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], Lyle J. Burdine[2], Zeljko Dvanajscak[3], Clarissa A. Cassol[3], Shree Sharma[3,*], Vijaya B. Kolachalama[1,4,*]

[1]Department of Medicine, Boston University Chobanian & Avedisian School of Medicine, Boston, MA, USA

[2]Department of Surgery, University of Arkansas for Medical Sciences, Little Rock, AR, US

[3]Arkana Laboratories, Little Rock, AR, USA

[4]Department of Computer Science, Boston University, Boston, MA, USA; Faculty of Computing & Data Sciences, Boston University, Boston, MA, USA

*S.S. and V.B.K. jointly supervised this work.

Corresponding author:

Vijaya B. Kolachalama, PhD, FAHA

72 E. Concord Street, Evans 636, Boston, MA, USA – 02118

Email: vkola@bu.edu; Phone: (+1) 617-358-7253; Twitter: vkola_lab

Abstract

The escalating incidence of kidney biopsies providing insufficient tissue for diagnosis poses a dual challenge, straining the healthcare system and jeopardizing patients who may require re-biopsy or face the prospect of an inaccurate diagnosis due to an unsampled disease. Here, we introduce a web-based tool that can provide real-time, quantitative assessment of kidney biopsy adequacy directly from photographs taken with a smartphone camera. The software tool was developed using a deep learning-driven automated segmentation technique, trained on a dataset comprising nephropathologist-confirmed annotations of the kidney cortex on digital biopsy images. Our framework demonstrated favorable performance in segmenting the cortex via 5-fold cross-validation (Dice coefficient: 0.788±0.130) (n=100). Offering a bedside tool for kidney biopsy adequacy assessment has the potential to provide real-time guidance to the physicians performing medical kidney biopsies, reducing the necessity for re-biopsies. Our tool can be accessed through our web-based platform: http://www.biopsyadequacy.org.

Keywords: Kidney biopsy, Biopsy adequacy, Machine learning

Introduction

Kidney biopsies are indispensable for diagnosing renal parenchymal diseases in both native and transplanted organs.[1] Traditionally, pathologists have conducted real-time assessments of biopsy adequacy using microscopes,[2] a practice vital for accurate histopathological evaluations. However, the process faces significant challenges: primarily staffing limitations in resource-constrained settings and the increasing involvement of radiologists in performing the biopsy procedures. Many healthcare institutions grapple with the inability to have a pathologist present on site, within the biopsy suite, for immediate adequacy assessment due to constraints in cost and staffing.[3] This shortfall has led to an escalation in kidney biopsies yielding incomplete diagnostic information due to insufficient tissue samples. Consequently, this increases the healthcare burden and poses substantial risks to patients, including the possibility of undergoing repeat biopsies or receiving potentially inaccurate diagnoses based on incomplete information. A recent study by Nissen and colleagues has shed light on this growing concern,[4] revealing that the rate of incomplete native renal biopsies has alarmingly increased from 2% in 2005 to 14% in 2020. This increase is notably correlated with a shift in biopsy procedures from nephrologists to radiologists,[5] particularly in the United States. In 2018, radiologists performed 95% of these biopsies, often opting for smaller-diameter needles (18g/20g), in stark contrast to only 5% in 2005. Such a transition in procedural practice has led to a decrease in critical biopsy components like the number of glomeruli per centimeter of core biopsy and the mean core width, especially when smaller needles are employed. While nephrologists are acutely aware of the necessity for an adequate amount of cortex containing glomeruli for optimal diagnostic yield, radiologists, who deal with a wide array of procedures, may have variable understanding of this crucial aspect. Therefore, they could significantly benefit from immediate, on-site guidance regarding biopsy adequacy. These developments underscore the urgent need for innovative, smartphone-powered technologies that can provide real-time assessment of biopsy adequacy.[6, 7] Such technologies can substantially reduce the risk to patients and improve the accuracy of diagnoses, particularly in settings where immediate access to onsite pathologists is not feasible.

We present a web-based software developed with a deep learning framework designed for segmenting digital kidney core biopsy images and determining the cortex percentage. This framework was trained on a dataset of digitized core biopsy images and leveraged a publicly available image segmentation architecture, the MedSAM model,[8] as the foundation for this deep learning framework (**Figure 1**). The kidney biopsy cores (16G) were obtained from the discarded kidneys using the iPhone 13 pro camera. The kidney biopsy images were meticulously annotated (20 images by 3 nephropathologists and 80 images by 1 nephropathologist) and evaluated by a consensus team of nephropathologists with a total experience of over 21 years between them (Mean±std: 7±3.27 years).

We evaluated our model's performance using 5-fold cross-validation and applied two key metrics - Intersection over Union (IoU) and Dice coefficient - which are standard in medical image segmentation evaluation. The model demonstrated impressive accuracy and precision across a set of 100 kidney biopsy images, achieving an IoU score of 0.665 (±0.144) and a Dice coefficient of 0.788 (±0.130). These metrics, IoU and Dice coefficient, gauge the similarity between the

expert-annotated and model-predicted cortex segments of kidney images, with their values ranging from 0 (indicating no overlap) to 1 (indicating perfect annotation). In **Figure 2**, we have illustrated the model's performance across five representative kidney biopsy images. This includes both the actual cortex regions, identified by medical experts, and the model's predictions, with white pixels indicating cortex areas. We also provided the actual and predicted cortex percentages for each image, representing the proportion of cortex pixels relative to the total pixels in the kidney biopsy image, on a scale from 0 to 1. Our results demonstrate that the model effectively handles kidney biopsy images with varying levels of cortex presence, including high (~100%), medium (~50%), and low (~25%) cortex levels. Further emphasizing the model's efficacy, we observed a significant correlation between the predicted and actual cortex percentages, validating the model's reliability in accurately annotating cortex regions in kidney biopsy images. This correlation is statistically substantiated with a Spearman correlation coefficient of 0.724 and a p-value of 1.632e-17.

To make our segmentation model widely accessible for onsite kidney biopsy adequacy estimation, we have developed it into a web-based software application (http://www.biopsyadequacy.org). Our tool offers a user-friendly interface that is readily accessible through any modern web browser such as Chrome or Firefox. To ensure security and manageability, the tool requires user registration. New users are guided to create an account with verifiable credentials, while returning users can simply log in with their existing credentials. Once registered and logged in, researchers have the capability to upload digitized images of biopsy cores in common formats such as JPG, PNG, and TIFF. After uploading an image, users are provided with a feature to delineate the region of interest, specifically the area of the biopsy within the image. This functionality enables users to crop the biopsy area in a square image format, excluding extraneous background elements, thereby focusing the analysis on the biopsy itself. This not only improves the quality of the output but also enhances overall user engagement with the tool. Upon confirming and submitting the uploaded image for analysis, the platform generates two key outputs: (1) An annotated image showing the cortex area within the core biopsy. (2) A calculated percentage representing the cortex area in relation to the entire core biopsy image. Our software is adept at handling various scenarios, including digitized biopsy images that may entirely lack cortex or contain the kidney cortex in full. In instances where the estimated cortex percentage is either 0 or 100, the software's response varies accordingly. If the predicted percentage is 100%, the software will display the percentage of the cortex, and the cortex visualization will encompass the entire biopsy image in the output. Conversely, if the estimation is 0%, then no cortex will be highlighted on the biopsy image. In this scenario, the displayed output will resemble the original image uploaded by the user, with no discernible cortex area. After seeing the displayed percentage of the cortex, the adequacy can be estimated by the user depending on their needs. For optimal diagnostic accuracy, we recommend that a biopsy is considered adequate when the model's estimation of the cortex area reaches or exceeds 70%.

For capturing optimal biopsy images, we recommend users to utilize the camera of a modern smartphone. These devices can produce high-resolution photographs, which are essential for the precise analysis of biopsy images. To enhance the suitability of these images for analysis by our

web-based tool, it is advisable to take photographs against a light, uniform background. Such a backdrop ensures that the digital image is well-prepared for effective processing, leading to more dependable outcomes from the analysis. We also encourage users to consider uploading multiple images of the same biopsy sample to our platform. Doing so can provide a more comprehensive estimate of adequacy, offering an averaged value that can enhance 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 one image at a time. As a result, users opting for multiple image uploads will need to manually record the values obtained from each image and calculate the average adequacy offline. This approach, although slightly more time-consuming, can provide a more nuanced and trustworthy assessment of the biopsy sample.

Our study has a few limitations. We trained and tested our model on a set of consensus annotations performed by a team of experienced nephropathologists practicing in the United States. For the tool to be generalizable across practices around the world, it would be important to train the model on a more diverse group of experts. As previously mentioned, the initial images used to train the model were from non-perfused kidneys, in which the distinction from cortex and medulla can be difficult due to lack of blood within glomeruli. Therefore, we plan to further optimize the model using images from biopsies from perfused kidneys, which will more accurately correspond to the real world intended use of the tool. Additionally, the protocol for digital image capture included selection of one smartphone, a relatively clean background, and a fixed distance between the object (i.e., core biopsy) and the camera. Future work will involve testing the robustness of the software framework on more variable real-world backgrounds and biopsy-to-camera distances, as well as other image capture technologies.

In conclusion, our innovative web-based tool provides a convenient and efficient solution for obtaining real-time, quantitative assessments of kidney biopsy adequacy. This tool is particularly valuable in resource-limited settings, where the lack of onsite pathologists can delay the immediate evaluation of biopsy adequacy. By simply taking a picture of the biopsy sample and accessing our tool via the internet, clinicians can instantly obtain an estimate of its adequacy. This immediate access to crucial information can significantly aid in effective clinical decision-making, enabling physicians to determine on the spot whether additional tissue samples are necessary. Although further research is required to fully understand its impact on patient outcomes and healthcare systems, our tool marks an important advancement in the integration of AI-assisted technologies into the management of kidney diseases. Its potential to streamline processes and improve diagnostic accuracy holds great promise for the future of medical care in nephrology.

Figure captions

Figure 1. Web-based tool for kidney biopsy adequacy assessment. Our web-based tool employs a machine learning model that has been trained on a dataset of kidney cortex annotations provided by nephropathologists. It comprises both a front-end and a back-end component. First, the users are given the option to upload a kidney biopsy image via the website. After the user uploads and submits the image for analysis, the back-end server processes it. The tool generates an image with the kidney cortex highlighted and calculates the percentage of cortex. The result is presented to the user, including the model-processed image showcasing the cortex area and the associated percentage value.

Figure 2. Sample kidney biopsy images along with their annotations and model predictions. The figure includes five representative cases, each illustrating the actual cortex regions, as determined by nephropathologists, and the corresponding cortex regions as predicted by our model. For each case, two critical pieces of information are provided: the actual cortex percentage and the predicted cortex percentage. These percentages indicate the proportion of cortex pixels relative to the total pixel count in each kidney biopsy image and are scaled on a range from 0 to 1. To quantitatively assess the model's accuracy in cortex annotation for each kidney biopsy image, we employed two essential segmentation metrics: the Intersection over Union (IoU) and the Dice coefficient. These metrics measure the similarity between the expert-annotated and model-predicted cortex regions, with a value range from 0 (indicating no similarity) to 1 (indicating perfect match in annotation).

Methods

Study population

We acquired 5 fresh, unutilized deceased donor kidneys from the University of Arkansas for Medical Sciences (UAMS) in Little Rock. From these kidneys, we performed a total of 747 core biopsies using a 16G needle. For the purposes of our study, we selected the first 500 images. Each biopsy core was photographed using a handheld iPhone 13 Pro camera. To ensure a clear focus on the kidney core, the camera was not mounted on a tripod but held approximately 2 inches from the tissue by hand. The biopsy cores were free of blood, and the background for each image was the standard absorbing paper typically used in pathology laboratories. Following the initial photography, each biopsy core underwent formalin fixation, paraffin embedding, and sectioning. From this process, we produced a single Periodic Acid–Schiff (PAS) stained slide for each specimen. These slides were then scanned at 20x magnification using our Olympus VS200 whole slide imaging system. The primary purpose of these high-resolution images was to establish a correlation between the histology of the PAS-stained slides and the gross morphology observed in the core biopsies. This comprehensive approach was designed to enhance our understanding of the relationship between the microscopic structure and the gross anatomical features of the kidney tissues.

Expert-driven annotations

The initial phase of our study involved uploading core biopsy images to PixelView (deepPath, Inc., Boston, MA), a web-based tool for digital pathology analysis. Three nephropathologists collaboratively annotated the first 20 images, establishing a consensus on the annotations. Based on this consensus, one nephropathologist proceeded to annotate the subsequent 80 images. The first set of 20 images underwent detailed annotation, focusing on distinct anatomical areas: the cortex, medulla, and fibrofatty tissue. Histology slides served as essential references during this process. However, the annotation task presented several challenges. A primary issue was the difficulty in distinguishing between the cortex and medulla in the gross images due to subtle color gradations. This problem was more noticeable in biopsies from deceased donors, where the kidney cores often had a shiny appearance, making the distinction less clear compared to those from fresh biopsies. Additionally, the transparency of some tissues allowed the background to be visible, complicating the distinction between cortex and medulla areas. A similar challenge was faced in differentiating the capsule from adjacent fibroconnective tissue. Considering these challenges and our goal of determining the biopsy's adequacy with absolute specificity, we decided to focus solely on annotating the cortex in the remaining images. This decision was informed by the clinical importance of the cortex in biopsy evaluations, as it is the primary area of interest during a biopsy. Therefore, it was concluded that providing percentages for all biopsy components was not essential for our analysis. A single nephropathologist, guided by the histology images, subsequently annotated only the cortex in the remaining 80 gross biopsy core images. In total, the annotations of all 100 images, including those from the initial collaborative effort and the subsequent individual annotations, were utilized for the development of our AI model. This approach in the annotation process ensured consistency, laying a foundation for the AI model's reliability in analyzing and interpreting kidney biopsy images.

Machine learning framework

We developed a deep learning framework for segmenting medical images using 100 gross biopsy images and their corresponding cortex annotations. This framework is based on the MedSAM model,[8] an advanced version of the Segment Anything Model (SAM) tailored for medical imagery.[9] MedSAM has been trained on a comprehensive dataset of 1,570,263 image-mask pairs across 10 imaging modalities, demonstrating superior precision in medical image segmentation tasks. Our process began by resizing each gross biopsy image and its annotation to a uniform size of 256x256 pixels. These resized image-annotation pairs were then fed into the MedSAM framework for segmentation analysis. For internal validation, we employed a 5-fold cross-validation method. This involved dividing the data into training (80%) and test (20%) sets, training the model on the training set, and then evaluating it with the test set. This cycle was repeated five times to ensure an unbiased evaluation of the model's performance and generalizability. Using the pre-trained MedSAM model as a foundation, we could exploit its extensive learning capabilities, resulting in more precise and efficient segmentation that meets the specific requirements of biopsy image analysis.

Performance metrics

We assessed the performance of the model by employing metrics such as Intersection over Union (IoU) and Dice coefficient to measure the congruence between the kidney cortex predictions generated by the model and expert annotations. These two metrics are used to evaluate the performance of object detection models by calculating the similarity between predicted segmented object and the ground truth object. Biopsy adequacy, defined as a percentage representing the cortex area relative to the overall biopsy area, was used for evaluation.

Software development

We engineered an advanced web-based application utilizing the robust capabilities of JavaScript and the Django framework,[10] which are integral to its front-end and back-end architecture, respectively. On the front-end, the website offers a user-centric interface where users can upload high-resolution images of kidney biopsy samples. The tool features an advanced cropping function that enables precise isolation of the main biopsy area, enhancing the accuracy and efficiency of the analysis by focusing on the most relevant section of the image. The front-end was developed using JavaScript, which is a versatile and widely used programming language, essential for adding interactive elements to web pages, and is supported by all modern web browsers, making it a foundational tool for web development. Once the user completes the cropping, the image is seamlessly transferred to the back-end server, built on the Django framework. Django, a high-level Python web framework, is renowned for its rapid development capabilities and strong emphasis on security, offering a built-in admin panel, excellent documentation, and a scalable architecture that makes it suitable for a wide range of web applications.

The server employs the machine learning framework (described above) to meticulously analyze the kidney biopsy image. This analysis involves identifying and highlighting the cortex parts of the kidney, which are crucial for assessing various renal conditions. The back-end server's

algorithm is designed to accurately calculate the proportion of the cortex in the biopsy image. Following the analysis, the server generates a detailed output image that visually highlights the cortex areas for easy interpretation. Alongside the visual output, the tool provides a quantitative assessment, displaying a numerical value that indicates the percentage of cortex regions relative to the entire area of the biopsy image. This quantitative analysis is essential for medical professionals in evaluating the extent of certain kidney conditions, such as glomerular diseases, which may affect the cortex region.

Computing infrastructure

We used PyTorch (v1.9.0) and a NVIDIA 2080Ti graphics card with 11 GB memory on a GPU workstation to implement the model. The training speed was about 0.05 iterations/s, and it took 3 hours to reach convergence. The inference time for a standard biopsy image of size 1 Mb is within a minute.

Disclosures

None of the authors report any conflicts of interest.

Acknowledgments

This work was supported by the NIH SBIR phase 1 grant (R43-DK134273). V.B.K. 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.

References

1. Dhaun N, Bellamy CO, Cattran DC, *et al.* Utility of renal biopsy in the clinical management of renal disease. *Kidney International* 2014; **85:** 1039-1048.

2. Walker PD, Cavallo T, Bonsib SM, *et al.* Practice guidelines for the renal biopsy. *Mod Pathol* 2004; **17:** 1555-1563.

3. Ferrer G, Andeen NK, Lockridge J, *et al.* Kidney Biopsy Adequacy. *American Journal of Surgical Pathology* 2019; **43:** 84-92.

4. Nissen CJ, Moreno V, Davis VG, *et al.* Increasing Incidence of Inadequate Kidney Biopsy Samples Over Time: A 16-Year Retrospective Analysis From a Large National Renal Biopsy Laboratory. *Kidney International Reports* 2022; **7:** 251-258.

5. Esposito V, Mazzon G, Baiardi P, *et al.* Safety and adequacy of percutaneous kidney biopsy performed by nephrology trainees. *BMC Nephrology* 2018; **19**.

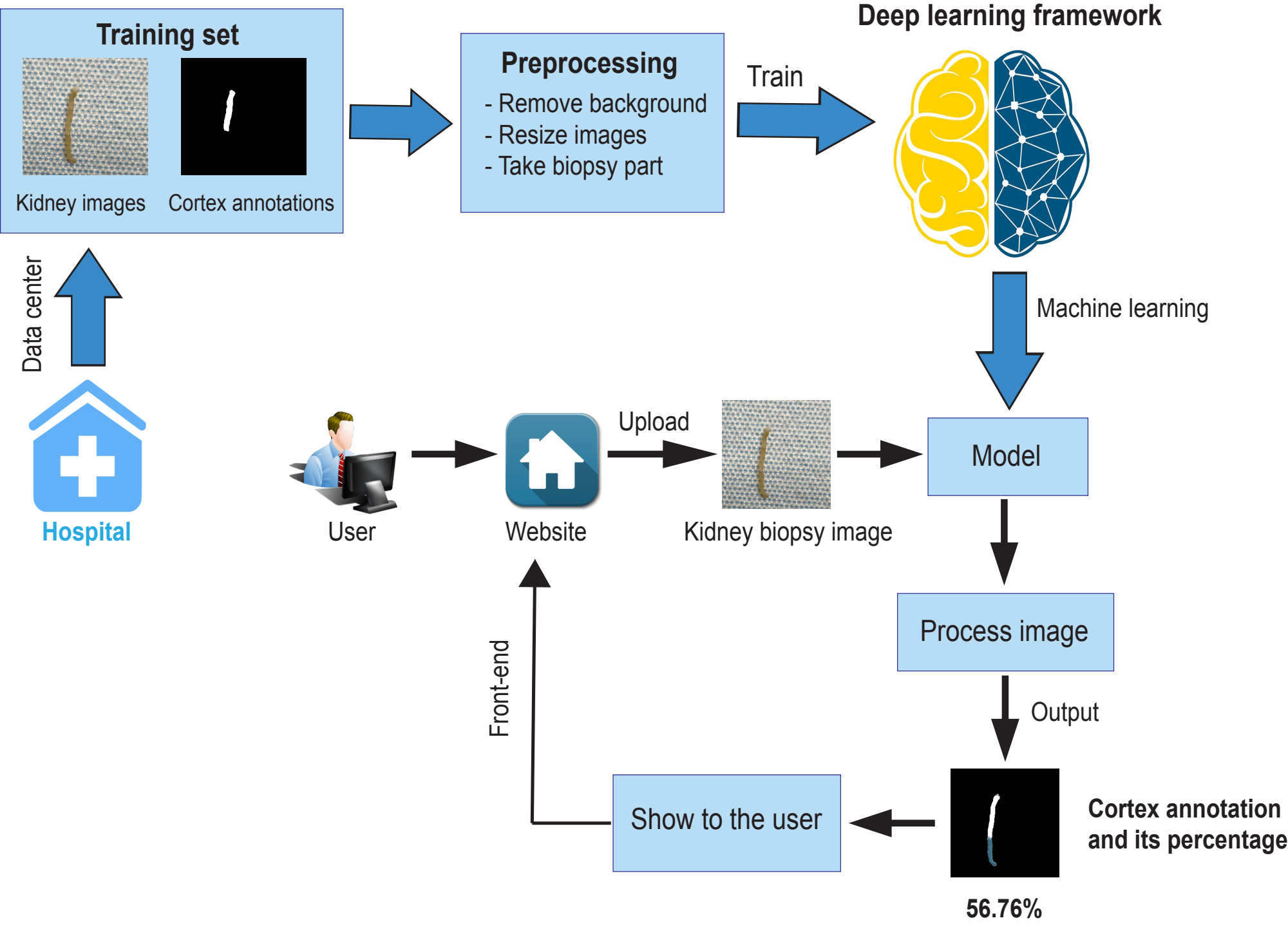
6. Sirithanaphol W, Incharoen N, Rompsaithong U, *et al.* Improvement of allograft kidney biopsy yield by using a handheld smartphone microscope as an on-site evaluation device. *Heliyon* 2021; **7**.

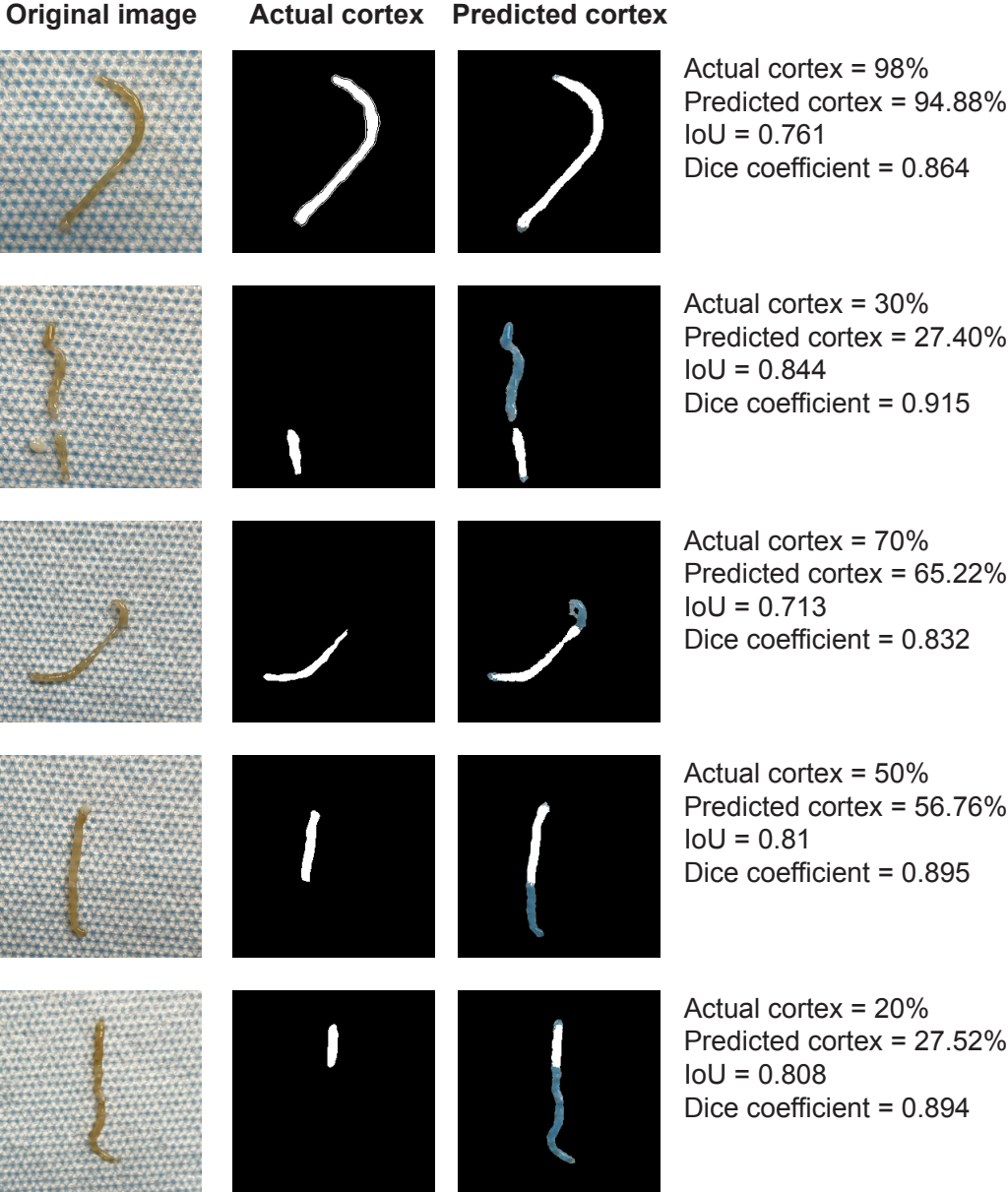
7. Singh G, Massak M, Czaplicki M, *et al.* Use of a Smartphone Camera at the Bedside to Assess Adequacy of Kidney Biopsies. *Journal of the American Society of Nephrology* 2021; **32:** 3024-3026.

8. Ma J, He Y, Li F, *et al.* Segment anything in medical images. *Nature Communications* 2024; **15**.

9. Kirillov A, Mintun E, Ravi N, *et al.*: Segment Anything. In *2023 IEEE/CVF International Conference on Computer Vision (ICCV)*, 2023, pp 3992-4003

10. Holovaty A, Kaplan-Moss J. *The Definitive Guide to Django*, 2009.


Training set
Kidney images
Cortex annotations
Preprocessing
- Remove background
- Resize images
- Take biopsy part
Train
Deep learning framework
Machine learning
Data center
Hospital
User
Website
Upload
Kidney biopsy image
Model
Process image
Output
Front-end
Show to the user
Cortex annotation and its percentage
56.76%


Original image
Actual cortex
Predicted cortex
Actual cortex = 98%
Predicted cortex = 94.88%
IoU = 0.761
Dice coefficient = 0.864
Actual cortex = 30%
Predicted cortex = 27.40%
IoU = 0.844
Dice coefficient = 0.915
Actual cortex = 70%
Predicted cortex = 65.22%
IoU = 0.713
Dice coefficient = 0.832
Actual cortex = 50%
Predicted cortex = 56.76%
IoU = 0.81
Dice coefficient = 0.895
Actual cortex = 20%
Predicted cortex = 27.52%
IoU = 0.808
Dice coefficient = 0.894