Supplementary Table of Contents

1. Supplementary term definitions

Glossary of terms	
Convolutional neural network (CNN)	A machine learning algorithm that automatically learns a chain of **convolutional filters** that match specific patterns in digital images.
Convolutional filter	A mathematical operation that compares the similarity of a pixel and its neighbors to a pre-specified template pattern, yielding greater values for stronger pattern matches.
Segmentation	The image analysis task of separating an object from the remainder of the image by identifying the pixels constituting its area.
DeepLab V2	A CNN which proposed the use of atrous convolutions (convolutions which take into account pixels beyond those directly neighboring the pixel of focus) and multi-resolution pooling (combining filtering results from multiple resolutions) to improve performance.
Encoder	The layers of the neural network that take the input image and produces a large set of features describing the image. This feature set can be used by the decoder to create a classification of every pixel in the image.
ResNet-50	A convolutional neural network model which proposed the use of residual connections, or the direct passing of information from earlier layers of a network to the later networks, in order to allow deeper, larger networks to be more efficiently trained.
Larger network architectures	CNN designs which use relatively greater numbers of convolutional filters than others designed to do the same task, typically by increasing the number of layers in the network.
Overfitting	Overfitting is when a supervised machine learning algorithm learns a model which is too specific to the data on which it was trained, thus not generalizing well to future data.
HSV	The HSV color space (Hue, Saturation, Value) is an alternative representation of the standard RGB (Red, Green, Blue) color model, which are both methods to represent digital images to a viewer. The HSV color space was designed to more closely reflect the way human vision perceives color making attributes.
Gaussian filter	An image filter whose output is a blurred version of the input image.
Thresholded	The image analysis task of assigning pixels in an image to the foreground or background based on their value, such that pixels below a certain value become the background and pixels above a certain value become the foreground.
Binary mask	A digital image which only contains only two possible values: 0 for the pixels identified as background, and 1 for pixels identified as foreground.
Sliding window technique	A method for chopping very large images into smaller tiles. First a fixed tile size is selected and placed at the upper left corner of the image to be chopped. The tile is slid across the rows and columns with

	a fixed step size, forming a grid over the large image. Each tile of the grid becomes an output image.
Image augmentation	A strategy to increase the amount of data that can be used to train a CNN when no new real data can be acquired. The goal is to make small manipulations to the real images in order to create slightly modified versions which can improve network learning. One example is flipping the images in the horizontal and vertical directions, which creates new perspectives of the input image without altering the underlying data structure.
Step	CNNs are trained iteratively. At each iteration, the CNN takes as input a certain number (i.e., batch size) of images and their labels, and makes internal adjustments to the parameters to optimally classify the input images. Each time the network makes an adjustment to the parameters is referred to as a training step.
Epoch	When the network has seen every unique training image one time.
Batch size	The number of training images the network uses to calculate parameter updates for each training iteration.
Initial learning rate	Learning rate refers to how large of updates are made to the network parameters at each iteration. Typical practice is to start with a larger learning rate and slowly decrease the learning rate value as the number of training steps increases. The learning rate that the network begins at is the initial learning rate.
Power	A specific learning rate policy which decreases the learning rate at each step according to the formula: $learning\ rate = initial\ learning\ rate * (1 - \frac{current\ step}{total\ steps})^{Power}$
Momentum	A technique to improve the convergence of CNN which helps to avoid converging on local minimum rather than global minima.
Weight decay	A strategy to train CNN's to be more generalizable. Weight decay penalizes the influence of each filter to prevent it from having too large of a contribution to the network output.
Initialized	CNN initialization is the task of setting the parameter values of the network before training begins. The parameters can be generated randomly or directly loaded from a set of parameters learned by another network with identical structure that was already trained for another prediction task (pretrained model).
Pretrained model	Repurposing parameter values learned by a network trained for a different image classification, to use as a starting point for new network learning. This greatly reduces the training time to convergence required by the network as opposed to randomly initializing the values.
Fully convolutional networks	A CNN which exclusively uses convolutions in its design, and directly produces a corresponding map of classification labels for every pixel in

	the input image. This allows the network to handle arbitrarily sized and shaped images at prediction time.
GPU	Graphical processing unit, also called a video card. A video card has a unique processing structure which is highly advantageous for training CNNs, greatly accelerating the speed at which they can be trained.
Ground truth	The "true" labels in an image classification task, for example, manual identification by a human expert. Serves as a foundation to compare computational performance against.
Class	The desired output targets of an image classification task (e.g., IFTA, glomerulus, sclerotic glomerulus).
Class imbalance	When one image class has many more potential training examples than another class. This proposes a challenge for neural networks because classes with greater number of samples are more likely to be selected in the training process and will disproportionally influence network learning, reducing performance for other classes.
Contingency matrix	A statistical table which displays the multivariate frequency distribution of two or more variables. In a binary classification sense, the contingency table is a 2x2 matrix with the four cells respectively representing the total number of true positives, false positives, true negatives, and false negatives.
Watershed algorithm	An algorithm that transforms a grayscale image by modeling it as a topographic map. The algorithm then discovers ridgelines in the topography which correspond to separations between different objects. Thus, the watershed algorithm is frequently used to separate overlapped or abutting objects in digital image processing tasks. One very common application of the watershed algorithm is the segmentation of overlapped histological nuclei.
Piecewise affine geometric distortion	A form of image augmentation where random regions of the image are warped to create small variation in the image structure.
LAB space	Another alternative color space, like HSV, used for displaying images, and is designed to approximate human vision. The L component of the image closely matches human perception of lightness (black vs white), the A component encodes green (-) to red (+), and the B encodes blue (-) to yellow (+).

2. *Collection of slides from each institution*

Supplementary Table 2. Number of biopsies from each institution used in each of our respective studies.

	substudy 1				*substudy 2*		*substudy 3*	
	Train	Test	External test	Type	*n*	Type	*n*	Type
I1	30	5	-	Post Tx	5	Post Tx (one year surveillance)	30	Post Tx (one year surveillance)
I2	22	4	-	23 DN, 3 control Nx	5	DN	15	DN early-advanced CKD
I3	10	3	-	DN	5	DN	31	DN early-advanced CKD
I4	-	-	20	DN	-	-	-	-
I5	10	3	-	DD	-	-	-	-
I6	7	2	-	DN	5	Post Tx indication biopsies <6mo	11	DN moderate-advanced CKD
Total	79	17	20		20		87	

I, Institution; Tx, transplant; DN, diabetic nephropathy, Nx, nephrectomy; DD, deceased donor; CKD, chronic kidney disease.

3. *Augmentation strategy*

Using the ImgAug python package, input images were flipped both vertically and horizontally (50% probability each), and locally distorted with random **piecewise affine geometric distortion** (scale 0.01 to 0.05). Images were also additionally shifted in the H component of the HSV space,[1] and gamma shifted in of the *L** component of the ***L*a*b**** space.[2] The first two techniques help the network to become invariant to orientation of structures, the third boosts invariance to structural deformation, and the last two help the network to be resilient against unseen stain variation. The color space shift amounts were selected from normal random distributions with $\mu = 0, \sigma = 0.025$ and $\mu = 1,\ \sigma = 0.05,$ respectively. When augmenting images, careful consideration must be made to the relative amount that each image class is augmented. CNNs typically perform poor on image classes where examples are low compared to the dataset size, and preferentially learn well on the examples that are plentiful. This was primarily only an issue for sclerotic glomeruli, but we applied an augmentation strategy scalable for all classes. First the number of input image patches containing each class type is tabulated. Second, an overall augmentation magnitude (α) is selected, and should be thought of as the maximum amount of times any single image can be augmented. The amount of images in each target class is converted to a fraction of the total image set, inverted, and scaled according to the augmentation magnitude. This forces more rare image classes to be preferentially augmented a higher number of times than more common classes, with a maximum number of augmentations for any one image equal to α. For our study, $\alpha = 6$.

References

1. Smith AR. Color gamut transform pairs. SIGGRAPH Comput Graph. 1978;12(3):12-9.
2. Jain AK. Fundamentals of digital image processing: Prentice-Hall, Inc.; 1989. 569 p.