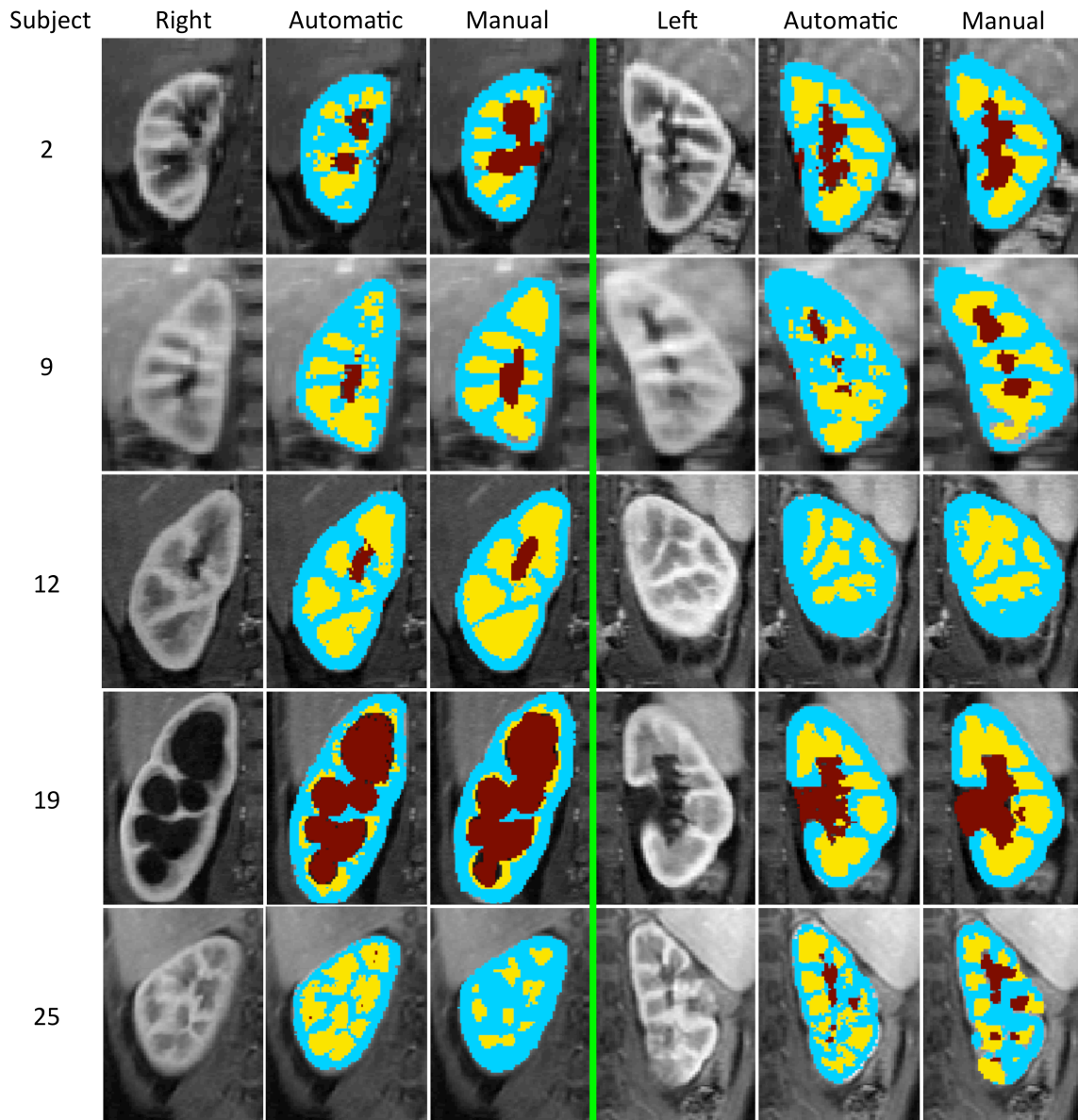


Supporting Document

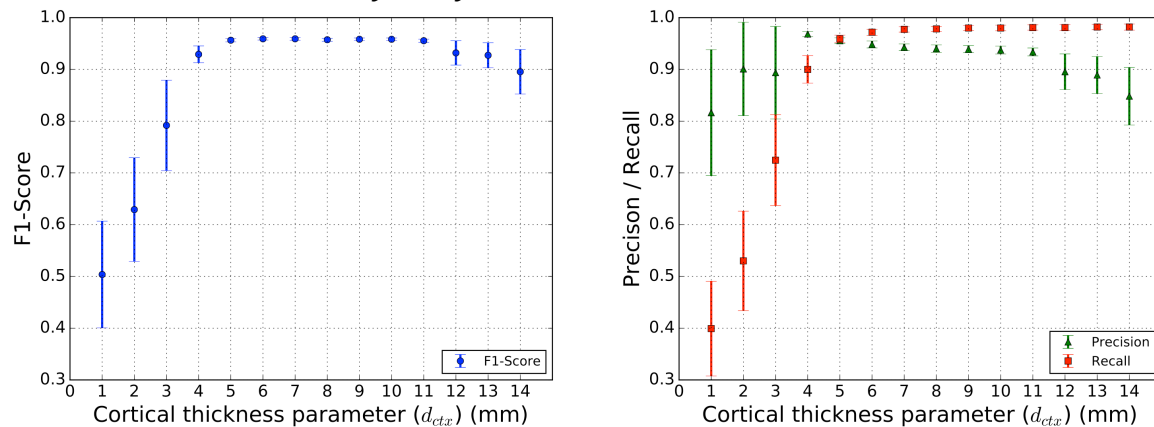
Segmentation Examples

Additional segmentation examples are shown in Supporting Figure S1.



Supporting Figure S1. The automatic segmentation results of five subjects. On each half the leftmost column is a selected temporal phase of the DCE-MRI image, the middle column is the output of the automatic segmentation algorithm and the rightmost column is the manual segmentation map (i.e. ground truth). Subjects 2 and 9 are from Group 1, 12 and 19 are from Group 2, and 25 is from Group 3.

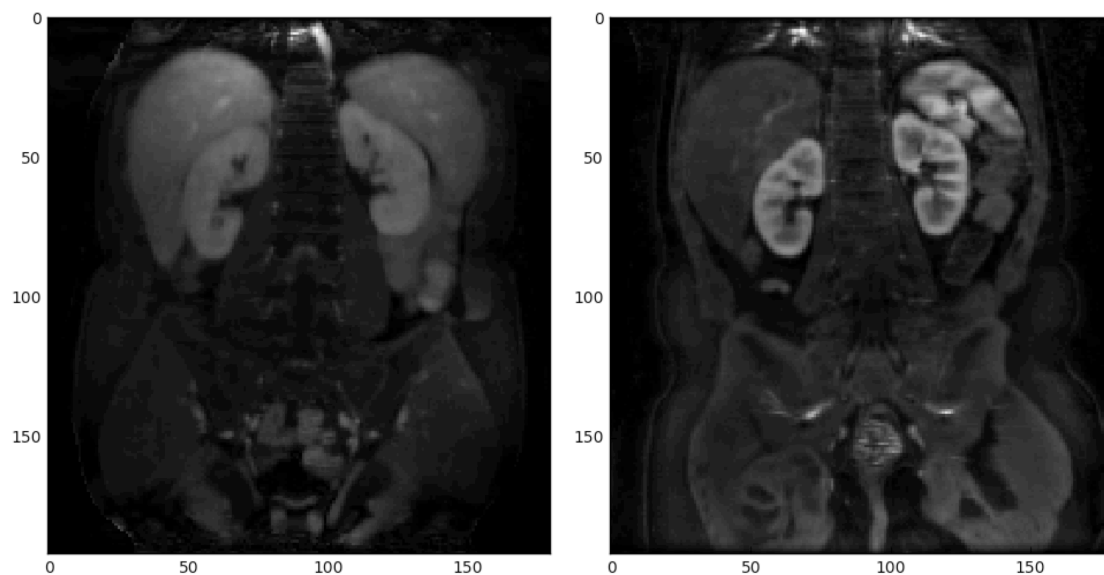
Cortical Thickness Sensitivity Analysis



Supporting Figure S2. Automatic segmentation performance with various cortical thickness (d_{ctx}) assumptions. The error bars show the standard error. The segmentation performance was not affected by the cortical thickness parameters between 5 mm and 11 mm.

Identifying Kidney Location

Supporting Figure S2 shows some of the cases where renal cortex detection is difficult.



Supporting Figure S3. Locating renal cortex from intensity images can be difficult. The left image is taken from Subject 4 at 45s after the initial enhancement of the renal cortex. Notice the interface between the liver and the renal cortex. These two tissues are hard to distinguish based on the intensity level alone. The image on the

right, taken from Subject 3 at 6s after the initial enhancement, has a cortical enhancement non-distinguishable from the splenic tissue. In both of these cases relying on the intensity of the cortical voxels prevent successful segmentation. Since the renal cortex does not always stand out as it does in some pristine datasets, we have decided to evaluate voxels based on their intensity-time curves. We targeted the medullary voxels, which have distinctly shaped signal time curves as seen in Fig. 2.

Feature Selection

The features used for random forest based cortical segmentation are listed below:

- Signal intensity values at $t = 0$ s, 30 s, 60 s, 90 s, 120s, 150s
- Depth of voxel in the renal tissue

The effectiveness of each feature was tested by removing the feature from the feature space and training the models with the remaining features. The decrease in the F1-score indicates the contribution of the feature to the classification task. In some cases the data acquisition may not be long enough to extend up to 150 seconds. For these short acquisition cases, separate models with fewer signal intensity features are trained and evaluated. These results are listed in Supporting Table S1.

Supporting Table S1 - The segmentation performance using different subsets of the feature space. $SI(t \leq T)$ indicates the signal intensity features at times up to and including T . For example $SI(t \leq 90s)$ represents four signal intensity features at $t = 0$ s, 30s, 60s, and 90s. $Depth$ is the depth of the voxel in renal tissue obtained from bulk kidney segmentation masks. The feature space used in the final model is displayed in bold.

Feature Space	Precision	Recall	F1-score
<i>Depth only</i>	0.69	0.79	0.74
<i>SI($t \leq 150$)</i>	0.82	0.76	0.79
<i>SI($t \leq 120$)</i>	0.80	0.76	0.78
<i>SI($t \leq 90$)</i>	0.78	0.73	0.75
<i>SI($t \leq 60$)</i>	0.73	0.72	0.72
<i>SI($t \leq 30$)</i>	0.69	0.73	0.71
<i>SI($t = 0$)</i>	0.36	0.22	0.27
<i>Depth + SI($t \leq 150$)</i>	0.85	0.91	0.88
<i>Depth + SI($t \leq 120$)</i>	0.84	0.91	0.87
<i>Depth + SI($t \leq 90$)</i>	0.85	0.90	0.87
<i>Depth + SI($t \leq 60$)</i>	0.83	0.88	0.85
<i>Depth + SI($t \leq 30$)</i>	0.83	0.87	0.85

<i>Depth + SI(t = 0)</i>	0.78	0.86	0.82
--------------------------	------	------	------

In addition to the features listed above, the following features were added to the feature space that resulted in no increase in F1-score:

- Dynamic range of the voxel ($SI_{\max} - SI_{\min}$)
- Time to 50% enhancement over baseline
- Time to 90% enhancement over baseline
- Local Binary Patterns ($r=1,2,3$ voxels at $t=0, 30, 120, 180$ seconds)

Addition of the dynamic range feature did not improve the result because it can be roughly described with $SI(0)$ and $SI(150)$. Similarly, time to 50% and 90% enhancement features are also redundant when the signal intensity levels are given at 30s intervals. The use of local binary patterns also did not provide any improvement. We believe the reason for this is the lack of texture in the cortex and medulla segments at the given resolution.