Automation of the kidney function prediction and classification through ultrasound-based kidney imaging using deep learning

Chin-Chi Kuo

Big Data Center, China Medical University Hospital, China Medical University, Taichung, Taiwan

Prediction of kidney function and chronic kidney disease (CKD) through kidney ultrasound imaging has long been considered desirable in clinical practice because of its safety, convenience, and affordability. However, this highly desirable approach is beyond the capability of human vision. We developed a deep learning approach for automatically determining the estimated glomerular filtration rate (eGFR) and CKD status. We exploited the transfer learning technique, integrating the powerful ResNet model pretrained on an ImageNet dataset in our neural network architecture, to predict kidney function based on 4,505 kidney ultrasound images labeled using eGFRs derived from serum creatinine concentrations. To further extract the information from ultrasound images, we leveraged kidney length annotations to remove the peripheral region of the kidneys and applied various data augmentation schemes to produce additional data with variations. Bootstrap aggregation was also applied to avoid overfitting and improve the model's generalization. Moreover, the kidney function features obtained by our deep neural network were used to identify the CKD status defined by an eGFR of <60 ml/min/1.73 m2. A Pearson correlation coefficient of 0.741 indicated the strong relationship between artificial intelligence (AI)- and creatinine-based GFR estimations. Overall CKD status classification accuracy of our model was 85.6% —higher than that of experienced nephrologists (60.3%–80.1%). Our model is the first fundamental step toward realizing the potential of transforming kidney ultrasound imaging into an effective, real-time, distant screening tool. AI-GFR estimation offers the possibility of noninvasive assessment of kidney function, a key goal of AI-powered functional automation in clinical practice.

Introduction

The main clinical application of kidney ultrasound imaging involves excluding reversible causes of acute kidney injury, such as urinary obstruction, or identifying irreversible chronic kidney disease (CKD) that precludes unnecessary workup such as kidney biopsy.1 Its noninvasiveness, low cost, lack of ionizing radiation, and wide availability make it an attractive option for frequent monitoring and follow-up of the longitudinal change in kidney length and sonographic characteristics of kidney cortex relevant to kidney functional change. However, the high subjective variability in image acquisition and interpretation makes it difficult to translate experience-based prediction into standardized practice, such as invasive serum creatinine measurement. Yet, noninvasive imaging techniques for organ functional and structural characterization have been increasingly investigated aiming to minimize the invasive approach in both diagnostic and screening settings. Lorenzo et al. has recently proposed a unique pediatric CKD care model balancing cost and minimizing invasiveness with the improvement of risk prediction by using kidney ultrasound imaging to predict the development of CKD and surgical outcomes in infants with hydronephrosis.2

Conventionally, nephrologists tend to use kidney length and volume and cortical thickness and echogenicity to evaluate the severity of kidney injury. Very short renal length (e.g., <8 cm), apparent white cortex, and contracted capsule contour, all indicate an irreversible kidney failing process with high specificity but limited sensitivity.3 Furthermore, whether these ultrasonographic parameters can be used to predict accurate estimated glomerular filtration rates (eGFRs) remains controversial. For instance, studies3,4,5,6,7,8,9,10,11,12,13,14 have reported that although kidney length is highly specific in detecting irreversible CKD, its correlation with eGFR was only weak to moderate, ranging no association to 0.66. Even if only studies using the conventional Modification of Diet in Renal Disease Study (MDRD) equation to estimate eGFR (MDRD-eGFR) are considered,15 the best correlation noted between kidney length and eGFR has been only 0.36.8,16 Similarly, a fair-to-moderate correlation of kidney volume and cortical echogenicity with eGFR has been reported.6,11,14,17,18 By contrast, cortical thickness seems to be better correlated with MDRD-eGFR than is kidney length, with a correlation coefficient as high as 0.85.3,8,9,10,11,12,13,16,18 However, the study reporting the correlation coefficient of 0.85 was obtained for a small sample of 42 adults with CKD without validation.8 Yapark et al.5 developed a CKD scoring system integrating three ultrasonographic parameters, namely kidney length, parenchymal thickness, and echogenicity, to improve the correlation; however, the correlation was moderate (r=0.587).5 Furthermore, the assigned score of each parameter remains subjective with undefined interobserver reliability.5,19

To overcome substantial interobserver variability in kidney ultrasound interpretation, machine learning provides a solid and objective foundation for analytic standardization to inform clinical decisions. Recent advances in image segmentation, classification, and registration through deep learning have considerably expanded the scope and scale of medical image analysis.20 Deep learning-oriented diagnostic applications may minimize unnecessary and invasive procedures, thus greatly improving the efficiency and sustainability of current health care systems. Moreover, with the phenomenal increase in computing performance, real-time computer-aided diagnosis may further change mobile telecare and telemedicine. In the current CKD care model, it remains controversial whether kidney function should be routinely screened in all asymptomatic adults.21 The most commonly used CKD screening tests include testing the urine for protein or testing the blood for serum creatinine; however, there is no conclusive evidence suggesting which screening test is more appropriate to the other in the context of routine screening. Developing an easily available and noninvasive image marker of kidney function using deep learning methods thus may provide a valuable complimentary tool for diagnosing CKD. To explore this possibility in clinical practice, we developed a deep learning algorithm based on both kidney ultrasound imaging and clinical data in a large registry-based CKD cohort.

Discussion

Kidney sonography has long been a convenient point-of-care diagnostic tool in nephrology. With the advancements in deep CNNs, artificial intelligence (AI) can be introduced for real-time interpretation of kidney sonography—an essential first step toward a wide telemedicine outreach for effectively screening CKD in a community setting. We attempt to use a deep learning algorithm to predict eGFR and CKD status in a study population with various degrees of CKD (stages 1–5). The proposed algorithm moderately predicts continuous eGFR. Furthermore, it can reliably determine whether eGFR is below 60 ml/min/1.73 m2, with an accuracy superior to that of senior nephrologists. Kidney function is particularly prone to irreversible decline after eGFR becomes <60 ml/min/1.73 m2. Thus, this algorithm is applicable because it helps optimize cost-effective CKD screening practices without laboratory testing, particularly in settings with limited health care resources. Notably, this algorithm provides a real-time diagnosis and patient referral. The present study also demonstrates the possible role of AI in turning conventional images into functional screening and diagnostic tools – this type of automation will be pervasive in the era of AI and Big Data. For instance, prior studies have applied AI in automatically identifying glomeruli to standardize renal biopsy interpretation,27,28,29 and even trying to predict kidney function.30 While the predictive accuracy for eGFR or CKD is not perfectly satisfied with current clinical practice at this stage, our proposed deep learning algorithm is to complement existing screening or case-finding instruments, rather than to replace them.

Surging CKD-related health care costs burden both developed and developing economies.31 In the United States, CKD prevalence is expected to increase by 16.7% by 2030. A study showed that adults aged 30–49 years without CKD at baseline had a residual lifetime incidence of CKD as high as 54%.32 Global trends in population aging may increase CKD prevalence because aging is a pertinent risk factor for CKD.33 The primary prevention of CKD through early detection is recommended particularly among high-risk patients with diabetes and hypertension.34 However, the screening relies on serum creatinine (invasive) and urine protein (noninvasive) level measurement, require blood and urine specimens to be analyzed by laboratory personnel by using laboratory equipment of appropriate quality, respectively. The mass screening for CKD in the general population by measuring serum creatinine levels is expensive in most health care systems because of the costs and invasiveness involved. Proteinuria-based screening, such as routine urinalysis, is more acceptable by the general population because it is noninvasive. Among the 10 studies enrolled in a recent systematic review of the cost-effectiveness of primary CKD screening, 8 used proteinuria-based screening, such as urine dipstick testing and protein-to-creatinine ratio measurement, reflecting the well perceived patient acceptance of noninvasive urine-based tests.34 However, the poor screening performance of routine urinalysis, with 93% specificity but only 11% sensitivity, in detecting early CKD among the general population reveals the need for new screening methods.35

The cost-effectiveness of mass screening in general population for CKD has long been debated.36,37 For instance, the American College of Physicians' 2013 clinical practice guideline for managing stage 1–3 CKD recommends against universal CKD screening among asymptomatic adults because evidence from randomized trials supporting the benefits of regularly screening for CKD is insufficient.38 By contrast, the American Society of Nephrology strongly recommends regular screening of CKD, given its clinical silence and preventable progression with relatively low cost of testing.39,40 More conclusive research is required to fill the practice gaps in key areas ranging from the identification of novel and cost-effective techniques to the development of systemic evaluation methods that are economically efficient for mass CKD screening.

Over the past decade, Taiwan has demonstrated the highest end-stage renal disease (ESRD) incidence and prevalence worldwide, and they are still increasing, despite the considerable amount of resources available for CKD care programs.41 Therefore, cost-effective universal screening for CKD may aid Taiwan. On the basis of our current study, we propose a two-stage CKD screening approach: Stage 1 comprises kidney ultrasound image screening using our AI-aided screening method, whereas stage 2 comprises serum creatinine quantification for identifying missed true positives. This two-stage screening model offers advantage in terms of logistics supportability (wide availability of ultrasound machines and pervasive Internet services in Taiwan) and sustainment with financial capability and affordability. This AI-aided model also provides a potential complementary care model to routine urinalysis or serum creatinine measurement for primary CKD screening. Conducting a comprehensive economic analysis to examine the cost-effectiveness of our proposed AI-aided screening model is beyond the scope of this study. Future studies should examine the economic viability of our model. Furthermore, our approach should be extended to mobile applications to augment its impact on health care efficiency and quality.

Ensuring a sufficient number of samples is a prerequisite for training a robust deep learning model. We evaluated how much performance improvement can be achieved by increasing the data size through the following steps of the experimental study: Use a sample 10% of the entire dataset without replacing the experimental data set. Train several ResNet models by using the experimental dataset under different random seeds, and average their testing performance to obtain a robust evaluation. Randomly add additional 10% of the entire dataset to the experimental dataset. Repeat Steps 2 and 3 until all data are added to the experimental dataset. We did not apply bootstrap aggregation in this experiment. The results are shown in Supplemental Fig. 2a: a clear declining trend in testing loss was noted when data size increased. Simultaneously, Pearson's correlation coefficient improved (Supplemental Fig. 2a). For instance, compared with 10% of the entire dataset, the testing loss using 50% of the entire dataset resulted in a twofold increase in performance and increase in correlation coefficient from 0.53 to 0.66. Therefore, the testing performance of our model may improve when more sonographic studies are available.

Relatively few sonographic studies in our dataset (15%) reported a normal eGFR of greater than 60 ml/min/1.73 m2. To resolve this data imbalance for CKD status classification, we reduced the weight of the samples using eGFR of <60 ml/min/1.73 m2 by a factor of 0.25 to balance their effects, classify the loss, and summarize the predictive performance based on unscaled data (Supplemental Table 6). The overall accuracy was comparable to the results based on scaled data, despite the inherent tradeoff between sensitivity and specificity. Both approaches can adequately be the first screening test in our proposed two-stage mass screening model. Because CKD and ESRD are highly prevalent in Taiwan, the best positive predictive power can be obtained by adjusting the algorithm targeting high specificity. This sensitivity experiment indicated the strategic flexibility of our deep neural network algorithm.

Numerous possibilities exist for functional AI-powered automation to support the efficiency of health care. Our proposed deep learning algorithm offers the possibility of noninvasive assessment of kidney function and represents a fundamental step for realizing the potential of transforming kidney ultrasound into an effective, real-time screening tool. With a diagnostic accuracy comparable to the predictions of experienced nephrologists, our CNN model has the potential to improve the cost efficiency of universal CKD screening, for instance, by selecting high-risk patients using kidney ultrasound in the first round of a two-stage screening model.