

Contents

Supplemental Methods.....	2
Supplemental Results	3
Supplemental Figure 1. Glomerular component maps for murine glomeruli.	4
Supplemental Figure 2. Deviation of network predictions as a function of dropped features.	5
Supplemental Figure 3. Memory as a function of run time for application of our algorithm in select cases.	6
Supplemental Table 1. Feature list and associated network impact scores.....	7
Supplemental Table 2. Linear weighted Cohen’s kappa and confidence limits for mouse experiments.....	4
Supplemental Table 3. Conditional probabilities for class assignment in mouse experiments given the ground truth class assignment.....	4

SUPPLEMENTARY METHODS

Glomerular feature extraction

Feature extraction is a very important component of image classification, and the two main pathways for feature generation are human-defined (hand-crafted) or computationally defined (via feature learning). Deep learning is a highly successful feature learning protocol which selects features that maximize classification accuracy. However, such features are poorly interpretable by humans, and their acceptability for clinical use is typically dictated by their intelligibility.¹ This is true because without intelligibility it is not knowable whether a model is making a decision because it has learned the proper problem representation or if it is simply exploiting artifacts of the data. To combine the strengths of both paradigms, we designed a hand-crafted set of features which target the pathological progression of glomerular structure in DN, and allowed a neural network to learn a non-linear representation of these features which determines the final diagnosis. A full list of exact feature names, precisely defined, in addition with their weighted importance to our neural network decisions from the Tervaert patient-level holdout analysis is available in supplementary Table 1 at the end of this document.

All glomerular feature extraction was done in MATLAB (MathWorks®, Natick, MA). Textural features were quantified once each for the classes of nuclei, PAS+, and lumina. The pixels contained in each different component segmentation were transformed from RGB to grayscale, then converted to a respective gray-level co-occurrence matrix. Entropy, energy, correlation, and homogeneity were extracted from the respective matrices².

Morphological features were calculated per individual object (such as each nucleus) within each component. These features included the mean, median, and mode of object area, and average convexity for identified compartmental objects. Summary statistics were computed along the glomerulus dimension (e.g., if a glomerulus contained 20 nuclei, the area of each nucleus was measured, and the mean, median, and mode of the respective set of 20 nuclei was computed).

Compartmental containment features defined the amount of one component object contained within the boundaries of another. For example, one measurement might be the measured amount of nuclear area contained within a detected glomerular lobule (a PAS+ object). Specifically, it is a ratio, where one part is the convex area of the containing component object, and the other part is the area of the contained component. This feature was calculated for every object in the PAS+ component and luminal component. However, it is not logical for nuclei to contain any other structures, so, a different form of containment was measured. Specifically, we measured the amount of pixels just outside the nuclear boundary which were classified as either luminal or PAS+, and took this amount and divided it by the total nuclear perimeter.

Inter-compartmental distance features were comprised of averaged distances between compartments and other identical glomerular compartments, or glomerular landmarks. Glomerular landmarks included the estimated glomerular centroid and the estimated glomerular boundary points. The following distances were calculated for each object of each glomerulus: 1) the pairwise distance between that object's centroid and identically labeled object's centroids, 2) the pairwise distance of object centroids from glomerular boundary, and 3) the pairwise distance between object centroids and the glomerular centroid. From these lists of distances, averages, minimums, and maximums were taken to create features.

Intra-compartmental distance features were quantified on distance transforms of component segmentations. They were another measurement on the morphology of each component. In particular, they were originally designed to target glomerular thickening in PAS+, but also are a general morphological feature for other structures. For each component, in addition to the whole glomerulus treated as one component, the distance transform was measured. The PAS+ distance transform was performed on the *PAS+ precursor mask*, because it more accurately reflected mesangium than the final mask, and we wanted to design a feature which was targeting mesangial expansion as directly as we could. Luminal distance transform was performed on the final luminal component mask, as was the nuclear distance transform. The glomerular distance transform was measured on the whole glomerular region, and was inverted, so that there was minimum value at the middle and maximum value at the glomerular edge. This was done so that the classification algorithm had a set of features that described potential spatial location information associated with each glomerular region. Histogram values were computed on each distance transform to create the final features, and the specific ranges of the histogram binning depended on the component of interest. For the PAS+ component, the bins went from 1 to 80 by steps of 2, with a wide final right bin edge of 2000. We denote this as [1:2:80,2000]. For the luminal component, the bins were specified as [1:1:60,2000]. For the nuclear component, [1:1:20,2000]. For the glomerular component, we first inverted the transform so that each pixel of the distance transform was describing the distance to the glomerular centroid. Then, on this component, the bins were specified as [2:25:600,20000]. These values were specifically selected to reflect the full range of values present in the distance transforms, and they would only be valid for other experiments if the image resolution was identical, (0.25 μm per pixel). For the PAS+ distance

transform only, we hand crafted some additional features, by thresholding the distance transform map at specific values (10 and 20), creating binary objects of those value ranges, and measuring morphological statistics on them (such as sum total, max, mean, and median of binary areas).

Watershed splitting

The DeepLab V2 network is a semantic segmentation network, which means classification is done at the pixel level. When pixels of two distinct objects with identical class are located next to each other in an image (such as nuclei, or glomeruli), the network may fail to identify the complete boundary between them, or worse, there may actually be no boundary between them. Therefore, we needed to split clustered nuclei which were output by the DeepLab V2 network. To do this, we used a distance transform-based³⁻⁵ marker controlled watershed.^{6, 7} Watershed has always been a go-to method for nuclear splitting, and has had significant success in histopathology as well.⁶ In our application of marker-controlled watershed, markers were estimated using an extended maximum transform. For this transform, the H-transform value was specified as 5% of the maximum value contained in a Gaussian-smoothed ($\sigma = 1,5 \times 5$) distance transform performed on the entire glomerular image. Gaussian smoothing was performed to smooth the distance transform values which resulted in improved detection of unique peak regions. Next, regions containing only one detected peak (single nuclei) were removed from the analysis, to reduce the amount of over-segmentations performed by the watershed algorithm. A distance transform of the remaining nuclei was taken. The skeleton by influence zones⁸ was calculated as the watershed ridgelines of this distance transform, and was imposed as a minima on the distance transform, along with the estimated peaks from before. The resultant watershed-split nuclei were recombined with the single nuclei, and morphologically opened with a disk of 1 pixel for smoothing to form the final detection image. Small regions less than 10 pixels in area were also removed.

STZ mouse model generation and analysis

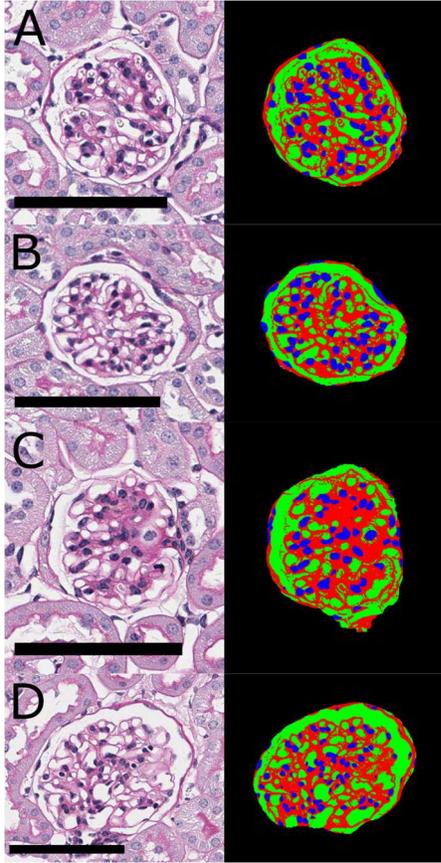
We chose to use a standard streptozocin (STZ) treated mouse model⁹ to develop murine renal tissue which reflected DN. C57BL/6J background mice (7 weeks old) were injected with STZ (50 mg/kg) for five consecutive days. They develop a mild form of diabetes mellitus type I, and after 25 weeks, mild-moderate DN. This strain of mouse was selected because they have kidneys which are resistant to the cytotoxic effects of STZ, allowing for more accurate depiction of DN related damages.¹⁰ Further, this strain of mice does not typically progress to severe disease phenotypes, a property we desired because we wanted a model which resembled the mild changes in early human disease stages where data is most scarce. Mice were sacrificed 15, 20, or 25 weeks after STZ treatment ($n = 5$ at each time point). Mice treated with STZ vehicle were used as control ($n = 3, 3, \text{ and } 4$, for mice sacrificed 15, 20, and 25 weeks after the STZ vehicle injection, respectively). After the experiment was performed, the animals were divided into three groups based on their fasting blood glucose levels, namely, control, mild DM, and moderate DM. Control mice were those who did not receive STZ treatment. Mild DM mice were mice that showed elevated fasting blood glucose >200 mg/dL which was not persistent when checked weekly 10 weeks post injection. Moderate DM mice were mice that showed elevated fasting blood glucose >200 mg/dL persistently when checked weekly starting 10 weeks post injection. Tissue sections were cut at 2 μm thickness, stained with periodic acid-Schiff (PAS) using hematoxylin counterstain, and brightfield imaged at 0.25 $\mu\text{m}/\text{pixel}$ resolution using a whole slide scanner (Aperio[®], Leica). All data was sourced and prepared within the laboratory of co-author Dr. Yacoub.

In regards to computation, a similar protocol was used as in the human analysis. Murine glomeruli were detected using our HAIL pipeline, and small errors in detection were corrected before feature extraction. Despite our human nuclear detection network having never been trained specifically on murine nuclei, the network was able to detect murine nuclei with surprisingly low error (supplementary Fig. 1). To detect PAS+ regions for murine PAS+ precursor masks, we used adaptive thresholding of the PAS+ stain deconvolution output. This was necessary because murine glomeruli are not as large as human ones, and thus, there was significantly less stained material, which reduced the effectiveness of the HSV transform method we used for human analysis. A reduced set of histogram bins was used to measure distance transform features (because the distances are smaller). The bin edges for the PAS+, luminal, nuclear, and glomerular compartments were [1:1:40,200], [1:1:40,200], [1:1:20,200], [2:10:300,2000], respectively. All remaining analysis was identical to the human protocols.

SUPPLEMENTARY RESULTS

Classification of murine STZ structure

To test the power of our computational method beyond strictly defined image classes, we applied our method to murine glomeruli from a mouse model of STZ-induced diabetes mellitus. In total, we classified $n = 25$ mouse kidneys by their DM severity, according to the classes control, mild, and moderate as defined in supplementary methods. In general,



Supplementary Figure 1. Glomerular component maps for murine glomeruli. A, B) Control murine glomeruli. C, D) Diabetic murine glomeruli. Scale bars indicate 100 μm .

our method required little modification to be applied to murine glomeruli. For PAS+ detection, we used color deconvolution instead of the HSV method used for human analysis. This was due to less PAS+ stain in murine structures, resulting in less saturation and darkness of stain. Interestingly, our nuclear segmentation network was capable of accurately identifying mouse nuclei without having been trained on any murine nuclei. This was likely possible for two reasons, the first of which is that the smallest human nuclei overlapped in size with mouse nuclei. Secondly, nuclei in mouse tissue sections had significantly reduced variability in presentation. This was because it is possible to maximally control mouse tissue image preparation, by discarding red blood cells via perfusion before extracting mouse kidneys, controlling section thickness to be 2 μm , and using the same core lab for tissue staining. Glomerular component maps for several murine glomeruli are shown in supplementary Fig. 1.

As in our main manuscript, we tested classification of mouse tissues using a baseline classifier selecting class randomly based on the label distributions and our RNN using mouse-level holdout as well as section-level holdout. Each of the 25 mice had a total of three sections cut, yielding a total of 75 WSI. The only classifier which achieved a significant level of agreement was the RNN classifier with split section holdout (supplementary Table 2). However, this task was much harder, as the classifiers were attempting to predict the level of diabetes rather than the amount of renal structural damage. The hardest class to identify was the moderate DM class, as can be observed in supplementary Table 3.

Supplementary Table 2. Linear weighted Cohen’s kappa and confidence limits for mouse experiments.

Comparison	Case number	Observed kappa	Lower 0.95 limit	Upper 0.95 limit
Mouse-level holdout	25	0.13	0.04	0.21
Section-level holdout	75	0.62	0.58	0.66
Baseline classifier	25	<0	n/a	n/a

Supplementary Table 3. Conditional probabilities for class assignment in mouse experiments given the ground truth class assignment.

Comparison	Case distribution	Control	Mild	Moderate
Mouse level holdout	10/8/7	0.41, [0.31, 0.51]	0.55, [0.43, 0.66]	0.09, [0.03, 0.18]
Section level holdout	30/24/21	0.82, [0.78, 0.86]	0.83, [0.78, 0.87]	0.42, [0.36, 0.49]
Baseline	10/8/7	0.38, [0.28, 0.48]	0.31, [0.21, 0.43]	0.34, [0.23, 0.47]

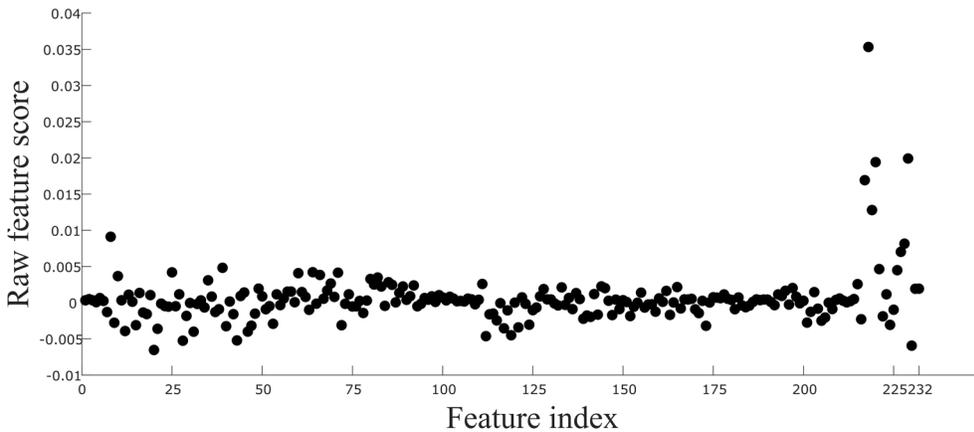
Glomerular component analysis and motivations

Several glomerular compartments can be difficult to distinguish from each other given that the only histological markers are PAS and hematoxylin. What is even more difficult is finitely selecting compartment boundaries which can be identified in all disease phenotypes. One such example might be the decoupling of glomerular basement membrane from mesangium, a task which can even be difficult in some samples via manual discrimination. However, we wondered whether this sort of precision was absolutely necessary for computational estimation of disease class.

Therefore, we explored simplification of the glomerular structure into three components. The identification of the three components can be distilled into three compressed steps: 1) predict nuclear locations with the DeepLab V2 network, 2) identify sparse locations of PAS+ objects and luminal objects using image thresholding, and 3) use naive Bayesian classification to refine the thresholding result in step 2. We chose to use a neural network to identify nuclei while using an unsupervised approach to identify the remaining glomerular components because of the difference in difficulty in detection and annotation of different glomerular compartments. At first, we wanted to use an unsupervised technique for all compartments, however, it soon became clear that human nuclear diversity in disease course is far too variable for typical techniques such as color deconvolution¹¹ to achieve high accuracy. Conveniently, however, glomerular nuclear annotations, though tedious to acquire, can easily be distinguished by untrained eye in most cases, and typically do not require domain experts for identification. On the other hand, other structures, such as the mesangium and membranous layers, can be very difficult to distinguish from each other by eye because the stain does not provide high contrast between these boundaries. Further, these structures have more ambiguous presentation in images than nuclei, requiring an expert eye. This makes annotation of these compartments on PAS-stained histological images very difficult, but identifying them as one contiguous component is much easier. For a similar reason, we grouped capillary lumina into the same class as the Bowman space because they are both similar in color profile, but the separation of them based on morphological properties becomes increasingly difficult depending on the disease state and section cut direction. In less severe phenotypes (Fig. 3A, 3C, 3D) it can be very obvious which regions are capillary lumina and which are Bowman regions, however, it is much more ambiguous in more severe disease phenotypes (Fig. 3B, 3D, 3F). This decision is further obfuscated by tangentially cut white regions which can be out-of-plane Bowman space folds, capillary lumen, or empty gaps embedded in fibrotic structures.

Feature relevance

Supplementary Fig. 2 demonstrates a graphical view of the raw values provided in supplementary Table 1 (feature ranking according to Tervaert classification). Values greater than zero indicate that the respective feature index (seen along the x-axis) was useful for the average diagnostic decision of the entire dataset, and values less than zero indicate



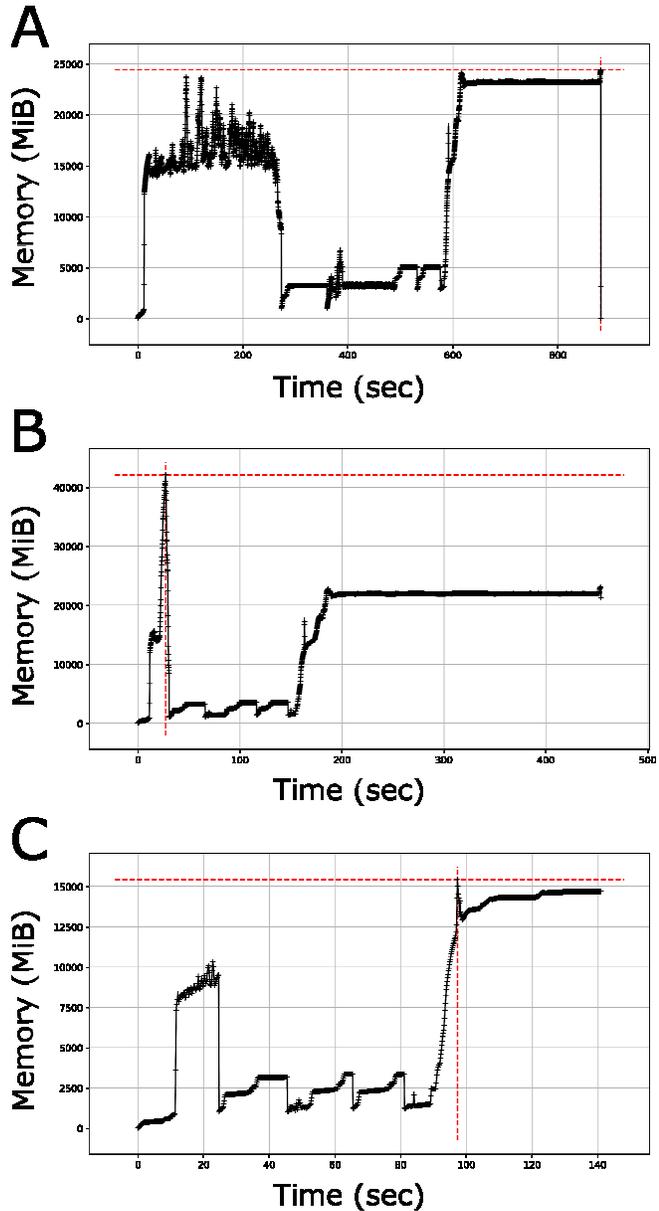
Supplementary Figure 2. Deviation of network predictions as a function of dropped features. Ticks along the x-axis represent the feature indices of each respective dropped feature according to supplementary Table 1. Ticks along the y-axis represent the raw amount the network prediction was deviated towards or away from the correct diagnoses, as compared to having no features dropped (see Eq. 1, main manuscript). A notable feature is that there are several features which strongly influence the network's decision in a positive way (help the network get the correct diagnosis), but none that influence the network's decision in a negative direction with the same magnitude (detract from the networks ability to get the correct diagnosis).

a feature which was less useful for the average diagnostic decision. Greater y-axis magnitudes indicate a stronger influence of that particular feature. It is clear from this plot that there were no features which significantly detracted from the networks performance (no features which have large negative magnitudes). It is also illustrated that the network was significantly more

dependent on color features (features indices 215-232) than any other feature class.

Complexity analysis

Supplementary Fig. 3A shows a plot of the time and memory for the full algorithm on the largest human WSI, containing ~3.5 billion pixels and containing a total of 63 glomeruli. Fig. 3B shows time versus memory for a mouse kidney section, containing ~438M pixels and 83 glomeruli. Fig. 3C shows time versus memory for the smallest human biopsy, containing ~470M pixels and 5 glomeruli. Red crosshairs indicate maximum memory usage. Plots were generated using the memory profiler python module.



Supplementary Figure 3. Memory as a function of run time for application of our algorithm in select cases. A) Largest human biopsy. B) Mouse whole-kidney section. C) Smallest human biopsy.

Supplementary Table 1. Feature list and associated network impact scores. The standardized column indicates values adjusted to have zero mean and unit standard deviation, the normalized column indicates values adjusted to be within the range 0 and 1. Abbreviations: M – morphological, C – containment, T – texture, Inter-C – inter-compartmental distance, Intra-C – intra-compartmental distance.

Index	Type	Name	Raw	Standardized	Normalized
1	M	Mean PAS+ object solidity	0.00033	-0.055	0.16
2	C	Mean lumina region contained in PAS+ object boundaries	0.00049	-0.014	0.17
3	C	Mean nuclear region contained in PAS+ object boundaries	0.00034	-0.054	0.16
4	M	Sum total PAS+ objects' areas	0.000068	-0.12	0.16
5	M	Mean of PAS+ objects' areas	0.00064	0.025	0.17
6	M	Median of PAS+ objects' areas	0.00025	-0.077	0.16
7	T	PAS+ textural contrast	-0.0013	-0.48	0.13
8	T	PAS+ textural correlation	0.0091	2.2	0.37
9	T	PAS+ textural energy	-0.0028	-0.86	0.09
10	T	PAS+ textural homogeneity	0.0037	0.81	0.24
11	M	Mean luminal object solidity	0.00033	-0.055	0.16
12	C	Mean PAS+ region contained in luminal object boundaries	-0.0039	-1.2	0.062
13	C	Mean nuclear region contained in luminal object boundaries	0.0011	0.15	0.18
14	M	Sum total luminal objects' areas	0.00013	-0.11	0.16
15	M	Mean of luminal objects' areas	-0.0031	-0.95	0.082
16	M	Median of luminal objects' areas	0.0013	0.21	0.19
17	T	Luminal textural contrast	-0.0014	-0.49	0.12
18	T	Luminal textural correlation	-0.0016	-0.55	0.12
19	T	Luminal textural energy	0.001	0.13	0.18
20	T	Luminal textural homogeneity	-0.0065	-1.8	0
21	C	Mean ratio of PAS+ pixels lying just outside nuclear perimeter to length of perimeter	-0.0036	-1.1	0.07
22	C	Mean ratio of luminal pixels lying just outside nuclear perimeter to length of perimeter	-0.00013	-0.17	0.15
23	M	Mean nuclear perimeter	-0.00047	-0.26	0.14
24	M	Sum total nuclear area	-0.00055	-0.28	0.14
25	M	Mean nuclear areas	0.0042	0.95	0.26
26	M	Mode nuclear areas	-0.00049	-0.27	0.14
27	T	Nuclear textural contrast	0.0012	0.16	0.18
28	T	Nuclear textural correlation	-0.0053	-1.5	0.031
29	T	Nuclear textural energy	-0.0018	-0.62	0.11
30	T	Nuclear textural homogeneity	-0.000041	-0.15	0.16
31	Inter-C	Mean distance of luminal object centroids from glomerular centroid	-0.004	-1.2	0.06
32	Inter-C	Mean of mean distances of luminal object centroids from glomerular boundary	-0.00017	-0.19	0.15
33	Inter-C	Mean of maximum distances of luminal object centroids from glomerular boundary	0.00032	-0.058	0.16

34	Inter-C	Mean of minimum distances of luminal object centroids from glomerular boundary	-0.00066	-0.31	0.14
35	Inter-C	Mean of mean distances of luminal object centroids from themselves	0.0031	0.66	0.23
36	Inter-C	Mean of maximum distances of luminal object centroids from themselves	0.00084	0.077	0.18
37	Inter-C	Mean of minimum distances of luminal object centroids from themselves	-0.0013	-0.48	0.13
38	Inter-C	Mean distance of PAS+ object centroids from glomerular centroid	-0.00091	-0.38	0.13
39	Inter-C	Mean of mean distances of PAS+ object centroids from glomerular boundary	0.0048	1.1	0.27
40	Inter-C	Mean of maximum distances of PAS+ object centroids from glomerular boundary	-0.0033	-0.99	0.078
41	Inter-C	Mean of minimum distances of PAS+ object centroids from glomerular boundary	0.00016	-0.099	0.16
42	Inter-C	Mean of mean distances of PAS+ object centroids from themselves	-0.0016	-0.55	0.12
43	Inter-C	Mean of maximum distances of PAS+ object centroids from themselves	-0.0052	-1.5	0.032
44	Inter-C	Mean of minimum distances of PAS+ object centroids from themselves	0.00092	0.098	0.18
45	Inter-C	Mean distance of nuclear object centroids from glomerular centroid	0.0014	0.23	0.19
46	Inter-C	Mean of mean distances of nuclear object centroids from glomerular boundary	-0.004	-1.2	0.061
47	Inter-C	Mean of maximum distances of nuclear object centroids from glomerular boundary	-0.0032	-0.97	0.08
48	Inter-C	Mean of minimum distances of nuclear object centroids from glomerular boundary	-0.0015	-0.54	0.12
49	Inter-C	Mean of mean distances of nuclear object centroids from themselves	0.0019	0.36	0.2
50	Inter-C	Mean of maximum distances of nuclear object centroids from themselves	0.00087	0.086	0.18
51	Inter-C	Mean of minimum distances of nuclear object centroids from themselves	-0.0009	-0.38	0.13
52	M	Total glomerular area	-0.00048	-0.27	0.14
53	M	Total PAS+ object number	-0.0029	-0.9	0.086
54	M	Total luminal object number	0.0012	0.16	0.18
55	M	Total nucleus number	-0.00033	-0.23	0.15
56	M/Intra-C	Sum of PAS+ distance transform values $0 < d \leq 10$	0.00059	0.013	0.17
57	M/Intra-C	Sum of PAS+ distance transform values $10 < d \leq 20$	0.0015	0.26	0.19
58	M/Intra-C	Sum of PAS+ distance transform values $20 < d \leq 1000$	0.0015	0.26	0.19
59	M/Intra-C	Maximum PAS+ distance transform value $10 < d \leq 20$	0.000062	-0.13	0.16
60	M/Intra-C	Number of connected objects with PAS+ distance transform value $0 < d \leq 10$	0.0041	0.92	0.25
61	M/Intra-C	Number of connected objects with PAS+ distance transform value $10 < d \leq 20$	0.0015	0.24	0.19
62	M/Intra-C	Mean of PAS+ distance transform values $0 < d \leq 10$	0.0008	0.066	0.18

63	M/Intra-C	Mean of PAS+ distance transform values 10<d<=20	-0.00099	-0.4	0.13
64	M/Intra-C	Median of PAS+ distance transform values 0<d<=10	0.0042	0.95	0.26
65	M/Intra-C	Median of PAS+ distance transform values 10<d<=20	-0.00014	-0.18	0.15
66	M/Intra-C	Mean area of objects with PAS+ distance transform value 0<d<=10	0.0038	0.85	0.25
67	M/Intra-C	Median area of objects with PAS+ distance transform value 0<d<=10	0.00054	0.00033	0.17
68	M/Intra-C	Maximum area of objects with PAS+ distance transform value 0<d<=10	0.0017	0.3	0.2
69	M/Intra-C	Mean area of objects with PAS+ distance transform value 10<d<=20	0.0027	0.55	0.22
70	M/Intra-C	Median area of objects with PAS+ distance transform value 10<d<=20	0.0008	0.067	0.18
71	Intra-C	Count of pixels with PAS+ distance transform value 1<d<=3	0.0041	0.94	0.26
72	Intra-C	Count of pixels with PAS+ distance transform value 3<d<=5	-0.0031	-0.95	0.082
73	Intra-C	Count of pixels with PAS+ distance transform value 5<d<=7	-0.00016	-0.18	0.15
74	Intra-C	Count of pixels with PAS+ distance transform value 7<d<=9	0.0012	0.16	0.18
75	Intra-C	Count of pixels with PAS+ distance transform value 9<d<=11	-0.00052	-0.28	0.14
76	Intra-C	Count of pixels with PAS+ distance transform value 11<d<=13	-0.00051	-0.27	0.14
77	Intra-C	Count of pixels with PAS+ distance transform value 13<d<=15	0.00025	-0.077	0.16
78	Intra-C	Count of pixels with PAS+ distance transform value 15<d<=17	-0.0014	-0.51	0.12
79	Intra-C	Count of pixels with PAS+ distance transform value 17<d<=19	0.00029	-0.065	0.16
80	Intra-C	Count of pixels with PAS+ distance transform value 19<d<=21	0.0033	0.71	0.23
81	Intra-C	Count of pixels with PAS+ distance transform value 21<d<=23	0.0025	0.51	0.22
82	Intra-C	Count of pixels with PAS+ distance transform value 23<d<=25	0.0035	0.76	0.24
83	Intra-C	Count of pixels with PAS+ distance transform value 25<d<=27	0.0022	0.44	0.21
84	Intra-C	Count of pixels with PAS+ distance transform value 27<d<=29	-0.00044	-0.26	0.15
85	Intra-C	Count of pixels with PAS+ distance transform value 29<d<=31	0.0028	0.59	0.22
86	Intra-C	Count of pixels with PAS+ distance transform value 31<d<=33	0.0025	0.5	0.22
87	Intra-C	Count of pixels with PAS+ distance transform value 33<d<=35	0.000045	-0.13	0.16
88	Intra-C	Count of pixels with PAS+ distance transform value 35<d<=37	0.0014	0.21	0.19
89	Intra-C	Count of pixels with PAS+ distance transform value 37<d<=39	0.0022	0.44	0.21
90	Intra-C	Count of pixels with PAS+ distance transform value 39<d<=41	0.0004	-0.038	0.17

91	Intra-C	Count of pixels with PAS+ distance transform value $41 < d \leq 43$	0.0009	0.092	0.18
92	Intra-C	Count of pixels with PAS+ distance transform value $43 < d \leq 45$	0.0024	0.48	0.21
93	Intra-C	Count of pixels with PAS+ distance transform value $45 < d \leq 47$	-0.0005	-0.27	0.14
94	Intra-C	Count of pixels with PAS+ distance transform value $47 < d \leq 49$	-0.00015	-0.18	0.15
95	Intra-C	Count of pixels with PAS+ distance transform value $49 < d \leq 51$	0.00064	0.025	0.17
96	Intra-C	Count of pixels with PAS+ distance transform value $51 < d \leq 53$	0.00041	-0.034	0.17
97	Intra-C	Count of pixels with PAS+ distance transform value $53 < d \leq 55$	0.00088	0.087	0.18
98	Intra-C	Count of pixels with PAS+ distance transform value $55 < d \leq 57$	0.0001	-0.11	0.16
99	Intra-C	Count of pixels with PAS+ distance transform value $57 < d \leq 59$	0.001	0.13	0.18
100	Intra-C	Count of pixels with PAS+ distance transform value $59 < d \leq 61$	0.00072	0.047	0.17
101	Intra-C	Count of pixels with PAS+ distance transform value $61 < d \leq 63$	0.00032	-0.058	0.16
102	Intra-C	Count of pixels with PAS+ distance transform value $63 < d \leq 65$	0.00082	0.072	0.18
103	Intra-C	Count of pixels with PAS+ distance transform value $65 < d \leq 67$	0.00056	0.0054	0.17
104	Intra-C	Count of pixels with PAS+ distance transform value $67 < d \leq 69$	0.0002	-0.09	0.16
105	Intra-C	Count of pixels with PAS+ distance transform value $69 < d \leq 71$	0.00025	-0.077	0.16
106	Intra-C	Count of pixels with PAS+ distance transform value $71 < d \leq 73$	0.00017	-0.098	0.16
107	Intra-C	Count of pixels with PAS+ distance transform value $73 < d \leq 75$	0.00056	0.0044	0.17
108	Intra-C	Count of pixels with PAS+ distance transform value $75 < d \leq 77$	0.00046	-0.022	0.17
109	Intra-C	Count of pixels with PAS+ distance transform value $77 < d \leq 79$	-0.00022	-0.2	0.15
110	Intra-C	Count of pixels with PAS+ distance transform value $79 < d \leq 2000$	0.0004	-0.037	0.17
111	Intra-C	Count of pixels with luminal distance transform value $1 < d \leq 2$	0.0026	0.53	0.22
112	Intra-C	Count of pixels with luminal distance transform value $2 < d \leq 3$	-0.0046	-1.3	0.046
113	Intra-C	Count of pixels with luminal distance transform value $3 < d \leq 4$	-0.0016	-0.56	0.12
114	Intra-C	Count of pixels with luminal distance transform value $4 < d \leq 5$	-0.0015	-0.53	0.12
115	Intra-C	Count of pixels with luminal distance transform value $5 < d \leq 6$	-0.0025	-0.78	0.098
116	Intra-C	Count of pixels with luminal distance transform value $6 < d \leq 7$	-0.000064	-0.16	0.15
117	Intra-C	Count of pixels with luminal distance transform value $7 < d \leq 8$	-0.0035	-1.1	0.071
118	Intra-C	Count of pixels with luminal distance transform value $8 < d \leq 9$	-0.0011	-0.42	0.13

119	Intra-C	Count of pixels with luminal distance transform value $9 < d \leq 10$	-0.0045	-1.3	0.049
120	Intra-C	Count of pixels with luminal distance transform value $10 < d \leq 11$	0.0000046	-0.14	0.16
121	Intra-C	Count of pixels with luminal distance transform value $11 < d \leq 12$	-0.0034	-1	0.075
122	Intra-C	Count of pixels with luminal distance transform value $12 < d \leq 13$	0.00074	0.05	0.17
123	Intra-C	Count of pixels with luminal distance transform value $13 < d \leq 14$	-0.00016	-0.18	0.15
124	Intra-C	Count of pixels with luminal distance transform value $14 < d \leq 15$	-0.003	-0.93	0.083
125	Intra-C	Count of pixels with luminal distance transform value $15 < d \leq 16$	-0.001	-0.41	0.13
126	Intra-C	Count of pixels with luminal distance transform value $16 < d \leq 17$	-0.00067	-0.32	0.14
127	Intra-C	Count of pixels with luminal distance transform value $17 < d \leq 18$	0.00088	0.087	0.18
128	Intra-C	Count of pixels with luminal distance transform value $18 < d \leq 19$	0.0019	0.35	0.2
129	Intra-C	Count of pixels with luminal distance transform value $19 < d \leq 20$	0.00046	-0.023	0.17
130	Intra-C	Count of pixels with luminal distance transform value $20 < d \leq 21$	0.00045	-0.025	0.17
131	Intra-C	Count of pixels with luminal distance transform value $21 < d \leq 22$	-0.0001	-0.17	0.15
132	Intra-C	Count of pixels with luminal distance transform value $22 < d \leq 23$	-0.00038	-0.24	0.15
133	Intra-C	Count of pixels with luminal distance transform value $23 < d \leq 24$	0.0021	0.41	0.21
134	Intra-C	Count of pixels with luminal distance transform value $24 < d \leq 25$	-0.0003	-0.22	0.15
135	Intra-C	Count of pixels with luminal distance transform value $25 < d \leq 26$	0.00065	0.027	0.17
136	Intra-C	Count of pixels with luminal distance transform value $26 < d \leq 27$	-0.00089	-0.37	0.13
137	Intra-C	Count of pixels with luminal distance transform value $27 < d \leq 28$	0.0013	0.2	0.19
138	Intra-C	Count of pixels with luminal distance transform value $28 < d \leq 29$	0.0005	-0.011	0.17
139	Intra-C	Count of pixels with luminal distance transform value $29 < d \leq 30$	-0.0022	-0.72	0.1
140	Intra-C	Count of pixels with luminal distance transform value $30 < d \leq 31$	-0.0018	-0.6	0.11
141	Intra-C	Count of pixels with luminal distance transform value $31 < d \leq 32$	-0.002	-0.65	0.11
142	Intra-C	Count of pixels with luminal distance transform value $32 < d \leq 33$	0.0012	0.17	0.18
143	Intra-C	Count of pixels with luminal distance transform value $33 < d \leq 34$	-0.0017	-0.57	0.12
144	Intra-C	Count of pixels with luminal distance transform value $34 < d \leq 35$	0.0023	0.44	0.21
145	Intra-C	Count of pixels with luminal distance transform value $35 < d \leq 36$	0.002	0.38	0.2
146	Intra-C	Count of pixels with luminal distance transform value $36 < d \leq 37$	0.00027	-0.072	0.16

147	Intra-C	Count of pixels with luminal distance transform value $37 < d \leq 38$	-0.0017	-0.59	0.12
148	Intra-C	Count of pixels with luminal distance transform value $38 < d \leq 39$	0.00042	-0.031	0.17
149	Intra-C	Count of pixels with luminal distance transform value $39 < d \leq 40$	-0.00089	-0.37	0.13
150	Intra-C	Count of pixels with luminal distance transform value $40 < d \leq 41$	0.00034	-0.052	0.16
151	Intra-C	Count of pixels with luminal distance transform value $41 < d \leq 42$	0.00009	-0.12	0.16
152	Intra-C	Count of pixels with luminal distance transform value $42 < d \leq 43$	-0.0018	-0.62	0.11
153	Intra-C	Count of pixels with luminal distance transform value $43 < d \leq 44$	-0.00056	-0.29	0.14
154	Intra-C	Count of pixels with luminal distance transform value $44 < d \leq 45$	-0.00001	-0.14	0.16
155	Intra-C	Count of pixels with luminal distance transform value $45 < d \leq 46$	0.0014	0.22	0.19
156	Intra-C	Count of pixels with luminal distance transform value $46 < d \leq 47$	-0.00067	-0.32	0.14
157	Intra-C	Count of pixels with luminal distance transform value $47 < d \leq 48$	-0.00024	-0.2	0.15
158	Intra-C	Count of pixels with luminal distance transform value $48 < d \leq 49$	-0.00022	-0.2	0.15
159	Intra-C	Count of pixels with luminal distance transform value $49 < d \leq 50$	-0.0012	-0.46	0.13
160	Intra-C	Count of pixels with luminal distance transform value $50 < d \leq 51$	0.0006	0.014	0.17
161	Intra-C	Count of pixels with luminal distance transform value $51 < d \leq 52$	0.00012	-0.11	0.16
162	Intra-C	Count of pixels with luminal distance transform value $52 < d \leq 53$	0.0016	0.28	0.2
163	Intra-C	Count of pixels with luminal distance transform value $53 < d \leq 54$	-0.0017	-0.58	0.12
164	Intra-C	Count of pixels with luminal distance transform value $54 < d \leq 55$	0.000029	-0.13	0.16
165	Intra-C	Count of pixels with luminal distance transform value $55 < d \leq 56$	0.0022	0.42	0.21
166	Intra-C	Count of pixels with luminal distance transform value $56 < d \leq 57$	-0.0008	-0.35	0.14
167	Intra-C	Count of pixels with luminal distance transform value $57 < d \leq 58$	0.00046	-0.022	0.17
168	Intra-C	Count of pixels with luminal distance transform value $58 < d \leq 59$	0.00044	-0.026	0.17
169	Intra-C	Count of pixels with luminal distance transform value $59 < d \leq 60$	0.00053	-0.0032	0.17
170	Intra-C	Count of pixels with luminal distance transform value $60 < d \leq 2000$	-0.00091	-0.38	0.13
171	Intra-C	Count of pixels with nuclear distance transform value $1 < d \leq 2$	-0.0014	-0.52	0.12
172	Intra-C	Count of pixels with nuclear distance transform value $2 < d \leq 3$	0.00026	-0.075	0.16
173	Intra-C	Count of pixels with nuclear distance transform value $3 < d \leq 4$	-0.0032	-0.97	0.08
174	Intra-C	Count of pixels with nuclear distance transform value $4 < d \leq 5$	0.000023	-0.14	0.16

175	Intra-C	Count of pixels with nuclear distance transform value $5 < d \leq 6$	0.00075	0.053	0.17
176	Intra-C	Count of pixels with nuclear distance transform value $6 < d \leq 7$	0.0007	0.04	0.17
177	Intra-C	Count of pixels with nuclear distance transform value $7 < d \leq 8$	0.00063	0.022	0.17
178	Intra-C	Count of pixels with nuclear distance transform value $8 < d \leq 9$	0.0011	0.16	0.18
179	Intra-C	Count of pixels with nuclear distance transform value $9 < d \leq 10$	0.00054	-0.001	0.17
180	Intra-C	Count of pixels with nuclear distance transform value $10 < d \leq 11$	0.00039	-0.039	0.17
181	Intra-C	Count of pixels with nuclear distance transform value $11 < d \leq 12$	-0.0009	-0.38	0.13
182	Intra-C	Count of pixels with nuclear distance transform value $12 < d \leq 13$	0.00072	0.046	0.17
183	Intra-C	Count of pixels with nuclear distance transform value $13 < d \leq 14$	-0.00039	-0.24	0.15
184	Intra-C	Count of pixels with nuclear distance transform value $14 < d \leq 15$	-0.00063	-0.3	0.14
185	Intra-C	Count of pixels with nuclear distance transform value $15 < d \leq 16$	-0.00041	-0.25	0.15
186	Intra-C	Count of pixels with nuclear distance transform value $16 < d \leq 17$	0.00011	-0.11	0.16
187	Intra-C	Count of pixels with nuclear distance transform value $17 < d \leq 18$	0.00046	-0.023	0.17
188	Intra-C	Count of pixels with nuclear distance transform value $18 < d \leq 19$	0.00036	-0.049	0.16
189	Intra-C	Count of pixels with nuclear distance transform value $19 < d \leq 20$	0.00044	-0.028	0.17
190	Intra-C	Count of pixels with nuclear distance transform value $20 < d \leq 2000$	0.00043	-0.029	0.17
191	Intra-C	Count of pixels with glomerular distance transform value $2 < d \leq 27$	0.000092	-0.12	0.16
192	Intra-C	Count of pixels with glomerular distance transform value $27 < d \leq 52$	-0.00034	-0.23	0.15
193	Intra-C	Count of pixels with glomerular distance transform value $52 < d \leq 77$	0.0011	0.16	0.18
194	Intra-C	Count of pixels with glomerular distance transform value $77 < d \leq 102$	0.00093	0.1	0.18
195	Intra-C	Count of pixels with glomerular distance transform value $102 < d \leq 127$	0.0016	0.29	0.2
196	Intra-C	Count of pixels with glomerular distance transform value $127 < d \leq 152$	-0.00026	-0.21	0.15
197	Intra-C	Count of pixels with glomerular distance transform value $152 < d \leq 177$	0.002	0.39	0.2
198	Intra-C	Count of pixels with glomerular distance transform value $177 < d \leq 202$	0.00082	0.073	0.18
199	Intra-C	Count of pixels with glomerular distance transform value $202 < d \leq 227$	-0.000087	-0.16	0.15
200	Intra-C	Count of pixels with glomerular distance transform value $227 < d \leq 252$	0.00026	-0.072	0.16
201	Intra-C	Count of pixels with glomerular distance transform value $252 < d \leq 277$	-0.0028	-0.86	0.09
202	Intra-C	Count of pixels with glomerular distance transform value $277 < d \leq 302$	-0.0013	-0.47	0.13

203	Intra-C	Count of pixels with glomerular distance transform value $302 < d \leq 327$	0.0015	0.24	0.19
204	Intra-C	Count of pixels with glomerular distance transform value $327 < d \leq 352$	-0.00083	-0.36	0.14
205	Intra-C	Count of pixels with glomerular distance transform value $352 < d \leq 377$	-0.0025	-0.79	0.097
206	Intra-C	Count of pixels with glomerular distance transform value $377 < d \leq 402$	-0.002	-0.67	0.11
207	Intra-C	Count of pixels with glomerular distance transform value $402 < d \leq 427$	-0.0000053	-0.14	0.16
208	Intra-C	Count of pixels with glomerular distance transform value $427 < d \leq 452$	-0.00088	-0.37	0.14
209	Intra-C	Count of pixels with glomerular distance transform value $452 < d \leq 477$	0.00038	-0.043	0.17
210	Intra-C	Count of pixels with glomerular distance transform value $477 < d \leq 502$	0.00058	0.0093	0.17
211	Intra-C	Count of pixels with glomerular distance transform value $502 < d \leq 527$	0.00028	-0.069	0.16
212	Intra-C	Count of pixels with glomerular distance transform value $527 < d \leq 552$	0.000046	-0.13	0.16
213	Intra-C	Count of pixels with glomerular distance transform value $552 < d \leq 577$	0.00021	-0.086	0.16
214	Intra-C	Count of pixels with glomerular distance transform value $577 < d \leq 20000$	0.0005	-0.011	0.17
215	Color	Mean of red values in PAS+ regions	0.0026	0.53	0.22
216	Color	Mean of green values in PAS+ regions	-0.0023	-0.74	0.1
217	Color	Mean of blue values in PAS+ regions	0.017	4.3	0.56
218	Color	Std of red values in PAS+ regions	0.035	9	1
219	Color	Std of green values in PAS+ regions	0.013	3.2	0.46
220	Color	Std of blue values in PAS+ regions	0.019	4.9	0.62
221	Color	Mean of red values in luminal regions	0.0046	1.1	0.27
222	Color	Mean of green values in luminal regions	-0.0019	-0.63	0.11
223	Color	Mean of blue values in luminal regions	0.0012	0.16	0.18
224	Color	Std of red values in luminal regions	-0.0031	-0.94	0.083
225	Color	Std of green values in luminal regions	-0.00097	-0.39	0.13
226	Color	Std of blue values in luminal regions	0.0045	1	0.26
227	Color	Mean of red values in nuclear regions	0.007	1.7	0.32
228	Color	Mean of green values in nuclear regions	0.0082	2	0.35
229	Color	Mean of blue values in nuclear regions	0.02	5	0.63
230	Color	Std of red values in nuclear regions	-0.0059	-1.7	0.014
231	Color	Std of green values in nuclear regions	0.0019	0.36	0.2
232	Color	Std of blue values in nuclear regions	0.0019	0.36	0.2

References

1. Caruana R, Lou Y, Gehrke J, Koch P, Sturm M, Elhadad N. Intelligible Models for HealthCare: Predicting Pneumonia Risk and Hospital 30-day Readmission. Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining; Sydney, NSW, Australia. 2788613: ACM; 2015. p. 1721-30.
2. Haralick RM, Shanmugam K, Dinstein I. Textural Features for Image Classification. *Ieee T Syst Man Cyb.* 1973;Smc3(6):610-21.
3. Maurer C. A Linear Time Algorithm for Computing Exact Euclidean Distance Transforms of Binary Images in Arbitrary Dimensions. *IEEE Trans Pattern Analysis and Machine Intelligence.* 2003;25(2):265-70.
4. Maurer CR, Qi RS, Raghavan V. A linear time algorithm for computing exact Euclidean distance transforms of binary images in arbitrary dimensions. *Ieee T Pattern Anal.* 2003;25(2):265-70.
5. Paglieroni DW. Distance Transforms - Properties and Machine Vision Applications. *Cvgip-Graph Model Im.* 1992;54(1):56-74.
6. Veta M, Huisman A, Viergever MA, van Diest PJ, Pluim JPW. Marker-Controlled Watershed Segmentation of Nuclei in H&E Stained Breast Cancer Biopsy Images. *I S Biomed Imaging.* 2011:618-21.
7. Meyer F, Beucher S. Morphological segmentation. *Journal of Visual Communication and Image Representation.* 1990;1(1):21-46.
8. Jain AK. *Fundamentals of digital image processing*: Prentice-Hall, Inc.; 1989. 569 p.
9. Tesch GH, Allen TJ. Rodent models of streptozotocin-induced diabetic nephropathy. *Nephrology (Carlton).* 2007;12(3):261-6.
10. Breyer MD, Bottinger E, Brosius FC, 3rd, Coffman TM, Harris RC, Heilig CW, et al. Mouse models of diabetic nephropathy. *J Am Soc Nephrol.* 2005;16(1):27-45.
11. Ruifrok AC, JOHnston DA. Quantification of histochemical staining of color deconvolution. *Anal Quant Cytol Histol.* 2001;23(4):291-9.