its automation is, consequently, key for being able to incorporate the clinical knowledge obtained into daily practice. However, manual segmentation is tedious and prone to error. Furthermore, semi-automated approaches with partial intervention of an operator are dependent on the specialist and require the evaluation of replicability and inter-observer variation. Thereupon, there is an increasing need to transform these techniques into fully-automated or at least less operator-dependent. According to this need, the literature shows a significant interest in the development of automated or semi-automated tools for the semantic segmentation of the kidney and its compartments. The development of machine learning (ML) and deep learning (DL) is increasing as a consequence of the successful solutions obtained in computer vision and related fields. In particular, models drawn from the field of DL are showing excellent performance ranging from automotive to industrial vision and this extends to medical imaging as well [13–15]. DL is part of ML, inspired by the way in which the brain carries out learning processes by imitating biological neurological networks [16]. It employs a series of layers with mathematically interconnected nodes. The weights of these nodes are adjusted from an optimization process during the course of a training stage to solve a specific problem (detection, segmentation, and classification, among others). To date, the applications of ML and DL models in renal MRI are scarce compared to the works found for computerized tomography (CT) [17,18] and DCE-MRI [19–21]. Some approaches that fully or partially automate the process of kidney segmentation and tissue differentiation on non-contrasted MRI have been published in recent years. In [22], an evaluation of the reproducibility of ASL kidney perfusion measurement is performed, based on semi-automatic segmentation of the whole kidney via intelligent scissor method and cortical and medullary differentiation by means of clusterization. In this work, high correlation of cortical, medullary and whole kidney perfusion is achieved, but it is limited to healthy PCASL renal data. Volumetric assessment of the kidney is performed in [23] based on thresholding and shape detection algorithms on non-contrasted T_1- and T_2-w MRI. In this work, healthy MRI data is used, and applied thresholds for intensity-value based separation highly depend on the dataset. Regarding the use of DL based techniques applied to renal ASL-MRI, a cascaded U-Net is proposed in [24,25] for renal cortical perfusion quantification on ASL and volumetric quantification on T_2- w MRI images, respectively. In addition, in [26], a Mask R-CNN is implemented for ASL and T_2^* image processing and enables multiclassification of
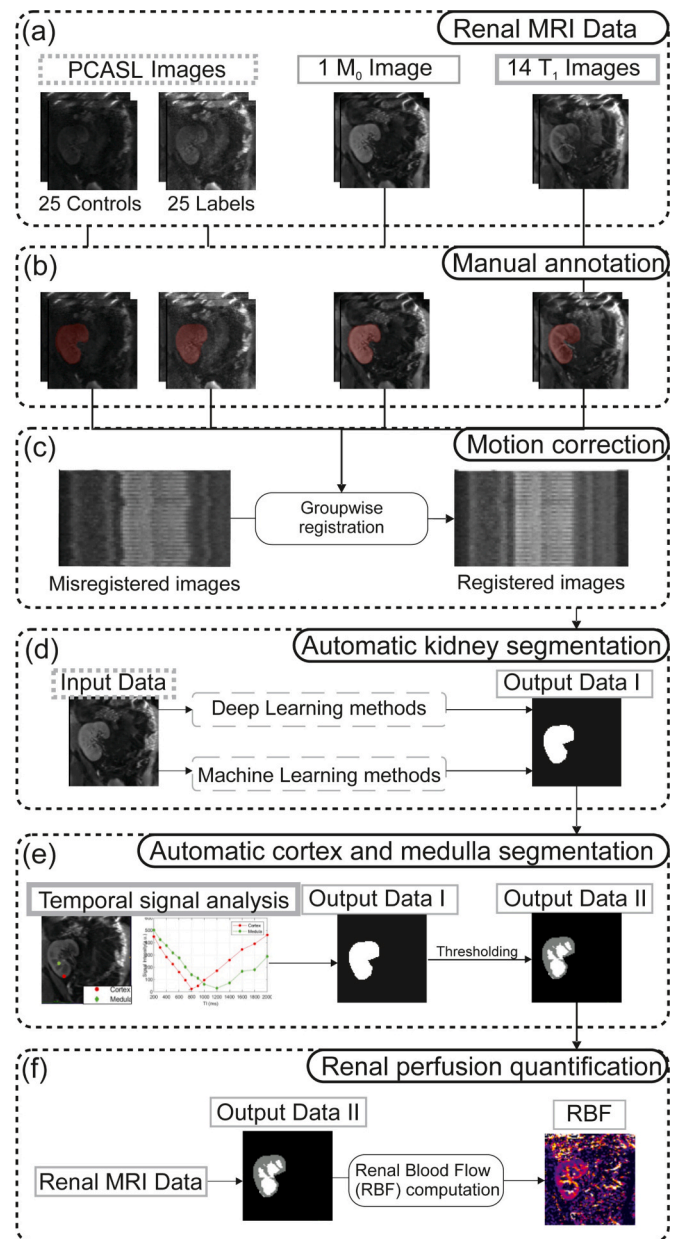

Fig. 1. Renal, cortical, and medullary tissue segmentation pipeline. (a) Renal MRI consists of ASL (PCASL), M_0 reference image and T_1-w image series. Manual annotation (b) has to be performed in order to generate GT masks and to train the implemented methods. Once data is labeled, MRI data should be motion corrected (c) prior to any segmentation approach. The main tasks in this work are the automatic kidney (d) and cortex and medulla segmentation (e). Finally, once tissues are segmented, RBF maps and renal perfusion values are estimated (f).

renal tissue and renal perfusion and oxygenation estimation. The present work aims to provide a fully-automated segmentation pipeline for the estimation of perfusion values on PCASL-MRI data using T_1-w MRI image series as support for the differentiation of cortex and medullary tissues. The principal objective of the work is the development of automatic segmentation algorithms for the evaluation of estimated RBF on transplanted kidneys. Applicability of purposed segmentation algorithm have also been tested and evaluated on healthy renal PCASL dataset and synthetic renal PCASL data. The main proposed system; first performs a preliminary whole kidney segmentation by ML/DL; secondly classifies the pixels within the kidney region into cortex and medulla classes; and finally estimates RBF on PCASL-MRI data for quantitative

Table 1
Renal MRI sequence parameters.

SE-EPI READOUT PARAMETERS	
TR/TE	5000/23
Flip angle	90
Partial Fourier	0.75
GRAPPA	2
FOV (mm2)	288 × 288
Acquisition matrix	96 × 96
Slice thickness (mm)	5
Voxel size (mm)	3 × 3 × 5
Number of slices	3
Slice gap (mm)	2.5
Phase oversampling (%)	25
Bandwidth (Hz/pixel)	1890
Fat supression	FAT SAT
Slice orientation	Coronal
Slice scan order	Anterior-Posterior
Breathing strategy	Free-breathing
PCASL LABELING PARAMETERS	
Labeling duration (s)	1.6
PLD (s)	1.2
Number of images	25 label, 25 control, 1 M_0
T_1-w PARAMETERS	
TIs (ms)	200, 300, 400, 500, 600, 700, 800, 900, 1000, 1200, 1400, 1600, 1800, 2000
Number of images	14

TR: Repetition Time. TE: Echo-Time. GRAPPA: GeneRalized Autocalibrating Partial Parallel Acquisition. FOV: field of view. PLD: post-labeling delay. TI: Inversion Time.

assessment.

2. Materials and methods

The proposed pipeline (Fig. 1) includes: (a) MRI acquisition, (b) manual annotation to generate ground truth (GT) regions for evaluation purposes, and (c-f) the image processing system. The latter, in turn, embraces: (c) motion compensation, (d) segmentation of kidney using DL/ML techniques, (e) cortical and medullary tissue differentiation using T_1-w image series and T_1-maps, and (f) renal perfusion estimation from MRI data.

2.1. Transplanted renal MRI data

This study was approved by the Ethics Research Committee at the University of Navarra. Written informed consent was obtained from all subjects before MRI evaluation. 16 (9 male, 7 female) renal transplant patients (mean age ± standard deviation (SD), 53.74 ± 14.65 years) participated in this study with the following inclusion criteria: adults (18+) and clinically stable, considered as patients with eGFR >50 ml/min/1.73m^2 that were transplanted more than a year before the study. Subjects were recruited by their referring nephrologist. A 3 T Skyra (Siemens, Erlangen, Germany) and 18-channel body-array coil was used to carry out renal MRI data acquisition. Perfusion images were acquired using a PCASL sequence with background suppression (BS) and spin-echo echoplanar readout (SE-EPI) [27]. Pre-saturation pulses were applied before the labeling pulses, and BS pulses were optimized to suppress the static signal to 10%. Voxel-wise mapping of T_1 relaxation time of the kidney was acquired using the classical inversion recovery (IR) scheme, where each repetition time contains a single 180° inversion pulse that is followed by a single readout after the inversion time (TI) [28]. Once full magnetization is achieved, the process is repeated for a number of 14 TIs (200, 300, 400, 500, 600, 700, 800, 900, 1000, 1200, 1400, 1600, 1800, and 2000 ms) to accurately sample the IR curve [28].

Subjects were scanned using the following readout: acquisition matrix, 96 × 96; field of view, 288 × 288 mm^2; voxel size, 3 × 3 × 5 mm^3; slice thickness, 5 mm; slice gap, 2.5 mm; number of slices, 3; slice orientation, coronal oblique or coronal-sagittal, and flip angle, 90°. The dataset contains data from 16 patients, each consisted of a reference image, named as M_0 image and 50 (25 control and 25 labels) PCASL images and 3 slices. Additionally, the T_1-w image series is obtained, consisting of 14 images and 3 slices per subject (see Fig. 1(a)). Renal MRI sequence parameters are shown in Table 1.

2.2. Motion correction

During the MRI acquisition, movement of the body or kidney itself can occur. For that reason, PCASL, M_0, and T_1-w images are collectively registered in Elastix [29], using BSpline non-rigid groupwise registration based on B-Spline stack transform and PCA2 metric [30]. The method aligns volumes on a slice-wise basis, using reduced dimension BSpline interpolator, stochastic gradient descent optimizer, and 200 iterations.

2.3. Automatic whole kidney segmentation

We compare three different DL- and ML-based approaches for kidney segmentation on PCASL images, namely, Supervised Descent Method (SDM) [31], U-Net [32], and Mask R-CNN [33]. The dataset is subdivided into training and testing using the Leave-One-Out method (patient-based). Training dataset refers to the sample of data that is used to fit the model and testing dataset refers to the sample of data used to provide an unbiased evaluation of the model. For model training and testing purposes, manual kidney labels (binary masks) are traced on ITK-SNAP software [34], encompassing renal area and excluding renal hilum over PCASL image series. This way, we obtained image/mask pairs belonging to the training dataset.

2.3.1. Supervised descent method

SDM is one of the leading cascaded regression approaches for face alignment [31]. It is based on histogram of gradient (HOG) extraction to encode local shape information from point locations x within the image. SDM aims at learning a series of descent directions and re-scaling factors such that it produces a sequence of updates ($x_{k+1} = x_k + \Delta x_k$) starting from x_0 (mean landmarks) that converges to x_* (labeled landmarks) in the training data [35]. In this work, kidney contour points are selected as landmarks and the following parameters are experimentally set: 40 landmarks (contour points), four weak regressors in cascade with radius patches with values of {12,12,10,10} and weight regularization term λ of 30. HOG features are extracted around specified point locations with {radius patch/2, radius patch/2} cell size, 2 × 2 block size, and 9 bins. The training set is created by corrupting GT masks with additional independent noise applied to each landmark, as data augmentation approach. In particular, 50 initializations (x_0) are included for each one of the images. For every image in the testing dataset, the contour is initialized, i.e. x_0, using the average contour obtained from the GT points of the images in the training dataset. The algorithm is implemented in Matlab on AMD Ryzen 93,900 × 12 Core-Processor, on parallel pools using 10 workers. Training took ~ 28 min.

2.3.2. U-Net

U-Net is a type of convolutional neural network (CNN) proposed in 2015 by Ronnerberger et al. [32] for biomedical image segmentation. In U-Net an structure defined as encoder is used to extract image features, consisting of a traditional stack of convolutional and max-pooling layers. Besides, a decoder is used to restore extracted features to the original image size and to output the final segmentation result. In this study, data augmentation is performed during model training, which is a common practice to artificially increase training dataset by generating new data points from existing data. A series of rotations, translations, and flips are applied to image/mask pairs to produce different anatomically feasible images. Besides, voxel intensities are normalized
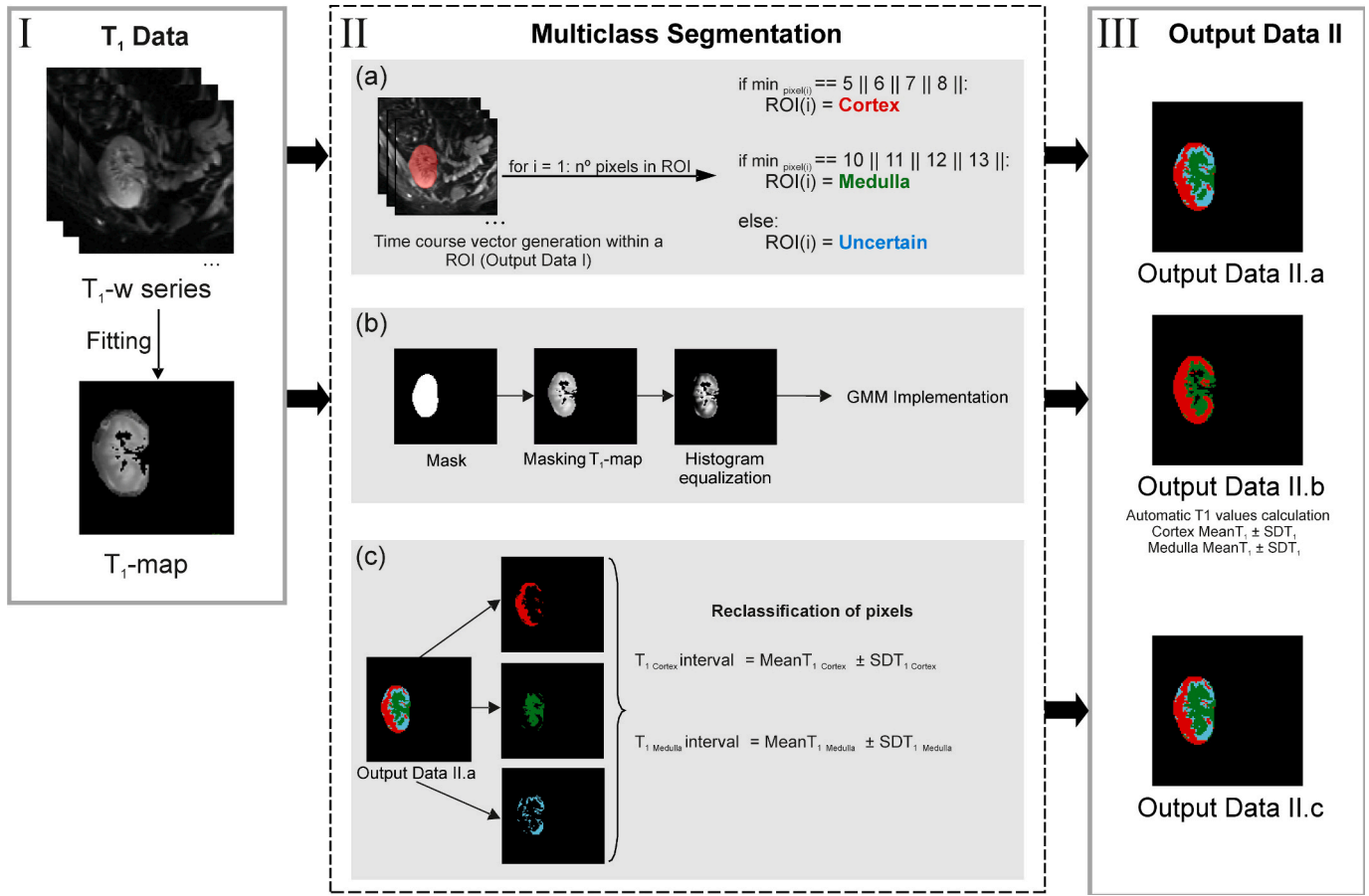
Fig. 2. Automatic cortex and medulla segmentation framework. T_1-w image series and fitted T_1-maps have been used as input data (I). II involves pixel-wise time course processing in which group-thresholds have been used for each of the classes of interest. Output Data II.a is the preliminary segmentation result and it is the input mask for IIb. An intermediate step is required for multiclass segmentation of the kidney within Output Data I region, based on GMM algorithm. In IIc, reclassification of Output Data.IIa is performed using automatically calculated T_1 values interval over segmented cortex and medulla areas. If any of the segmented class of IIa belongs to a T_1 values interval corresponding to other class, this class is reclassified. The final segmentation result is Output Data II.c.

between 0 and 255 to achieve higher contrasted images. Our model is trained on 637 volumes using the Leave-One-Out method, validated on 113 (15% of the training dataset) and tested on 50, which correspond to unique patient data. Validation dataset refers to the sample of data used to provide an unbiased evaluation of the model fit on the training dataset and the tuning of the hyperparameters of the model. Besides, in order to handle class imbalance between the foreground (kidney) and the background classes, sample weights (inversely proportional to the frequency of respective classes, increasing the importance of less prevalent class that is the kidney) are introduced in the Dice loss function of the network (0.9 for kidney and 0.1 for background), that is minimized via Adam optimizer and learning rate of 10^{-4}, with a batch size of 16 and 100 epochs. The training and testing of the model is implemented on Python 3.8 using Tensorflow as backend on GPU NVIDIA GeForce RTX 3090. Model is trained on the labeled training set (image/mask pairs) using manually generated GT masks (see Fig. 1(b)) and is then tested using previously unseen data (testing dataset). Segmentation results are post-processed to keep the biggest connected component to be identified as the kidney. Training took ~ 80 min.

2.3.3. Mask R-CNN approach

Mask R-CNN [33] is a two-stage CNN oriented to object detection and instance segmentation. We found applications of the Mask R-CNN model in other fields of medical imaging different from renal, such as pancreas CT imaging in combination with U-Net [36] and in chest X-Ray imaging for Covid-19 detection [37]. The first stage of the Mask R-CNN

model consists of the backbone and Regional Proposal Network (RPN) [38]. The backbone extracts multilevel image features and the RPN takes those features maps and outputs candidate bounding boxes, i.e. ROIs. Then, to unify the size of these ROIs and to get feature maps for the last section of the network the *RoiAlign* layer [33] is applied. Finally, the model provides a *Box Head* and a *Mask Head*. From the *Box Head* the network obtains the class and a refined bounding box of each object and the *Mask Head* generates a binary mask applying a Fully Convolutional Network (FCN). Mask R-CNN has a multi-stage loss function formed by a classification loss, bounding box loss, and mask loss. During training, data augmentation is used. Rotations, flips, and translation are applied to image/mask pairs. The model is slice-wise trained on 600 volumes, validated on 150 and tested on 50. As the backbone, we selected Feature Pyramid Networks + *Resnet50* [39] and chose the pretrained parameters on the COCO 2016 challenge dataset [40]. The model is trained for 150 epochs using supervised gradient descent optimizer and learning rate of 10^{-4} to minimize the loss function. In addition, the RPN needs to initialize the anchor scales and ratios. In our case, anchor scales and ratios are set to 0.5, 1, 2 and 32, 64, 128, 256, and 512, respectively. We used Python 3.8 and TensorFlow on GPU NVIDIA GeForce RTX 3090 for model training and testing. Training took ~ 120 min.

2.3.4. Evaluation

The performance of the approaches is evaluated using the Dice similarity coefficient (DSC), that measures the similarity between automatic and GT segmentation masks. A value of 0 indicates no overlap

whereas a value of 1 indicates perfect similarity. It is defined as $DSC = 2 \bullet |G \cap S| / (|G| + |S|)$ (%), where G represents the GT mask, and S represents corresponding automatically predicted mask. Statistical analysis is performed using Minitab, version 19.1.0.1 and DSC values are compared between methods (SDM, U-Net and Mask R-CNN) using repeated measures ANOVA with Tukey pairwise comparison, where $p < 0.05$ is considered significant.

2.4. Automatic cortex and medulla segmentation

The segmentation of the internal renal compartments of cortex and medulla is useful for functional and morphological follow-up of the allograft. However, it is challenging due to the limited resolution of the study images and the inter-patient variability [41]. The main renal regions are: the renal cortex, the medulla, and the renal pelvicalyceal system. The cortex is a connective and granular tissue due to the presence of nephrons. The glomerulus and convoluted tubules of the nephron are located in the cortex of the kidney, while the collecting ducts are located in the pyramids of the kidney's medulla. The medulla is the innermost region of the kidney, organized into pyramid-like structures and contains the majority of the length of nephrons, which are the main functional components of the kidney that take care of filtering fluid from blood [42]. The pelvicalyceal system is the central part of the kidney and connects it to the rest of the organism. It is continuous to the ureter and collects the urine as it is produced [42].

In the proposed work, we have implemented an automatic approach for cortical and medullary tissue differentiation, in which T_1-w image series are used. T_1-maps are calculated on a pixel-by-pixel basis for the kidney region (obtained from the best segmentation result from the automatic kidney segmentation, Output Data I in Fig. 1)(d)) by fitting the classic IR mapping scheme for M_0 and T_1-w image [28], where the signal value at κ^{th} scan, S_κ is modelled as.

$$S_k = M_0 \left(1 - 2exp^{-TI_k/T_1} \right).$$

where M_0 is the equilibrium magnetization (Fig. 1(a)), T_1 is the pixel-by-pixel fitted T_1 value, and TI_κ is the inversion recovery time at κ^{th} IR scan. We proposed two methods for cortical and medullary tissue differentiation, namely, gaussian mixture model (GMM) method and T_1 temporal analysis-based method (TA), on T_1-maps and T_1-w image series (Fig. 2I), respectively. Output Data I in Fig. 1(d) serves as an input for pixel-wise tissue differentiation. Two custom scripts have been written in Matlab version R2021a and in Python version 3.8, respectively. Overall representation of the multiclass segmentation framework is shown in Fig. 2.

2.4.1. Segmentation of renal compartments with GMM on T_1 maps

A Gaussian Mixture is a function that assumes an underlying finite mixture of Gaussians, each of them identified by $\kappa \, \varepsilon \, \{1, …, K\}$, being K the number of clusters of the dataset [43]. Each Gaussian density function is parameterized by the mean vector μ, and covariance matrix, Σ, and its parameters are estimated using the iterative expectation-maximization algorithm [44]. The GMM is a finite mixture probability distribution model [44].

In this work, GMM has been implemented to segment cortical and medullary tissues based on Gaussian distribution of the T_1-map within the renal region (Fig. 2IIb). Image features (in this case, the T_1 values) are represented as vectors in an n-dimensional space, which can be modelled by a mixture of Gaussian densities [44]. We initialize the algorithm with $K = 3$ (background, cortex, and medulla) and previously segmented kidney (Output Data I in Fig. 1(d)) as a mask over T_1-map. As a pre-processing step, histogram equalization has been used (Fig. 2II(b)) in order to adjust the gray intensity level of the pixels within the kidney region. After fitting the data, each pixel is labeled as background, cortex, or medulla and segmentation result Output Data II.b is generated (Fig. 2III).
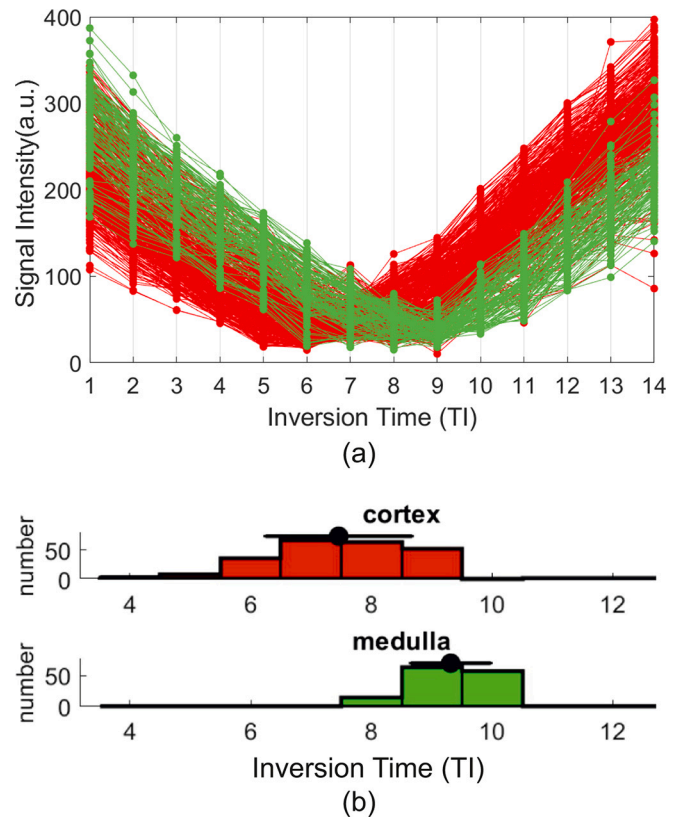

Fig. 3. An example of supervised analysis of time-intensity curve distribution for cortical and medullary tissues. (a) Pixel- wise time-intensity curves for cortex and medulla regions. Each curve corresponds to the temporal signal intensity (T_1 value) along TI_κs (TI at κ^{th} IR scan). (b) Temporal distribution of the null points of time-intensity curves for cortical and medullary tissues.

2.4.2. Segmentation of renal compartments based on temporal analysis of T_1-w image series

This segmentation approach consists of the segmentation of cortical and medullary tissue using a multi-stage algorithm based on: pixel-wise time-intensity curves from IR images (Fig. 2II(a)) and reclassification of pixels based on previously calculated T_1 tissue value intervals (Fig. 2II (c)). The assignation of signal time-courses to voxels within renal mask has been reported to be used in [19–21] over DCE-MRI images aiming at renal perfusion assessment. Prior to the segmentation pipeline, the temporal evolution of manually annotated cortex and medulla classes is studied, over previously motion-corrected T_1-w image series. It is important to emphasize that T_1 longitudinal relaxation time is tissue specific and is reported to be higher in medullary tissue than in cortical tissue [45–48]. In Fig. 3 there is an example of the pixel-based time series analysis, in which 234 pixels belong to the cortex class and 137 to the medulla class. Horizontal axis represents the TI_κ ($\kappa = \{1,14\}$) and vertical axis represents the T_1 value of each pixel in each TI_κ or image from the T_1-w image series. Note that, generally, cortical tissue attains its null point before the medulla does. For the unsupervised segmentation of both tissues, previously calculated kidney region (Output Data I in Fig. 1(d)) has been used as reference mask in order to construct pixel-wise time-intensity vectors. Based on this, a two-step classification approach is performed. As depicted in Fig. 2IIa, in the first stage, each pixel in the masked image is classified as cortex if its null point is found at $5 \leq \kappa \leq 8$ TIs and as medulla if its null point is found at $10 \leq \kappa \leq 13$ TIs. The pixels, that are not classified in any of these compartments, are assigned an uncertain class (Output Data II.a in Fig. 2). In a second stage, through the intermediate GMM step (see Section 2.4.1), patient-based T_1 mean values are automatically calculated for medulla ($T_{1Medulla}$) and cortex ($T_{1Cortex}$) tissues and a pixel-wise reclassification is performed.
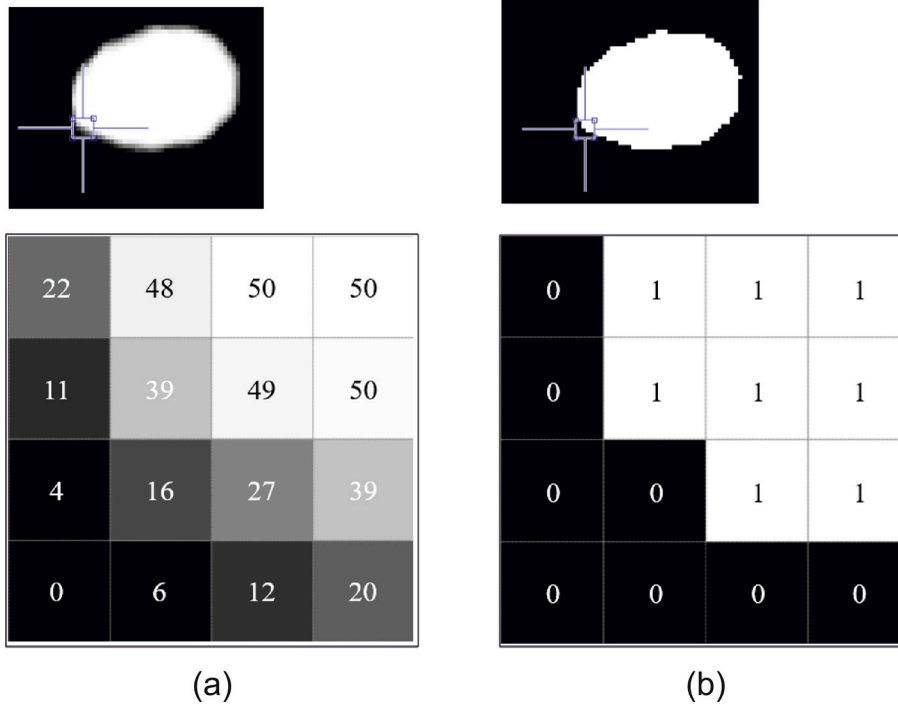
Fig. 4. An example of binarized mask of probabilistic map. (a) Result of the summation of the overall result series of an individual set and slice-wise. (b) Binarized map obtained from (a), with a threshold value of 25 counts out of 50.

We use $K = 3$ (background, cortex, and medulla). This way, pixels in the cortical region of Output Data II.a in Fig. 2 are reclassified as medulla if any of their values belong to the interval values of the medulla, defined within the range (mean ± SD) $T_{1Medulla}$. In the same way, pixels in the medulla region are reclassified as cortex if they belong to the interval values of the cortex, defined within the range (mean ± SD) $T_{1Cortex}$. For the case of the region classified as uncertain, pixels are reclassified as cortex or medulla if any of them belong to aforementioned cortex and medulla values. This stage is considered a refinement process. Final results are shown in Output Data II.c (Fig. 2III).

2.4.3. Evaluation

For the purpose of instance segmentation evaluation, a set of standard metrics has been used: the DSC, precision (PC), recall (RC), and F-measure (FM). G represents the GT mask and S represents corresponding automatically predicted mask. They both contain 3 classes: background, cortex, and medulla. Segmentation evaluation does not include the uncertain class. Recall is the ratio between the true segmentation, $S \cap G$ and all positive elements G, defined as $RC = |S \cap G|/|G|$ (%). The precision is the ratio between the true segmentation $|S \cap G|$ and the automatic segmentation S, defined as $PC = |S \cap G|/|S|$ (%). F-measure (FM) consists of the harmonic mean of precision and recall, where β is the scaling of these two metrics [49]. In this work, a β value of 2 was defined, in order to raise the importance of recall. It is calculated as $FM\_\beta = ((1 + \beta^2) \bullet (PC \bullet RC))/((\beta^2 \bullet PC + RC))$ (%). Statistical analysis is performed using Minitab, version 19.1.0.1 and differences between segmentation methods (GMM and TA) are compared with paired *t*-test, where $p < 0.05$ is considered significant.

2.5. Renal perfusion estimation

The final objective of the proposed work is the estimation of perfusion values as biomarkers of renal function. To this end, the perfusion-weighted image (PWI) is to be calculated by simply subtracting the label from the control image, where the signal intensity is proportional to perfusion [50]. Mean cortical signal (gray-scale intensity values) was

measured in the cortical region of each subtracted label-control pairs, using the mask obtained from the cortical segmentation done in Section 2.4. Label-control pairs are considered outliers and consequently discarded if their mean cortical signal is >2 SD away from the global cortical mean [27]. Finally, mean PWI is calculated. RBF maps are computed using the single compartment model [28,27] as

$$RBF\left(\frac{\frac{ml}{min}}{100g}\right) = \frac{6000 \bullet \lambda \bullet I_{PW} \bullet exp^{PLD/T_{1Blood}}}{2 \bullet \alpha \bullet T_{1Blood} \bullet I_{M_0} \bullet \left(1 - exp^{-\tau/T_{1Blood}}\right)}$$

where I_{PW} is the pixel-wise signal value of the mean PWI, I_{M_0} is the pixel-wise signal value of the M_0 image; λ is the tissue-blood-water partition coefficient, assumed 0.9 ml/g; α is the labeling efficiency, set to 0.65 considering a PCASL efficiency of 0.74 and the effect of BS pulses [51]; τ is the labeling duration set to 1.6 s; T_{1Blood} is the arterial blood T_1 value, assumed as 1.65 s, and PLD is the post-labeling delay, set to 1.2 s.

2.5.1. Evaluation

In order to evaluate the effect of automatic cortical and medullary segmentation in the estimation of renal perfusion, median and SD RBF values in ml/100 g/min per patient and slice are estimated using GT and predicted cortical and medullary segmentations. The perfusion estimation discrepancy between GT labeling and automatic labeling is measured in terms of relative error (%), using GT and automatically segmented cortex and medulla masks over calculated RBF maps.

3. Results

The proposed model is evaluated in two steps. First, whole kidney segmentation results and segmentation of renal compartments is evaluated in terms of similarity metrics, comparing with GT masks. Second, the estimation of renal perfusion values is evaluated.

Table 2
Segmentation results of the whole kidney.

Model	DSC (Mean $\pm$ SD) (%)
SDM	84.40 $\pm$ 11.89
U-Net	87.87 $\pm$ 1.30
Mask R-CNN	93.90 $\pm$ 2.00

3.1. Segmentation of the whole kidney

For the evaluation of whole kidney segmentation, a probabilistic map extracted from predicted results is generated. The map is constructed through the sum (Fig. 4(a)) of binary predicted result series, slice-wise and patient-wise. From this map, a binary mask is produced as a unique segmentation result applying a threshold value of 25 counts out of 50 (Fig. 4(b)).

Numerical DSC results for whole kidney segmentation are shown in Table 2. In general, we can observe that DL based approaches tend to be better for whole kidney segmentation task compared to ML based image processing technique. Furthermore, Mask R-CNN presents a higher similarity score with >90% of success. Fig. 5 depicts an example of automated kidney segmentations for two patients, over an original PCASL control image. Statistical analysis showed significant differences on DSC means between methods ($p < 0.05$), with Tukey pairwise testing showing statistical differences between all combinations of methods except between DSC values of U-Net and SDM. In Fig. 6 a cumulative
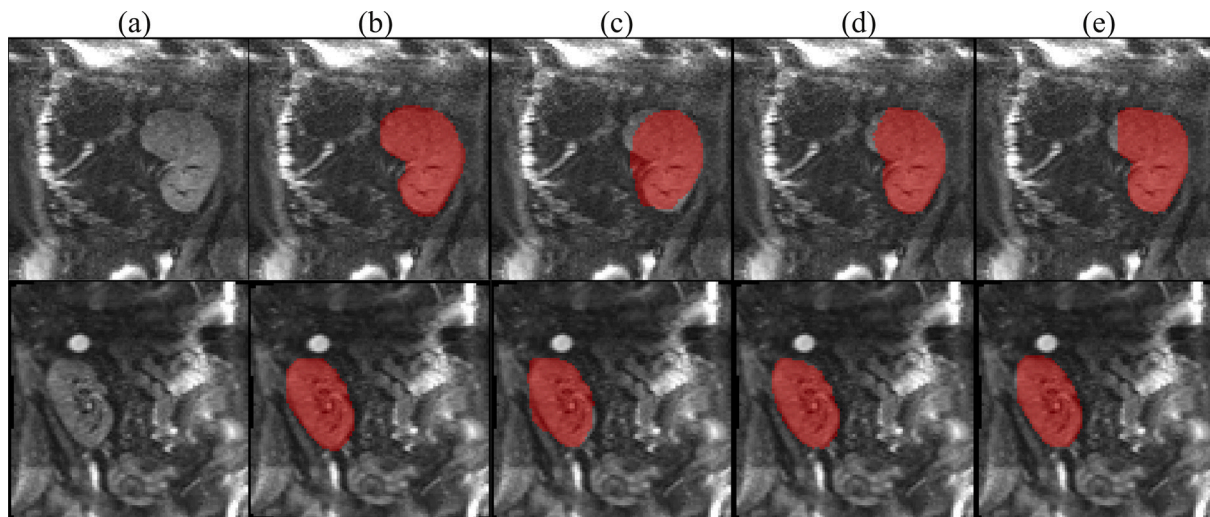

Fig. 5. Examples of automated kidney segmentations over original PCASL control image for two patients. The red label represents the kidney class. (a) Original PCASL control image. (b) GT mask. (c) Binarized summation of SDM results. (d) Binarized summation of U-Net results. (e) Binarized summation of Mask R-CNN results. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)
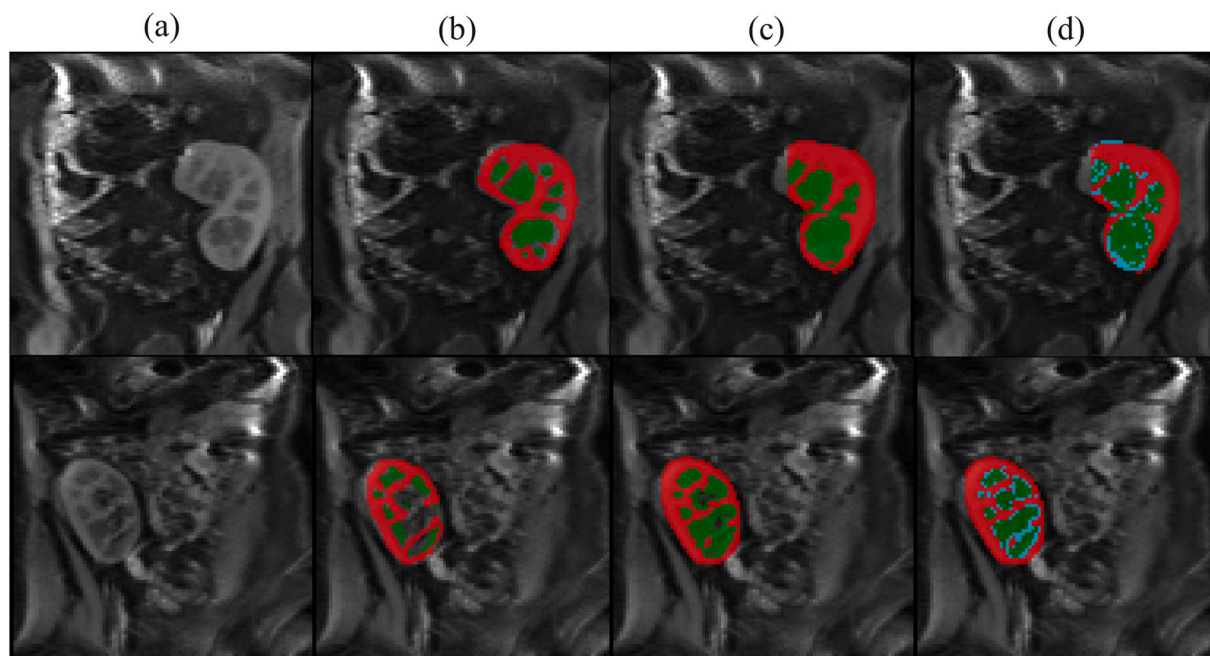

Fig. 6. Cumulative distribution function of DSC calculated on binarized probability map. The y-axis shows the probability that calculated metrics will take the values equal to or less than values in x-axis. In this case, the y-axis represents the number of images of the whole dataset measured in percentage and x-axis represents the DSC, ranging from 0 to 1. 1 indicates perfect match between the GT and predicted masks obtained through our models and 0 indicates non-similarity between the GT and predicted masks.
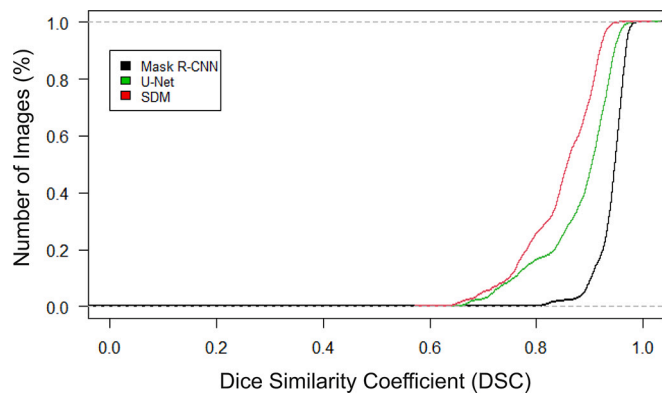
Fig. 7. Comparison between GMM and TA segmentation methods. The boxplots illustrate comparative (a) recall, (b) precision, (c) F-measure and (d) DSC. For each segmentation method evaluation metrics of cortex (in red), medulla (in green) and weighted (in magenta) class are calculated. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Table 3
Segmentation results of renal compartments.

Model	Tissue	Recall	Precision	F-measure	DSC
GMM	Cortex	83.19 ± 9.74	96.32 ± 3.93	85.29 ± 8.34	88.86 ± 6.02
	Medulla	91.99 ± 8.40	67.74 ± 17.83	84.10 ± 11.06	76.31 ± 140.8
	Weighted	85.03 ± 8.27	89.79 ± 3.39	85.03 ± 8.10	86.01 ± 6.75
TA	Cortex	91.37 ± 13.51	94.51 ± 4.95	91.56 ± 12.32	92.07 ± 10.32
	Medulla	85.82 ± 11.53	82.35 ± 17.55	83.85 ± 10.73	82.41 ± 12.50
	Weighted	89.66 ± 9.99	91.85 ± 4.89	89.47 ± 10.61	89.70 ± 10.23

distribution function (CDF) of DSC values is shown. For the Mask R-CNN approach, the probability of DSC to be <90% is around 8%, whereas the probability of DSC to be <80% decreases until 1%. For the U-Net and SDM approaches, the probabilities of DSC values of being <90% highly increase up to 40% and 75%, and the probabilities of DSC values of being <80% decrease to 15% and 25%, respectively.

3.2. Segmentation of cortical and medullary tissues

Multiclass segmentation results are compared in terms of RC, PC, FM, and DSC with the reference GT masks (Fig. 7), comparing the segmentation results obtained through GMM and TA for cortex and medulla tissues. Table 3 shows segmentation results for renal compartments. Moreover, in order to counteract the problem of class imbalance between cortex and medulla samples, the weighted evaluation metric ensures that RC, PC, FM, and DSC of the predictions are not inflated due to classes of high-frequency (cortex) that dominates over others (medulla), in terms of number of pixels for each class. For weighted class, achieved RC is $85.03 \pm 8.27\%$ and $89.66 \pm 9.99\%$, PC is $89.79 \pm 3.39\%$ and $91.85 \pm 4.89\%$, FM is $85.03 \pm 8.10\%$ and $89.47 \pm 10.61\%$ and DSC is $86.01 \pm 6.75\%$ and $89.70 \pm 10.23\%$, for GMM and TA method, respectively. In Fig. 8, there is an example of multiclass segmentation for two patients over 14th T_1-w image. Regarding the segmentation of cortical tissue, our sample data supports the notion that the averaged paired difference between GMM and TA methods does not equal zero ($p < 0.05$). Specifically, the TA score mean is greater than the GMM score mean ($p < 0.05$), for all the implemented similarity metrics. In the case of medullary tissue segmentation, the mean score of GMM and TA segmentation methods did statistically differ ($p < 0.05$) and TA showed greater mean score ($p < 0.05$) compared to the mean score of the GMM segmentation method. For both GMM and TA methods, the segmentation results of the cortex present higher averaged similarity scores (above 80%) with lower standard deviation values compared to those corresponding to the medulla tissue (above 75%). For tissue-weighted results, high segmentation metrics above 80% have been achieved in both GMM and TA approaches. The implementation of GMM explained in Section 2.4.1 also serves as a T_1 value calculation method over detected cortical and medullary regions, with no statistical difference ($p > 0.05$) with respect to T_1 (mean $\pm$ SD) values calculated through GT annotations in T_1-map. T_1 mean $\pm$ SD values of 1304 ± 72 ms and 1572 ± 62 ms are obtained from automatically segmented cortex and medulla masks, respectively.

With the objective of temporal analysis-based segmentation validation, pixel-wise temporal evolution analysis has been performed on phantom images. Validation of segmentation framework has been done on T_2 contrast spheres values at 3.0 T measured at 20 °C performed at NIST. An image series consisting of 14 TIs have been used as input data for the validation algorithm. The acquisition matrix of images was 96×102. From T_2 contrast spheres values at 3.0 T, T_2–3 and T_2–5 values have been selected; corresponding to 1901.28 ms and 1197.57 ms, respectively. We considered those values the closest to the real T_1 values of medulla and cortex, respectively. Ahead of image processing, phantom images have been masked within a manually selected region encompassing T_2–3 and T_2–5 spheres and surrounding areas. Temporal pixel-wise vectors have been classified as cortex if its null point is found at $5 \leq \kappa \leq 8$ TIs and as medulla if its null point is found at $9 \leq \kappa \leq 12$ TIs. The segmentation results showed that the proposed temporal information-based method obtains DSC above 95% with respect to the ground truth masks.

3.3. Estimation of renal perfusion

For the evaluation of calculated RBF maps (see Eq. (2)), cortical and medullary perfusion values have been computed using GT and predicted multi-class masks. Should be pointed that outlying points on median cortical and medullary perfusion values were discarded. Fig. 9 contains scatter plots that show the relationship between perfusion values obtained through segmented masks from GMM (Fig. 9a) and TA (Fig. 9b) methods, and GT perfusion values. All the pairwise comparisons show a positive association, because as the GT perfusion value increases, so does the predicted result and linear relationship, as the GT perfusion variable increases by approximately the same rate as the predicted variables change by one unit. For perfusion values obtained from masks from GMM and TA methods, in the case of cortical median and SD values, a strong positive linear relationship between them is shown with a correlation coefficient (r) above 96% and 65%, respectively. For medullary median values, lower r is observed and SD values achieved values above 76%. In addition, a similar median RBF is found for predictions (GMM and TA) and GT with cortical perfusion values of 154 ± 90 ml/min/100 g, 153 ± 87 ml/min/100 g, and 162 ± 70 ml/min/100 g, respectively; and medullary perfusion values of 100 ± 84 ml/min/100 g, 69 ± 74 ml/min/100 g, and 67 ± 62 ml/min/100 g, respectively. Moreover, the cortical perfusion value discrepancy is 4.90% and 6.78%, for GMM and TA methods and the medullary discrepancy is 45.54% and 18.31%, for GMM and TA methods, respectively.

4. Applicability of proposed method for different datasets

Mask R-CNN segmentation of whole kidneys was tested on two distinct datasets: renal PCASL studies on healthy subjects and synthetic renal PCASL of XCAT phantom. As the proposed model was initially trained with transplant patients (one kidney per image and different acquisition plane), both new renal datasets were pre-processed in order to fit into the characteristics of previously designed model. A gaussian filter was used to smooth the half of the image, in order to segment one
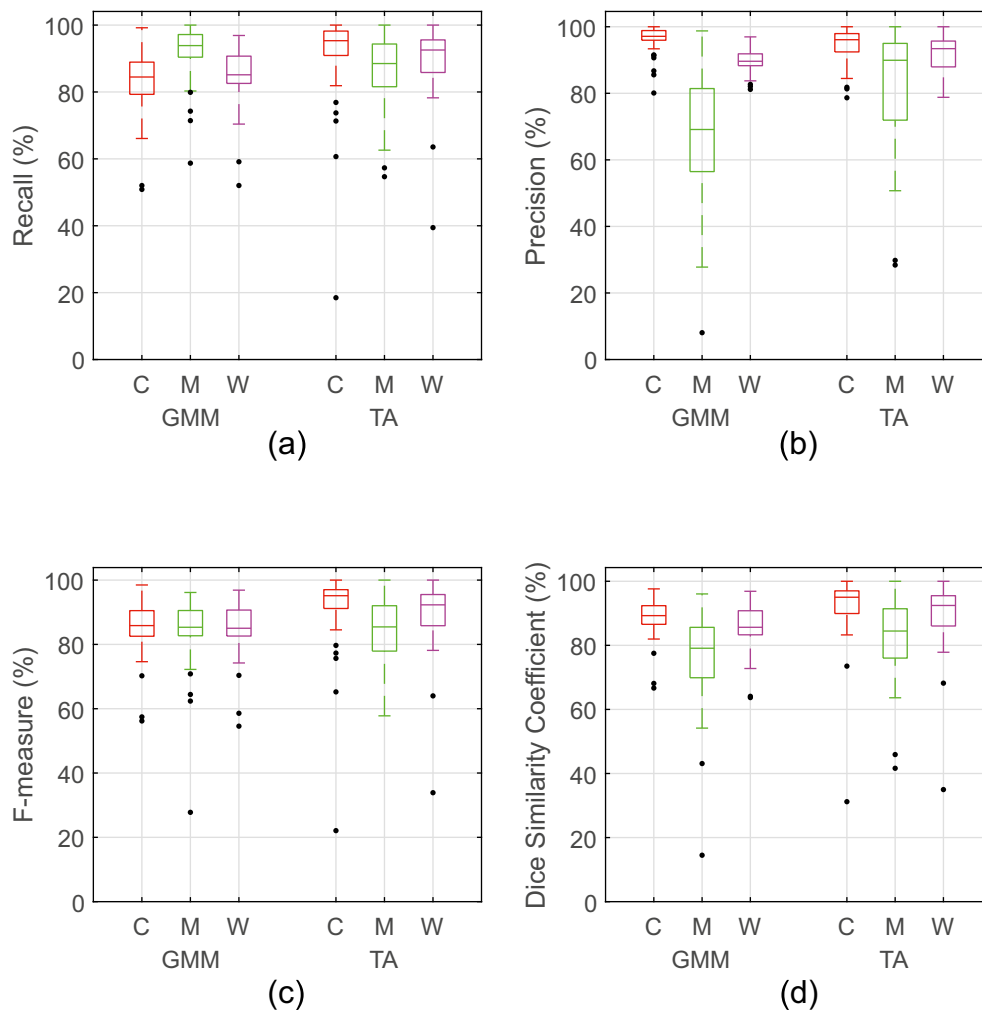
Fig. 8. Examples of automated cortex and medulla segmentations for two patients. The red and green labels correspond to cortex and medulla classes, respectively and the lighted blue label in (c) and (d) represents the uncertain class. (a) Original T_1-w image. (b) GT mask for cortex and medulla classes. (c) Output results obtained from GMM method (Output Data II.b in Fig. 1). (d) Output results obtained from TA method (Output Data II.c in Fig. 1). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)
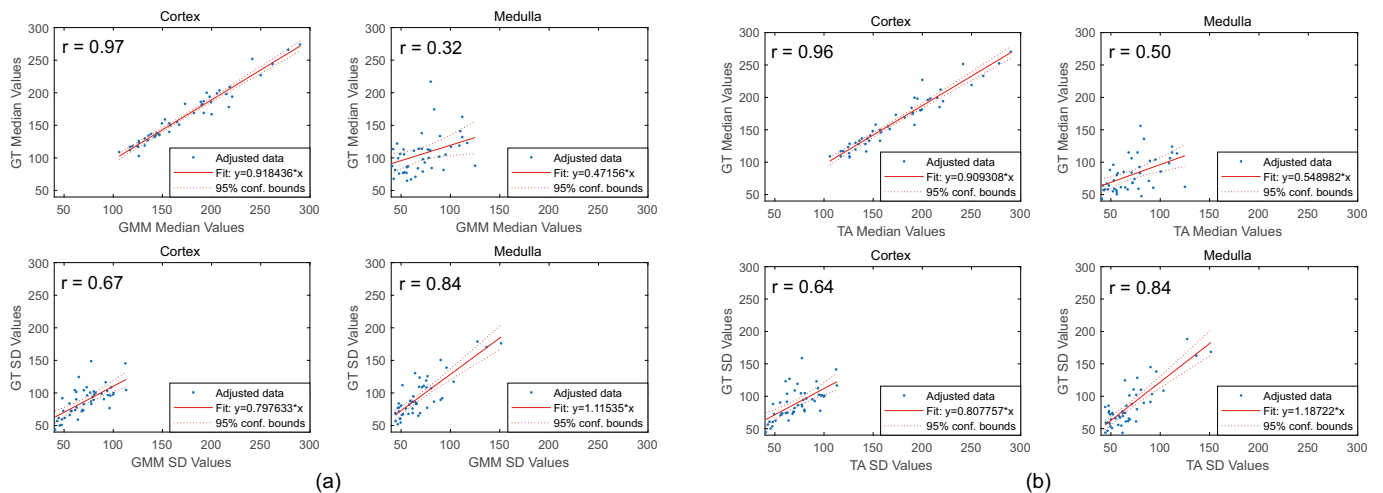

Fig. 9. Scatter plot of pairwise comparison between GT and proposed methods. Each dot on the scatterplot represents a slice of one individual from the data set (3 slices each). (a) Pairwise comparison between renal perfusion GT and obtained GMM values. (b) Pairwise comparison between renal perfusion GT and obtained TA values.
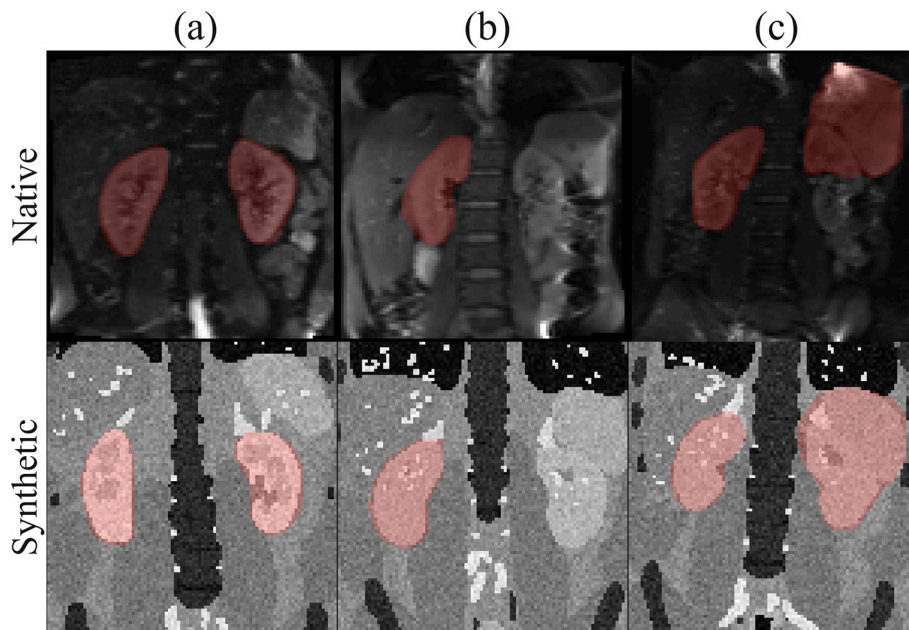
Fig. 10. Examples of automated renal segmentations on native and synthetic ASL. (a) Example of complete segmentation of both kidneys. (b) Example of incomplete segmentation, only the right kidney of the patient is segmented. (c) Example of over-segmentation of right kidney of the patient.

Table 4
Visual assessment of segmentation results on native and synthetic PCASL.

	Total coverage (%)			Incomplete coverage (%)			No detection (%)		
	Total	Right	Left	Total	Right	Left	Total	Right	Left
Native	65	64	36	27	19	81	8	37	63
Synthetic	61	73	27	37	13	87	2	36	64

kidney each time the image is tested. Finally, binary result masks were gathered in order to compare full renal coverage. The assessment of segmentation performance on test images was visually done, based on a 3-level quantization of the coverage of the binary mask within the renal region. Total coverage indicates that the kidney is well delimited within renal region, Incomplete coverage indicates that the kidney has not been well delimited (over segmented or under-segmented) and No detection indicates that the kidney is not detected at all by the model.

- Healthy renal PCASL studies. Data acquisition of renal PCASL study on healthy subjects was performed on the 3D Skyra and 18-channel body-array coil, using the same readout as in the transplanted renal PCASL dataset (see Section 2.1). The study was approved by the Ethics Research Comittee at the University of Navarra and written consent was obtained from all the subjects. 14 healthy patients covering the age range from 22 to 32 participated in the study. One study per patient was used in this experiment. In total, 808 renal images were tested (1616 kidneys). GT masks were not provided.
- Synthetic renal PCASL using body models from the XCAT phantom [52]. Data acquisition simulating in vivo acquisitions were generated using body models from the XCAT phantom, that provides anatomical structures by five models (77, 80, 92, 93 and 108). Perfusion was added using the general kinetic model and literature values for tissue properties were applied (relative proton density ρ, T_1, T_2 values at 3 T) [53]. In this work, 5 PCASL datasets with healthy perfusion where used for testing purposes, and consisted of a M_0 and 25 control and labels pairs and one slice. In total, 255 renal images were used (510 kidneys). Should be noted that GT masks were not be provided due to license restrictions, but we would like to thank Brumer, I. from University of Heidelberg for providing DSC metrics obtained by the comparison of original GT masks and our segmentation results.

Examples of automated renal segmentations on native and synthetic ASL data are shown in Fig. 10. Segmentation results on both datasets are shown in Table 4. Even if, in overall, high renal coverage is achieved, difficulties on left kidney differentiation are found in both datasets. This is due to the low differentiation between renal and spleen tissue, that causes the binary mask to extend all over the area of kidney and spleen. Should be noted that in case the operator aims at precise renal analysis, some post-processing needs to be done. More than the 60% of the total number of kidneys are well delimited within renal area and in <10% of the cases no kidney is detected. The remaining percentage corresponds to incomplete renal detection, under- or over-segmented. In the case of synthetic data, 87% of the incomplete result masks correspond to the over-segmentation of the left kidney and just the 13% is due to the slight under- or over-segmentation of the right renal area. In the case of synthetic dataset, just the 2% of the total number of kidneys is not detected at all. Results on native ASL data showed that the 81% of the incomplete result masks correspond to the over-segmentation of the left kidney and just the 19% is due to the incompleteness of the right kidney. In the case of synthetic PCASL data, median DSC values of 0.70 ± 0.16 for the left kidney, and 0.89 ± 0.02 for the right kidney were calculated.

5. Discussion

In this work, a fully automatic method for perfusion estimation in renal MRI data is proposed. The segmentation of tissues consisted of a two-step pipeline, where, first, the segmentation of the whole kidney is done on a PCASL MRI dataset and then, cortical and medullary classification is performed using the temporal information of T_1-w IR images. The proposed approach leads to a renal perfusion estimation. Regarding the automatic whole kidney segmentation, proposed method implements both, machine (SDM) and deep (U-Net and Mask R-CNN) learning

Table 5

Summary of ML and DL approaches for kidney segmentation and final volumetric and perfusion quantification on non-contrast-enhanced MRI. The work in [24] is the only one with which we can make a fair comparison since it uses the same type of images as in our study.

Literature reference	Aim	Methods	MRI modality	Dataset (subjects)	Kidney condition	Tissue	Results (% DSC, when provided)
[24]	RBF quantification	Automatic (cascaded U-Nets)	ASL-MRI	10 (train) 4 (validation)	Healthy	Cortex	78.00 ± 4.00 75.00 ± 3.00
[22]	RBF quantification	Semiautomatic (intelligent scissors and k-means)	ASL-MRI	14	Healthy	Whole Cortex Medulla	Not provided
[26]	RBF oxygenation and quantification	Automatic (Mask R-CNN)	ASL-MRI	Up to 1000 images	Not defined	Whole Cortex Medulla	Visually assessed
[23]	TKV measurement	Automatic (thresholding and shape detection)	T_1-w MRI T_2-w MRI	12	Healthy	Whole Cortex Medulla Pelvis	93.64 ± 1.31 80.82 ± 3.22 72.59 ± 4.30 71.87 ± 6.91
[25]	TKV measurement	Automatic (U-Net)	T_1-w MRI	30	Healthy and CKD		93.00 ± 1.00
Proposed	RBF quantification	Automatic (ML/DL approaches and temporal analysis)	ASL-MRI	16	TK	Whole co	93.90 ± 2.00 92.42 ± 10.29 82.83 ± 12.52

RBF: renal blood flow. TKV: total kidney volume. DSC: Dice similarity coefficient. TK: transplanted kidney.

based algorithms using the PCASL-MRI dataset. In particular, the model based on Mask R-CNN network presents outstanding results, obtaining averaged DSC values above 93%, outperforming the current state-of-the-art. The segmentation performance highly depends on the intensity range of the images. Our method uses M_0, control and label images together as the training dataset, considering that their gray-scale intensity ranges are located in different values. Even if intensity rescale is applied, the heterogeneity of image intensity should be considered when testing with new data. Nevertheless, we have demonstrated that whole kidney segmentation network also works with healthy synthetic PCASL renal data, with acceptable segmentation accuracy, denoting that the model is not limited to a unique dataset. Unlike transplanted kidney data, both healthy and synthetic renal PCASL data contained two kidneys (left and right) and that supposed some post-processing steps. The segmentation performance of the model could be increased by adding the native and synthetic kidney data to the training set. Concerning the automatic segmentation of cortex and medulla using the data in the T_1-w image series, two different methods with varying complexity are evaluated. Although GMM method obtains an acceptable classification performance, the combination of GMM provided intervals with the temporal information derived from the T_1-w image series presents the best results. As expected, the results obtained for the cortex are better than the ones extracted for the medulla compartment. The segmentation performance of medullary tissue shows higher dissimilarities between manually drawn labels and automatically achieved ones. This discrepancy is mainly caused by the mislabelling (under-segmentation) of the medulla region, which tends to be less precise than the segmentation of the cortex due to the low-contrasted interface between the tissues and the partial volume effects in MRI images. Proposed method also generates an uncertain class mask in areas where the differentiation between cortex and medulla pixels is not clear, that could be processed in further steps to complete cortical and medullary masks, and indeed, the estimation of perfusion values. Regarding the estimation of renal perfusion, our work demonstrates that multiclass segmentations do have an effect on cortical and medullary RBF estimation. The medullary discrepancy is considerably higher in the case of the GMM, due to low-contrasted interface and partial volume effects on the T_1 maps, compared to the T_1-w image series temporal evolution, that shows higher contrast between tissues and enables temporal differentiation. Regardless, the perfusion values obtained for cortex and medulla compartments are considered to be within the acceptable ranges according to the literature [22,24,26,27]. It should be pointed out that there were limitations to the proposed renal perfusion estimation method evaluation. The real value of renal blood flow was unknown, as we had no other measurement

technique with which to compare the estimation results. In the future, the accuracy and reliability of this tool should be validated, which will require the estimation of perfusion GT values using other techniques, such as renal scintigraphy [54] or contrast enhanced first pass perfusion MRI [55].

Finally, we have indirectly compared the segmentation results of our approach with five state-of-the-art methods for kidney segmentation in non-contrasted MRI data (see Table. 5) in terms of DSC. Compared to the approach presented in [24], the cortical DSC achieved by our method is ~18% higher. In addition, this approach exclusively focuses on cortical perfusion values and does not segment the medullary tissue, therefore we cannot compare our medullary segmentation results with those obtained in [24]. In [22], Hammon et al. implemented a semiautomatic approach for whole kidney, cortex and medulla segmentation and posterior renal perfusion evaluation. Besides the fact that it is not an automatic method, the segmentation results presented are not in terms of DSC and it is not an automatic approach. In [26], as in our work, a Mask R-CNN network is implemented for cortical and medullary tissue segmentation for renal blood flow estimation. In this case, visual evaluation of results is performed and none DSC are given. Furthermore, renal volumetric analysis is aimed at [23] using T_1- and T_2-w images that provides higher tissue differentiation compared to our study images. In comparison, slightly better DSC is achieved by our method for whole kidney segmentation and ~10% higher cortical and medullary DSCs. Finally, Daniel et al. [25] achieved a slightly lower DSC for whole kidney segmentation using T_2-w images for volumetric analysis purposes.

6. Conclusions

The fully automatic framework described in this work represents a significant advance for the state-of-the-art that will allow researchers and specialists to reduce the effort required to manually segment the kidney and its compartments and, therefore, automate the process of kidney perfusion estimation. The segmentation software is publicly accessible.

Ethics approval statement

This work involved human subjects in its research. The study was approved by the Ethics Research Committee at the University of Navarra on the September 27, 2018 (number 2018.121). Written informed consent was obtained from all subjects.

Declaration of Competing Interest

None.

Acknowledgments

References

[1] Jiang Kai, Lerman Lilach O, et al. Am J Nephrol 2019;49(2):111–3. Publisher: Karger Publishers.

[2] Collins Allan J, et al. US renal data system 2013 annual data report. Am J Kidney Dis January 2014;63(1 Suppl):A7.

[3] Hariharan Sundaram, Israni Ajay K, Danovitch Gabriel. Long- term survival after kidney transplantation. N Engl J Med August 2021;385(8):729–43. https://doi.org/10.1056/NEJMra2014530. Publisher: Massachusetts Medical Society _eprint.

[4] Echeverria-Chasco Rebeca, Martin-Moreno Paloma L, Garcia-Fernandez Nuria, Vidorreta Marta, Aramendia-Vidaurreta Verónica, Cano David, et al. Multiparametric renal magnetic resonance imaging: a reproducibility study in renal allografts with stable function. NMR Biomed. 2022:2.

[5] Michael J Field, Harris David C, Pollock Carol A. 5 - Glomerular filtration and acute kidney injury. In: Field Michael J, Harris David C, Pollock Carol A, editors. The renal system (second edition). Churchill Livingstone; January 2010. p. 57–67.

[6] Alpert Nathaniel M, Rabito Carlos A, Correia John A, Babich John W, Littman Bruce H, Tompkins Ronald G, et al. Mapping of local renal blood flow with PET and H(2)(15)O. J Nucl Med April 2002;43(4):470–5.

[7] Anderson H, Yap JT, Wells P, Miller MP, Propper D, Price P, et al. Measurement of renal tumour and normal tissue perfusion using positron emission tomography in a phase II clinical trial of razoxane. Br J Cancer July 2003;89(2):262–7.

[8] Lee Sang-Kwon, Jang Youjung, Jung Jin-Woo, Je Hyejin, Choi Jihye. Comparison of renal blood flow using maximum slope- based computed tomography perfusion and ultrasound flow probe in healthy dogs. Front Vet Sci 2020;7:541747.

[9] Jaschke W, Cogan MG, Sievers R, Gould R, Lipton MJ. Measurement of renal blood flow by cine computed tomography. Kidney Int April 1987;31(4):1038–42.

[10] Odudu Aghogho, et al. Arterial spin labelling MRI to measure renal perfusion: A systematic review and statement papervol. 33; 2018ii15–21.

[11] David C Alsop, et al. Recommended implementation of arterial spin labelled perfusion MRI for clinical applications: a consensus of the ISMRM perfusion study group and the european consortium for ASL in dementia. Magn Reson Med 2015;73(1):102–16.

[12] Alnazer Israa, et al. Recent advances in medical image processing for the evaluation of chronic kidney disease. Med Image Anal Apr 2021;69:101960.

[13] Ker Justin, Wang Lipo, Rao Jai, Lim Tchoyoson. Deep learning applications in medical image analysis. In: IEEE access. vol. 6; 2018. p. 9375–89. Conference Name: IEEE Access.

[14] Luckow Andre, et al. Artificial intelligence and deep learning applications for automotive manufacturing. In: 2018 IEEE international conference on big data (big data); December 2018. p. 3144–52.

[15] Shah Het, Desai Shivani, Vyas Tarjni, Nair Anuja, Degadwala Sheshang. Insights of deep learning applications. In: 2021 5th International Conference on Electronics, Communication and Aerospace Technology (ICECA); December 2021. p. 1355–60.

[16] Goodfellow Ian, Bengio Yoshua, Courville Aaron. Deep learning. MIT Press; 2016. http://www.deeplearningbook.org.

[17] Isensee Fabian, Maier-Hein Klaus H. An attempt at beating the 3d u-net. 2019.

[18] Zhang Yao, et al. Cascaded volumetric convolutional network for kidney tumor segmentation from ct volumes. 2019.

[19] Klepaczko Artur, Eikefjord Eli, Lundervold Arvid. Healthy kidney segmentation in the dce-mr images using a convolutional neural network and temporal signal characteristics. Sensors 2021;21(20).

[20] Chevaillier B, Ponvianne Y, Collette JL, Mandry D, Claudon M, Pietquin O. Functional semi-automated segmentation of renal DCE-MRI sequences. In: 2008 IEEE international conference on acoustics, speech and signal processing; March 2008. p. 525–8.

[21] Yang Xin, Le Minh Hung, Cheng Kwang-Ting Tim, Sung Kyung Hyun, Liu Wenyu. Renal compartment segmentation in DCEMRI images. Med Image Anal August 2016;32:269–80.

[22] Hammon Matthias, et al. Reproducibility of kidney perfusion measurements with arterial spin labeling at 1.5 Tesla MRI combined with semiautomatic segmentation for differential cortical and medullary assessment. Medicine March 2016;95(11). e3083.

[23] Will Susanne, Martirosian Petros, Würslin Christian, Schick Fritz. Automated segmentation and volumetric analysis of renal cortex, medulla, and pelvis based on non-contrast-enhanced T1- and T2-weighted MR images. Magma (New York, NY) October 2014;27(5):445–54.

[24] Bones Isabell K, Bos C, Moonen C, Hendrikse J, van Stralen M. Workflow for automatic renal perfusion quantification using asl-mri and machine learning. Magn Reson Med 2022;87(2):800–9.

[25] Alexander J Daniel, et al. Automated renal segmentation in healthy and chronic kidney disease subjects using a convolutional neural network. Magn Reson Med 2021;86(2):1125–36. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/mrm.28768.

[26] Laursen Mathilde Overgaard, Köylü Büsra, Haddock Bryan, Sorensen John Aa. Kidney segmentation for quantitative analysis applying MaskRCNN architecture. In: 2021 IEEE Symposium Series on Computational Intelligence (SSCI); December 2021. p. 1–6.

[27] Echeverria-Chasco Rebeca, et al. Optimization of pseudo-continuous arterial spin labeling for renal perfusion imaging. Magn Reson Med March 2021;85(3):1507–21.

[28] Nery Fabio, et al. Consensus-based technical recommendations for clinical translation of renal ASL MRI. Magma (New York, NY) February 2020;33(1):141–61.

[29] Klein S, Staring M, Murphy K, Viergever MA, Pluim J. elastix: a toolbox for intensity-based medical image registration. IEEE Trans Med Imaging Jan 2010;29(1):196–205.

[30] Huizinga Wyke, et al. PCA-based groupwise image registration for quantitative MRI. Med Image Anal Dec 2015;29.

[31] Lou Jianwen, Cai Xiaoxu, Wang Yiming, Yu Hui, Canavan Shaun. Multi-subspace supervised descent method for robust face alignment. Multimed Tools Appl December 2019;78(24):35455–69.

[32] Ronneberger Olaf, Fischer Philipp, Brox Thomas, Navab Nassir, Hornegger Joachim, Wells William M, et al. U-Net: convolutional networks for biomedical image segmentation. In: Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015. Series Title: Lecture Notes in Computer Science, vol. 9351. Cham: Springer International Publishing; 2015. p. 234–41.

[33] He Kaiming, Gkioxari Georgia, Dollár Piotr, Girshick Ross. Mask r-cnn. In: 2017 IEEE International Conference on Computer Vision (ICCV); 2017. p. 2980–8.

[34] ITK-SNAP Home. http://www.itksnap.org/pmwiki/pmwiki.php.

[35] Xiong Xuehan, De la Torre Fernando. Supervised descent method and its applications to face alignment. In: 2013 IEEE conference on computer vision and pattern recognition; June 2013. p. 532–9. Portland, OR, USA. [IEEE].

[36] Dogan Ramazan Ozgur, Dogan Hulya, Bayrak Coskun, Kayikcioglu Temel. A two-phase approach using mask R-CNN and 3D U-Net for high-accuracy automatic segmentation of pancreas in CT imaging. Comput Methods Programs Biomed August 2021;207:106141.

[37] Podder S, Bhattacharjee S, Roy A. An efficient method of detection of COVID-19 using mask R-CNN on chest X-ray images. AIMS Biophys 2021;281–290.

[38] Ren Shaoqing, He Kaiming, Girshick Ross, Sun Jian. Faster r-cnn: towards real-time object detection with region proposal networks. IEEE Trans Pattern Anal Mach Intell 2017;39(6):1137–49.

[39] Lin Tsung-Yi, Dollár Piotr, Girshick Ross, He Kaiming, Hariharan Bharath, Belongie Serge. Feature pyramid networks for object detection. In: 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR); 2017. p. 936–44.

[40] Lin Tsung-Yi. et al.. Microsoft coco: Common objects in context. 2014.

[41] Tang Yucheng, et al. Renal cortex, medulla and pelvicaliceal system segmentation on arterial phase CT images with random patch-based networks. In: Medical imaging 2021: Image processing. vol. 11596. SPIE; Feb 2021. p. 379–86.

[42] Kidneys | SEER training. https://training.seer.cancer.gov/anatomy/urinary/components/kidney.html.

[43] Oscar Contreras Carrasco. Gaussian mixture models explained. https://towardsdatascience.com/gaussian-mixture-modelsexplained-6986aaf5a95; February 2020.

[44] Xuan Guorong, Zhang Wei, Chai Peiqi. EM algorithms of Gaussian mixture model and hidden Markov model. In: Proceedings 2001 International Conference on Image Processing (Cat. No.01CH37205). vol. 1; October 2001. p. 145–8.

[45] Huang Yin, et al. Measurement and comparison of T1 relaxation times in native and transplanted kidney cortex and medulla. J Magn Reson Imaging May 2011;33(5):1241–7.

[46] Semelka RC, Corrigan K, Ascher SM, Brown JJ, Colindres RE. Renal corticomedullary differentiation: observation in patients with differing serum creatinine levels. Radiology January 1994;190(1):149–52.

[47] Vivian S Lee, et al. What causes diminished corticomedullary differentiation in renal insufficiency? J Magn Reson Imaging 2007;25(4):790–5. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/jmri.20878.

[48] Cedric M, de Bazelaire J, Duhamel Guillaume D, Rofsky Neil M, Alsop David C. MR imaging relaxation times of abdominal and pelvic tissues measured in vivo at 3.0 T: preliminary results1. Radiology 2004;230. Radiological Society of North America.

[49] Fernandez-Moral Eduardo, Martins Renato, Wolf Denis, Rives Patrick. A new metric for evaluating semantic segmentation: leveraging global and contour accuracy. In: IEEE intelligent vehicles symposium (IV), pages 1051–1056, Changshu, June 2018; 2018 [IEEE].

[50] Min-Chi Ku, María A, Fernández-Seara Frank Kober, Niendorf Thoralf. Noninvasive renal perfusion measurement using Arterial Spin Labeling (ASL) MRI: basic concept. Methods Mol Biol (Clifton, N.J.) 2021;2216:229–39.

[51] Vidorreta Marta, Wang Ze, Rodríguez Ignacio, Pastor María A, Detre John A, Fernández-Seara María A. Comparison of 2D and 3D single-shot ASL perfusion fMRI sequences. NeuroImage February 2013;66:662–71.

[52] Brumer Irene. Synthetic renal ASL MRI [data]. Technical report. 2022.

[53] Brumer Irene, Bauer Dominik F, Schad Lothar R, Zöllner Frank G. Synthetic arterial spin labeling MRI of the kidneys for evaluation of data processing pipeline. Diagnostics August 2022;12(8):1854.

[54] Peters AM. Scintigraphic imaging of renal function. Exp Nephrol 1998 Sep-Oct;6 (5):391–7. https://doi.org/10.1159/000020547. PMID: 9730654.

[55] Martin DR, Sharma P, Salman K, Jones RA, Grattan-Smith JD, Mao H, et al. Individual kidney blood flow measured with contrast-enhanced first-pass perfusion MR imaging. Radiology 2008 Jan;246(1):241–8. https://doi.org/10.1148/radiol.2461062129. 18096538.