KI REPORTS
KIReports.org

Pilot Study of a Web-Based Tool for Real-Time Adequacy Assessment of Kidney Biopsies

Meysam Ahangaran[1], Emily Sun[1], Khang Le[1], Jiawei Sun[1], William M. Wang[1], Tian Herng Tan[1], Lingkai Yin[1], Lyle J. Burdine[3], Zeijko Dvanajscak[4], Clarissa A. Cassol[4], Shree Sharma[4,5] and Vijaya B. Kolachalama[1,2,5]

[1]Department of Medicine, Boston University Chobanian and Avedisian School of Medicine, Boston, Massachusetts, USA; [2]Department of Computer Science and Faculty of Computing & Data Sciences, Boston University, Boston, Massachusetts, USA; [3]Department of Surgery, University of Arkansas for Medical Sciences, Little Rock, Arkansas, USA; and [4]Arkana Laboratories, Little Rock, Arkansas, USA

Correspondence: Vijaya B. Kolachalama, Department of Medicine, Boston University Chobanian & Avedisian School of Medicine, 72 E. Concord Street, Evans 636, Boston, Massachusetts 02118, USA. E-mail: vkola@bu.edu

[5]SS and VBK contributed equally to this work.

Received 2 April 2024; revised 22 May 2024; accepted 10 June 2024; published online 20 June 2024

Kidney Int Rep (2024) **9,** 2809–2813; https://doi.org/10.1016/j.ekir.2024.06.019
KEYWORDS: biopsy adequacy; kidney biopsy; kidney cortex; machine learning

INTRODUCTION

Kidney biopsies are essential for diagnosing kidney disorders in both native and transplanted kidneys.[1,2] However, the increasing incidence of biopsies yielding insufficient tissue for diagnosis presents a dual challenge: it strains the health care system and endangers patients, who may require rebiopsy or face the risk of inaccurate diagnoses due to unsampled disease.[3] Traditionally, real-time assessment of biopsy adequacy has been performed by pathologists using microscopes.[4] Unfortunately, due to cost and staffing constraints, many institutions cannot have a pathologist present at the time of biopsy. As a result, the incidence of kidney biopsies yielding incomplete diagnostic information due to insufficient quantity of tissue has surged, imposing a substantial burden on the health care system, and placing patients at risk of undergoing repeat biopsies or receiving inaccurate diagnoses.

A recent study by Nissen and colleagues examined the evolution of the native renal biopsy miss rate over 15 years, revealing a significant increase from 2% in 2005 to 14% in 2020.[5] Notably, this rise is linked to radiologists performing 95% of biopsies in 2018, primarily with smaller-diameter needles (18G/20G), compared to just 5% in 2005. This shift in the performance of the biopsy procedure from nephrologists to radiologists correlates with the rising rates of biopsy inadequacy, underscoring the urgency of addressing this critical issue.[6] The impact of this transition is evident in the reduction of critical components such as the number of glomeruli per centimeter of core biopsy and mean core width, especially when smaller needles are used. These developments highlight the compelling necessity for novel technologies capable of addressing this issue in real-time, mitigating the risks posed to patients, and enhancing the accuracy of diagnoses.

RESULTS

We present findings from a pilot study of a web-based software developed with a deep learning framework designed for real-time, quantitative assessment of kidney biopsy adequacy directly from smartphone photographs. The software highlights the cortex areas and determines the cortex percentage relative to the entire biopsy. This framework was trained on a dataset comprising nephropathologist-confirmed annotations of the kidney cortex on 100 digital biopsy images (Supplementary Figure S1). It leverages the segment anything in medical images model,[7] a publicly available image segmentation architecture, as the foundation for this deep learning framework (Figure 1). The kidney biopsy cores (16G) were obtained from discarded kidneys and photographed using an iPhone 13 Pro camera. The biopsy images were meticulously annotated by nephropathologists (20 images by 3 nephropathologists and 80 images by 1 nephropathologist) and evaluated

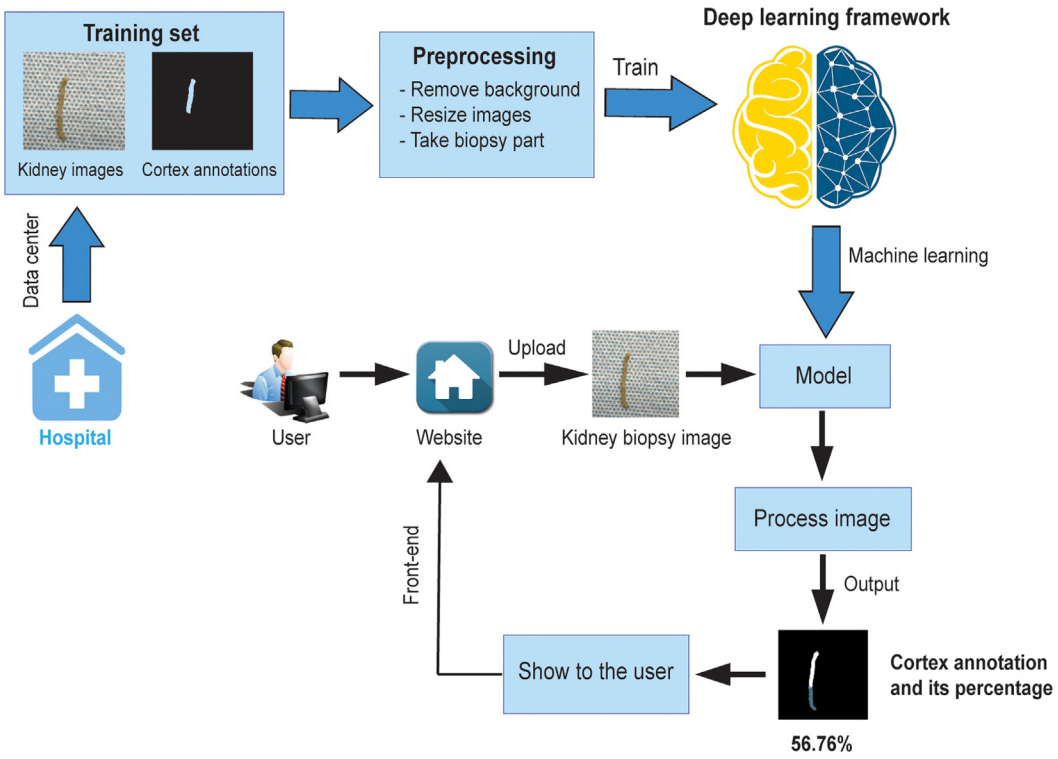

Figure 1. Web-based tool for kidney biopsy adequacy assessment. The depicted web-based tool integrates a machine learning model trained on a dataset of kidney cortex annotations provided by nephropathologists. It comprises both front-end and back-end components. Users are prompted to upload a kidney biopsy image through the website's interface. Upon submission, the back-end server processes the uploaded image, generating an annotated image highlighting the kidney cortex and providing the corresponding percentage. The processed result, including the annotated image and percentage value, is then presented to the user.

by a consensus team of nephropathologists with a combined total experience of over 21 years (Mean ± SD: 7 ± 3.27 years).

To evaluate the model's performance, we employed a 5-fold cross-validation method using the 2 most used metrics for the assessment of medical image segmentation tasks: Intersection over Union[S3] and Dice Similarity Coefficient[S4] (Supplementary Methods). The model demonstrated robust performance, yielding high-precision outputs on 100 kidney biopsy images with an Intersection over Union of 0.6319 ± 0.266 and a Dice Similarity Coefficient of 0.7309 ± 0.2721.

In Figure 2, we present the model's results on 5 representative kidney biopsy images. In these images, the white pixels in the actual cortex images indicate the cortex areas as annotated by experts, whereas the white pixels in the predicted cortex images represent the cortex areas identified by the model. The corresponding actual and predicted cortex percentages are calculated as the ratio of cortex pixels to the total number of kidney biopsy pixels, ranging from 0 to 1. The Intersection over Union and Dice Similarity Coefficient metrics measure the similarity between the actual and predicted cortex segments, with values ranging from 0 (no intersection) to 1 (perfect annotation).

The results indicate that the model performs effectively across various levels of cortex presence in biopsy images, including high (∼100%), medium (∼50%), and low (∼20%) cortex levels. The mean square error for predicting cortex percentage across all 100 kidney images was 0.2689, with a mean and SD of 0.18 ± 0.2 (Supplementary Figures S2 and S3). Detailed results, including all predicted and actual cortex percentages and their corresponding Intersection over Union and Dice Similarity Coefficient scores, are available in Supplementary Data S1.

In addition, the model exhibited a high correlation between predicted and actual cortex percentages, demonstrating its reliability in cortex annotation of kidney biopsy images. The Spearman correlation was 0.5629, the Pearson correlation was 0.5357, and the coefficient of determination (R^2) was 0.1548 (Supplementary Table S1). These metrics underscore the model's strong performance and potential utility in real-world kidney biopsy adequacy assessments.

Overall, the performance metrics and correlation analyses underscore the efficacy of our deep learning model in accurately identifying and quantifying cortex areas in kidney biopsy images. The model's consistent performance across varying levels of cortex presence and its high correlation with expert annotations

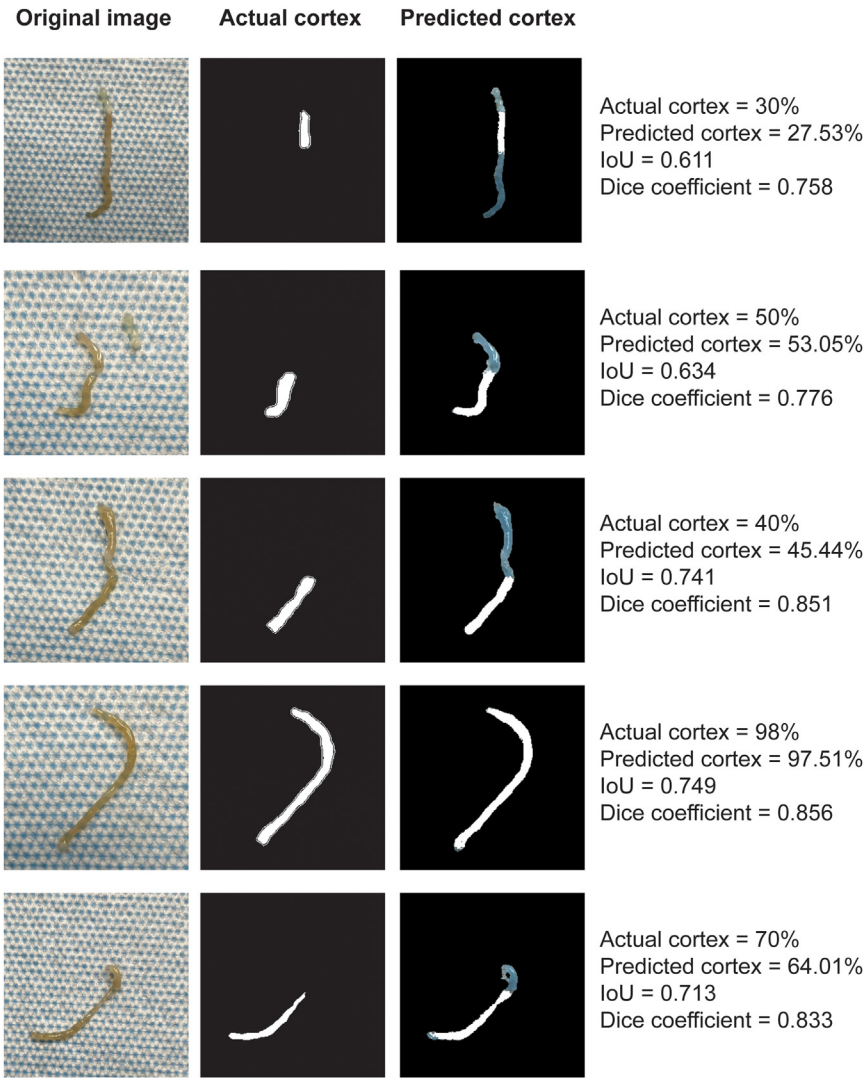

Figure 2. Representative kidney biopsy images with annotations and model predictions. Cortex-annotated images produced by the model alongside their corresponding ground truth annotations for 5 illustrative kidney biopsy cases. Actual and predicted cortex percentages are provided adjacent to each image, representing the proportion of cortex pixels relative to the entire kidney image. In addition, the precision of the model's cortex annotations for each biopsy image is evaluated using 2 segmentation metrics. DSC, Dice Similarity Coefficient; IoU, Intersection over Union;.

highlight its potential as a reliable tool for real-time biopsy adequacy assessment.

DISCUSSION

To ensure that our segmentation model is widely accessible, we have developed it into a user-friendly, web-based application [Details on the web-based tool are available here: https://github.com/vkola-lab/kir2024]. The tool features an intuitive interface compatible with modern web browsers such as Google Chrome and Firefox. To ensure security and manageability, the application requires user registration. New users are guided through an account creation process with verifiable credentials, whereas returning users can easily log in with their existing credentials. Once registered and logged in, users can upload digitized square images of biopsy cores in common image formats such as JPG, PNG, and TIFF. The application provides a feature for delineating the region of interest, allowing users to crop the biopsy area in a square image format and exclude extraneous background elements. This functionality focuses the analysis on the biopsy itself, improving the quality of the output, and enhancing user engagement.

Upon confirming and submitting an image for analysis, the platform generates 2 key outputs as follows: (i) an annotated image highlighting the cortex area within the core biopsy, and (ii) a calculated percentage representing the cortex area in relation to the entire core biopsy image. We denote the computed cortex area as an indicator of kidney biopsy adequacy. The software is

capable of handling various scenarios, including digitized biopsy images that may entirely lack cortex or contain the full kidney cortex. In cases where the estimated biopsy adequacy is either 0% or 100%, the software's response varies accordingly. If the predicted adequacy is 100%, the software displays the percentage of the cortex, with the cortex visualization encompassing the entire biopsy image in the output. Conversely, if the estimated adequacy is 0%, no cortex is highlighted on the biopsy image, and the displayed output resembles the original image uploaded by the user, with no discernible cortex area.

To capture optimal biopsy images, we recommend users use the camera of a modern smartphone, which can produce high-resolution photographs essential for precise analysis. For the best results with our web-based tool, it is advisable to take photographs against a light, uniform background. This backdrop ensures the digital image is well-prepared for effective processing, leading to more reliable outcomes. We also encourage users to consider uploading multiple images of the same biopsy sample. This practice can provide a more comprehensive estimate of adequacy by offering an averaged value, thereby enhancing the user's confidence in the software's accuracy. However, it is important to note that the current version of our tool supports uploading only 1 image at a time. Consequently, users opting for multiple image uploads will need to manually record the values obtained from each image and calculate the average adequacy offline. Although this approach is slightly more time-consuming, it can yield a more nuanced and trustworthy assessment of the biopsy sample.

An important limitation of our model is the restricted scope of the training dataset, which comprises images obtained from only 5 deceased donor kidneys. The images were captured against a uniform background material and with the same smartphone model (iPhone 13 Pro), which does not reflect the variability encountered in different laboratory settings. For example, backgrounds may include blood, fat, and capsular tissue, and other laboratories might utilize distinctly colored backgrounds. Furthermore, the current model's training is limited to images taken with 16G needles, potentially limiting its applicability across different clinical environments. Given these constraints, it is crucial to emphasize that further validation of this tool with a more diverse set of biopsy samples (i.e., 18G needles and smaller cores), and other conditions is essential before it can be reliably implemented in clinical practice. No external validation on real kidney biopsies has been performed to date, and this tool should not be used for clinical decision-making without comprehensive evaluation across varied and realistic biopsy scenarios.

In addition, our tool currently does not consider the total tissue volume, focusing instead on estimating the percentage of cortex area relative to the overall biopsy area from a 2-dimensional image. Although this method provides a useful proxy, it may not fully capture the complexity of 3-dimensional tissue structures. Future versions of the tool, potentially leveraging advanced smartphone imaging technologies, may be able to capture the 3-dimensional shape of the tissue, thus enabling accurate estimation of tissue volume.

The model's current methodology emphasizes the cortical region and calculates the cortex percentage relative to the entire biopsy sample. However, we acknowledge that sampling cortex without glomeruli can occur, and such biopsies may still be deemed inadequate. Future iterations will aim to incorporate advanced image analysis techniques capable of identifying and counting glomeruli within the cortical region, focusing on perfused glomeruli, which are clinically significant. Moreover, the tool's effectiveness on parenchyma with advanced chronicity needs further investigation. The performance of the tool can vary depending on the extent of tissue damage and the specific characteristics of the chronic condition. Validation studies focusing on more advanced chronicity are needed to fully assess and optimize the tool's performance in these scenarios.

In conclusion, our solution offers a user-friendly, web-based tool capable of providing real-time quantitative assessments of kidney biopsy adequacy. Although further evaluation is required to fully ascertain its impact on patient outcomes and the health care system, this tool represents an important advancement in the application of artificial intelligence–assisted technologies for kidney disease management.

DISCLOSURE

VBK is on the scientific advisory board for Altoida Inc. and serves as a consultant to AstraZeneca. SS has served on advisory board of Novartis. CAC is on Aurinia Advisory Board for Volclosporin and served as a consultant for Novartis Iptacopan Educational Steering Committee. All the other authors declared no competing interests.

ACKNOWLEDGMENTS

This work was supported by the NIH SBIR phase 1 grant (R43-DK134273). VBK also acknowledges support from the following sources: the Karen Toffler Charitable Trust, the American Heart Association (20SFRN35460031), the National Institutes of Health (RF1-AG062109, R01-HL159620, and R21-CA253498), and the National Institute on Aging's Artificial Intelligence and Technology Collaboratories (AITC) for Aging Research program (P30-AG073014).

SUPPLEMENTARY MATERIAL

Supplementary File (PDF)

Supplementary Methods.

Supplementary References.

Supplementary Data S1 (CSV)

Figure S1. Density plot illustrating the distribution of actual cortex percentages across all 100 kidney biopsy images, employing Kernel Density Estimation (KDE) methodology.

Figure S2. Scatter plot depicting the relationship between predicted and actual cortex percentages. The Mean Square Error (MSE) for cortex prediction is calculated as 0.2689, with the red line representing the estimated regression line.

Figure S3. Density plot illustrating the distribution of prediction error values, employing Kernel Density Estimation (KDE) methodology.

Table S1. Performance analysis of the proposed method for predicting cortex annotation of kidney biopsy images, showcasing metrics including Mean Square Error (MSE), Prediction error, Intersection over Union (IoU), Dice Similarity Coefficient (DSC), Spearman and Pearson correlations, each accompanied by their respective *P*-values.

STROBE Statement.

REFERENCES

1. Whittier WL, Gashti C, Saltzberg S, Korbet S. Comparison of native and transplant kidney biopsies: diagnostic yield and complications. *Clin Kidney J*. 2018;11:616–622. https://doi.org/10.1093/ckj/sfy051
2. Schnuelle P. Renal biopsy for diagnosis in kidney disease: indication, technique, and safety. *J Clin Med*. 2023;12:6424. https://doi.org/10.3390/jcm12196424
3. Marinaki S, Kapsia E, Liapis G, et al. Clinical impact of repeat renal biopsies in patients with lupus nephritis: renal biopsy is essential especially later in the course of the disease. *Eur J Rheumatol*. 2020;7:2–8. https://doi.org/10.5152/eurjrheum.2019.18146
4. Walker PD, Cavallo T, Bonsib SM. Ad Hoc Committee on renal biopsy guidelines of the renal pathology society. Practice guidelines for the renal biopsy. *Mod Pathol*. 2004;17:1555–1563. https://doi.org/10.1038/modpathol.3800239
5. Nissen CJ, Moreno V, Davis VG, Walker PD. Increasing incidence of inadequate kidney biopsy samples over time: a 16-year retrospective analysis from a large national renal biopsy laboratory. *Kidney Int Rep*. 2022;7:251–258. https://doi.org/10.1016/j.ekir.2021.11.026
6. Ferrer G, Andeen NK, Lockridge J, et al. Kidney biopsy adequacy: a Metric-based study. *Am J Surg Pathol*. 2019;43:84–92. https://doi.org/10.1097/PAS.0000000000001102
7. Ma J, He Y, Li F, Han L, You C, Wang B. Segment anything in medical images. *Nat Commun*. 2024;15:654. https://doi.org/10.1038/s41467-024-44824-z