





Expert-level segmentation using deep learning for volumetry of polycystic kidney and liver

Tae Young Shin^{1,2,*} , Hyunsuk Kim^{3,*} , Joong-Hyup Lee¹ , Jong-Suk Choi² , Hyun-Seok Min⁴ ,
Hyungjoo Cho⁴ , Kyungwook Kim⁵ , Geon Kang² , Jungkyu Kim², Sieun Yoon⁵ , Hyungyu Park²,
Yeong Uk Hwang⁶ , Hyo Jin Kim⁷ , Miyeun Han⁷ , Eunjin Bae⁸ , Jong Woo Yoon³ , Koon Ho Rha⁹ ,
Yong Seong Lee¹⁰ 

¹Synergy A.I. Co.Ltd., Chuncheon, ²Department of Urology, Hallym University Chuncheon Sacred Heart Hospital, Hallym University College of Medicine, Chuncheon, ³Department of Internal Medicine, Division of Nephrology, Hallym University Chuncheon Sacred Heart Hospital, Hallym University College of Medicine, Chuncheon, ⁴Tomocube, Inc., Daejeon, Korea, ⁵Schulich School of Medicine & Dentistry, The University of Western Ontario, London, ON, Canada, ⁶Department of Radiology, Inje University Ilsan Paik Hospital, Goyang, ⁷Department of Internal Medicine, Pusan National University Hospital, Busan, ⁸Department of Internal Medicine, Gyeongsang National University Changwon Hospital, Gyeongsang National University College of Medicine, Changwon, ⁹Department of Urology, Urological Science Institute, Yonsei University College of Medicine, Seoul, ¹⁰Department of Urology, Hallym University Sacred Heart Hospital, Hallym University College of Medicine, Anyang, Korea

Purpose: Volumetry is used in polycystic kidney and liver diseases (PKLDs), including autosomal dominant polycystic kidney disease (ADPKD), to assess disease progression and drug efficiency. However, since no rapid and accurate method for volumetry has been developed, volumetry has not yet been established in clinical practice, hindering the development of therapies for PKLD. This study presents an artificial intelligence (AI)-based volumetry method for PKLD.

Materials and Methods: The performance of AI was first evaluated in comparison with ground-truth (GT). We trained a V-net-based convolutional neural network on 175 ADPKD computed tomography (CT) segmentations, which served as the GT and were agreed upon by 3 experts using images from 214 patients analyzed with volumetry. The dice similarity coefficient (DSC), interobserver correlation coefficient (ICC), and Bland–Altman plots of 39 GT and AI segmentations in the validation set were compared. Next, the performance of AI on the segmentation of 50 random CT images was compared with that of 11 PKLD specialists based on the resulting DSC and ICC.

Results: The DSC and ICC of the AI were 0.961 and 0.999729, respectively. The error rate was within 3% for approximately 95% of the CT scans (error < 1%, 46.2%; 1% ≤ error < 3%, 48.7%). Compared with the specialists, AI showed moderate performance. Furthermore, an outlier in our results confirmed that even PKLD specialists can make mistakes in volumetry.

Conclusions: PKLD volumetry using AI was fast and accurate. AI performed comparably to human specialists, suggesting its use may be practical in clinical settings.

Keywords: Artificial intelligence; Polycystic kidney diseases; Tomography

This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/4.0>) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

Received: 13 March, 2020 • **Revised:** 3 June, 2020 • **Accepted:** 23 June, 2020 • **Published online:** 27 October, 2020

Corresponding Author: Yong Seong Lee  <https://orcid.org/0000-0003-0611-4554>

Department of Urology, Hallym University Sacred Heart Hospital, Hallym University College of Medicine, 22 Gwanpyeong-ro 170beon-gil, Dongan-gu, Anyang 14068, Korea

TEL: +82-33-240-5000, FAX: +82-33-255-6244, E-mail: novavia@hallym.or.kr

*These authors contributed equally to this study and should be considered co-first authors.

INTRODUCTION

Autosomal dominant polycystic kidney disease (ADPKD) is the most common life-threatening hereditary kidney disease, and 50% of ADPKD patients aged 60 years or older progress to end-stage renal disease [1,2]. Furthermore, up to 94% of ADPKD patients aged over 35 years present with multiple liver cysts. As patients age, kidney and liver volume steadily increase, causing not only a decline of renal function, but also associated symptoms and complications that severely impair quality of life (such as pain, infection, and malnutrition [3-5]).

Volumetry of the kidneys and liver is one of the most important biomarkers for quantifying the severity of disease and the effectiveness of treatments [6-8]. Recently, various clinical trials have been conducted to evaluate disease-modifying agents. However, the major “rate-limiting factor” is the inefficient process of measuring volume. Two-dimensional stereology has been used for volumetry [9]; this method is based on multiple tomographic images, obtained through

modalities such as computed tomography (CT) and magnetic resonance imaging (MRI), and requires the manual annotation of the kidney and liver regions in each image slice.

Since volumetry is a laborious operation, data acquisition inhibits reproducibility and the degree to which volumetry can be used productively. Although we need adequate amounts of reproducible data for academic advances in the understanding of this important condition, the difficulties in acquiring volumetric data lead to other problems. First, if volumetry is the most important biomarker of disease severity, we must determine whether the measurements are mathematically rigorous. In this field, paradoxically, since volumetric data are so precious, even if the measurements might not be mathematically perfect, we have had no choice but to match measurements statistically by minimizing inter- or intraobserver variability to overcome human error. Although efforts have been made to supplement accuracy and reproducibility, it is still difficult to verify each volume measurement. Furthermore, it is unclear whether the intraobserver variability and interobserver variability are

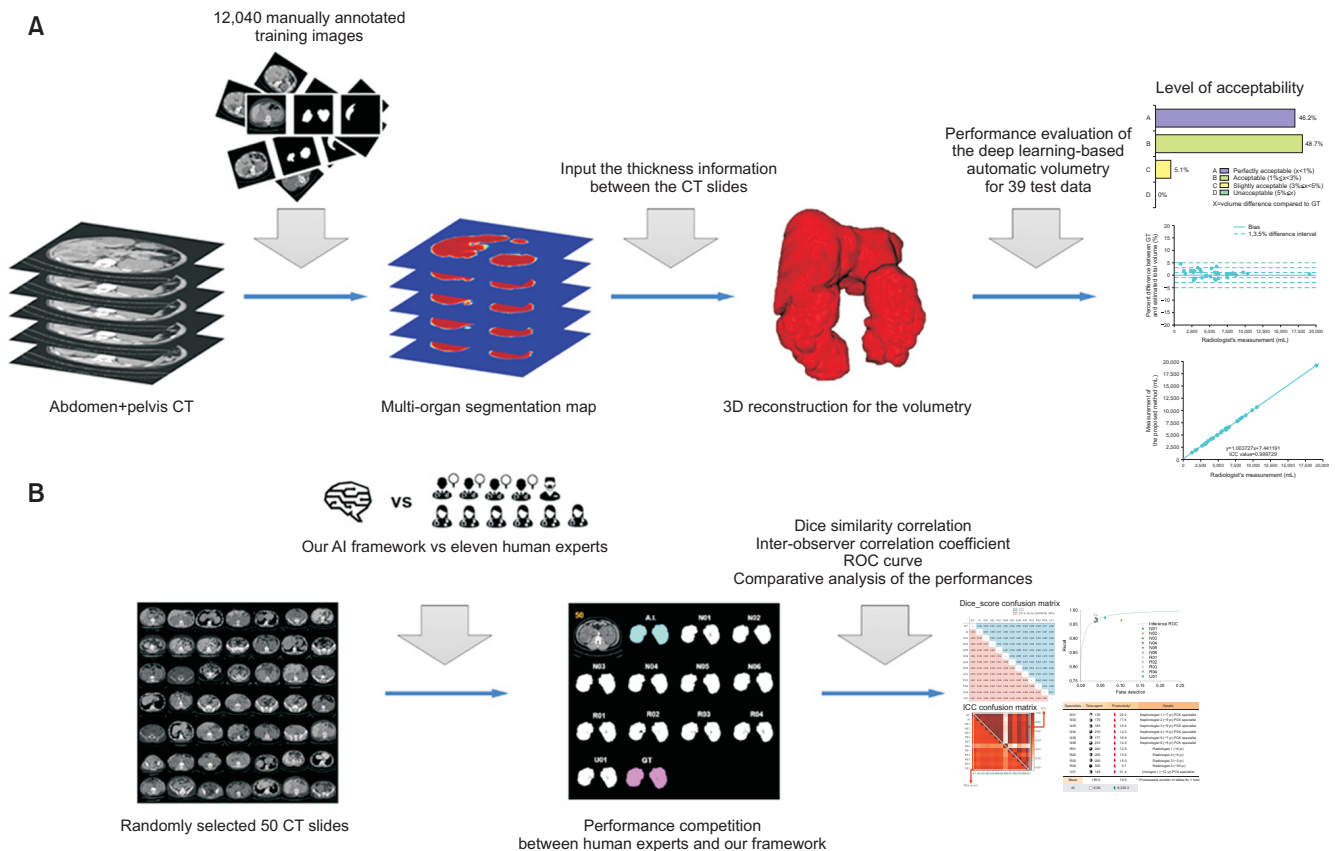


Fig. 1. Sequential experiments to evaluate the performance of our framework for automatic segmentation and volumetry. (A) The first phase illustrates the process of multiorgan segmentation. The volumetric performance of our framework on 39 CT scans (3,302 image slices) in the validation set is analyzed in Fig. 3. (B) In the second phase of the experiment, the performance of our framework on 50 randomly selected PKLD image slices was compared to that of 11 PKLD experts. The results of the comparative analysis are illustrated in Fig. 3 and Fig. 4. CT, computed tomography; GT, ground-truth; ICC, interobserver correlation coefficient; AI, artificial intelligence; PKLD, polycystic kidney and liver disease.

Table 1. The participating hospitals, the CT data characteristics, and the model names of the CT scanners

Data characteristics at CT	Chuncheon	Pyeongchon	Kangdong	Kangnam	Dongtan	Total
Training (n)						
Noncontrast	24	45	35	30	19	153
Contrast	5	6	3	2	6	22
Test (n)						
Noncontrast	3	9	6	7	4	29
Contrast	1	3	0	2	4	10
Model names of the CT scanners in the participating hospitals	SOMATOM Definition Flash (Siemens, Malvern, PA, USA)	SOMATOM Sensation 64 (Siemens)	SOMATOM Smile (Siemens)	Brilliance CT 64-Channel (Philips)	Brilliance Big Bore (Philips)	
		SOMATOM Definition Edge (Siemens)	MX8000 IDT (Philips, Andover, MA, USA)	MX 8000 IDT (Philips)		
		SOMATOM Definition Flash (Siemens)		SOMATOM Sensation 64 (Siemens)		

CT, computed tomography.

evaluated for each volume measurement. Second, because acquiring moderate to severe liver volumetry is especially difficult, the level of academic knowledge regarding volume progression and treatment of polycystic liver disease (PLD) remains limited.

The clinical applications of deep learning algorithms are rapidly increasing. In particular, the level of research in the field of medical image segmentation has dramatically improved. However, owing to the morphological complexity of the kidneys and liver, automatic segmentation of the polycystic states of these organs remains one of the most challenging aspects of medical image processing. Herein, we present results showing the powerful performance of our framework for multiorgan segmentation of polycystic kidney and liver disease (PKLD).

MATERIALS AND METHODS

1. Two phases of experiments

In the initial experiment (Fig. 1), volumetry was produced, which was trained with 175 cases. Table 1 presents information on the training data. The five participating hospitals utilized seven different types of CT scanners, which collected data in various formats. The first experiment was conducted to determine the accuracy of the liver and kidney volumes derived from 39 sets of test data (Fig. 1A). To illustrate the agreement between the ground-truth (GT), which was determined by human specialists, and the calculated volume based on deep-learning-based automatic segmentation, Bland–Altman analysis and interobserver correlation coefficients (ICCs) were used (Fig. 2A, B). The level of data acceptability in terms of error was evaluated in accordance with the clinical features of PKLD (Fig. 2C).

The second experiment was designed to compare the segmentation performance of our framework with that of 11 human PKLD experts (Fig. 1B). This experiment was conducted to evaluate whether our framework could overcome the disadvantages of humans and show stable performance in randomized conditions. We randomly selected 50 image slices from the test data. The sample size was calculated using G*Power 3.1.5 (Franz Faul, Universität Kiel, Kiel, Germany). A simulation showed that the standard deviation of difference could be assumed to be 70. Using a power (1-b error probability) of 0.95, an assumed effect size of 0.5, and a significance level of 0.05, the size of the sample was calculated to be 45 cases or more. Receiver operating characteristic (ROC) curves (Fig. 3A) were used to compare performance, and heatmap visualizations were developed (Fig. 4B) based on the ICC and the dice similarity coefficient (DSC) to com-

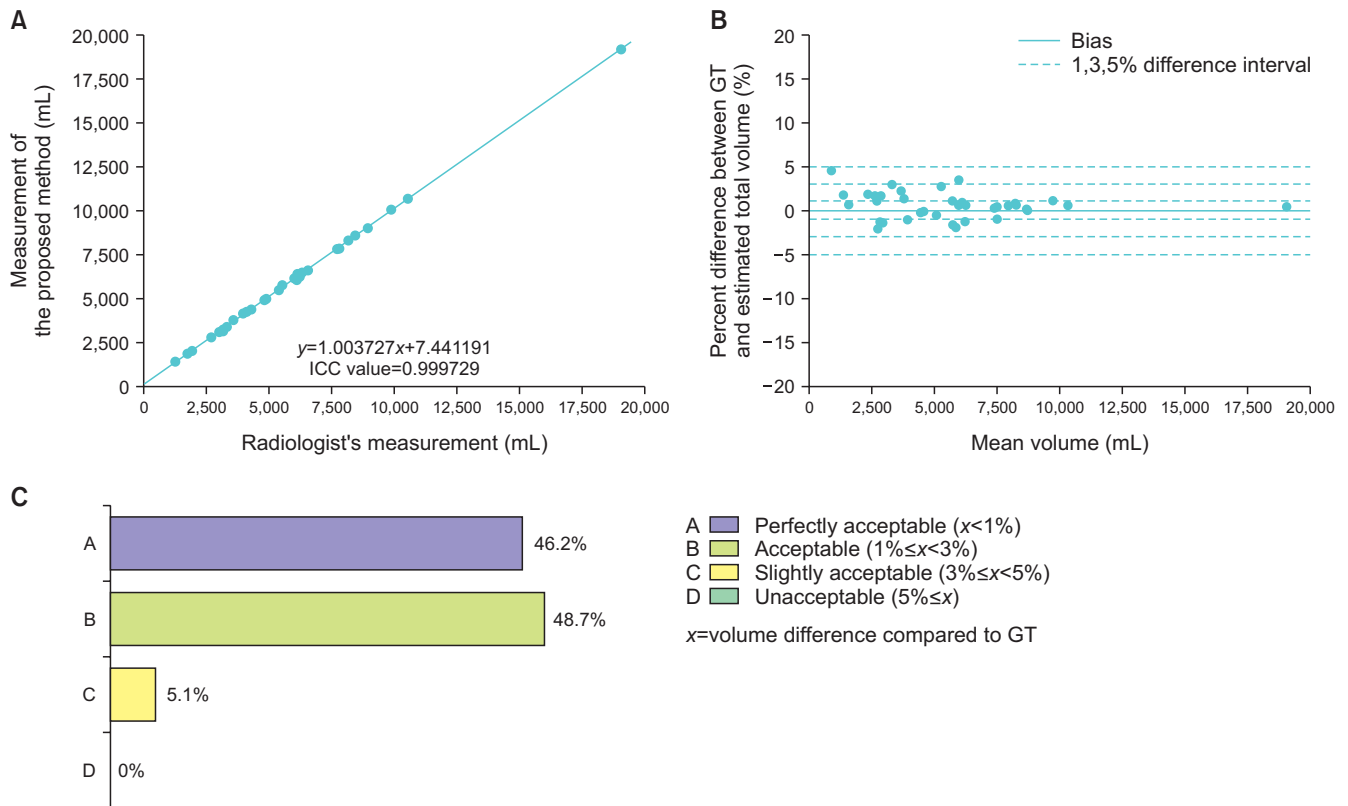


Fig. 2. Performance evaluation for the volume calculations based on automatic segmentation using our framework. (A) Interobserver correlation coefficients, (B) Bland–Altman analysis, and (C) levels of acceptability classified as level A (perfectly acceptable), B (acceptable), C (slightly acceptable), and D (unacceptable). ICC, interobserver correlation coefficient; GT, ground-truth.

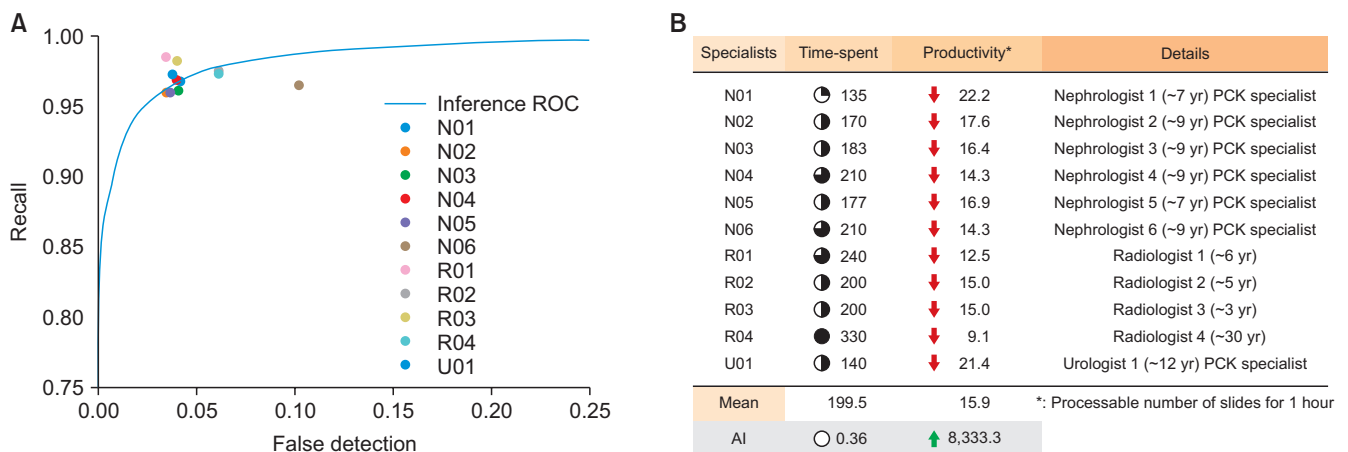


Fig. 3. The comparative analysis of performance between our framework and specialists. (A) Performance on the second-phase experiment of the independent test set of 50 randomly selected CT image slices. The receiver operating characteristic (ROC) diagram shows the segmentation accuracy of our framework versus all 11 experts. The blue ROC curve was created by sweeping a threshold over the inference of our framework for the ground-truth. (B) Table presenting the results of the performance comparison, the time spent for 50 image slices, the processable number of image slices in 1 hour, and the clinical experience of each specialist. N01-06, nephrologists; R01-04, radiologists; U01, urologist; PCK, polycystic kidney disease.

pare the correlations between the specialists.

2. Datasets

Axial CT images were acquired from 214 patients with

a diagnosis of ADPKD. In the entire group, 206 cases (96.3%) had liver cysts, and 39 cases (18.2%) had moderate to severe PLD. We used a training set of roughly 80% of the dataset (175 cases) and a validation set of roughly 20%. Each case

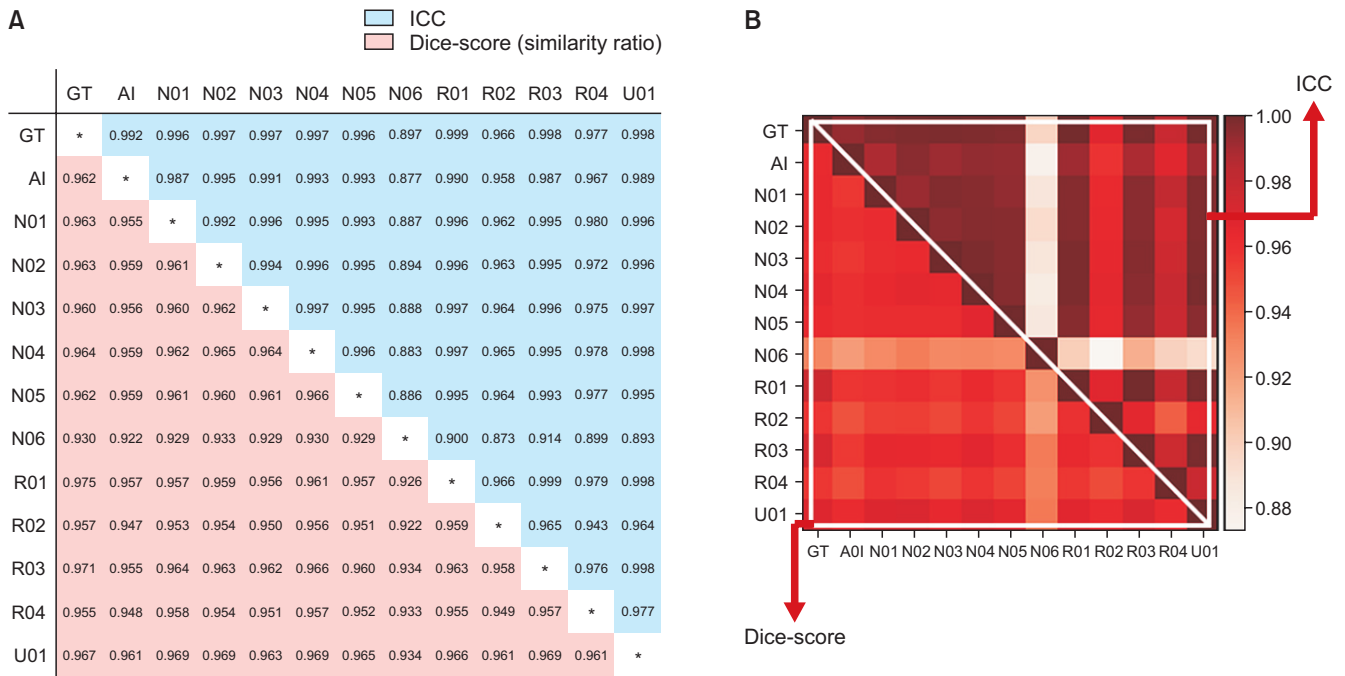


Fig. 4. Comparison of segmentation performance with that of human experts. (A) The details of the dice similarity coefficient (lower left, red) and the interobserver correlation coefficient (right upper, blue) are listed for each specialist. (B) Heatmap table for clearer comparisons and visibility. GT, ground-truth; AI, artificial intelligence; N01-06, nephrologists; R01-04, radiologists; U01, urologist; ICC, interobserver correlation coefficient.

had a different number of slices, ranging from 29 to 256. The training and test data were sourced from seven CT scanners used in the past two decades at five hospitals to ensure a representative diversity of image formats and characteristics. In the model training process, we sought to improve the quality of the data by working with three PKLD specialists to conduct at least five rounds for data refinement per case to reduce the ambiguity of the labels. The total number of images was 12,040, and the number of images with the target organs was 3,302. Analyzing the error rate for test data, 95% confidence intervals and mean bias were 0.47 and 0.70. The ratios of noncontrast and contrast data in the training and test datasets were 87.4:12.6 and 74.4:25.6, respectively.

3. Convolutional neural network architecture

In a deep learning network, complex mathematical operations are performed to achieve a specific goal based on the input data. In other words, the network learns millions of weights (parameters) to reach its intended aim. A suitable deep learning method for image learning is the convolutional neural network (CNN), which is composed of several layers that learn hierarchical features without relying on handcrafted features. Taking images as input, to implement tasks such as segmentation, it is necessary to acquire a pixel-wise classification from the CNN. As shown in Fig. 5, we applied a network based on a V-net (SYN-PCK-AN.

v1.00; Synergy A.I. Co. Ltd, Chuncheon, Korea) for ADPKD segmentation to use information from successive frames [10]. However, unlike the original V-net, our model used 16 image slices of three-dimensional CT image chunks as the input and a pixel-wise sigmoid operation, instead of softmax, as the output for the binary classification. To achieve stable and fast learning, batch normalization was used to properly initialize every convolution layer.

4. The training process of our model

To alleviate overfitting and to achieve good generalizations, we applied three different augmentation methods to the training set: the random crop method, the random scaled method, and cutout augmentation [11,12] with 80% probability. We applied a random-position zero mask to images with sizes ranging from L/5 to L/4, where L is the length of the image. Each input was 16 sequential slices selected from the training cases. We used a sampling size of 8 (batch-size) for each iteration, binary cross entropy loss as our cost function, and the Adam optimizer for weight updating. All experiments were performed using the Pytorch framework.

The output of prediction consisted of foreground pixels (kidneys and liver) and background pixels, where pixels with a probability higher than the threshold were regarded as foreground pixels. The selected threshold was based on the analysis of the ROC space. We also computed the precision,

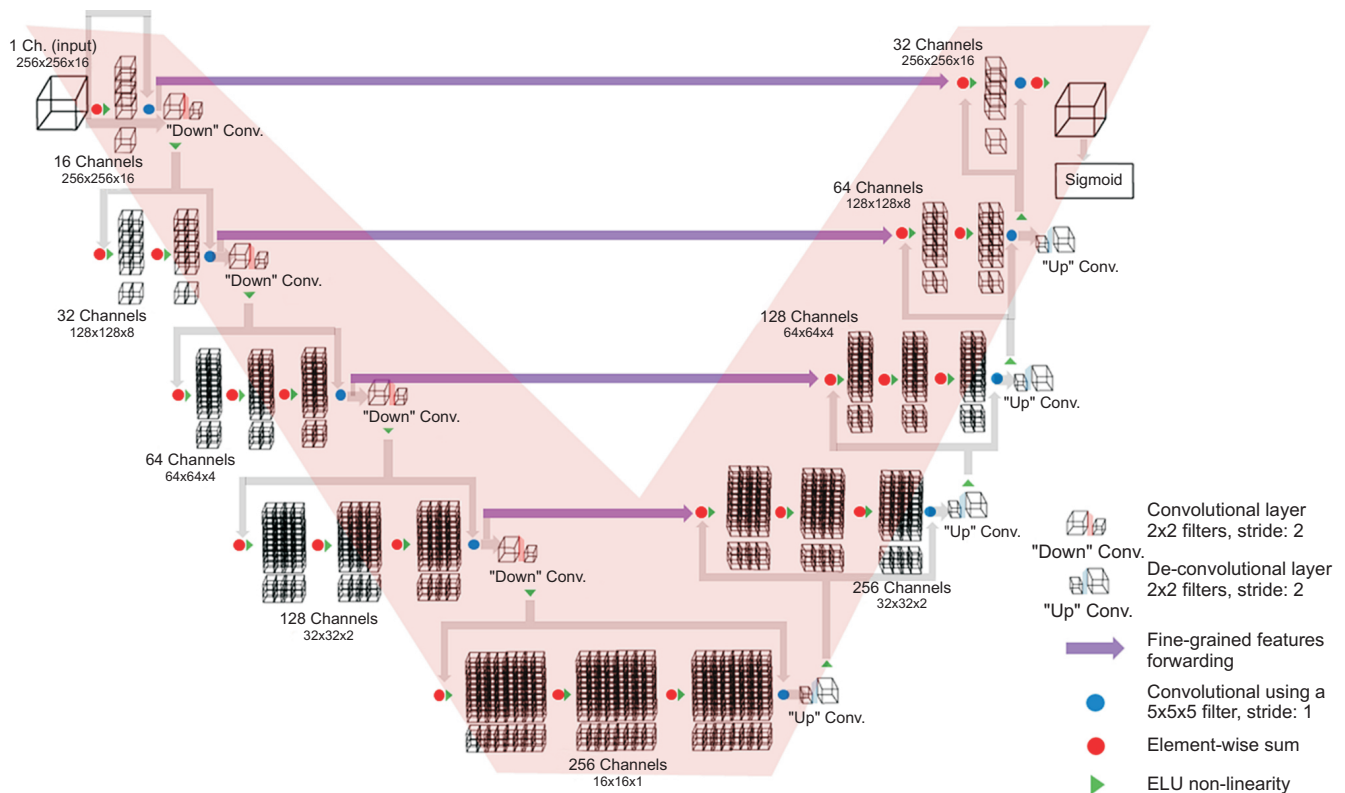


Fig. 5. Schematic diagram of the V-net architecture of our framework. Our custom implementation processes three-dimensional data by performing volumetric convolutions. Conv., convolutional.

recall, and dice scores.

5. Total kidney volume computation

All CT datasets were manually segmented by two clinical experts and one trained staff member to obtain GT annotations of the kidneys and liver. The total volume of the kidney and liver was computed as the product of the number of foreground pixels multiplied by the pixel spacing in the x and y directions and the corresponding slice thickness.

6. Statistical analyses

To evaluate the performance of our automated segmentation method, the DSC was used as a statistical validation metric to assess the accuracy of spatial overlap between the predicted and true manual segmentation labels. The ICC was used to evaluate the reliability and reproducibility of the automated method with respect to the standard manual method. Furthermore, Bland–Altman analysis was used to assess the agreement between the values estimated from the automated segmentation method and those obtained from the corresponding manual segmentation. Both absolute and relative differences were computed using Bland–Altman plots.

7. Ethical approval

This study was approved by the Institutional Review Board of Chuncheon Sacred Heart Hospital (IRB approval number: 2017-11-106). All research was performed in accordance with relevant guidelines/regulations, and informed consent was obtained from all participants.

RESULTS

1. Phase I: volume calculation accuracy

Fig. 2A depicts the scatter plot between the volume based on GT and the volume based on the automatic segmentation of our framework. The accuracy of the volume calculations in our framework was demonstrated by a DSC of 0.961 and an ICC of 0.999729 (Fig. 2A, $p < 0.00001$). In the Bland–Altman analysis (Fig. 2B), the differences between the GT-based volume and the inference-based volume were converted to percentage values. In terms of the level of acceptability, an unacceptable case was defined as 0% (Fig. 2C). In our first experiment, 94.9% of the volume measurements made using our framework showed ‘perfectly acceptable’ or ‘acceptable’ accuracy, confirming the powerful promise of this framework to ensure data integrity through a simple review and modification process by experts. Table 2 shows the fair accu-

Table 2. Comparative analysis of PKLD volume calculations between ground-truth and AI-driven volumetry

Volumetry	Manual (ground-truth)		AI-driven		p-value by t-test	
	Training data	Test data	Training data	Test data	Training data	Test data
TKV (mL)	3,596.4±456.7	4,562.7±547.2	3,524.5±463.2	4,555.5±539.6	<0.001	<0.001
htTKV (mL/m)	2,247.7±367.3	2,887.8±385.5	2,245.6±352.2	2,897.1±374.2	<0.001	<0.001

Values are presented as mean±standard deviation.

PKLD, polycystic kidney and liver disease; AI, artificial intelligence; TKV, total kidney volume; htTKV, height-adjusted total kidney volume.

racy of volume calculation by artificial intelligence (AI).

2. Phase II: comparison of segmentation performance with human experts

All 11 human experts had more than 3 years of professional experience and were designated to perform manual annotations on 50 randomly selected image slices. We placed no time limitations, and requested them to review the results that they produced. Notable findings are shown in Fig. 4A. Except for N06, the DSCs and ICCs of 10 specialists and our framework had a fairly close match to GT, ranging from 0.957 to 0.975 and from 0.966 to 0.999, respectively. The performance of our framework roughly corresponded to the average performance of the 10 specialists, with a DSC and ICC of 0.962 and 0.992, respectively. However, the performance of N06 was a significant outlier, as shown in Fig. 4B. In this case, some of the annotated photographs were mixed and badly processed, resulting in accidental errors. Although N06 is a specialist in ADPKD with 9 years of experience, the results of N06 showed a DSC of 0.930 and an ICC of 0.897 compared to GT, whereas the mean DSC and ICC of the others were 0.992 and 0.964, respectively. This is an excellent example illustrating the human errors that can occur when creating and analyzing clinical cohorts, and it is one reason thorough censorship is required. Fig. 3A illustrates the segmentation accuracy of our framework through an ROC curve, which indicates the noninferiority of its performance compared to the manual annotation of other PKLD specialists. In addition to its expert-level performance, the productivity of our framework was 8,333.3 image slices/hour, which is over 500 times the experts' average of 15.9 image slices/hour (Fig. 3B). Assuming an average of 60 image slices per patient, our framework showed a overwhelming speed advantage of 216 slices/patient, compared to 12,960 slices/patient by the human doctors.

DISCUSSION

Recently, several techniques for automated segmentation have been proposed to overcome the limitations of manual annotation, including stereology and semi-automatic seg-

mentation [13-18]. Due to marked anatomical variations, the automatic segmentation results of PKLD in previous research have left much to be desired. About a decade ago, when there were no expectations for deep-learning-based automatic segmentation, a prospective cohort of volumetry (CRISP) was launched to evaluate ADPKD progression and to pursue objective volume measurements by a number of clinicians. To minimize the human factors in the cohort, the proportion of true values was calculated on the basis of comparisons with the GT (the volume of phantom kidneys), and inter- and intraclass reliability coefficients (Pearson correlations) were reported. The other clinical trials of ADPKD, including ALADIN, DIPAK 1, ELATE, and PREVENT-ADPKD [19-22], structured their study designs to minimize unavoidable human factors in volume measurements of target organs. However, we could find no detailed protocols for minimizing human factors in calculating volumetry in a large-scale cohort [23] or the major clinical trials [6,24-26].

If deep-learning-based automatic segmentation can guarantee a certain level of performance, we could utilize the consistent output of deep-learning-based automatic segmentation based on a fixed algorithm and explore its potential to overcome human factors. Technically, for deep-learning-based automatic segmentation, the intra- and interobserver variabilities are zero for a given algorithm. As such, the statistical assessment of the agreement of results would no longer be necessary, biomarkers would be estimated with mathematical rigor, and researchers could concentrate on the effectiveness of various management strategies based on objective output.

This AI-assisted segmentation platform, which showed expert-level segmentation accuracy, has the advantage of minimizing intra- and interobserver variability, especially when comparing volumes before and after drug treatment. This automatic system will dramatically increase the production of reliable data with mathematical rigor, and will significantly contribute to the evaluation and development of appropriate therapeutic agents for this intractable disease.

For a total of 214 cases, 3,302 images were produced by the ADPKD specialists. These images were divided into 175

training data and 39 test data, the AI was trained using the training data, and the model performance was evaluated through internal validation. Then, the objective performance of the model was evaluated using the test data, which was a completely independent dataset. In a previous paper, the AI was trained using 244 CT cases [19,27].

The significance of using the combined volume of the liver and the kidneys (total abdominal volume, or TAV) is shown in the current paper. This combined volume is known to be associated with malnutrition and abdominal symptoms. For example, Suwabe et al. [28] showed that height-adjusted total kidney volume (htTAV) is correlated with abdominal distension. In addition, Ryu et al. [5] reported a significant relationship between htTAV and malnutrition. In the case of drugs (e.g., somatostatin analogues) that affect both total liver volume and TKV (total kidney volume), the combined volume can be used as an effective biomarker [21].

Although a previous study on volumetry in a large-scale ADPKD cohort using MRI has recently been reported, we chose to use CT data to develop our segmentation framework. In fact, the use of MRI is not straightforward in many countries, including Korea. MRI has several disadvantages that make it unsuitable for volumetry studies. First, MRI is expensive and image acquisition takes too long. Particularly for patients with severe organomegaly, volumetry is routinely required in clinical practice. However, because of the symptoms of mass effects (e.g., dyspnea and pain), patients often have difficulties maintaining a supine position for a long time and cannot hold their breath for the length required for routine MRI settings. MRI data acquired under these circumstances have fundamental drawbacks that make accurate volume measurements impossible. Second, MR images from various institutions are difficult to collect in a consistent format. Data inconsistency is the predominant reason for which a considerable amount of data is discarded, and this issue complicates multicenter study protocols.

There are a few limitations to the current study. We had to follow the traditional way of setting the GT, using training data refinement to reduce the ambiguity of the labels at least five times per case with three PKLD specialists. Still, there is no innovative way for identifying the 'real' GT except by making calculations after organ resection. A further limitation is that we only analyzed CT images; however, considering the high expense and difficulties in the clinical utilization of MRI discussed above, CT imaging could be a reasonable solution for volumetry. CT is also more useful than MRI for observing complications of ADPKD, such as stones and calcifications. CT scans often require the injection of contrast medium, which may have nephrotoxic effects,

because manual tracing of the kidney and liver is particularly challenging in CT data acquired without the injection of contrast medium. However, more than 85% of our cases (182/214) had noncontrast imaging, and our AI model showed a very high performance for these cases. This finding underscores the potential value of this technique, as it does not require the use of contrast. Noncontrast CT provides a relatively good image compared with an MR image.

In addition to the potential of this approach for research purposes, rapid communication of volumetric information to patients can promote democratization of information in the clinical setting, in which information asymmetry is a severe problem. We believe that the remarkable outcomes of the current study will be an important milestone on the path toward the full-scale application of AI in our clinical field.

CONCLUSIONS

In PKLD, volumetry using AI showed overwhelming speed compared with humans. Moreover, AI was not inferior to human specialists in terms of accuracy and performance. Considering that even PKLD specialists can make mistakes, AI might be used effectively in clinical practice.

CONFLICTS OF INTEREST

The authors have nothing to disclose.

ACKNOWLEDGMENTS

This research was supported by a grant of the National Research Foundation of Korea (NRF-2017R1C1B1008260, NRF-2016R1D1A1B03934173) and the Hallym University Research Fund 2017 (HURF-2017-36), and the Young Investigator Research Grant from the Korean Society of Nephrology (2019).

AUTHORS' CONTRIBUTIONS

S.T.Y., K.H., and M.H.S. conceived this study and drafted the manuscript. L.J.H., C.J.S., and C.H. helped collect the clinical samples. K.K., K.G., K.J., Y.S., and P.H. helped to perform volumetry. K.H., H.Y.U., and S.T.Y. determined the ground-truth. K.H.J., H.M., B.E., and Y.J.W. participated in the experiment competing with AI. K.H., R.K.H., L.Y.S., and S.T.Y. contributed to revision and discussion of this study.

REFERENCES

- Harris PC. Autosomal dominant polycystic kidney disease: clues to pathogenesis. *Hum Mol Genet* 1999;8:1861-6.
- Bae KT, Zhu F, Chapman AB, Torres VE, Grantham JJ, Guay-Woodford LM, et al. Magnetic resonance imaging evaluation of hepatic cysts in early autosomal-dominant polycystic kidney disease: the Consortium for Radiologic Imaging Studies of Polycystic Kidney Disease cohort. *Clin J Am Soc Nephrol* 2006;1:64-9.
- Chauveau D, Fakhouri F, Grünfeld JP. Liver involvement in autosomal-dominant polycystic kidney disease: therapeutic dilemma. *J Am Soc Nephrol* 2000;11:1767-75.
- Kim H, Park HC, Ryu H, Kim K, Kim HS, Oh KH, et al. Clinical correlates of mass effect in autosomal dominant polycystic kidney disease. *PLoS One* 2015;10:e0144526.
- Ryu H, Kim H, Park HC, Kim H, Cho EJ, Lee KB, et al. Total kidney and liver volume is a major risk factor for malnutrition in ambulatory patients with autosomal dominant polycystic kidney disease. *BMC Nephrol* 2017;18:22.
- Torres VE, Chapman AB, Devuyst O, Gansevoort RT, Grantham JJ, Higashihara E, et al. Tolvaptan in patients with autosomal dominant polycystic kidney disease. *N Engl J Med* 2012;367:2407-18.
- Serra AL, Poster D, Kistler AD, Krauer F, Raina S, Young J, et al. Sirolimus and kidney growth in autosomal dominant polycystic kidney disease. *N Engl J Med* 2010;363:820-9.
- Walz G, Budde K, Mannaa M, Nürnberger J, Wanner C, Sommerer C, et al. Everolimus in patients with autosomal dominant polycystic kidney disease. *N Engl J Med* 2010;363:830-40.
- Bae KT, Tao C, Zhu F, Bost JE, Chapman AB, Grantham JJ, et al. MRI-based kidney volume measurements in ADPKD: reliability and effect of gadolinium enhancement. *Clin J Am Soc Nephrol* 2009;4:719-25.
- Milletari F, Navab N, Ahmadi S. V-net: fully convolutional neural networks for volumetric medical image segmentation [abstract]. 2016 Fourth International Conference on 3D Vision (3DV); 2016 Oct 25-28; Stanford, CA, United States. Piscataway (NJ): IEEE; 2016. p. 565-571.
- Chung M, Lee J, Lee M, Lee J, Shin YG. Deeply self-supervising edge-to-contour neural network applied to liver segmentation. arXiv. 180800739 [Preprint]. 2018 [cited 2019 Dec 2]. Available from: <https://arxiv.org/abs/1808.00739>.
- DeVries T, Taylor GW. Improved regularization of convolutional neural networks with cutout. arXiv. 170804552 [Preprint]. 2017 [cited 2019 Dec 2]. Available from: <https://arxiv.org/abs/1708.04552>.
- Ruskó L, Bekes G, Fidrich M. Automatic segmentation of the liver from multi- and single-phase contrast-enhanced CT images. *Med Image Anal* 2009;13:871-82.
- Masoumi H, Behrad A, Pourmina MA, Roosta A. Automatic liver segmentation in MRI images using an iterative watershed algorithm and artificial neural network. *Biomed Signal Process Control* 2012;7:429-37.
- Huynh HT, Karademir I, Oto A, Suzuki K. Computerized liver volumetry on MRI by using 3D geodesic active contour segmentation. *AJR Am J Roentgenol* 2014;202:152-9.
- Turco D, Valinoti M, Martin EM, Tagliaferri C, Scolari F, Corsi C. Fully automated segmentation of polycystic kidneys from noncontrast computed tomography: a feasibility study and preliminary results. *Acad Radiol* 2018;25:850-5.
- Sharma K, Rupprecht C, Caroli A, Aparicio MC, Remuzzi A, Baust M, et al. Automatic segmentation of kidneys using deep learning for total kidney volume quantification in autosomal dominant polycystic kidney disease. *Sci Rep* 2017;7:2049.
- Kim Y, Bae SK, Cheng T, Tao C, Ge Y, Chapman AB, et al. Automated segmentation of liver and liver cysts from bounded abdominal MR images in patients with autosomal dominant polycystic kidney disease. *Phys Med Biol* 2016;61:7864-80.
- Caroli A, Perico N, Perna A, Antiga L, Brambilla P, Pisani A, et al. Effect of longacting somatostatin analogue on kidney and cyst growth in autosomal dominant polycystic kidney disease (ALADIN): a randomised, placebo-controlled, multicentre trial. *Lancet* 2013;382:1485-95.
- Meijer E, Drenth JP, d'Agnolo H, Casteleijn NE, de Fijter JW, Gevers TJ, et al. Rationale and design of the DIPAK 1 study: a randomized controlled clinical trial assessing the efficacy of lanreotide to Halt disease progression in autosomal dominant polycystic kidney disease. *Am J Kidney Dis* 2014;63:446-55.
- Chrispijn M, Drenth JP. Everolimus and long acting octreotide as a volume reducing treatment of polycystic livers (ELATE): study protocol for a randomized controlled trial. *Trials* 2011;12:246.
- Wong ATY, Mannix C, Grantham JJ, Allman-Farinelli M, Badve SV, Boudville N, et al. Randomised controlled trial to determine the efficacy and safety of prescribed water intake to prevent kidney failure due to autosomal dominant polycystic kidney disease (PREVENT-ADPKD). *BMJ Open* 2018;8:e018794.
- Chapman AB, Torres VE, Perrone RD, Steinman TI, Bae KT, Miller JB, et al. The HALT polycystic kidney disease trials: design and implementation. *Clin J Am Soc Nephrol* 2010;5:102-9.
- Torres VE, Chapman AB, Devuyst O, Gansevoort RT, Perrone RD, Dandurand A, et al. Multicenter, open-label, extension trial to evaluate the long-term efficacy and safety of early versus delayed treatment with tolvaptan in autosomal dominant polycystic kidney disease: the TEMPO 4:4 Trial. *Nephrol Dial*

- Transplant 2017;32:1262.
25. Torres VE, Chapman AB, Devuyst O, Gansevoort RT, Perrone RD, Koch G, et al. Tolvaptan in later-stage autosomal dominant polycystic kidney disease. *N Engl J Med* 2017;377:1930-42.
 26. Ruggenti P, Gentile G, Perico N, Perna A, Barcella L, Trillini M, et al. Effect of sirolimus on disease progression in patients with autosomal dominant polycystic kidney disease and CKD stages 3b-4. *Clin J Am Soc Nephrol* 2016;11:785-94.
 27. Kline TL, Korfiatis P, Edwards ME, Blais JD, Czerwiec FS, Harris PC, et al. Performance of an artificial multi-observer deep neural network for fully automated segmentation of polycystic kidneys. *J Digit Imaging* 2017;30:442-8.
 28. Suwabe T, Ubara Y, Mise K, Kawada M, Hamanoue S, Sumida K, et al. Quality of life of patients with ADPKD-Toranomon PKD QOL study: cross-sectional study. *BMC Nephrol* 2013;14:179.