Received: 10 March 2021 | Revised: 31 August 2021 | Accepted: 1 September 2021

DOI: 10.1002/mrm.29016

TECHNICAL NOTE

Magnetic Resonance in Medicine

Workflow for automatic renal perfusion quantification using ASL-MRI and machine learning

Isabell K. Bones[1] | **Clemens Bos**[1] | **Chrit Moonen**[1] | **Jeroen Hendrikse**[2] | **Marijn van Stralen**[1]

[1]Center for Image Sciences, University Medical Center Utrecht, Utrecht, The Netherlands

[2]Department of Radiology, University Medical Center Utrecht, Utrecht, The Netherlands

Correspondence
Isabell K. Bones, Center for Image Sciences, University Medical Center Utrecht, Heidelberglaan 100, 3584 CX Utrecht, The Netherlands.
Email: i.k.bones@umcutrecht.nl

Funding information
Applied and Engineering Sciences with project number 14951, which is (partly) financed by the Netherlands Organization for Scientific Research (NWO)

Purpose: Clinical applicability of renal arterial spin labeling (ASL) MRI is hampered because of time consuming and observer dependent post-processing, including manual segmentation of the cortex to obtain cortical renal blood flow (RBF). Machine learning has proven its value in medical image segmentation, including the kidneys. This study presents a fully automatic workflow for renal cortex perfusion quantification by including machine learning-based segmentation.

Methods: Fully automatic workflow was achieved by construction of a cascade of 3 U-nets to replace manual segmentation in ASL quantification. All 1.5T ASL-MRI data, including M_0, T_1, and ASL label-control images, from 10 healthy volunteers was used for training (dataset 1). Trained cascade performance was validated on 4 additional volunteers (dataset 2). Manual segmentations were generated by 2 observers, yielding reference and second observer segmentations. To validate the intended use of the automatic segmentations, manual and automatic RBF values in mL/min/100 g were compared.

Results: Good agreement was found between automatic and manual segmentations on dataset 1 (dice score = 0.78 ± 0.04), which was in line with inter-observer variability (dice score = 0.77 ± 0.02). Good agreement was confirmed on dataset 2 (dice score = 0.75 ± 0.03). Moreover, similar cortical RBF was obtained with automatic or manual segmentations, on average and at subject level; with 211 ± 31 mL/min/100 g and 208 ± 31 mL/min/100 g ($P < .05$), respectively, with narrow limits of agreement at −11 and 4.6 mL/min/100 g. RBF accuracy with automated segmentations was confirmed on dataset 2.

Conclusion: Our proposed method automates ASL quantification without compromising RBF accuracy. With quick processing and without observer dependence, renal ASL-MRI is more attractive for clinical application as well as for longitudinal and multi-center studies.

KEYWORDS
automatic ASL quantification, automatic segmentation, machine learning, RBF, renal ASL MRI

1 | INTRODUCTION

Functional renal imaging is an emerging field in medical research, where renal perfusion has proven its value as a potential biomarker for renal health.[1,2] Methods to accurately, non-invasively, image global renal perfusion would be of benefit for clinical routine by providing repeatable measures of global kidney function.

One promising, non-invasive method, arterial spin labeling (ASL), has already shown value in research settings.[1,2] It uses magnetically labeled blood water as an endogenous tracer, alleviating the use of contrast agents and gives the possibility to capture regional renal function of both kidneys at the same time. Nevertheless, ASL quantification involves many manual interactions, making it time consuming and observer dependent, limiting ASL clinical practicability and usage. It is recommended to report ASL MRI renal blood flow (RBF) exclusively for cortical voxels,[3] because medullary ASL perfusion is low,[3,4] and the renal pelvis contains urine, which should not contribute to renal perfusion measurements either. Segmentation of the cortex, however, requires manual interaction.[5-7] Automating the ASL quantification pipeline, excluding manual segmentation, would remove an important obstacle for wide clinical adoption and use in large multi-parametric MRI studies.

So far, at most semi-automatic approaches for renal ASL processing have been reported.[5-11] Here, segmentation of the kidney on anatomic images is performed manually, and automatic subsequent intensity-based cortical voxel extraction from T_1-maps often requires manual adaptation.[12-16]

In the field of machine learning, the potential for automatic medical image segmentation in many different organs,[17] including the kidney,[18] has recently been shown. Studies based on CT and MRI images have shown that especially convolutional neural networks[18-22] can accurately segment the entire kidney, extract its compartments, and even distinguish tumor tissue.

The aim of our study is to propose a clinically viable workflow for fully automatic renal ASL quantification from ASL MRI datasets by replacing the manual segmentation steps with fully automatic segmentation based on 3D U-nets. Performance of the cross-validated U-nets was demonstrated on another independent dataset. Finally, feasibility of the obtained automatic segmentations for renal ASL quantification is evaluated based on the accuracy of cortical RBF.

2 | METHODS

This study was approved by the local institutional review board. Written informed consent was obtained from all subjects before the examination.

2.1 | ASL quantification pipeline

The proposed quantification pipeline aims to obtain a purely cortical RBF from the ASL input data, being ASL label control pairs, M_0-images and T_1-maps (see Figure 1). It includes motion correction, segmentation, and ASL calculation steps, similar to established ASL quantification in the brain.[23]

2.1.1 | ASL data

All 1.5T ASL-MRI data were acquired at $3 \times 3 \times 6$ mm with a single-shot gradient echo EPI 2D readout in 7 oblique coronal slices with a 1-mm slice gap in each subject. The ASL protocol consisted of balanced pseudo-continuous ASL (pCASL) label (label duration = 1500 ms, post labeling delay = 1500 ms) and control pairs with 10 repetitions, an M_0-image with 3 repetitions, and inversion recovery images at 11 inversion times to map T_1. The labeling plane was carefully planned to avoid susceptibility artefacts that could influence pCASL labeling efficiency[24] and to avoid undesired labeling of the kidneys by making sure they would not move into the labeling slab during respiration. This was achieved by staying well below the diaphragm while placing the label slab as high as possible inside the FOV. Background suppression (BGS) was applied for ASL to decrease the influence of (physiological) noise and achieved by 2 hyperbolic secant inversion pulses. Two datasets, dataset 1 with 10 healthy subjects (age 23-38 years, 6 men) and dataset 2 with 4 different healthy subjects (age 24-28 years, 4 men), were used in this study, containing the same ASL sequences scanned at the same MRI scanner. Dataset 1 was used for methods development, dataset 2 for validation of the methods. Quality of all data with regard to image artifacts was visually assessed.

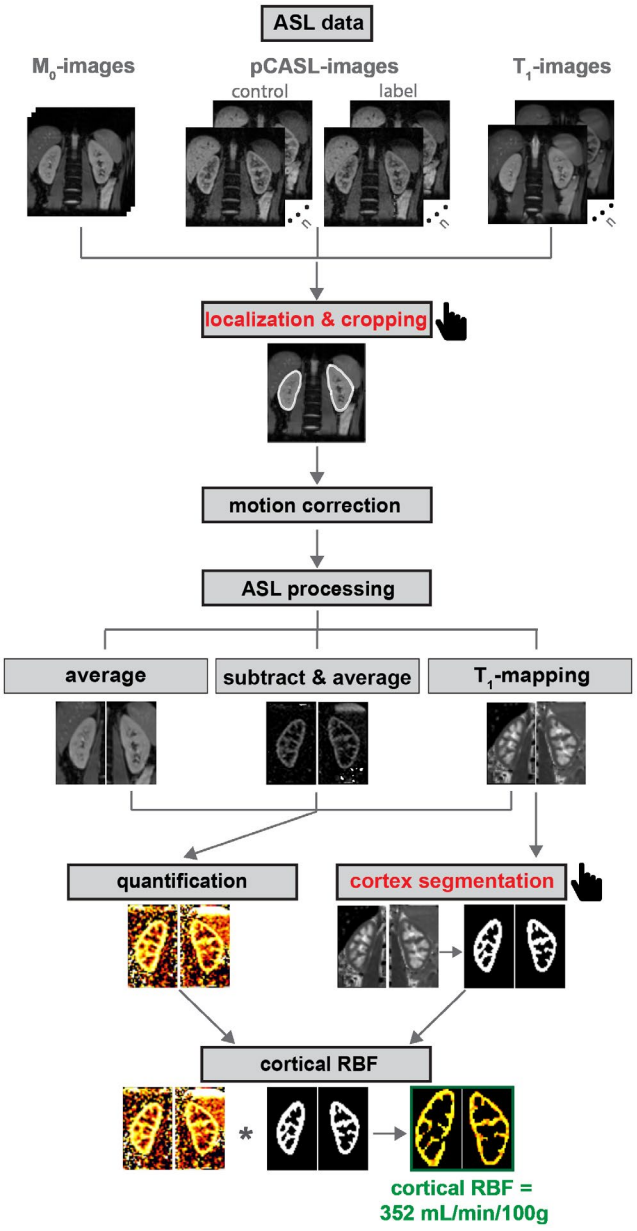

FIGURE 1 Common ASL quantification pipeline with single-slice example images. Steps requiring manual interaction are highlighted in red. Automation of those manual steps using machine learning is illustrated in Figure 2. Note that with this design the low signal ASL source data (because of background suppression) is not directly used in the segmentation task, but the segmentation information is transferred via registration

2.1.2 | Localization and cropping

Before motion correction, the kidneys were localized by a coarse manual delineation on an M_0-image, providing their location and extent. For each kidney, the image was then automatically cropped to a smaller, but sufficiently large region, to cover the kidney and allow for respiratory introduced displacement of the kidneys between images. Cropping allows for kidney wise motion correction, because kidneys may move differently,[3,25,26] without distraction by motion of background structures.

2.1.3 | Motion correction

Motion correction of cropped ASL, M_0, and T_1 images was performed collectively using Elastix,[27]

per subject, per kidney, per slice. Contrast differences between the sequences were accounted for using a principal component analysis-based (PCA) group-wise metric.[28] All images were registered collectively to a common space. We used adaptive stochastic gradient descent optimization in 250 iterations with 1000 random samples each.

2.1.4 | Cortex segmentation

Cortex segmentation is achieved via an intermediate step that segments the whole kidney outer borders. Manual segmentations of the whole kidney and cortical region for training were drawn slice-wise using ITK-SNAP[29,30] on the M_0-image (anatomic image), excluding the renal pelvis (collecting system and large vessels) (Supporting Information Figure S1). Care was taken to exclude partial volume voxels of the cortex. For cortical kidney segmentation, the whole kidney segmentation was superimposed on the T_1-map (Supporting Information Figure S1), providing corticomedullar contrast, allowing cortex segmentation. This manual process results in reference cortex and non-cortex masks, together covering the whole kidney, taking ~20 min per subject.

Two observers with experience in renal ASL data processing performed this process for all subjects of dataset 1. Dataset 2 was only segmented by 1 observer.

2.1.5 | ASL processing

After motion correction M_0-images are averaged, the ASL label and control pairs are subtracted and averaged, and the T_1-images are fitted to a mono exponential function to produce a T_1-map. Together, the T_1-map, subtraction map, and M_0-image are used for ASL signal quantification.

2.1.6 | Quantification and cortical RBF extraction

RBF was estimated using Buxton's general kinetic model[31] for continuous ASL at a single time point, defining the subtraction (ΔM) as:

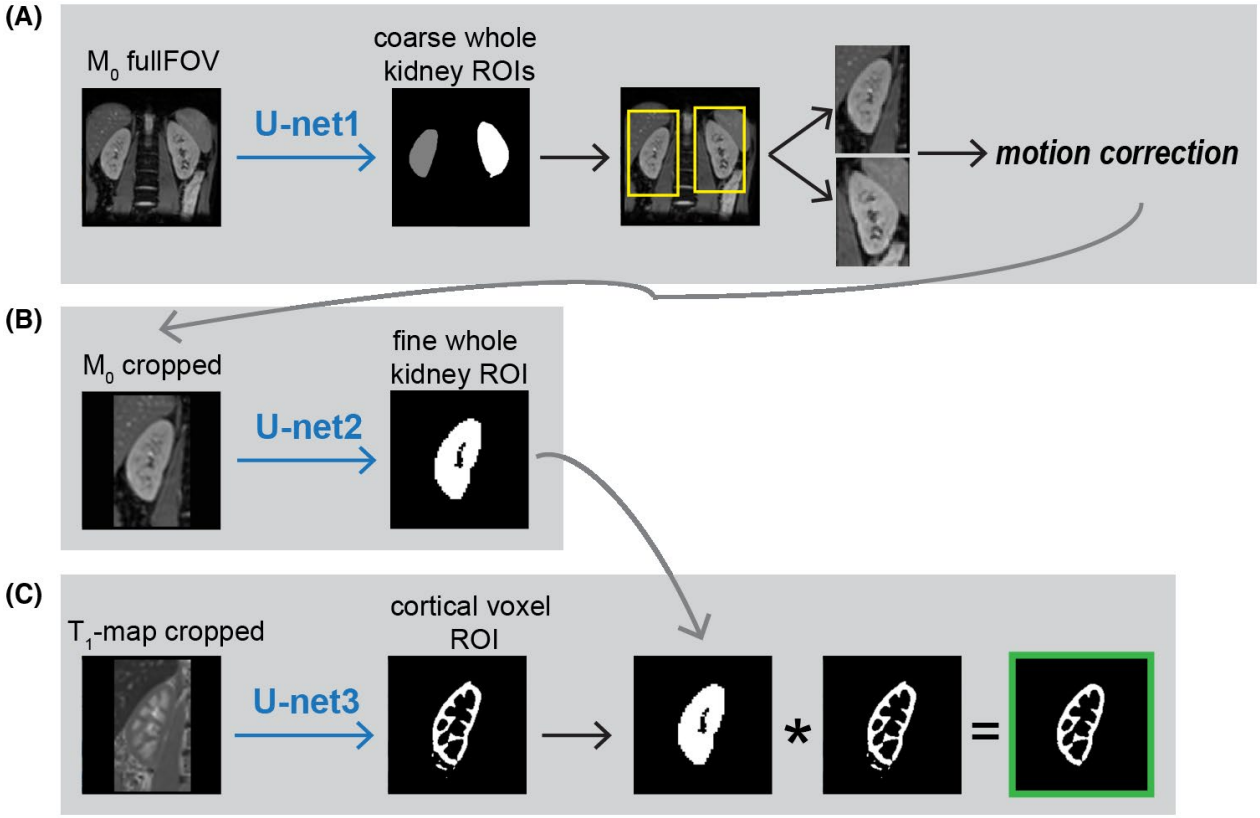

FIGURE 2 Schematic representation of our segmentation cascade for kidney localization and cortical voxel extraction to fully automatize renal ASL quantification. In Figure 1, those are steps that require manual interaction otherwise. (A) U-net1: kidney localization for image cropping. (B) U-net2: fine whole kidney segmentation. (C) U-net3: cortical segmentation. In a last step, U-net2 and U-net3 segmentations are multiplied to remove eventual erroneous cortical predictions outside of the kidney

$$\Delta M = \frac{M_0 \cdot \alpha_{BGS}^{n} \cdot RBF \cdot T_{1t'} \cdot 2\alpha_{pCASL} \cdot e^{-ATT/T_{1b}} \cdot e^{-(PLD-BD-ATT)/T_{1t'}} \cdot \left(1 - e^{-BD/T_{1t'}}\right)}{6000 \cdot \lambda}. \quad (1)$$

The apparent tissue relaxation $T_{1t'}$ is given by $1/\left(1/T_{1t} + RBF/6000\lambda\right)$, with T_{1t} being the tissue T_1. For the blood partition coefficient λ 0.9 mL/g was used.[24] α_{pCASL}, the label pulse inversion efficiency, was set to 0.85.[24] Post-labeling-delay (PLD) and bolus duration (BD) were 1500 ms each, and the arterial transit time (ATT) was assumed to be 750 ms.[32] BGS inversion efficiency α_{BGS} was assumed to be 0.95, for each of the n BGS pulses applied. T_1 of arterial blood (T_{1b}) at 1.5T was set to 1350 ms.[24]

Voxelwise fitting results in a perfusion map with RBF values in mL/100 g/min. Finally, the cortical segmentation is used to obtain a purely cortical RBF value.

2.2 | Automatic segmentation cascade

2.2.1 | Pre-training image processing

All processing before training was performed using MeVisLab (MeVis Medical Solutions AG, Bremen, Germany). For each dataset, 5 central slices were selected from 7 coronal slices because the outer slices in some subjects did not contain kidney tissue. Slices were appended to allow for 3D training. Data was resampled in-plane to 1 × 1 mm voxels and padded to a square image. M_0-images were normalized to 0 mean and unit standard deviation (SD). T_1-maps consisted of quantitative values and were not scaled or normalized to preserve inter-subject variation.

2.2.2 | Cascade

To fully automate renal ASL MRI quantification, a cascade consisting of 3 subsequent 3D U-nets was implemented to replace the manual tasks: (1) localization and cropping (U-net1), (2) whole kidney segmentation (U-net2), and (3) cortical voxel extraction within the whole kidney (U-net3). The cascade is illustrated in Figure 2A-C. (1) Pre-registration, U-net1 performs a coarse segmentation

of each kidney on the full FOV M_0-images (256 × 256 after padding) to localize and separate the kidneys. These kidney locations are then used for cropping all ASL-data of 1 volunteer. (2) After motion correction, U-net2 predicts fine whole kidney segmentations based on cropped M_0-images (160 × 160 after padding), and (3) U-net3 extracts cortical segmentations from cropped T_1-maps (160 × 160 after padding). Resulting predictions from U-net2 and U-net3 are voxel-wise multiplied to remove potential erroneous cortical predictions outside of the kidney. The resulting cortical segmentation was used for ASL quantification.

2.2.3 | Architecture

All U-nets were implemented and trained using Keras. Supporting Information Figure S3 illustrates the detailed U-net[33] architecture, with 5 resolution levels and residual blocks per level. The final layer is a 1 × 1 convolution that maps each feature to a class using sigmoid activation. For U-net1 multiple classes were predicted (background, left, and right kidney) using a soft-max activation.

2.2.4 | Training and hyper-parameters

The 3 U-nets in the segmentation cascade were independently trained on dataset 1 in a cross-validation setup, using the masks of observer 1 as the reference. The following hyperparameters were optimized over all folds with dataset 1 per U-net. For that matter, we explored the use of batch sizes of (1, 2, 3, and 4), epochs between 100 and 700 as well as dynamic or fixed learning rate of (0.005, 0.01, and 0.02) and (0.001-0.005), respectively. A NAdam optimizer was used. Per epoch, all data was seen.

U-net1: a dynamic learning rate of lr = 0.01 (momentum = 0.8, decay = lr/epochs) was used to optimize the generalized dice loss,[34] with a batch size of 2 for 200 epochs reaching convergence.
U-net2: the loss was composed as the sum of soft dice loss, binary cross entropy and volume difference, and optimized with a learning rate of 0.001, using a batch size of 3 for 350 epochs.
U-net3: the loss was composed as the sum of soft dice loss, binary cross entropy and volume difference, and optimized with a dynamic learning rate of lr = 0.01 (momentum = 0.8, decay = lr/epochs), using a batch size of 3 for 150 epochs.

Training and testing in cross-validation of the cascade took 55 min using an NVIDIA GeForce 1060 GPU with 6GB RAM.

2.2.5 | Segmentation post-processing

For U-net1 and U-net2 connected component analysis was used to remove small segmentation outliers. For U-net3 this was not necessary. We masked the cortical segmentation from this network with the whole kidney segmentation from U-net2.

2.3 | Evaluation of cascade performance and inter-observer variability

Cross-validation results as well as inter-observer variability of cortical segmentations from dataset 1 were reported as well as the cascade performance on an independent dataset 2.

2.3.1 | Segmentation performance

Network performance evaluation was based on the dice score (DS) (%), Hausdorff distance (HD) (mm) and volumetric difference (VD) (%) between automatic cortical segmentations and manual segmentations. The DS measures the volumetric overlap between 2 segmentations (Equation 1), whereas the HD is a contour distance measure, especially penalizing false positive segmentation outliers (Equation 2). The volumetric difference measures volume bias (i.e., over or under segmentation) (Equation 3).

$$\boldsymbol{DS} = \left(\frac{2 \cdot |ref \cap pred|}{|ref| + |pred|}\right) \cdot 100, \tag{2}$$

$$\boldsymbol{HD}(REF, PRED) = \max\{\min\{d(ref, pred)\}\} \tag{3}$$

with

$$ref \in REF \; pred \in PRED$$

$$\boldsymbol{VD}(REF, PRED) = (V_{PRED} - V_{REF})/V_{REF} \tag{4}$$

with

$$V = voxelSize \cdot voxels$$

2.3.2 | Evaluation of segmentation performance on ASL perfusion quantification

To study the influence of the accuracy of automatic segmentation of cortical segmentations for renal ASL quantification, therefore their ultimate usage, the RBF in mL/100 g/min per subject was quantified using manual and automatic cortical segmentations.

All statistical analyses were performed using GraphPad Prism 8 version 8.0.1 (244) for Windows (GraphPad Software, San Diego, CA). Differences in cortical RBF were

tested between predictions and reference (observer 1) as well as observer 2 and reference using paired *t* tests with a significance level α of 0.05. Moreover, Bland-Altman analyses were performed to investigate the agreement of subject level cortical RBF gained from ASL quantification using predictions, reference, and observer 2 cortical segmentations; bias is reported together with the standard error of the mean (SEM) of the differences next to the 95% confidence intervals.

3 | RESULTS

Automated ASL quantification pipeline was obtained by replacing the manual steps (highlighted in red) by the U-nets from the proposed cascade in Figure 2. Automatic cortical segmentations of 2 kidneys took <1 s.

3.1 | Evaluation of cascade performance and inter-observer variability

3.1.1 | Segmentation performance

A summary of segmentation performance results using different cortical segmentations on dataset 1 and dataset 2 is given in Table 1. On average, cross-validation on dataset 1 yielded a DS of 0.78 ± 0.04, a HD of 6.3 ± 1.2 mm, and a VD of −9.6% ± 5.4%. Small SDs indicate consistent performance among subjects. Visual illustration of final cascade performance is given in Figure 3 and Supporting Information Figure S2.

Between the 2 manual observers (observer 2-observer 1), we found a comparable DS of 0.77 ± 0.02 and a HD of 8.5 ± 4.0 mm and a larger VD of 27.7% ± 5.6% (Table 1).

Testing the trained cascade on an independent dataset 2 yielded DS of 0.75 ± 0.03, HD of 7.0 ± 1.7 mm, and VD of −20.0% ± 5.6% (Table 1). Those results are in line with cross-validation as well as inter-observer variability stated above on dataset 1.

3.1.2 | Influence of segmentation performance on ASL perfusion quantification

An example of how the segmentations are used for cortical extraction in the ASL quantification pipeline is illustrated in Figure 3. Automatic cortical segmentations in dataset 1 yielded cortical RBF values, which were in line with those obtained using the manual reference, on average as well as on individual level (Figure 4A).

Similar group average RBF was found for predictions and reference with 211 ± 31 mL/min/100 g and 208 ± 31 mL/min/100 g, respectively. Bland-Altman analysis on RBF values of automatic and manual segmentations exposed a bias of −3.2 mL/min/100 g with a SEM of ±1.3 mL/min/100 g (Figure 4B). This bias was small, but significant ($P = .032$). Consistency across subjects was underlined by small limits of agreement at −11 and 4.6 mL/min/100 g, which was similar to the intra-observer limits of agreement of 3.8 and 23 mL/min/100 g and small compared with the "effect size", that is, the RBF variation between subjects (±31 mL/min/100 g).

Inter-observer variability was slightly but significantly higher on average and also on individual level (Figure 4C); with a significantly lower average cortical RBF of 195 ± 27 mL/min/100 g using the segmentations of observer 2, ($P < .001$) resulting in a larger bias of 14 mL/min/100 g in the Bland-Altman analysis. This was consistent between subjects as indicated by small limits of agreement at 3.8 and 23 mL/min/100 g (Figure 4D) in comparison with the variance between subjects.

The RBF accuracy with automated segmentations was confirmed on the independent dataset 2, resulting in an accurate cortical RBF, which was slightly but significantly higher than based on manual segmentations: 262 ± 33 mL/min/100 g versus 248 ± 32 mL/min/100 g, respectively ($P < .05$). Supporting Information Figure S4 provides individual RBF values per subject of dataset 2.

TABLE 1 Performance evaluation metrics DS, HD, and VD averaged over all included subjects in the dataset

Dataset		DS	HD (mm)[a]	VD (%)
1	Reference vs prediction	0.78 (0.04)	6.3 (1.2)	−9.6 (5.4)
	Reference vs observer 2	0.77 (0.02)	8.5 (4.0)	27.7 (5.6)
2	Reference vs prediction	0.75 (0.03)	7.0 (1.7)	−20.0 (5.6)

First row: on dataset 1 used for training and cross-validation. Comparison for reference vs automatic prediction as well as reference versus second manual observer (observer 2). Second row: on independent dataset 2. Comparison for reference versus automatic prediction. Standard deviation of evaluation metrics between subjects displayed in brackets.

Abbreviations: DS, dice score; HD, Hausdorff distance; VD, volume difference.

[a]Note the low original acquired image resolution of 2.54 × 2.54 × 6 mm.

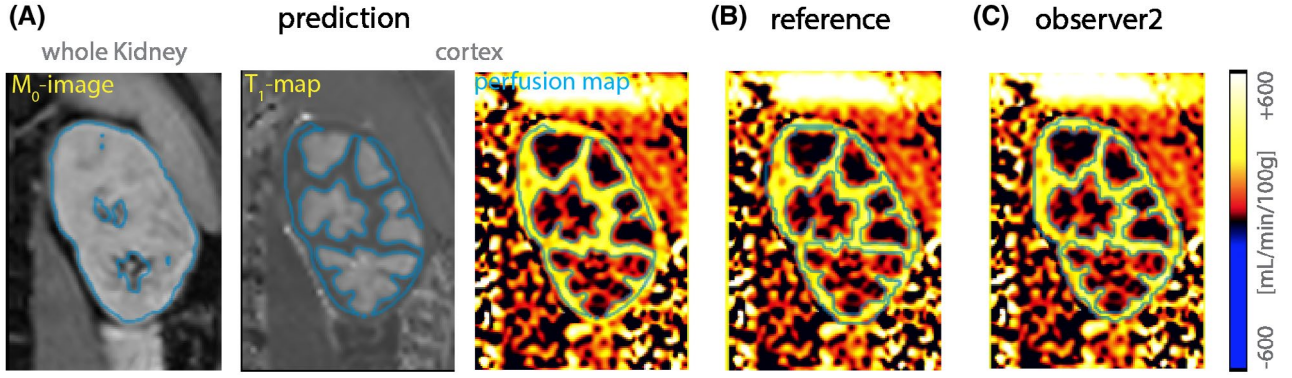

FIGURE 3 Single slice segmentation example. Segmentations are displayed in blue contours. (A) M_0-image with whole kidney contours as a result from U-net2. T_1-map and perfusion map with cortical contours as a result from U-net3, corrected with U-net2 output. (B) Reference cortical contours manually drawn by observer 1. (C) Cortical contours manually drawn by observer 2. Good agreement between the 3 different cortical contours can be seen; the bright cortical perfusion signal is captured by all contours, assuring accurate mean cortical RBF calculation

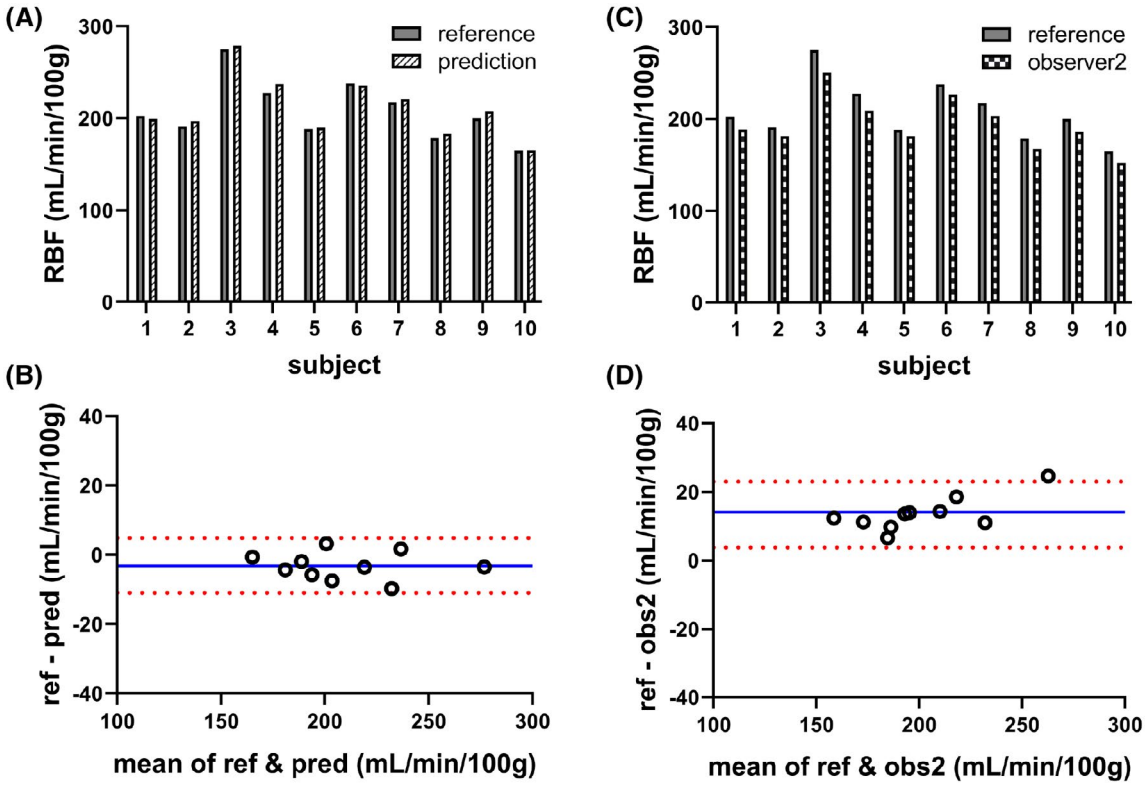

FIGURE 4 (A) Cortical RBF per subject quantified with either the reference (gray) or the prediction (striped). (B) Bland-Altman plot of cortical RBF values resulting from ASL analysis using the reference (ref) and predicted (pred) cortical segmentation. Solid blue line: mean difference, dotted red lines: 95% limits of agreement. (C) Cortical RBF per subject quantified with either the reference (gray) or the second observer (hatched). (D) Bland-Altman plot of cortical RBF values resulting from ASL analysis using the reference (ref) and observer 2 (obs2) cortical segmentation

4 | DISCUSSION

In this study, we demonstrate the feasibility of fully automatic renal cortex segmentation using machine learning for automatic ASL quantification.

To the best of our knowledge, there are no previous studies reporting the influence of segmentation accuracy on quantified cortical RBF. Certainly, creating a ground truth segmentation with high accuracy is difficult because of the large ASL voxel size. With the current study

we have shown that differently generated segmentations, manually or automatically, do have an effect on the quantified cortical RBF; which was small, −3.2 mL/100 g/min (equivalent to 1.5%), but still significant. This remains small although in comparison to pathology induced cortical RBF changes, for example, the 28% difference observed in patients with diabetic nephropathy compared to healthy controls.[5] Additionally, we found a larger inter-observer variability with 14 mL/100 g/mL in cortical RBF in our study, which is equivalent to 7%. At the same time, previous reports on ASL scan re-scan variability state RBF coefficients of variation in healthy subjects in the range of 4%-13%.[2,6,10,14,35,36] With fully automatic processing this variation may be reduced, giving promising ground for more reproducible renal ASL.

Moreover, this small, but still significant difference in RBF was found between predictions (211 ± 31 mL/min/100 g) and reference (208 ± 31 mL/min/100 g) data. This shows the substantial biological variation across subjects but also suggests segmentation consistency between observers, which is additionally supported by small SD in segmentation evaluation metrics of this study (Table 1).

We found DS of 0.78 and 0.77 between manual and automatic segmentations and between observers, respectively, which is somewhat low as compared to other studies with high resolution imaging.[21] DS values tend to be lower for non-convex thin objects such as the renal cortex, as well as for images with a limited resolution with respect to the size of the object such as this renal imaging data, which results in a large portion of partial volume voxels. With that in mind, a DS of 0.78 is a quite satisfactory result. Moreover, relating the DS to the inter-observer DS of 0.77, we qualify 0.78 as good because it makes the automatic method a good alternative to manual segmentation, especially because of the limited influence of the segmentation differences on the desired perfusion measurement.

The HD in the renal cortex was mostly determined by a small number of pixels, which hardly contributed to the mean RBF. Hence, for agreement in determining cortical RBF, it is expected that DS is a more predictive measure.

In the end, the prediction cascade has been trained with the reference segmentations and should therefore generate similar results, without significant difference. This could be caused by the connected component analysis that we applied as a processing step of the generated whole kidney segmentations (U-net1 and U-net2) to remove small errors in segmentations in the periphery, but is not included during training. Future work could include connected component analysis within training for optimal usage of a loss function that is volume aware.

Our method is entirely based on the structural imaging of the kidney, and independent of the functional imaging. Other methods on functional MRI (e.g., DCE MRI) use the functional/temporal information to distinguish renal tissues, which might introduce an undesired dependence of the segmentation on kidney function.[37,38] However, in cases where T_1 of cortex and/or medulla is decreasing because of disease,[39] our cascade network could drop in performance and additional training will be necessary.

The proposed segmentation method may be used on ASL datasets acquired using a different readout sequence, (e.g., 3D GRASE or 2D spin echo EPI). Additional preprocessing steps or retraining may be required to adapt the method accordingly, especially when applying it to data from scanners at different field strength. Intensity inhomogeneities may occur on higher field strengths because of field inhomogeneities, often hampering automatic image segmentation because of diminishing tissue contrast. It has been demonstrated that simulation of those artifacts and their inclusion for training improves automatic image segmentation,[40] which could be a valuable addition when extending our method to 3T images. Further interesting future work could be to investigate convergence during training with increasing number of training data.

To evaluate the actual clinical performance of our proposed automatic method, a larger subject group, including patients is needed. Future studies should focus on increasing generalizability of the presented method by training on large amounts of data from different scanners and sites.

5 | CONCLUSION

Our proposed framework automates crucial steps in renal ASL quantification, removing observer dependence and increasing time efficiency and scalability. With this, we show the potential of automatically predicted segmentations to take away an important barrier for adoption of non-invasive quantitative renal ASL-MRI perfusion imaging in clinical practice.

ACKNOWLEDGMENTS

This work is part of the research program Applied and Engineering Sciences with project number 14951, which is (partly) financed by the Netherlands Organization for Scientific Research (NWO). We thank MeVis Medical Solutions AG (Bremen, Germany) for providing MeVisLab medical image processing and visualization environment, which was used for image analysis.

CONFLICT OF INTEREST

Marijn van Stralen: Co-founder and shareholder of MRIguidance B.V.

ORCID

Isabell K. Bones https://orcid.org/0000-0002-7916-5013

REFERENCES

1. Selby NM, Blankestijn PJ, Boor P, et al. Magnetic resonance imaging biomarkers for chronic kidney disease: a position paper from the European Cooperation in Science and Technology Action PARENCHIMA. *Nephrol Dial Transplant*. 2018;33(suppl_2):ii4-ii14.
2. Odudu A, Nery F, Harteveld AA, et al. Arterial spin labelling MRI to measure renal perfusion: a systematic review and statement paper. *Nephrol Dial Transplant*. 2018;33(suppl_2): ii15-ii21.
3. Nery F, Buchanan CE, Harteveld AA, et al. Consensus-based technical recommendations for clinical translation of renal ASL MRI. *Magn Reson Mater Phys Biol Med*. 2020;33:141-161.
4. Wu WC, Su MY, Chang CC, Tseng WYI, Liu KL. Renal perfusion 3-T MR imaging: a comparative study of arterial spin labeling and dynamic contrast-enhanced techniques. *Radiology*. 2011;261:845-853.
5. Mora-Gutiérrez JM, Garcia-Fernandez N, Slon Roblero MF, et al. Arterial spin labeling MRI is able to detect early hemodynamic changes in diabetic nephropathy. *J Magn Reson Imaging*. 2017;46:1810-1817.
6. Gillis KA, McComb C, Patel RK, et al. Non-contrast renal magnetic resonance imaging to assess perfusion and corticomedullary differentiation in health and chronic kidney disease. *Nephron*. 2016;133:183-192.
7. Robson PM, Madhuranthakam AJ, Smith MP, et al. Volumetric arterial spin-labeled perfusion imaging of the kidneys with a three-dimensional fast spin echo acquisition. *Acad Radiol*. 2016;23:144-154.
8. Nery F, Gordon I, Thomas D. Non-invasive renal perfusion imaging using arterial spin labeling MRI: challenges and opportunities. *Diagnostics*. 2018;8:2.
9. Wang J, Zhang Y, Yang X, et al. Hemodynamic effects of furosemide on renal perfusion as evaluated by ASL-MRI. *Acad Radiol*. 2012;19:1194-1200.
10. Hammon M, Janka R, Siegl C, 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. *Med (United States)*. 2016;95:1-9.
11. Cai Y-Z, Li Z-C, Zuo P-L, et al. Diagnostic value of renal perfusion in patients with chronic kidney disease using 3D arterial spin labeling. *J Magn Reson Imaging*. 2017;46:589-594.
12. Buchanan C, Cox E, Francis S. Evaluation of 2D imaging schemes for pulsed arterial spin labeling of the human kidney cortex. *Diagnostics*. 2018;8:43.
13. Bones IK, Franklin SL, Harteveld AA, et al. Influence of labeling parameters and respiratory motion on velocity-selective arterial spin labeling for renal perfusion imaging. *Magn Reson Med*. 2020;84:1919-1932.
14. Cox EF, Buchanan CE, Bradley CR, et al. Multiparametric renal magnetic resonance imaging: validation, interventions, and alterations in chronic kidney disease. *Front Physiol*. 2017;8:1-15.
15. Gardener AG, Francis ST. Multislice perfusion of the kidneys using parallel imaging: image acquisition and analysis strategies. *Magn Reson Med*. 2010;63:1627-1636.
16. Breidthardt T, Cox EF, Squire I, et al. The pathophysiology of the chronic cardiorenal syndrome: a magnetic resonance imaging study. *Eur Radiol*. 2015;25:1684-1691.
17. Hesamian MH, Jia W, He X, Kennedy P. Deep learning techniques for medical image segmentation: achievements and challenges. *J Digit Imaging*. 2019;32:582-596.
18. Isensee F, Jaeger PF, Kohl SAA, Petersen J, Maier-Hein KH. nnU-Net: a self-configuring method for deep learning-based biomedical image segmentation. *Nat Methods*. 2021;18:203-211.
19. Hou X, Xie C, Li F, Nan Y. Cascaded semantic segmentation for kidney and tumor. In Conference paper: 2019 kidney tumor segmentation challenge: KiTS19. Published online. 2019;2-6.
20. Zhang Y, Wang Y, Hou F, et al. Cascaded volumetric convolutional network for kidney tumor segmentation from CT volumes. *arXiv*. 2019:1-7.
21. Couteaux V, Si-Mohamed S, Renard-Penna R, et al. Kidney cortex segmentation in 2D CT with U-Nets ensemble aggregation. *Diagn Interv Imaging*. 2019;100:211-217.
22. Heller N, Isensee F, Maier-Hein KH, et al. The state of the art in kidney and kidney tumor segmentation in contrast-enhanced CT imaging: results of the KiTS19 challenge. *Med Image Anal*. 2021;67:101821.
23. Mutsaerts HJMM, Petr J, Groot P, et al. ExploreASL: An image processing pipeline for multi-center ASL perfusion MRI studies. *Neuroimage*. 2020;219:1053-8119.
24. Alsop DC, Detre JA, Golay X, et al. Recommended implementation of arterial spin-labeled 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;73:102-116.
25. Siva S, Pham D, Gill S, et al. An analysis of respiratory induced kidney motion on four-dimensional computed tomography and its implications for stereotactic kidney radiotherapy. *Radiat Oncol*. 2013;8:1.
26. Song R, Tipirneni A, Johnson P, Loeffler RB, Hillenbrand CM. Evaluation of respiratory liver and kidney movements for MRI navigator gating. *J Magn Reson Imaging*. 2011;33:143-148.
27. Klein S, Staring M, Murphy K, Viergever MA, Pluim J. elastix: a toolbox for intensity-based medical image registration. *IEEE Trans Med Imaging*. 2010;29:196-205.
28. Huizinga W, Poot D, Guyader J-M, et al. PCA-based groupwise image registration for quantitative MRI. *Med Image Anal*. 2016;29:65-78.
29. Yushkevich PA, Piven J, Hazlett HC, et al. User-guided 3D active contour segmentation of anatomical structures: Significantly improved efficiency and reliability. *NeuroImage*. 2006;31:1116-1128.
30. ITK-SNAP Home Accessed March 8, 2021. http://www.itksnap.org/pmwiki/pmwiki.php
31. Buxton RB, Frank LR, Wong EC, Siewert B, Warach S, Edelman RR. A general kinetic model for quantitative perfusion imaging with arterial spin labeling. *Magn Reson Med*. 1998;40:383-396.
32. Roberts DA, Detre JA, Bolinger L, et al. Renal perfusion in humans: MR imaging with spin tagging of arterial water. *Radiology*. 1995;196:281-286.
33. Ronneberger O, Fischer P, Brox T. U-net: convolutional networks for biomedical image segmentation. *Lect Notes Comput Sci*. 2015;9351:234-241.
34. Sudre CH, Li W, Vercauteren T, Ourselin S, Jorge Cardoso M. Generalised dice overlap as a deep learning loss function for highly unbalanced segmentations. *Lect Notes Comput Sci (including Subser Lect Notes Artif Intell Lect Notes Bioinformatics)*. 2017;10553 LNCS:240-248.

35. Artz NS, Sadowski EA, Wentland AL, et al. Arterial spin labeling MRI for assessment of perfusion in native and transplanted kidneys. *Magn Reson Imaging*. 2011;29:74-82.
36. Lanzman RS, Wittsack H-J, Martirosian P, et al. Quantification of renal allograft perfusion using arterial spin labeling MRI: initial results. *Eur Radiol*. 2010;20:1485-1491.
37. Yoruk U, Hargreaves BA, Vasanawala SS. Automatic renal segmentation for MR urography using 3D-GrabCut and random forests. *Magn Reson Med*. 2018;79:1696-1707.
38. Li S, Zöllner FG, Merrem AD, et al. Wavelet-based segmentation of renal compartments in DCE-MRI of human kidney: initial results in patients and healthy volunteers. *Comput Med Imaging Graph*. 2012;36:108-118.
39. Wolf M, de Boer A, Sharma K, et al. Magnetic resonance imaging T1- and T2-mapping to assess renal structure and function: a systematic review and statement paper. *Nephrol Dial Transplant*. 2018;33:ii41-ii50.
40. Khalili N, Lessmann N, Turk E, et al. Automatic brain tissue segmentation in fetal MRI using convolutional neural networks. *Magn Reson Imaging*. 2019;64:77-89.

SUPPORTING INFORMATION

Additional supporting information may be found in the online version of the article at the publisher's website.

FIGURE S1 Manual segmentation steps based on M_0-image and T_1-map. For the whole kidney segmentation, the M_0-image (anatomic image) was used and all kidney voxels were labeled, excluding the renal pelvis (collecting system and large vessels). Care was taken to exclude partial volume voxels of the cortex with the surrounding background tissue or with the renal pelvis, respectively. To obtain the cortical kidney segmentation, the whole kidney segmentation was superimposed on the T_1-map, which offers good corticomedullary contrast. On the T_1-map, cortical voxels differentiate from the medulla with a brighter signal, based on which cortical voxels were labeled

FIGURE S2 (A) Five slices of the left kidney of 1 subject with high cascade network performance. Good accordance between reference and prediction are seen (dice score = 0.81). (B) An example for poorer network performance with under-segmentation (dice score = 0.66). Orange arrows point out areas where cortical regions are incorrectly not labeled. (C) Example of cascade performance on the independent dataset 2 (dice score = 0.81). Good agreement between the automatic and manual segmentation can be seen

FIGURE S3 Architecture illustration. Each network consists of a contraction and expansion path. In the contraction path, features are encoded by 5 layers with each containing 2 convolution-normalization activation blocks. In the expansion path, up sampling is followed by double convolutions. Arrows denote operations and number of feature-maps per layer is displayed above the blocks

FIGURE S4 Cortical RBF for the 4 subjects of dataset 2, quantified with either the reference or the prediction from the trained automatic segmentation cascade. Reference is represented in gray, prediction in striped bars. Using the prediction yields slightly higher RBF than using the reference

How to cite this article: Bones IK, 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;87:800–809. https://doi.org/10.1002/mrm.29016