Fully automated classification of glomerular lesions in lupus nephritis

Brandon Ginley, Kuang-Yu Jen, Avi Rosenberg, Giovanni Maria Rossi, Sanjay Jain, et al.

Brandon Ginley, Kuang-Yu Jen, Avi Rosenberg, Giovanni Maria Rossi, Sanjay Jain, Pinaki Sarder, "Fully automated classification of glomerular lesions in lupus nephritis," Proc. SPIE 11320, Medical Imaging 2020: Digital Pathology, 113200Y (16 March 2020); doi: 10.1117/12.2548528

SPIE.

Event: SPIE Medical Imaging, 2020, Houston, Texas, United States

Fully automated classification of glomerular lesions in lupus nephritis

Brandon Ginley,[1] Kuang-Yu Jen,[2] Avi Rosenberg,[3] Giovanni Maria Rossi,[3,4] Sanjay Jain,[5] and Pinaki Sarder[1,6,7,*]

Departments of [1]Pathology and Anatomical Sciences, [6]Biostatistics, and [7]Biomedical Engineering, University at Buffalo–The State University of New York, Buffalo, New York; [2]Department of Pathology and Laboratory Medicine, University of California, Davis Medical Center, Sacramento, California; [3]Department of Pathology, Johns Hopkins University School of Medicine, Baltimore, Maryland; [4]U.O. Nefrologia, Azienda Ospedaliero-Universitaria di Parma, Dipartimento di Medicinae Chirurgia, Università di Parma; [5]Division of Nephrology, Department of Medicine, Washington University School of Medicine, St. Louis, Missouri;

[*]Address all correspondence to: Pinaki Sarder
Tel: 716-829-2265; E-mail: pinakisa@buffalo.edu

ABSTRACT

Systemic lupus erythematosus is a disease in which the immune system attacks host tissues. One organ commonly attacked is the kidney, in which case the resultant acute and chronic damages are called lupus nephritis. The accumulated damage can result in renal failure. The percutaneous renal biopsy is invaluable to the assessment of the disease and its therapeutic response. A large portion of the pathological assessment is done by histological analysis of the biopsied tissue with light microscopy. Computational models can alleviate a portion of expert disagreement by providing unified, reproducible quantifications of digitized image structures. In this work, we perform fully automated whole slide segmentation of glomeruli from Periodic Acid-Schiff (PAS), hematoxylin and eosin, silver, and trichrome stained lupus nephritis biopsies. The automatically extracted PAS glomeruli are quantified by a set of 285 hand-crafted features designed specifically to target glomerular lesions in lupus nephritis. These features are fed in sequence to a recurrent neural network architecture which views multiple glomerular features from a single biopsy, and outputs a continuous diagnostic value representative of classes II-V of the scheme by Weening et al[1]. On 82 whole slide images taken from 65 patients, compared to renal pathologist annotations and using only the PAS stain, the network achieved a Cohen's kappa of 0.42 with 95% confidence interval [0.32, 0.52] to render the correct class chosen from II-V, and 0.56, 95% CI [0.43, 0.69] to render an additional class V diagnosis when required.

Key-words (10): digital diagnosis, digital pathology, lupus nephritis, glomerulus, feature extraction, disease classification, silver, trichrome, PAS, H&E

1. INTRODUCTION

Systemic lupus erythematosus (SLE) is an autoimmune disease in which a person's own immune system attacks and damages healthy tissues and organs. SLE can attack a wide range of places in the body such as joints, skin, muscles, bones, blood, the heart, the kidneys, and eyes. One particularly unfavorable organ to have involved is the kidney, in which case the resulting disease manifestation is termed lupus nephritis (LN). SLE has a prevalence in the US of 5-80 per 100,000, with differences depending on race and gender[2]. Among adults diagnosed with SLE, approximately 50% will develop LN[2]. If left unchecked, LN can progress to renal failure. Thankfully, however, because of early recognition and management of SLE, end stage renal failure due to LN occurs in less than 5% of cases[3, 4], and 10-year survival improves from 46% to 95% if disease remission is achieved[2]. The major diagnosis of LN is done by percutaneous renal biopsy; the observed histological changes within glomeruli help guide treatment of patients and are correlated with patient outcome. In typical clinics, the diagnosis is rendered based on a consensus classification system that incorporates information from several

Medical Imaging 2020: Digital Pathology, edited by John E. Tomaszewski, Aaron D. Ward, Proc. of SPIE Vol. 11320, 113200Y · © 2020 SPIE
CCC code: 1605-7422/20/$21 · doi: 10.1117/12.2548528

different histological stains, immunofluorescent interrogation, and possibly ultrastructural information from electron microscopy. Histological stains are used to highlight commonly affected disease compartments and structures, which are then interpreted by the on duty pathologist. Many parts of the examination are at least partially subjective, which limits the degree to which clinicians with can agree on the diagnostic findings. Further, some types of interpretation, such as the enumeration of the number of nuclei in a particular region of interest, is exhausting, tedious, and repetitive, increasing the likelihood of errors. Computational algorithms, on the other hand, do not fatigue and can be made completely objective. A common ground of objective quantifications can help improve inter-rater agreement, increase quality of diagnostic information, and expedite research discoveries.

Typical clinics now favor the use of whole slide images (WSIs) to assess histologically stained renal biopsies. A WSI is a digitized image of an entire microscopy slide, captured in many small acquisition patches stitched together, and rendered at multiple resolutions for viewing ease. In our past work, we developed and released a complete pipeline for classifying Periodic Acid-Schiff (PAS) WSIs of diabetic nephropathy (DN) biopsies according to their glomerular characteristics[5]. Our digital DN classification agreed with three renal pathologists with moderate to substantial Cohen's kappa (on par with the inter-rater agreement of the pathologists). We wanted to explore if the algorithms which were derived in that work could be repurposed to classify glomeruli from LN biopsies. As a pilot study to answer this question, we first fine-tuned our pre-existing glomerular segmentation models to detect glomeruli from the four typical histological stains used in LN assessment with high accuracy. The glomeruli from the PAS stain (total n = 82 WSIs taken from 59 LN patients and 6 control tissues) were further compartmentalized and structure quantified via 285 individual glomerular features. The glomerular features in this work are an expansion of a past published feature set by adding information on the radial and angular location of detected glomerular components with respect to the estimated glomerular centroid. We found that using this technique we are able to classify PAS stained LN WSIs into classes II-V with Cohen's kappa 0.42 and 95% confidence interval [0.32, 0.52]. The network also decides if a concurrent diagnosis of class V should be rendered along with the first diagnosis, for which the network agreed with renal pathologist annotations with kappa 0.56 and 95% confidence interval [0.43, 0.69]. The capability of the network to achieve fair to moderate agreement using only a fraction of what the true diagnosis is based on (i.e., up to four histological stains, immunofluorescence, and electron microscopy) is interesting and warrants further investigation. Future work will expand glomerular structure quantification to support measurement of structure from silver, trichrome, hematoxylin and eosin (H&E), immunofluorescence (IF), and electron microscopy.

2. METHODS

2.1 Image data

Biopsy samples from human LN patients were acquired from two different sources. The first source of LN data, from which 19 samples were taken, is the archival records at the University of California at Davis Medical Center. The second source of LN data, from which 40 samples were taken, was the Johns Hopkins Renal Pathology archival records. Control samples consisted of nephrectomy tissues from renal cell carcinoma cases, cut at a level where histological changes were indistinguishable from normal tissue. These tissues were collected and validated by the Kidney Translational Research Center at Washington University School of Medicine directed by co-author Dr. Jain. Human data collection procedure followed a protocol approved by the Institutional Review Board at University at Buffalo.

2.2 Imaging and data preparation

WSIs of biopsied LN renal tissue were collected, sliced at 2 µm, and stained with PAS at their corresponding source institution. The images were also captured with the whole slide scanners resident to the corresponding source institutions. Control data were stained and scanned at the University at Buffalo. All images were captured at 40x apparent magnification resulting in 0.25 µm / pixel resolution.

2.3 Ground-truth annotation

Ground-truth annotation of LN disease class was performed by co-authors Dr. Jen and Dr. Rosenberg according to the 2003 ISN/ RPS classification[1].

2.4 Glomerular segmentation

Glomerular segmentation was performed by fine-tuning our previously published network models for segmentation of diabetic glomeruli using the DeepLab V2[6, 7] semantic segmentation network[5]. Specifically, we used our openly available pipeline for whole-slide semantic segmentation, HAIL (human artificial intelligence loop)[8], to manage the fine-tuning. HAIL automates the training of semantic segmentation models from WSIs and also projects network predictions back onto

the WSI. The previous models were trained only on the PAS stain, however, current clinical diagnosis of LN relies on up to three additional histological stains, possibly also with immunofluorescent and ultramicroscopic images. Therefore, for this study, we expanded our segmentation model to detect glomeruli from LN biopsies stained with PAS, H&E, trichrome, and PAS-silver stains. We fine-tuned our segmentation model on 80 WSIs in total, 20 selected randomly from each of the four stains. Before the quantification and classification stages, any minor errors in glomerular boundary prediction are corrected directly on the WSI, to maximize the precision of glomerular feature quantification.

2.5 Glomerular component analysis

To simplify computational detection of PAS glomerular structure in widely varying phenotypes, we reduced glomerular structure to three components based on their appearance in a PAS stained tissue: 1) nuclei, 2) PAS-positive components consisting of mesangium, glomerular basement membranes, and Bowman capsule, and 3) a luminal component consisting of Bowman space and capillary lumina. Nuclei are identified using a separate trained model of the DeepLab V2 semantic segmentation network published in the same previous work[5]. A two-step process is used to separate PAS-positive and luminal components. First, PAS candidate pixels are roughly estimated via automatic thresholding[9, 10] of the color-deconvolution[11] output for PAS stain. Luminal candidate pixels are selected via automated thresholding of the lightness component of the *L*a*b** color space. The two sets of candidate pixels are used to train a naive Bayesian classifier which can predict the classification of the remaining unlabeled pixels based on their RGB value.

2.6 Glomerular feature definition

Description of all 285 glomerular features used in this work is intractable, therefore, we defer to our previously published manuscript which has extensive description of 232 of these features[5], which were initially conceptualized for use in classification of DN biopsies. Briefly, these features measure the following characteristics: the mean and standard deviation of the red, green, and blue channels of the three glomerular components; the textural energy, homogeneity, contrast, and correlation of the three glomerular components; the average distance between glomerular objects and themselves, the glomerular boundary, as well as distance from dissimilar compartments; morphological measurements of each glomerular component; distance transform measurements of object thickness; count of glomerular component detected objects. These original features were upgraded with 53 new features specifically designed to target lesions associated with LN. The added features are a histogram of coordinate locations of identified glomerular components. The coordinates are referenced from a polar plane centered on the estimated glomerular centroid location. For PAS+, luminal, and nuclear components, histograms of the radial coordinates were measured with bin edges starting at 0, incrementing by 100 until 1000, with an additional right edge bin at 1300 to capture outlying pixels. For the nuclear component only, a histogram of the angular coordinates was taken on 20 even intervals between $-\pi$ and π. These features were selected to help the classification network understand the concept of segmental and diffuse lesion presentations, and the spatial distribution of lesions.

2.7 Recurrent biopsy classification

Our classification technique involves a custom designed RNN which accepts as input a sequence of glomerular features extracted from a single biopsy, and outputs two continuous diagnostic classification values. The network architecture accepts an input vector of 285 features, which is immediately connected to a dense layer with 285 outgoing units. This dense layer is connected to a long short term memory unit (LSTM[12]) with 50 hidden units and 10 time steps. This recurrent layer is connected to a second recurrent layer with 10 hidden units and 10 time steps, with 50% dropout between the recurrent layers. The second LSTM outputs to a dense layer with two outgoing units. One of the outgoing units is optimized to minimize the L_1 loss of the predicted value from the diagnostic class converted to a number between 0 and 6. The other outgoing unit is optimized to minimize L_1 loss of the predicted absence (valued 0) or presence (valued 1) of class V. An extended description is available in[5]. To assess classifier performance, the average of 10 cross validation folds was used.

3. RESULTS

3.1 Glomerular segmentation

The most widely used schemes for LN diagnosis rely on glomerular findings, and thus, automation of glomerular detection is a high priority. We have re-trained our convolutional neural network (CNN) model for DN glomerulus segmentation and trained it to segment LN glomeruli from four unique stains: H&E, PAS, trichrome, and silver/PAS-silver. The biggest noticeable flaws in the semantic segmentation were inability to split conjoined glomeruli (glomeruli abutting in the image space) and a lowered detection rate of glomerulosclerosis in H&E sections (due to poor contrast with the surrounding tissue, making even detection by eye difficult in some cases). At the end of one epoch of training, however, the network models reached a high level of segmentation accuracy on all of the individual stain types, as can be seen from the holdout-

data predictions shown in Figure 1. The level of segmentation accuracy was deemed sufficient based on the amount of manual labor required to "perfect" the glomerular boundaries before the quantification step. The holdout predictions in Figure 1 demonstrate fewer than 3 false positive glomeruli on all four items, and less than 3 glomerular boundaries which were insufficient for quantification (all were under-segmented regions where the full boundary was not captured). Thus, this minimal amount of error in network prediction is trivial to correct on the WSI so long as the number of cases is manageable (this manual curation strategy would poorly scale to a dataset of 100,000 patients, for example).

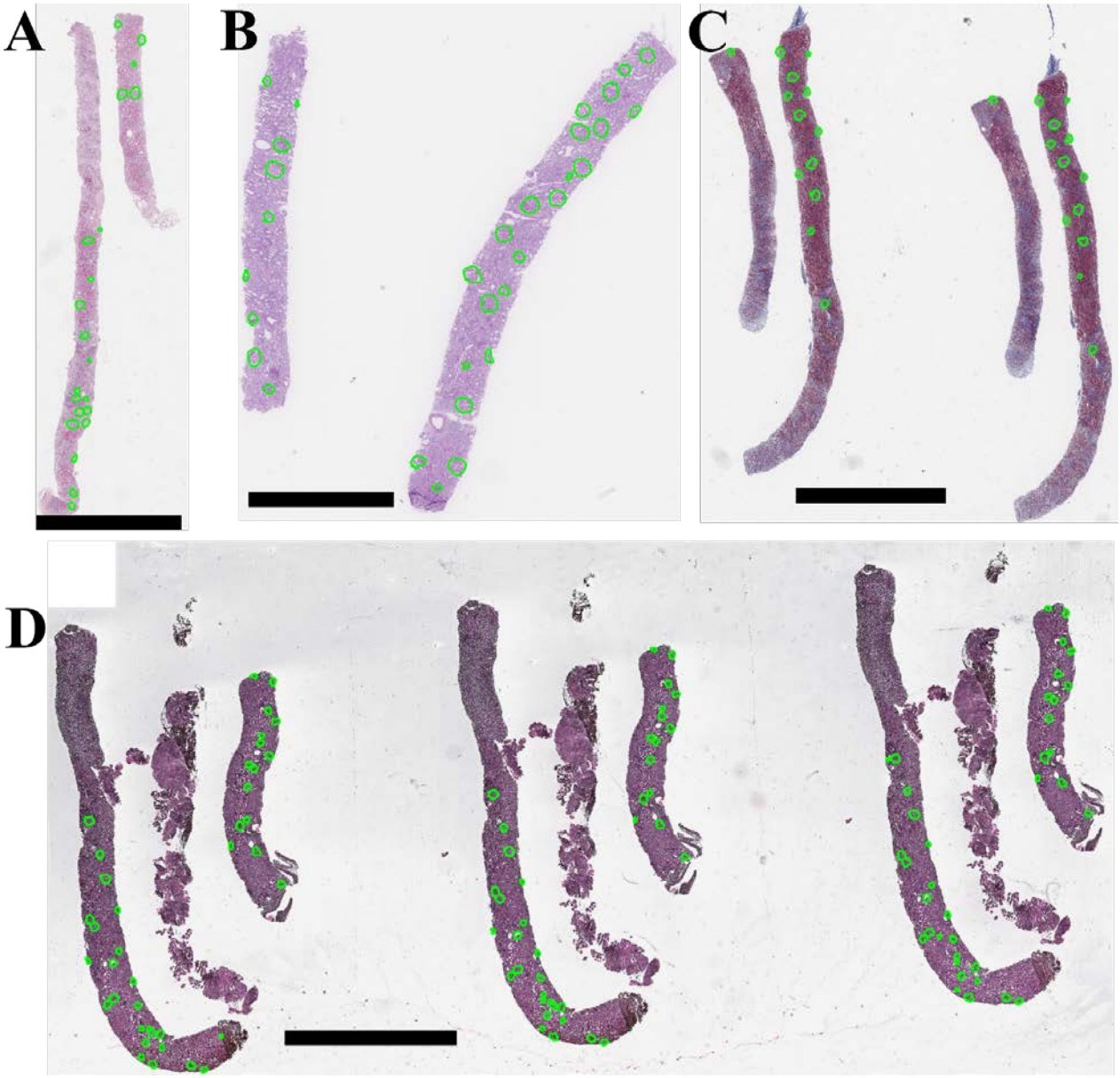

Figure 1. Automated segmentation of glomeruli in LN biopsies. A) H&E stained biopsy. B) PAS stained biopsy. C) Trichrome stained biopsy. D) PAS-silver stained biopsy.

3.2 Glomerular component analysis

To simplify the computational detection of PAS glomerular structure in extensively varying phenotypes, we reduced glomerular structure to three components: PAS-stained structures (basement membranes, mesangium, capillary loops), the unstained structures (capillary lumina and Bowman space), and hematoxylin stained structures (nuclei). We have shown

in our past publication that this reduction of glomerular structure facilitates complete detection of glomerular structures in a widely varying spectrum of glomerular phenotypes and lesions, yet still retains classification power. An example of glomerular component detection is shown in Figure 2.

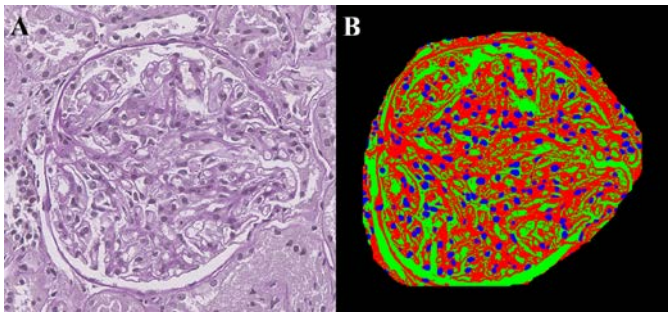

Figure 2. Detection of glomerular components. A) Original PAS glomerulus. B) Segmented glomerulus image showing luminal component in green, PAS+ component in red, and nuclei in blue.

3.3 Glomerular feature definition

We expanded the 232 features published in our past work[5] with 53 additional features, for total 285. Previous features measure color, texture, morphology, intra-compartment distance, inter-compartment distance, and compartment containment for each of the respective glomerular components. The new additional features quantified the approximate distribution of angular and radial coordinates of detected components referencing the estimated glomerular centroid. This is accomplished by first extracting detected glomerular component pixel locations and converting them to a polar coordinate system based on the estimated glomerular centroid. Then, the distribution of the glomerular components in polar coordinates is measured by taking even histogram bins across the full range of potential angular and radial possibilities. These features were selected so that the classifying network will be able to quantify differences between segmental and diffuse lesions. An example of how the polar features might be important can be understood with an example, in the context of crescentic lesions. Crescentic lesions would demonstrate an increased density of nuclear and PAS pixels at far radial coordinates and within a restricted set of angular coordinates. Similarly, these features would help to localize other lesion forms such as segmental and global sclerosis.

3.4 Recurrent biopsy classification

Our previously published recurrent glomerular classification pipeline[5] utilized a recurrent neural network (RNN) that takes as input the hand-crafted glomerular features from a single biopsy and provides as output a single continuous number reflective of the diagnostic classification. This was a good strategy for DN, where the class stages are continuous. In LN, however, the diagnosis is not just a single value. In LN, there are six classes, i.e., I, II, III, IV, V, and VI. However, class V can be diagnosed on its own or can be superimposed on top of another class (for example, one patient's diagnosis could be IV + V). Because of this difference, we altered our RNN to provide two diagnostic values. The first value can have range between 0 and 6, with 0 representing control cases, 1 representing LN class I, 2 representing LN class II, etc. The second output can take on values between 0 and 1, 0 representing the absence of a concurrent class V diagnosis and 1 representing presence. We trained this new RNN using a 10-fold cross validation split on 82 WSIs from 65 patients. Of these patients, 6 were control, 21 had a diagnosis of class II, 17 class III, 12 class IV, and 9 had an exclusive diagnosis of class V. Within the 65 patients, 9 had a diagnosis of class V superimposed on another class. For classification of biopsies according to the primary classification of class II-V, the network agreed with annotations made by two renal pathologists with Cohen's kappa 0.42 and 95% confidence interval [0.32, 0.52]. With regards to identifying a secondary diagnosis of class V, the network agreed with renal pathologist annotations with kappa 0.56 and 95% confidence interval [0.43, 0.69]. This level of performance may seem on the lower end of the agreement scale, however, it is important to keep in mind that the original diagnosis (ground truth) was rendered off multiple stain types and immunofluorescent findings, not the PAS stain alone.

4. DISCUSSION

We show fully automated detection of glomeruli from multiple histological stains commonly used to assess LN. CNN's yield near adequate performance for quantification studies, at least on a restricted set of regions of interest which can be captured with high performance. However, having data from only around 60 unique patients, the network was unable to generalize to perfectly detect glomeruli on all holdout cases. It is possible that increasing the training data set size by several hundred patients may eliminate many of the prediction errors encountered in this work. However, as it stands, this method would not be applicable to thousands of cases for quantification, because it would be infeasible to check every case for small errors. A secondary network trained to detect outputs which are artifacted/imperfect may be a more trivial solution. We further show that a computational approach is capable of diagnosing glomeruli from LN using only information contained in a PAS stain. Only the PAS-stained glomeruli were used for digital classification, because the

digital glomerular features we designed were created specifically for glomerular structure in the PAS stain, and it will be a very extensive work to expand the feature list to multiple stain types. Thus, for this pilot study, we aimed to assess the predictive value of the PAS stain alone. In future works, additional glomerular stains will be included in classification. We expect that the addition of further structural modalities of information will greatly boost the classifier's ability to provide the correct diagnosis.

5. CONCLUSION & FUTURE WORK

Additional further works should look toward moving beyond traditional diagnostic and prognostic schemes, and use progression-driven data analysis to identify unique sub-groups of patients which have high risk or therapeutic response. Computational analytics and modern artificial intelligence will play a critical role in the extraction of high volume datasets and the investigation of clinical data at massive scale.

6. ACKNOWLEDGMENT

This work was supported by the faculty startup funds from the Jacobs School of Medicine and Biomedical Sciences, University at Buffalo; Buffalo Blue Sky grant, University at Buffalo; NIDDK CKD Biomarker Consortium grant U01 DK103225; and NIDDK Kidney Precision Medicine Project grant U2C DK114886. The authors would like to acknowledge Felicia Yen and Diane Salmon for their assistance in selecting cases for analysis in this work. The authors would also like to acknowledge the support of the histology and imaging cores from University of California at Davis, Johns Hopkins University, Washington University in St. Louis, and the University at Buffalo. We thank NVIDIA Corporation for the donation of the Titan X Pascal GPU used for this research (NVIDIA, Santa Clara, CA).

REFERENCES

1. Weening, J.J., V.D. D'Agati, M.M. Schwartz, S.V. Seshan, C.E. Alpers, G.B. Appel, et al., *The classification of glomerulonephritis in systemic lupus erythematosus revisited.* Kidney Int, 2004. **65**(2): p. 521-30.
2. Almaani, S., A. Meara and B.H. Rovin, *Update on Lupus Nephritis.* Clin J Am Soc Nephrol, 2017. **12**(5): p. 825-835.
3. Ward, M.M., *Changes in the incidence of end-stage renal disease due to lupus nephritis, 1982-1995.* Arch Intern Med, 2000. **160**(20): p. 3136-40.
4. Somers, E.C., W. Marder, P. Cagnoli, E.E. Lewis, P. DeGuire, C. Gordon, et al., *Population-based incidence and prevalence of systemic lupus erythematosus: the Michigan Lupus Epidemiology and Surveillance program.* Arthritis Rheumatol, 2014. **66**(2): p. 369-78.
5. Ginley, B., B. Lutnick, K.Y. Jen, A.B. Fogo, S. Jain, A. Rosenberg, et al., *Computational Segmentation and Classification of Diabetic Glomerulosclerosis.* J Am Soc Nephrol, 2019.
6. Chen, L.C., G. Papandreou, I. Kokkinos, K. Murphy and A.L. Yuille, *DeepLab: Semantic Image Segmentation with Deep Convolutional Nets, Atrous Convolution, and Fully Connected CRFs.* Ieee Transactions on Pattern Analysis and Machine Intelligence, 2018. **40**(4): p. 834-848.
7. Wang, Z. and S. Ji, *Smoothed Dilated Convolutions for Improved Dense Prediction*, in *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*. 2018, ACM: London, United Kingdom. p. 2486-2495.
8. Lutnick, B., B. Ginley, D. Govind, S.D. McGarry, P.S. LaViolette, R. Yacoub, et al., *An integrated iterative annotation technique for easing neural network training in medical image analysis.* Nature Machine Intelligence, 2019. **1**(2): p. 112-119.
9. Otsu, N., *A Threshold Selection Method from Gray-Level Histograms.* IEEE Transactions on Systems, 1976. **9**(1).
10. Bradley, D. and G. Roth, *Adaptive Thresholding using the Integral Image.* Journal of Graphics Tools, 2007. **12**(2): p. 13-21.
11. Ruifrok, A.C. and D.A. Johnston, *Quantification of histochemical staining by color deconvolution.* Analytical and Quantitative Cytology and Histology, 2001. **23**(4): p. 291-299.

12. Hochreiter, S. and J. Schmidhuber, *Long short-term memory*. Neural Computation, 1997. **9**(8): p. 1735-1780.