Artificial Intelligence (AI)-Driven Estimation of Total Kidney Volume (TKV) in ADPKD

Elham Yousef Kalafi PhD, Anand Kadumberi MSc, Andrew Michalski Ph.D., Rohit Sood MD., PhD.

Background

ADPKD characteristics:

- Fluid-filled cysts
- Bilateral renal enlargement
- Eventual renal failure

Regulatory guidance:

- US FDA supports using imaging-based Total Kidney Volume (TKV) assessment as a qualified prognostic biomarker in clinical trials

AI-driven methods:

- Enable automated kidney segmentation
- Provide high repeatability
- Ensure reproducibility

The study objective:

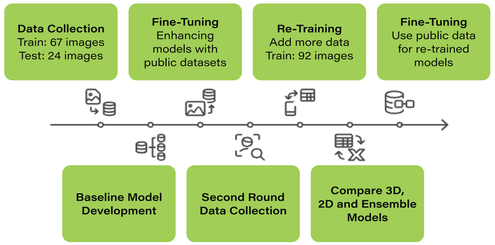
- Develop an AI-driven methodology to estimate TKV on MRI

Method

A retrospective cohort of MRI scans from ADPKD patients enrolled in the CRISP study was analyzed (T1w and T2w images). One reader (15 yrs experience) annotated the T2w images.

Data Preparation: 1) registering T1w images rigidly to corresponding T2w images, 2) splitting the dataset into training, validation and test sets.

Model Development:



Public Datasets: MRI (AMOS, CHAOS, Totalsegmentator, CKD) and CT (AbdomenCT-1K).

Evaluation Metrics: The performance of all trained models was evaluated on the held-out test set using Dice Similarity Coefficient (DSC),Volume Error (VE), Hausdorff Distance (HD) and Intersection over union (IoU).

Results

Best Initial Training (67 images, 24 test cases)

Total Kidney Performance (Initial Test Set):

- DSC: 0.843
- Hausdorff: 39.1 mm
- IoU: 0.745
- Volume Error: 155,135 mm³

Expanded Training (92 images, hold-out test set)

3D Model (T2_3D):

- DSC: 0.897, IoU: 0.822, Volume Error: 133,531 mm³

2D Model (best performance) (T2_2D):

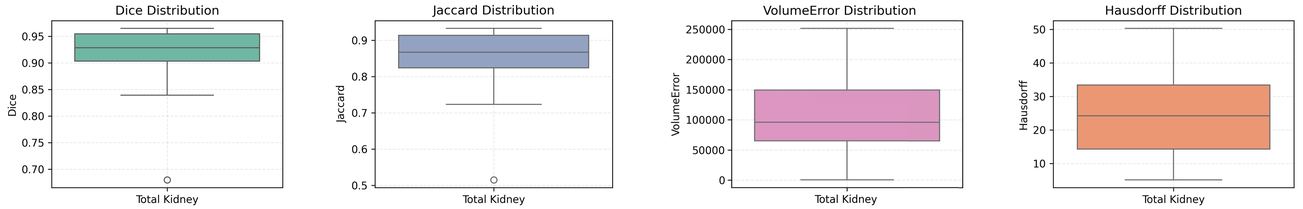
- DSC: 0.917, IoU: 0.852, Volume Error: 106,522 mm³

Ensemble Model (T2_Ensemble):

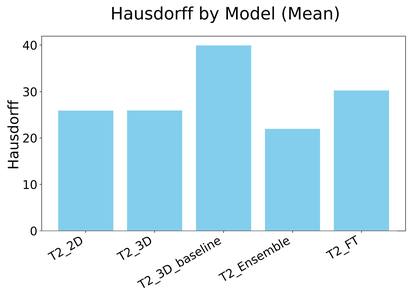
- DSC: 0.920, IoU: 0.856, Volume Error: 113,495 mm³
- Hausdorff: 21.1 mm (improved boundary accuracy)

Fine-Tuning (T2_FT):

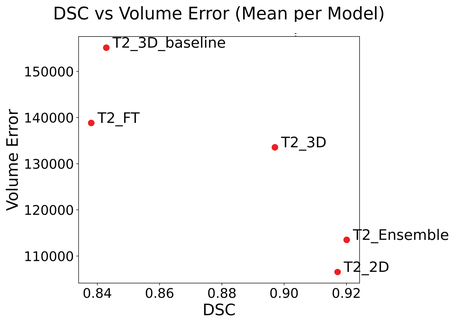
- DSC: 0.84, IoU: 0.75, Volume Error: 131,158 mm³
- Performance did not surpass the 2D/Ensemble models, but remained consistent across cases.



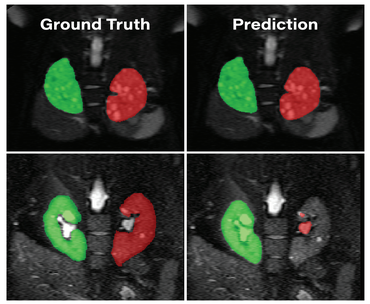
Boxplots illustrating the performance of T2_2D model for TKV across metrices.



Comparison of the Hausdorff Distance metric across all evaluated segmentation models to assess boundary accuracy.



Comparison of models performance showing Dice Similarity Coefficient (DSC) versus Volume Error across different models.



Overlay examples of the best prediction (DSC 0.96) and less accurate prediction (DSC 0.68).

Conclusions

- AI-driven models achieved **high accuracy in estimating Total Kidney Volume (TKV)** in ADPKD patients, with DSC approaching **0.92** and strong volumetric agreement.
- Expanding the dataset from **67 to 92 training images** substantially improved performance, particularly with **ensemble modeling**, reducing both segmentation error and Hausdorff distance.
- Although **fine-tuning did not improve accuracy**, the results suggest **better generalizability** across held-out test sets, an important property for clinical translation.
- These findings support the feasibility of deploying AI-based tools for **reliable, automated TKV estimation** in ADPKD clinical and research settings.
- The models do better with larger kidney volumes, evident from the Pearson and Spearman correlation coefficients.