#####

[**Table of contents:**](#_heading=h.gjdgxs)1

[**Appendix S1.** MRI acquisition parameters.](#_heading=h.1fob9te)2

[**Appendix S2.** Flowchart of patient inclusion/exclusion.](#_heading=h.3znysh7)6

[**Appendix S3.** Details on model training and performed experiments.](#_heading=h.2et92p0)7

[**Table S4.** Detailed cohort characteristics for training/test split of datasets A and B.](#_heading=h.tyjcwt)9

[**Table S5.** Results of automated kidney and liver segmentation in test datasets A and B, for all the sequences.](#_heading=h.3dy6vkm)10

[**Table S6.** Patient stratification according to Mayo height-adjusted TKV risk classes for automated versus reference methods of kidney volumetry in the test set A.](#_heading=h.1t3h5sf)11

[**Table S7.** Comparison of differences in kidney and liver volume between baseline and follow-up for manual tracing and automated method.](#_heading=h.4d34og8)12

**Table S8.** TKV growth by the Mayo imaging class for patients with longitudinal data.

*(n – number of subjects, TKV – total kidney volume)* **13**

#####

##### **Appendix S1.** MRI acquisition parameters.

**Table S1.A**. MRI acquisition parameters for sequences from dataset A.

|  | **T2 axial** | **T2 SPIR axial** | **T2 Mapping axial** | **T2w coronal** |
| --- | --- | --- | --- | --- |
| Sequence type | TSE | TSE | GraSE | TSE |
| TE (ms) | 100 | 80 | 6.8 x 26 echoes | 100 |
| TR (ms) | 3508 | 2639 | 1007 | 977 |
| EPI factor | - | - | 7 | - |
| SENSE factor | 2 | 2 | 2 | 2 |
| Matrix size (pixels) | 440 x 440 | 328 x 328 | 320 x 238 | 252 x 238 |
| FOV (mm) | 375 x 375 | 375 x 375 | 400 x 304 | 457 x 375 |
| No. slices | 50 | 50 | 15 | 30 |
| Slice thickness (mm) | 5 | 5 | 5 | 5 |
| Gap (mm) | 0 | 0 | 5 | 0 |
| Spatial resolution (mm) | 0.85 x 0.85 x 5 | 1.15 x 1.15 x 5 | 1.25 x 1.25 x 5 | 1.8 x 1.5 x 3 |
| FA | 90° | 90° | 90° | 90° |
| Scan duration (min) | 2:58 | 2:36 | 3:00 | 1:14 |

Abbreviations: SPIR - Spectral Presaturation with Inversion Recovery, TE - Time to echo, TR - Repetition time, EPI factor - Echo-planar imaging factor, SENSE factor - Sensitivity encoding factor, FOV - Field of view, FA - Fractional anisotropy, TSE - Turbo spin echo, GRaSE - Gradient and spin echo

**Table S1.B**. MRI acquisition parameters for sequences from dataset B.

|  | **TRUFI cor** | **HASTE cor** |
| --- | --- | --- |
| Sequence type | GRE | TSE |
| TE (ms) | 2 | 70-190\* |
| TR (ms) | 7 | max.1900 |
| Matrix size (pixels) | 256 x 256 | 256 x 256 |
| FOV (mm) | 350 x 350 | 350 x 350 |
| Slice thickness (mm) | 4 | 4 |

\*Depending on the scanner model

Abbreviations: TRUFI - True fast imaging with steady-state free precession, HASTE - Half-Fourier-acquired single-shot turbo spin-echo, cor - coronal, GRE - Gradient echo, TSE - Turbo spin echo, TE - Time to echo, TR - Repetition time, FOV - Field of view

**Table S1.C** MRI acquisition parameters for common sequences in dataset C.

| **Manufacturer** | **Scanner model** | **Field strength** | **Sequence** | **TE (ms)** | **TR (ms)** | **Spatial resolution** | **Matrix size (pixels)** |
| --- | --- | --- | --- | --- | --- | --- | --- |
| SIEMENS | Aera | 1.5T | HASTE ax | 91 | 1200 | 1.64 x 1.64 x 5 | 256 x 256 |
| HASTE cor | 91 | 1400 | 1.25 x 5 x 1.25 | 320 x 320 |
| HASTE fs ax | 96 | 1600 | 1.88 x 1.88 x 6 | 320 x 260 |
| T2 ax | 121 | 1520 | 1.48 x 1.48 x 5 | 256 x 256 |
| T2 cor | 88 | 3730 | 1.41 x 5 x 1.41 | 320 x 260 |
| BLADE fs ax | 93 | 4300 | 1.31 x 1.31 x 5 | 320 x 320 |
| Avanto | 1.5T | HASTE ax | 91 | 1350 | 0.74 x 0.74 x 5 | 512 x 360 |
|  | HASTE cor | 92 | 1000 | 1.41 x 1.41 x 5 | 320 x 320 |
|  | HASTE fs ax | 91 | 1350 | 0.74 x 0.74 x 5 | 512 x 360 |
|  | T2 ax | 92 | 2750 | 1.19 x 1.19 x 5 | 320 x 260 |
|  | T2 cor | 92 | 2200 | 1.19 x 5 x 1.19 | 320 x 288 |
|  | T2 fs ax | 91 | 3430 | 1.19 x 1.19 x 5 | 320 x 260 |
|  | TRUFI ax | 1.56 | 3.75 | 1.48 x 1.48 x 5 | 256 x 192 |
| SymphonyTim | 1.5T | HASTE ax | 96 | 1200 | 0.63 x 0.63 x 5 | 640 x 448 |
| HASTE cor | 92 | 1500 | 0.74 x 6 x 0.74 | 512 x 512 |
| TRUFI ax | 1.89 | 4.72 | 0.99 x 0.99 x 6.5 | 384 x 288  |
| TRUFI cor | 2.12 | 651 | 0.82 x 4 x 0.82 | 512 x 512 |
| Amira | 1.5T | HASTE ax | 91 | 1300 | 1.48 x 1.48 x 5 | 256 x 192 |
| HASTE cor | 91 | 1300 | 1.56 x 5 x 1.56 | 256 x 256 |
| STIR ax | 96 | 1100 | 1.56 x 1.56 x 5 | 156 x 208 |
| PHILIPS | Ingenia | 1.5T | T2 ax | 80 | 1200 | 0.71 x 0.71 x 5 | 720 x 720 |
|  |  |  | T2 SPIR ax | 70 | 1000 | 0.80 x 0.80 x 5.5 | 560 x 560 |
|  |  |  | T2 cor | 80 | 1200 | 1.0 x 6 x 1.0 | 448 x 448 |
|  | Intera | 1.5T | T2 ax | 80 | 1200 | 0.69 x 0.69 x 5.0 | 640 x 640 |
|  |  |  | T2 SPIR ax | 80 | 1200 | 0.69 x 0.69 x 5.0 | 640 x 640 |
|  |  |  | T2 cor | 90 | 1542 | 0.71 x 5.0 x 0.71 | 576 x 576 |
|  | Achieva | 1.5T\* | T2 ax | 80 | 395 | 0.98 x 0.98 x 5.0 | 384 x 384 |
|  |  |  | T2 SPIR ax | 80 | 539 | 0.67 x 0.67 x 5.0 | 640 x 640 |
|  |  |  | T2 cor | 80 | 461 | 0.98 x 5.0 x 0.98 | 384 x 384 |
|  |  |  | BTFE ax | 1454 | 2.91 | 1.46 x 1.46 x 5.5 | 256 x 256 |
|  | Achieva dStream | 1.5T | T2 ax | 100 | 3887 | 0.85 x 0.85 x 5.0 | 528 x 528 |
|  |  |  | T2 SPIR ax | 80 | 3444 | 0.80 x 0.80 x 5.0 | 560 x 560 |
|  | Ingenia S | 1.5T | T2 ax | 100 | 3481 | 0.88 x 0.88 x 5.0 | 512 x 512 |
|  |  |  | T2 SPIR ax | 80 | 1630 | 0.88 x 0.88 x 6 | 512 x 512 |
|  |  |  | T2 cor | 80 | 1200 | 0.98 x 6 x 0.98 | 512 x 512 |
|  | Panorama HFO | 1.5T | T2 ax | 100 | 3640 | 0.73 x 0.73 x 5.0 | 512 x 512 |
|  |  |  | T2 SPIR cor | 100 | 1959 | 0.91 x 6.0 x 0.91 | 384 x 384 |
| GE | Signa HDxt | 1.5T | FIESTA ax | 1672 | 3777 | 0.76 x 0.76 x 6.0 | 512 x 512 |
|  |  |  | FIESTA cor | 1632 | 3711 | 0.82 x 6.0 x 0.82 | 512 x 512 |
|  | Optima MR450w | 1.5T | PROPELLER fs ax | 121808 | 25000 | 0.82 x 0.82 x 3.0 | 512 x 512 |
|  |  | SSFSE cor | 140 | 1402 | 0.66 x 0.66 x 6.0 | 512 x 512 |

\*Twenty-six patients scanned with 1.5T, three patients scanned with 3T

Abbreviations: TE - Time to echo, TR - Repetition time, fs - fat-saturated, ax - axial, cor - coronal, HASTE - Half-Fourier-acquired single-shot turbo spin-echo, TRUFI - True fast imaging with steady-state free precession, STIR - Short Tau Inversion Recovery, SPIR - Spectral Presaturation with Inversion Recovery, BTFE - Balanced Turbo Field Echo, FIESTA - Fast Imaging Employing Steady-state Acquisition, PROPELLER/BLADE - Periodically Rotated Overlapping Parallel Lines with Enhanced Reconstruction, SSFSE - Single-shot fast spin-echo

##### **Appendix S2.** Flowchart of patient inclusion/exclusion.



##### **Appendix S3.** Details on model training and performed experiments.

For every training batch, at least a third of sample patches had to contain the foreground class. Patch sizes were adapted to GPU memory so that batch size is at least 2 (**Table S3.A**). Network depth and pooling operations along each axis were adapted accordingly to guarantee an effective receptive field size at least as large as the patch size. The training was terminated automatically when the learning rate was smaller than 10-6 and the exponential moving average of the validation loss did not improve by more than 5 x 10-3 over the last 60 epochs. As postprocessing, all but the largest connected component for each class were removed.

**Table S3.A**. Network training hyperparameters selected for the segmentation task.

| **Feature/Model** | **2D axial U-Net** | **3D axial U-Net** | **2D coronal U-Net** | **3D coronal U-Net** |
| --- | --- | --- | --- | --- |
| Batch size | 16 | 2 | 32 | 2 |
| Patch size (pixels) | 448 x 448 | 256 x 224 x 32 | 320 x 320 | 224 x 224 x 40 |
| No of poolings per axis | [6, 6] | [3, 6, 5] | [6, 6] | [3, 5, 5] |

The best hyperparameters were selected for each of the four U-Net configurations in the training dataset.

##### **Table S4.** Detailed cohort characteristics for training/test split of datasets A and B.

|  | **Dataset A** | **Dataset B** |
| --- | --- | --- |
|  | **training axial** | **test axial** | **test coronal\*** | **training coronal** | **test coronal** |
| Patients in total | 101 | 25 | 101 | 226 | 68 |
| Number of MR series | 403 | 149 | 159 | 589 | 201 |
| Age | 42.9 ± 12.1 | 44.0 ± 12.9 | 42.8 ± 12.3 | n.a.**十** |
| Male | 45 | 10 | 44 |
| Female | 56 | 15 | 57 |
| Height (m) | 1.76 ± 0.11 | 1.75 ± 0.12 | 1.76 ± 0.11 |
| Weight (kg) | 84.4 ± 20.2 | 81.3 ± 18.9 | 82.9 ± 19.5 |
| eGFR, CKD-EPI, ml/min per 1.73 m2 | 72.6 ± 28.6 | 73.5 ± 27.6 | 72.8 ± 27.8 |
| Mayo class: |  |  |  |
| 1A - 1B | 29 (29%) | 8 (32%) | 28 (28%) |
| 1C - 1E | 72 (71%) | 17 (68%)  | 73 (72%) |
| Kidney volume (ml): |  |  |  |
| * Total
 | 1648 ± 1047 | 1473 ± 772  | 1609 ± 990 | 2324 ± 1637 | 2009 ± 1310 |
| * Left
 | 826 ± 544 | 771 ± 392 | 816 ± 509 | 1302 ± 855 | 1027 ± 650 |
| * Right
 | 821 ± 627 | 703 ± 397 | 792 ± 603 | 1221 ± 816 | 982 ± 699 |
| Liver volume (ml) | 2679 ± 1508 | 2138 ± 1421 | 2411 ± 1448 | 2704 ± 1685 | 2409 ± 1448 |

\* includes all patients from dataset A, who have coronal images

**十** n.a., not available (details on this cohort can be found in [1](https://www.zotero.org/google-docs/?vPjfqu)**)**

##### **Table S5.** Results of automated kidney and liver segmentation in test datasets A and B, for all the sequences.

| **Cohort** | **Organ** | **Dice** | **Jaccard** | **Sensitivity** | **Precision** | **Hausdorff** | **MSD\*** |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Dataset A - test | Total kidney | 0.94 [0.89-0.97] | 0.88 [0.80-0.94] | 0.93 [0.89-0.96 | 0.95 [0.89-0.97] | 10.5 [5.9-18.6] | 0.55 [0.35-1.33] |
|  | Left kidney | 0.94 [0.89-0.97] | 0.89 [0.81-0.94] | 0.94 [0.89-0.97] | 0.95 [0.89-0.97] | 9.0 [5.0-14.1] | 0.59 [0.35-1.16] |
|  | Right kidney | 0.94 [0.89-0.98] | 0.88 [0.80-0.94] | 0.93 [0.88-0.96] | 0.95 [0.88-0.98] | 8.1 [5.0-15.13] | 0.54 [0.33-1.34] |
| Dataset B - test | Total kidney | 0.97 [0.94-0.98] | 0.93 [0.89-0.95] | 0.97 [0.94-0.98] | 0.96 [0.94-0.98] | 10.0 [6.3-18.5] | 0.36 [0.27-0.56] |
|  | Left kidney | 0.97 [0.94-0.98] | 0.93 [0.88-0.96] | 0.97 [0.94-0.98] | 0.97 [0.93-0.98] | 8.2 [5.0-15.0] | 0.35 [0.26-0.54] |
|  | Right kidney | 0.97 [0.94-0.98] | 0.93 [0.89-0.95] | 0.97 [0.94-0.98] | 0.96 [0.93-0.98] | 8.9 [5.0-16.0] | 0.46 [0.26-0.57] |
|  | Liver | 0.96 [0.94-0.97] | 0.93 [0.88-0.95] | 0.96 [0.92-0.98] | 0.96 [0.94-0.98] | 14.1 [8.1-23.3] | 0.46 [0.36-0.73] |
| Dataset C | Total kidney | 0.92 [0.75-0.97] | 0.84 [0.60-0.95] | 0.92 [0.73-0.99] | 0.92 [0.76-0.98] | 12.0 [7.0-31.1] | 0.76 [0.28-2.95] |
|  | Left kidney | 0.92 [0.73-0.97] | 0.85 [0.57-0.95] | 0.92 [0.75-0.98] | 0.93 [0.74-0.98] | 10.1 [5.5-24.8] | 0.72 [0.28-2.71] |
|  | Right kidney | 0.91 [0.69-0.97] | 0.84 [0.53-0.94] | 0.92 [0.69-0.97] | 0.92 [0.72-0.98] | 10.4 [5.9-28.3] | 0.76 [0.29-3.06] |

 \* Mean surface distance

Values are reported as median and 10. - 90. percentile range.

##### **Table S6.** Patient stratification according to Mayo height-adjusted TKV risk classes for automated versus reference methods of kidney volumetry in the test set A.

| **Automated volumetry** | **Manual volumetry** |
| --- | --- |
| **A** | **B** | **C** | **D** | **E** |
| **A** | 0 | 0 | 0 | 0 | 0 |
| **B** | 0 | 7 | 0 | 0 | 0 |
| **C** | 0 | 0 | 7 | 0 | 0 |
| **D** | 0 | 0 | 0 | 8 | 0 |
| **E** | 0 | 0 | 0 | 1\* | 2 |

\*Only one patient was reclassified into another Mayo risk group.

##### **Table S7.** Comparison of differences in kidney and liver volume between baseline and follow-up for manual tracing and automated method.

|  | **Dataset A - test** |  | **Dataset B - test** |
| --- | --- | --- | --- |
| **Organ** | **Manual** | **Automated** | **p-value**\* | **Organ** | **Manual** | **Automated** | **p-value**\* |
| Total kidney (n = 25) |  |  |  |  Total kidney (n = 68) |  |  |
| * change in ml
 | 15 ± 85  | 15 ± 84 | 1.0 |  | 324 ± 385 | 324 ± 369 | 0.96 |
| * change in %
 | 0.8 ± 4.9 | 1.1 ± 4.9 | 0.87 |  | 15.7 ± 17.9 | 15.5 ± 15.9 | 0.96 |
| Left kidney |  |  |  |  Left kidney |  |  |
| * change in ml
 | 4 ± 53 | 7 ± 52  | 0.32 |  | 162 ± 188 | 162 ± 179 | 0.59 |
| * change in %
 | 0.2 ± 5.8 | 0.8 ± 5.7 | 0.26 |  | 15.2 ± 19.5 | 15.2 ± 17.1 | 0.75 |
| Right kidney |  |  |  |  Right kidney |  |  |
| * change in ml
 | 11 ± 41 | 6 ± 42 | 0.44 |  | 163 ± 209 | 161 ± 201 | 0.45 |
| * change in %
 | 1.5 ± 5.0 | 1.2 ± 4.7 | 0.50 |  | 16.4 ± 17.3 | 15.9 ± 15.5 | 0.62 |
| Liver (n = 29) |  |  |  |  Liver (n = 68) |  |  |
| * change in ml
 | 205 ± 276 | 145 ± 217 | 0.09 |  | -60 ± 404 | -44 ± 352 | 0.63 |
| * change in %
 | 7.8 ± 11.0 | 5.1 ± 8.7 | 0.17 |  | -2.5 ± 15.1 | -2.1 ± 14.6 | 0.73 |

\*calculated with Wilcoxon signed-rank test

For dataset A, the follow-up examination was on average 53 weeks after baseline, for dataset B – 132 weeks after baseline (average annual TKV growth of 0.8% in dataset A and 6.2% in dataset B, more details in the manuscript). Notably, 50% of the patients in the DIPAK1 study (Dataset B) were treated with Lanreotide explaining the decrease in liver volume.

**Table S8.** TKV growth by the Mayo imaging class for patients with longitudinal data. *(n – number of subjects, TKV – total kidney volume)*

|  | **Dataset A** | **Dataset B** |
| --- | --- | --- |
| **MIC** | **n** | **Manual TKV growth (ml)** | **Automated TKV growth (ml)** | **n** | **Manual TKV growth (ml)** | **Automated TKV growth (ml)** |
| 1A | 0 | - | - | 5 | 27 [12 - 44] | 37 [19 - 41] |
| 1B | 16 | 27 [12 - 44] | 37 [19 - 41] | 48 | 38 [-1 - 96] | 31 [-17 - 60] |
| 1C | 23 | 38 [-1 - 96] | 31 [-17 - 60] | 93 | 11 [-58 - 158] | 28 [-24 - 89] |
| 1D | 22 | 11 [-58 - 158] | 28 [-24 - 89] | 67 | 96 [64 - 126] | 68 [34 - 97] |
| 1E | 4 | 96 [64 - 126] | 68 [34 - 97] | 47 | 27 [12 - 44] | 37 [19 - 41] |

**References**

[1. Meijer E, Visser FW, van Aerts RMM, Blijdorp CJ, Casteleijn NF, D’Agnolo HMA, et al.: Effect of Lanreotide on Kidney Function in Patients With Autosomal Dominant Polycystic Kidney Disease: The DIPAK 1 Randomized Clinical Trial. *JAMA* 320: 2010–2019, 2018](https://www.zotero.org/google-docs/?Thq14k)