Article:
Taylor, J., Thomas, R., Metherall, P. et al. (15 more authors) (2023) An artificial intelligence generated automated algorithm to measure total kidney volume in ADPKD. Kidney International Reports. ISSN 2468-0249

https://doi.org/10.1016/j.ekir.2023.10.029

An Artificial Intelligence generated Automated Algorithm to measure Total Kidney Volume in ADPKD

Jonathan Taylor, Richard Thomas, Peter Metherall, Marieke van Gastel, Emilie Cornec-Le Gall, Anna Caroli, Monica Furlano, Nathalie Demoulin, Olivier Devuyst, Jean Winterbottom, Roser Torra, Norberto Perico, Yannick Le Meur, Sebastian Schoenherr, Lukas Forer, Ron T. Gansevoort, Roslyn J. Simms, Albert CM. Ong

PII: S2468-0249(23)01571-1

DOI: https://doi.org/10.1016/j.ekir.2023.10.029

Reference: EKIR 2526

To appear in: *Kidney International Reports*

Received Date: 19 June 2023

Revised Date: 26 October 2023

Accepted Date: 30 October 2023

Please cite this article as: Taylor J, Thomas R, Metherall P, van Gastel M, Cornec-Le Gall E, Caroli A, Furlano M, Demoulin N, Devuyst O, Winterbottom J, Torra R, Perico N, Le Meur Y, Schoenherr S, Forer L, Gansevoort RT, Simms RJ, Ong AC, An Artificial Intelligence generated Automated Algorithm to measure Total Kidney Volume in ADPKD, *Kidney International Reports* (2023), doi: https://doi.org/10.1016/j.ekir.2023.10.029.

An Artificial Intelligence generated Automated Algorithm to measure Total Kidney Volume in ADPKD

Jonathan Taylor[1], Richard Thomas[1], Peter Metherall[1], Marieke van Gastel[2], Emilie Cornec-Le Gall[3], Anna Caroli[4], Monica Furlano[5], Nathalie Demoulin[6], Olivier Devuyst[6], Jean Winterbottom[7,8], Roser Torra[5], Norberto Perico[4], Yannick Le Meur[9], Sebastian Schoenherr[10], Lukas Forer[10], Ron T Gansevoort[2], Roslyn J Simms[7,8,*], Albert CM Ong[7,8,*]

[1]3D Lab, Medical Imaging Medical Physics, Sheffield Teaching Hospitals NHS Foundation Trust, Sheffield, UK
[2]Department of Nephrology, University Medical Centre Groningen, Groningen, The Netherlands
[3]University Brest, Inserm, UMR 1078, GGB, CHU Brest, F-29200 Brest, France
[4]Istituto di Ricerche Farmacologiche Mario Negri IRCCS, Bergamo, Italy
[5]Inherited Kidney Disorders, Nephrology Department, Fundació Puigvert, IIB Sant Pau, Universitat Autònoma de Barcelona, Barcelona, Spain
[6]Cliniques Universitaires Saint-Luc, UCLouvain Medical School, Brussels, Belgium
[7]Academic Nephrology Unit, Division of Clinical Medicine, School of Medicine and Population Health, Faculty of Health, University of Sheffield, Sheffield, UK
[8]Sheffield Kidney Institute, Sheffield Teaching Hospitals NHS Foundation Trust, Sheffield, UK
[9]University Brest, Inserm, UMR 1227, LBAI, CHU Brest, F-29200 Brest, France
[10]Institute of Genetic Epidemiology, Department of Genetics and Pharmacology, Medical University of Innsbruck, Austria

* Joint corresponding authors

Correspondence:
Roslyn Simms or Albert CM Ong
Academic Nephrology Unit, Division of Clinical Medicine, School of Medicine and Population Health, University of Sheffield, Beech Hill Road, Sheffield S10 2RX, United Kingdom
Telephone: 0114 215 9542
Fax: 0114 271 1863
Email: r.simms@sheffield.ac.uk; a.ong@sheffield.ac.uk

Keywords: ADPKD, total kidney volume, machine learning, artificial intelligence, magnetic resonance imaging

Abstract

Introduction

Accurate tools to inform individual prognosis in patients with autosomal dominant polycystic kidney disease (ADPKD) are lacking. Here, we report an artificial intelligence (AI) generated method for routinely measuring total kidney volume (TKV).

Methods

An ensemble U-net algorithm was created using the nnUNet approach. The training and internal cross-validation cohort consisted of all 1.5T MRI data acquired using 5 different MRI scanners (454 kidneys, 227 scans) in the CYSTic consortium which was first manually segmented by a single human operator. As an independent validation cohort, we utilised 48 sequential clinical MRI scans with reference results of manual segmentation acquired by 6 individual analysts at a single centre. The tool was then implemented for clinical use and its performance analysed.

Results

The training / internal validation cohort was younger (mean age 44.0 vs 51.5 years) and the female-male ratio higher (1.2 v 0.94) compared to the clinical validation cohort. The majority of CYSTic patients had *PKD1* mutations (79%) and typical disease (Mayo Imaging Class 1, 86%). The median DICE score on the clinical validation dataset between the algorithm and human analysts was 0.96 for left and right kidneys with a median TKV error of -1.8%. The time taken to manually segment kidneys in the CYSTic dataset was 56 (±28) min whereas manual corrections of the algorithm output took 8.5 (±9.2) min per scan.

Conclusions

Our AI-based algorithm demonstrates performance comparable to manual segmentation. Its rapidity and precision in real-world clinical cases demonstrate its suitability for clinical application.

Introduction

Autosomal dominant polycystic kidney disease (ADPKD) is the most common inherited kidney disease, characterised by the progressive development and growth of kidney cysts which results in kidney enlargement and kidney failure in 50% of affected patients by 60 years [1]. The clinical course of ADPKD is however highly variable between individuals even if renal outcomes can be stratified based on the causative gene and variant type [2]. The longitudinal CRISP (Consortium for Radiologic Imaging Studies of PKD) studies identified that prior to the decline in kidney function, total kidney volume (TKV) is increased and predictive of an eGFR < 60ml/min/1.73m^2 [3]. TKV has since been approved as a prognostic imaging biomarker by the European Medicines Agency (EMA) in 2015 and Food and Drug Administration (FDA) in 2016. As there is now an effective treatment to slow disease progression, Tolvaptan [4, 5] , the timely identification of patients at risk of rapid progression to kidney failure is vital to optimise and personalise patient care [6]. Nonetheless, a major challenge to the use of TKV in clinical practice has been the difficulty of accurately segmenting the kidneys and the significant human operator time (45-90min per patient) required of skilled, experienced staff to measure TKV.

In a previous study, we reported the development of a rapid, semi-automated, open access TKV tool to facilitate the wider adoption of TKV measurements into clinical practice [7]. Here we report a new rapid, high performance, artificial intelligence (AI) segmentation tool developed using MRI scans acquired from 4 European centres (the CYSTic consortium) [8] (**Table 1**). Validation of the algorithm in a second non-overlapping ADPKD clinical cohort analysed by multiple operators confirms its suitability for routine clinical practice. Following clinical implementation, additional analysis demonstrates the significant time savings that could be achieved through adoption of the AI approach.

Methods

Patient recruitment and centre participation

The inclusion and exclusion criteria for entry into the International Consortium to build a longitudinal observational cohort of patients with ADPKD (CYSTic consortium) have been recently reported [8]. Over 450 patients were initially recruited from six expert centres across Europe (Belgium, France, Italy, Netherlands, Spain, and UK) with baseline clinical data recorded including HR-QoL (KDQoL-SFv1.3 questionnaire), abdominal MRI for TKV measurements and DNA for genotyping. Each study centre consented to transfer their data to a cloud-based web platform incorporating a study-specific electronic database (Askimed) (https://www.askimed.com). The study was approved by a Regional Ethics Committee (18/EE/0247) and by the study sponsor, Sheffield Teaching Hospitals NHS Foundation Trust. Ethics approval was also obtained by each participating centre within their own country.

Technical development

The general approach taken is summarised in **Figure 1**. The training and internal validation set consisted of all 1.5T MRI scans (n=227, 454 kidneys) from the CYSTic consortium [8] excluding cases where the kidney was not completely included in the field of view, as identified through visual analysis, or where scan quality was affected by artefacts to such an extent that manual segmentations could not be confidently drawn (3.4%, n=8). Each kidney was manually segmented according to a standard operating procedure by a single operator (RT) with over 6 years of performing TKV measurements, using MIM Maestro software (v6.9.3) and a Huion pen display tablet.

Clinical MRI cases used as an independent validation dataset (n=48) were collected from the imaging archives at Sheffield Teaching Hospitals, excluding Sheffield CYSTic patients. All scans were manually segmented, again using MIM software, but with a standard mouse. Clinical cases are routinely processed by multiple different trained operators working in the 3DLab and there were six different individuals that had performed the TKV measurements. These operators had a range of experience levels (processing between 9 and 53 clinical cases each). Patient and acquisition details for the different datasets are summarised in **Table 1**.

The nnUnet algorithm [9] was selected for training an automated segmentation tool. This approach is well-established, showing high performance in multiple, varied segmentation applications [10]. In addition, nnUnet has been successfully applied in other studies where a mixed training cohort from separate scanners has been used [11].

All images and kidney contours were first converted from dicom to nifti format using the python package medio (v0.4.0). Algorithm performance was improved when using one kidney label category rather than two (i.e. left and right kidneys labelled with the same value). The label map images were therefore binary.

Image data was bias-corrected using the SimpleITK N4 bias field correction algorithm [12]. The internal validation images were used for 5-fold cross-validation, with each fold stratified to control for biases between centres (80% of the data from each centre was allocated to within-fold algorithm training and 20% for testing). Data was shuffled between folds such that each individual case was used for testing only once across the 5 folds. Cross-validation was repeated using the Sheffield CYSTIc cases only. Further details of the methodology can be found in the **Supplementary material**.

Finally, the ensemble of algorithms trained during cross validation were applied to the clinical validation dataset.

Clinical implementation

The AI tool was implemented clinically as a remote DICOM service in the 3D laboratory at Sheffield in August 2022, setup to trigger automatically whenever a new MRI image was acquired. The tool generates a segmentation mask for each image, which is then viewed and edited as required by a trained operator in MIM software. The time taken to manually load, edit and finalise the kidney segmentation mask is automatically registered in a database along with TKV values for both the unedited and edited segmentations.

All available records (n=33) were extracted from the database in May 2023 for analysis. Recorded times for AI segmentation editing were compared to processing time figures for the original manual processing technique obtained for the Sheffield CYSTic patients (n=64).

Comparison with other software

Algorithm performance was compared against another recently reported deep learning method, ADPKD-net [13]. This software package was downloaded from Docker Hub (https://hub.docker.com/repository/docker/piotrekwoznicki/adpkd-net) and the cases from the clinical validation dataset were processed through the software, one at a time, using the default parameters. TKV results were collated and compared to those achieved through manual segmentation and with our new algorithm.

Results

The average time taken to manually segment each case in the internal validation dataset (both kidneys) was 54 minutes (SD of 31 minutes). Intra-operator variability for manual segmentation was low, with a mean difference in TKV measurements between repeat manual segmentations of 2.1% ± 2.7% (left kidney) and 1.6% ±1.7% (right kidney). The internal validation data contained a range of different appearances, with 22 cases having a right kidney-liver border that was visually classed as being difficult to differentiate.

The internal cross-validation showed high DICE scores with low percentage volume differences between the new AI-derived TKV data and manual results (**Table S1**). Separating the results from different centres (**Figure 2**), there was a small bias in improved performance towards the Sheffield and Groningen datasets, possibly due to the use of similar MRI scanners and acquisition sequences. However, the Mayo classification categories which is based on

height-adjusted TKV, had no impact on TKV accuracy (**Figure 3**), indicating good performance across a range of kidney volumes and shapes.

Application of the full automated algorithm to the clinical validation dataset showed similar close agreement between the results for automated segmentation and manually segmented TKV despite being analysed by 6 different operators (**Table S2, Figure 4**). Some examples of automated segmentation from the clinical validation dataset are shown in **Figure S1**. The performance of the algorithm on the clinical validation dataset was largely unchanged when trained with Sheffield CYSTic data only (**Table S3**).

Analysis of outliers (5.7%) with discordance between the automated and manually measured TKV (DICE<0.92) showed that cysts in close proximity to the liver border (either originating in the liver or kidney) were the most common visual feature associated with reduced performance (**Table 2**, **Figure 5**).

Next, we tested the performance of the tool for routine TKV analysis after implementation in a hospital laboratory setting by analysts experienced in manual kidney segmentation (**Table 3**). Compared to historical data from the Sheffield CYSTic patients, the time taken for manual correction of the AI segmentations was 8.5 (±9.2) min v 56 (±28) min for fully manual processing. Mean volume differences between AI-TKV and after manual editing were -2.0 (±4.0) % and -1.3 (±3.5) % for the RK and LK respectively.

Finally, processing the clinical validation dataset through the recently reported ADPKD-net algorithm (**Figure 6**) showed a general overestimate of TKV, with greater overestimates seen for larger kidneys. Visual analysis of ADPKD-net outputs suggests that the overestimate is largely due to the inclusion of the renal pelvis in segmentations (which is routinely excluded at Sheffield) and by other published methods [14].

Discussion

We have created a new automated segmentation algorithm derived from a large European dataset of MRI images of ADPKD kidneys to accurately and rapidly measure TKV. It performed accurately on a wide range of kidney volumes (0.1L to 4.4L) and anatomical shapes (Mayo Class 1 and 2) [15]. Measured TKV errors for the algorithm were of similar magnitude to intra-operator variability results and to inter-operator results reported previously [7] implying that the algorithm has reached human levels of performance.

Internal cross-validation results were consistently high across different centres despite the lack of any specific domain adaptation steps employed. Comparison of the performance on the clinical validation cohort between the algorithm trained on the full CYSTic cohort, and that trained with Sheffield CYSTic patients only (**Tables S2-3, figure 4**) showed that the inclusion of patient data from different scanners and different populations was not detrimental to performance. This suggests that the algorithm is not biased towards a particular sub-population within the CYSTic training cohort.

Mayo class 2 ADPKD cases are often not included in automated segmentation research. In this study, 32 (14%) class 2 patients were part of the internal validation / training cohort but cross-validation results demonstrated that they were not associated with inferior performance for TKV measurement. This provides reassurance that the algorithm would be robust enough to analyse TKV in atypical cases without pre-selection.

We utilised a well-established technique to generate a segmentation algorithm based on the U-net [9]. Other published algorithms based on similar U-net technology have also demonstrated high performance in the segmentation of healthy, chronic kidney disease and ADPKD kidney images [16-18] increasing confidence that the algorithm presented here is likely to be effective. Indeed, the ADPKD-net algorithm that was selected as a comparator in this study also used the same baseline architecture [13]. Nonetheless, the results from the ADPKD-net algorithm demonstrated a general overestimate of TKV on the clinical validation dataset due to the inclusion of the renal pelvis. This part of the kidney is not traditionally included in TKV segmentations [14] and is not included in local routine measurements. Therefore, our developed algorithm is likely to be more consistent with general accepted practice.

It should be noted that other organs such as the liver can be affected by ADPKD, but these areas are excluded by our trained algorithm. Further work is being undertaken to specifically target polycystic livers. In addition, the algorithm is designed to work with data acquired in the same way as that of the CYSTic cohort (i.e. coronal Steady State Free Procession type sequences) [7, 8]. This type of acquisition is widely adopted in other ADPKD research [14] but is not universally used in clinic and therefore our algorithm will not be applicable across all centres.

Our new automated algorithm demonstrates high precision compared to manual TKV segmentation and performs reliably in most patients with ADPKD, with a range of kidney volumes, shapes and coexisting polycystic liver disease. The mean processing time for manual segmentation by an experienced operator was approximately 1 hour per case. Use of

the algorithm in clinical practice does not completely remove the need for clinical staff from the TKV measurement process; a trained clinical observer (such as a radiologist or radiographer) is always required to review AI generated results. However, the algorithm required minimal manual edits and changes to the generated contours, reducing the average processing time per case to 9 minutes. Finally, its accuracy when validated in real-world clinical datasets demonstrates that such AI tools can provide a reliable means of measuring TKV in routine practice by reducing the barriers of analyst time and experience.

Disclosure statement

The authors report no disclosures relevant to the study

Acknowledgements

The study was funded by grants from the Sheffield Hospitals Charitable Trustees, Sheffield Kidney Research Foundation and the PKD Charity UK. We thank Jim Wild and Wendy Tindale for advice and support. JT, PM, ACMO and RJS are members of INSIGNEO (Institute for In-silico Medicine) at the University of Sheffield. We are grateful for the generous participation of many patients and their referring physicians within the CYSTic consortium.

Supplementary material
Detailed methods (PDF)

Table S1 Internal cross-validation results summary (5 folds) of the algorithm (PDF)

Table S2 Performance of algorithm on an independent clinical dataset, trained on the full CYSTic algorithm (PDF)

Table S3 Performance of algorithm on an independent clinical dataset, trained on Sheffield-CYSTic only algorithm (PDF)

Figure S1 Examples of algorithm results from the clinical external validation dataset (PDF)

Supplementary information is available at KI Report's website.

References

1. Ong AC, Devuyst O, Knebelmann B*, et al.* Autosomal dominant polycystic kidney disease: the changing face of clinical management. *Lancet* 2015; **385:** 1993-2002.

2. Cornec-Le Gall E, Audrezet MP, Chen JM*, et al.* Type of PKD1 mutation influences renal outcome in ADPKD. *J Am Soc Nephrol* 2013; **24:** 1006-1013.

3. Chapman AB, Bost JE, Torres VE*, et al.* Kidney volume and functional outcomes in autosomal dominant polycystic kidney disease. *Clinical journal of the American Society of Nephrology : CJASN* 2012; **7:** 479-486.

4. Torres VE, Chapman AB, Devuyst O*, et al.* Tolvaptan in patients with autosomal dominant polycystic kidney disease. *The New England journal of medicine* 2012; **367:** 2407-2418.

5. Torres VE, Gansevoort RT, Czerwiec FS. Tolvaptan in Later-Stage Polycystic Kidney Disease. *N Engl J Med* 2018; **378:** 489-490.

6. Chebib FT, Torres VE. Assessing Risk of Rapid Progression in Autosomal Dominant Polycystic Kidney Disease and Special Considerations for Disease-Modifying Therapy. *Am J Kidney Dis* 2021; **78:** 282-292.

7. Simms RJ, Doshi T, Metherall P*, et al.* A rapid high-performance semi-automated tool to measure total kidney volume from MRI in autosomal dominant polycystic kidney disease. *Eur Radiol* 2019; **29:** 4188-4197.

8. Winterbottom J, Simms RJ, Caroli A*, et al.* Flank pain has a significant adverse impact on quality of life in ADPKD: the CYSTic-QoL study. *Clin Kidney J* 2022; **15:** 2063-2071.

9. Isensee F, Jaeger PF, Kohl SAA*, et al.* nnU-Net: a self-configuring method for deep learning-based biomedical image segmentation. *Nat Methods* 2021; **18:** 203-211.

10. Antonelli M, Reinke A, Bakas S*, et al.* The Medical Segmentation Decathlon. *ArXiv210605735 Cs Eess* 2021.

11. Full PM, Isensee F, Jager PF*, et al.* Studying Robustness of Semantic Segmentation under Domain Shift in cardiac MRI. *ArXiv201107592 Cs Eess* 2020.

12. Beare R, Lowekamp B, Yaniv Z. Image Segmentation, Registration and Characterization in R with SimpleITK. *J Stat Softw* 2018; **86**.

13. Woznicki P, Siedek F, van Gastel MDA*, et al.* Automated Kidney and Liver Segmentation in MR Images in Patients with Autosomal Dominant Polycystic Kidney Disease: A Multicenter Study. *Kidney360* 2022; **3:** 2048-2058.

14. Kline TL, Edwards ME, Korfiatis P*, et al.* Semiautomated Segmentation of Polycystic Kidneys in T2-Weighted MR Images. *AJR Am J Roentgenol* 2016; **207:** 605-613.

15. Irazabal MV, Rangel LJ, Bergstralh EJ*, et al.* Imaging classification of autosomal dominant polycystic kidney disease: a simple model for selecting patients for clinical trials. *J Am Soc Nephrol* 2015; **26:** 160-172.

16. Kline TL, Korfiatis P, Edwards ME*, et al.* Performance of an Artificial Multi-observer Deep Neural Network for Fully Automated Segmentation of Polycystic Kidneys. *Journal of digital imaging* 2017; **30:** 442-448.

17. van Gastel MDA, Edwards ME, Torres VE*, et al.* Automatic Measurement of Kidney and Liver Volumes from MR Images of Patients Affected by Autosomal Dominant Polycystic Kidney Disease. *J Am Soc Nephrol* 2019; **30:** 1514-1522.

18. Daniel AJ, Buchanan CE, Allcock T*, et al.* Automated renal segmentation in healthy and chronic kidney disease subjects using a convolutional neural network. *Magn Reson Med* 2021; **86:** 1125-1136.

Figure legends

Figure 1 Schematic of the development of the new algorithm through testing, internal and clinical validation phases**.**

Figure 2 5-fold internal cross validation results summary, separated according to study centre (BER = Bergamo, BRE = Brest, GRO = Groningen, SHE = Sheffield). Left and right kidneys were labelled separately.

Figure 3 Comparison of volume results obtained from manual contouring on training data vs AI tool in 5-fold internal cross-validation. Results for right or left kidneys, Mayo class 1 and 2 are displayed separately.

Figure 4 Comparison of volume results obtained from manual contouring on clinical validation dataset vs AI tool (algorithm trained using the full internal dataset). Left and right kidneys were labelled separately.

Figure 5 Example of a large kidney cyst (top) or liver boundary cyst (bottom) leading to under-segmentation by the algorithm (left original image, right image with algorithm segmentation overlaid).

Figure 6 Comparison of volume results obtained from manual contouring on clinical validation dataset vs the ADPKD-net algorithm. Left and right kidneys were labelled separately.

Tables

Table 1 Patient characteristics and MRI acquisition details for training and internal validation (CYSTic) and clinical validation datasets. Note that genotype and Mayo classification information were not available for all patients in the clinical validation set.

	Training and internal validation dataset (CYSTic)				**Clinical validation**
Study centre	**Groningen**	**Sheffield**	**Bergamo**	**Brest**	
Mean age (SD)	43.3 (12.8)	43.7 (14.7)	43.8 (11.2)	46.8 (13.4)	51.5 (5.6)
Sex	M=34, F=44	M=30, F=34	M=19, F=22	M=20, F=24	M=17, F=16
Genotype PKD1 (%)	PKD1=44 (78.2%)	PKD1=50 (78.1%)	PKD1=27 (65.9%)	PKD1=43 (97.7%)	
Mayo classification	Class 1 = 72, Class 2A = 6	Class 1 = 51, Class 2A = 10, Class 2B = 3	Class 1 = 33, Class 2A = 8	Class 1 = 39, Class 2A = 5	
Scanner	Siemens Avanto, Avantofit, Aera	Siemens Avanto	GE Optima MR450W	GE Optima MR450W	Siemens Avantofit
Selected sequence	TRUFI	TRUFI	3D FIESTA	3D FIESTA	TRUFI
Total scans	78	64	41	44	48[#]

[#]15 patients had >1 scan

Table 2 Visual analysis of cases where autosegmentation performance was reduced (DICE < 0.92)

	Internal Cross-validation	**Clinical validation**
Image or segmentation appearance associated with reduced algorithm performance	**Number (%)**	**Number (%)**
Autosegmentation under or over segments liver-kidney border cysts	5 (2.2)	2 (4.2)
Partial autosegmentation of a single large kidney cyst	3 (1.3)	0
Autosegmentation includes kidney tissue that is uncertain from visual analysis	3 (1.3)	0
Autosegmentation includes renal pelvis	0	1 (2.1)
Human segmentation error	1 (0.4)	0
Autosegmentation is overly smooth between slices, does not follow sharply changing kidney geometry	1 (0.4)	0

Table 3 Clinical implementation of the AI tool for routine TKV analysis

	Method	
	AI-assisted (n=33 clinical cases)	**Manual (n=64 Sheffield cases from CYSTIC cohort)**
Mean time to process	8.5 mins (SD 9.2 mins)	56 mins (SD 28 mins)

	Mean volume difference: AI TKV measurement minus human-edited AI TKV measurement
R (ml)	-5.3 (SD 8.3)
L (ml)	-2.2 (SD 15.6)
R (%)	-2.0 (SD 4.0)
L (%)	-1.3 (SD 3.5)

Figure 1

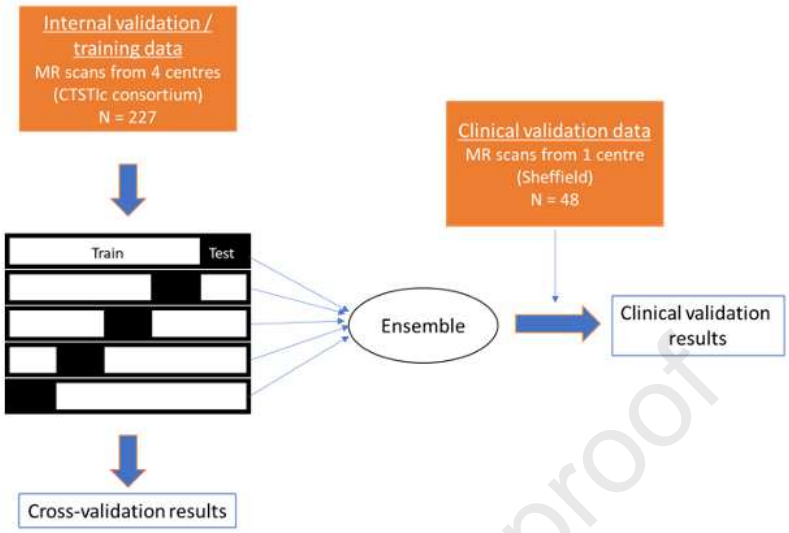

Figure 2

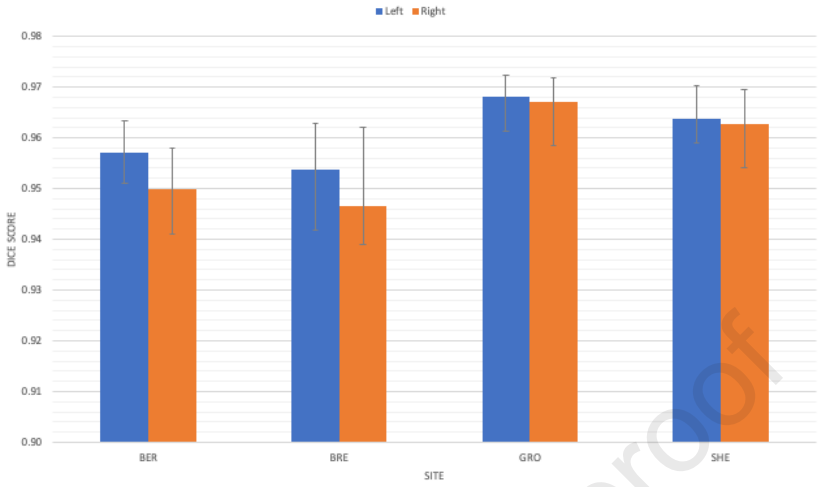

Figure 3

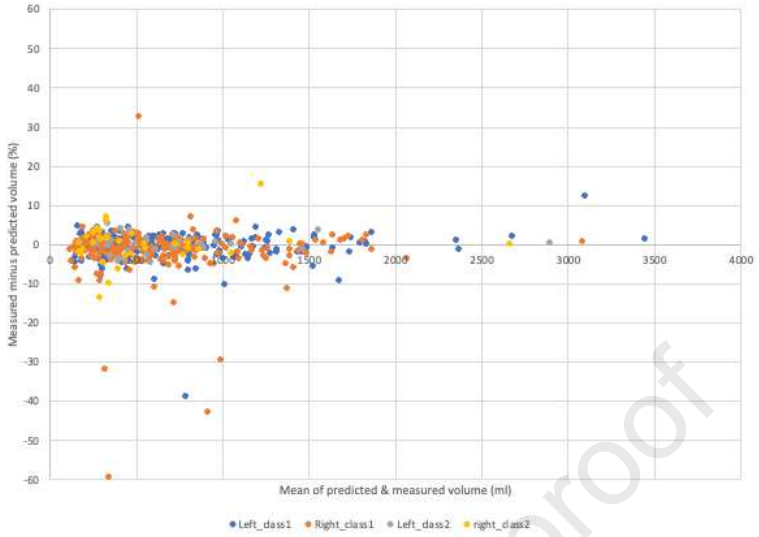

Figure 4

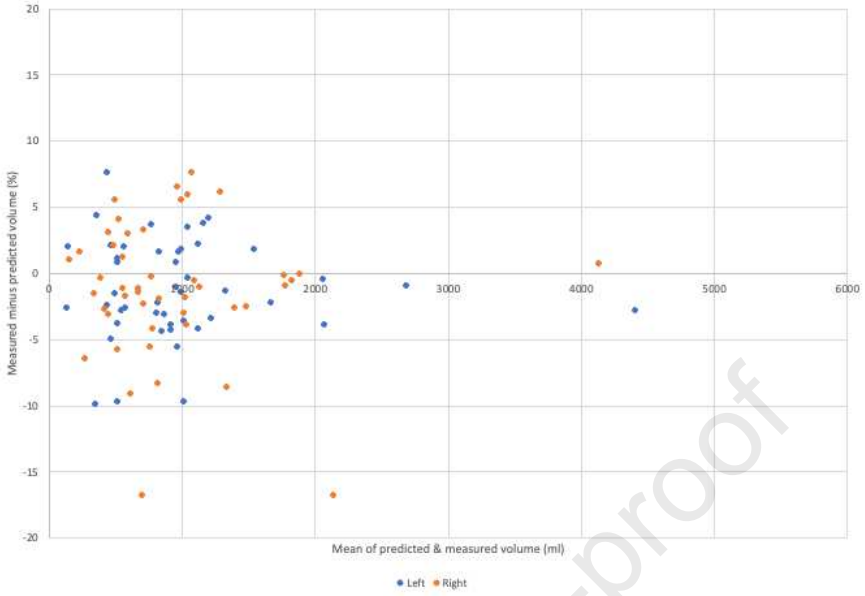

Figure 5

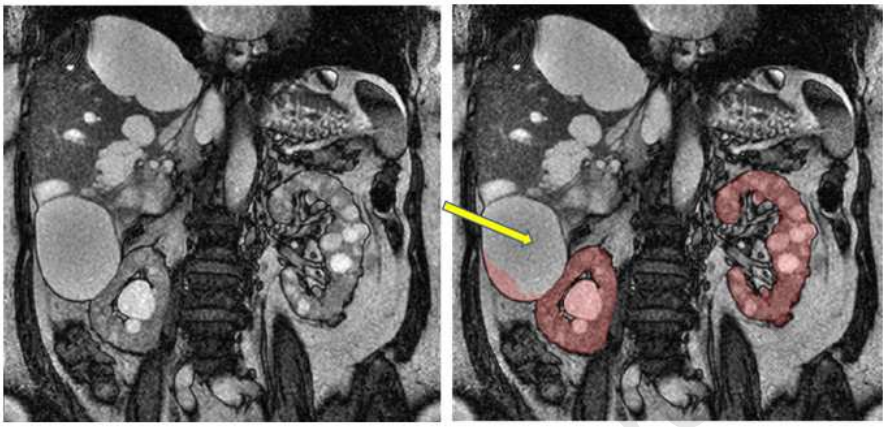

Figure 6