

PROCEEDINGS OF SPIE

SPIDigitalLibrary.org/conference-proceedings-of-spie

Probabilistic modelling of diabetic nephropathy progression

Samuel Border, Kuang-Yu Jen, Washington LC dos-Santos, John Tomaszewski, Pinaki Sarder

Samuel Border, Kuang-Yu Jen, Washington LC dos-Santos, John Tomaszewski, Pinaki Sarder, "Probabilistic modelling of diabetic nephropathy progression," Proc. SPIE 11320, Medical Imaging 2020: Digital Pathology, 1132014 (16 March 2020); doi: 10.1117/12.2549171

SPIE.

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

Probabilistic modeling of Diabetic Nephropathy progression

Samuel Border¹, Kuang-Yu Jen², Washington LC dos-Santos³, John Tomaszewski¹,
Pinaki Sarder^{1,*}

¹Department of Pathology and Anatomical Sciences, University at Buffalo

²Department of Pathology, University of California at Davis

³Department of Pathology, Federal University of Bahia

*Address all correspondence to: Pinaki Sarder

Tel: 716-829-2265; E-mail: pinakisa@buffalo.edu

ABSTRACT:

Diabetic Nephropathy (DN) progression is stratified into several stages with different levels of proteinuria, albuminuria, and physical characteristics as observed by pathologists. These physical changes are primarily visible within a patient's glomeruli which function as filtration units for blood returning for oxygenation. As DN stage increases, it is possible to observe the thickening of the glomerular basement membrane, expansion of the mesangium, and development of nodular sclerosis. Classification of different stages of DN by pathologists is based on semi-qualitative assessments of these characteristics on an individual glomerulus basis. Being able to probabilistically infer stage membership of individual glomeruli based on a combination of easily observable and hidden image features would be an invaluable tool for furthering our understanding of the drivers of DN progression. Markov Particle filters, included in the *bnlearn* package in R, were used to query a Bayesian Network (BN) constructed using the structural Hill-Climbing algorithm on a set of glomerular features. These features included both traditional characteristics such as glomerular area and number of mesangial nuclei as well as more abstract features derived from Minimum Spanning Trees (MST) to quantify spatial distribution of mesangial nuclei. Our results using images from multiple institutions suggest that these abstract features exercise a variable influence on DN stage membership over the course of disease progression. Further research incorporating clinical data will give nephrologists a "white box" visual of quantitative factors present in DN patients.

Keywords: Bayesian network, diabetic nephropathy, probabilistic graphical modelling, minimum spanning trees (MST)

1. INTRODUCTION

Understanding how individual factors influence the chances of disease progression is one of the many ways in which we can understand underlying cellular mechanisms that contribute to worsening health outcomes. In the case of diabetic nephropathy (DN), failure to respond to advancement in a timely manner can increase a patient's chances of kidney failure and end-stage renal disease (ESRD). This is especially difficult in the case of kidney disease due to the fact that early stages exhibit no outward symptoms and are often undetected. Around 42% of cases of ESRD have a co-diagnosis of DN in the United States^[1]. Classification of DN is traditionally performed by needle biopsy analysis by a pathologist on an individual glomerulus basis, which is slow and susceptible to inter-observer variability^[2]. Our lab recently developed a recursive neural network based method to conduct automatic diagnosis of DN on a whole biopsy^[3]. In this work we developed an alternative method based on graph theory to visualize features which contribute to DN staging. Further, in order to better understand the causes for DN progression, statistical representations of glomerular image features were generated and tested with simulated data for probability of higher stages membership. Quantitative hidden patterns describing individual glomerulus images were calculated and input into a structural Hill-Climbing algorithm to generate a Bayesian Network (BN) from which conditional probability of DN stage is determined. Features used in the BN included both readily visible features such as glomerular cross sectional area and

the number of nuclei present in the mesangium, and features calculated from Minimum Spanning Trees (MST) to quantify nuclear packing. MSTs are graphical method defined as a set of straight lines connecting a set of points in such a way that all of the points are connected, no closed loops are formed, and the sum of the total edge lengths is minimized^[4]. In the context of digital pathology image analysis, MSTs are used to quantitatively understand spatial distribution of sub-visual cellular compartments in an image^[5]. Our results suggest that our set of feature values was proficient at classifying the simulated data belonging to samples from more advanced stages. Adding further features related to clinical data will provide a machine learning framework that is more relatable to medical professionals and will aid them in the early diagnosis of patients with kidney disease.

2. METHODS

Human data: Biopsy samples from human DN patients with chronic kidney disease stages II and III were collected from Kidney Translational Research Center at Washington University School of Medicine via collaborator Dr. Sanjay Jain and from Vanderbilt University Medical Center (VUMC) via collaborator Dr. Agnes Fogo. Total $n = 21$ DN biopsy cases and $n = 8$ control biopsy cases were used. Control cases were obtained from nephrectomies of renal cell carcinoma patients, and tissue samples with no apparent abnormalities were used. Additional $n = 9$ images of DN biopsies were received from Dr. Washington Luis Conrado of the Universidade Federal da Bahia (UFBA). Human data collection procedure followed a protocol approved by the Institutional Review Board at all participating institutions.

Imaging and data preparation: Tissue staining and imaging are conducted as discussed in our other work^[6].

Ground-truth annotation and segmentation: Individual glomeruli were segmented from whole slide kidney biopsy images using our developed human-artificial-intelligence-loop method^[7]. Ground-truth annotations of DN structural disease state was performed by the co-author Dr. Kuang-Yu Jen.

Feature generation: From the segmented images, all nuclei within each glomerulus were isolated using a combination of color deconvolution and intensity thresholding^[8]. Manual segmentation was used to exclude nuclei located along the endothelium and not allow an expansive Bowman's space to artificially inflate glomerular area measurements. The remaining nuclei were used as root nodes for constructing MSTs on a per-glomerulus basis^[9]. The output of MST generation is a set of each of the edge weights between specific nuclei nodes from which feature values are calculated.

Non-MST Features: The first set of features calculated from the image set are those not specifically related to MSTs. These are the number of nuclei and the glomerular area. The number of nuclei present within the mesangium, excluding those present along the endothelium, are an indicator of the degree of cellularity within that glomerulus. Mesangial hypercellularity and expansion are known to be general indicators of glomerular injury and traditionally are associated with the progression of DN between stages I and II^[2].

Nuclei Spread: To quantify the degree in which the nuclei are spread over the cross-sectional area of the glomerulus, we used the MST edge length and eccentricity. To better understand the total distribution of edge lengths for each glomerulus, both the mean and standard deviation of edge lengths were recorded and used for comparisons. Eccentricity in this case refers to the graph eccentricity, i.e., the graph distance between one node and the furthest node away. In the same way as with the edge lengths, we recorded both the mean and standard deviation of graph eccentricity as well as the maximum node eccentricity, known as the graph diameter.

Nuclei Proximity: In order to capture each node's connectedness/clustering with nearby nodes, we calculated betweenness centrality, degrees per node, and leaf fraction. Betweenness centrality is a measure of node importance to the overall structural integrity of the graph^[10]. To meaningfully capture the distinctive differences between different MSTs originated from glomeruli with different DN stages, we chose to record both the maximum rank as well as the number of nodes within 10% of the maximum rank. In this way we can determine if there is an uneven distribution of nodes within one graph or if they are uniformly spread out. A node's degree is determined by the number of branches that extend from it. For example, a node that has one branch going into it and then another branch going to another node would have a degree of two. If a graph has a degree per node value that is close to two we can infer that the nodes are fairly sparse. A node that has only one branch entering it is known as a terminal or 'leaf' node. The

leaf fraction is calculated by taking the number of leaf nodes present in the graph divided by the total number of nodes^[11]. A low leaf fraction would be indicative of a densely packed MST.

MST Shape: The final feature values calculated from the MSTs were used to distinguish between multiple different shapes and structures of different MSTs. We employed a metric proposed by Rainbolt *et al.* for analyzing Large Hadron Collider (LHC) output^[5]. The normalized edge length is calculated by computing the natural logarithm of each edge length and dividing it by the average edge length of each graph. Again, to preserve the differences that are present in the shape of the distribution, both the mean and standard deviation of the normalized edge length are saved as individual feature values.

Bayesian Networks: Bayesian Networks are a subset of probabilistic graphical models used to represent conditional dependencies of either discrete or continuous variables describing a dataset^[12]. By describing the relationships between individual feature values, it is possible to probabilistically infer missing values. The organization of nodes, or feature variables, in Bayesian Networks can be determined through a variety of different methods focusing on generating a model that most accurately describes the dataset used to create it. In this study, we employed a score-based structural Hill-Climbing algorithm included in the ‘*bnlearn*’ package for R^[13]. This approach generates a set of models, calculates the Bayesian Information Criterion (BIC) score, and then uses gradient descent to find the structure with the lowest score. The resulting graph contains all of the nodes and edges necessary to most accurately represent the input data. Conditional probability queries of this model are computed by generating individual samples that represent the posterior probability of certain observations and randomly sampling from these according to their individual likelihood weighting, this process is known as applying a Particle Filter^[14]. It is especially useful when the underlying posterior probabilities of a system are unknown and we must make use of noisy observations.

3. RESULTS

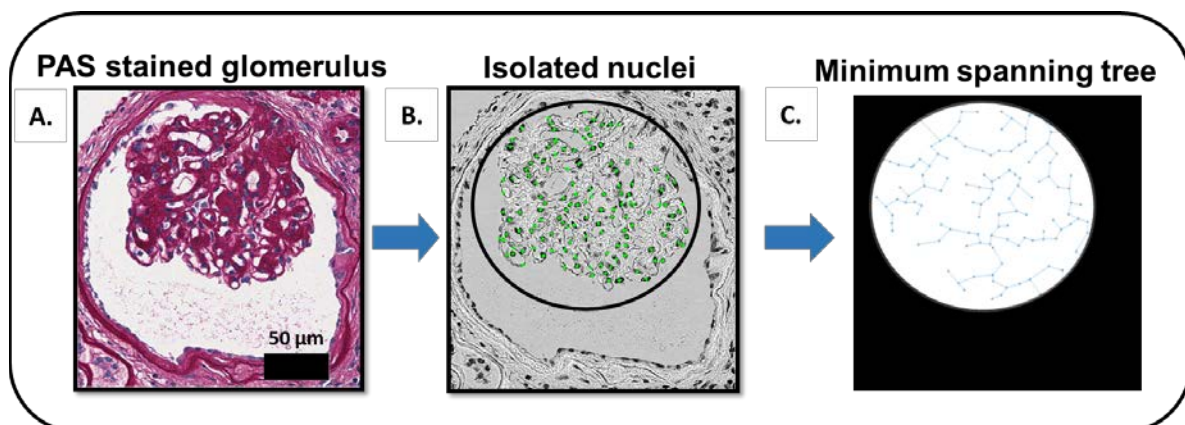


Figure 1: MST Construction Pipeline. (A) Initial PAS stained images stain channels separated using color deconvolution. (B) Nuclei isolated and input into Prim’s Algorithm as graph nodes. (C) MST used to calculate feature values.

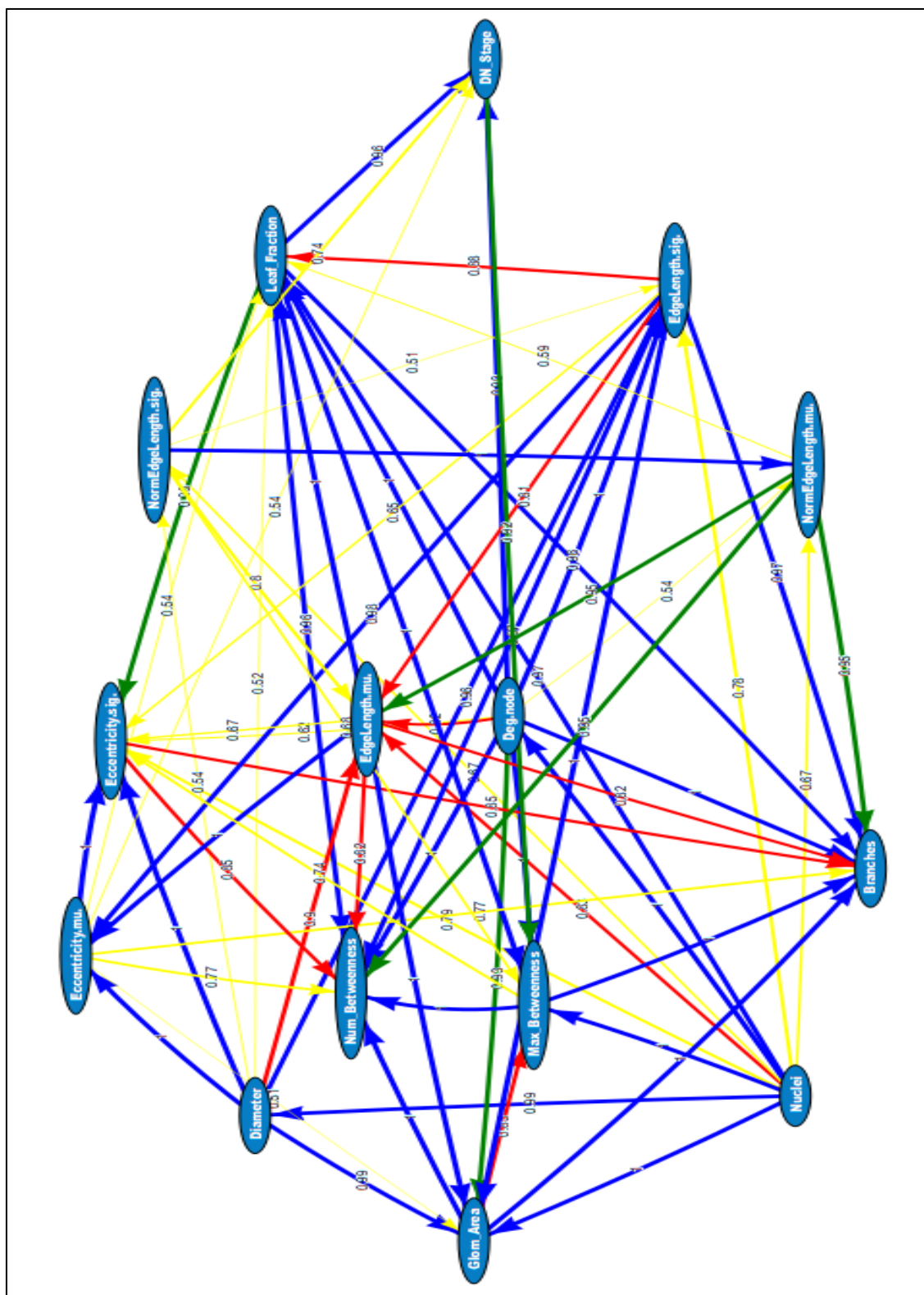


Figure 2: Bayesian Network with Separated Feature Categories. BN constructed using structural Hill-Climbing method in *bnlearn* using feature values from 795 glomerular MSTs. Color of arrows corresponds to normalized edge weight. (Blue = 0.95-1.0, Green = 0.9-0.95, Red = 0.8-0.9, Yellow=0.5-0.8).

Data from feature vectors was used to generate a BN using the method described above and the resulting network is shown in Figure 2. Each of the nodes were modeled as Gaussian distributions of the full range of values for each feature. This is opposed to employing discretization techniques to represent the feature ranges as a series of levels. Edges present in this network range in strength from 0.5-1.0 as all edges below 0.5 were removed from the completed graph. This way, only the most significant interactions were preserved for querying. Querying in this network was performed using *bnlearn*'s built-in conditional probability querying function which uses Monte Carlo particle filters to randomly sample from the provided distribution and return the likelihood of the event given a set of evidence ranges.

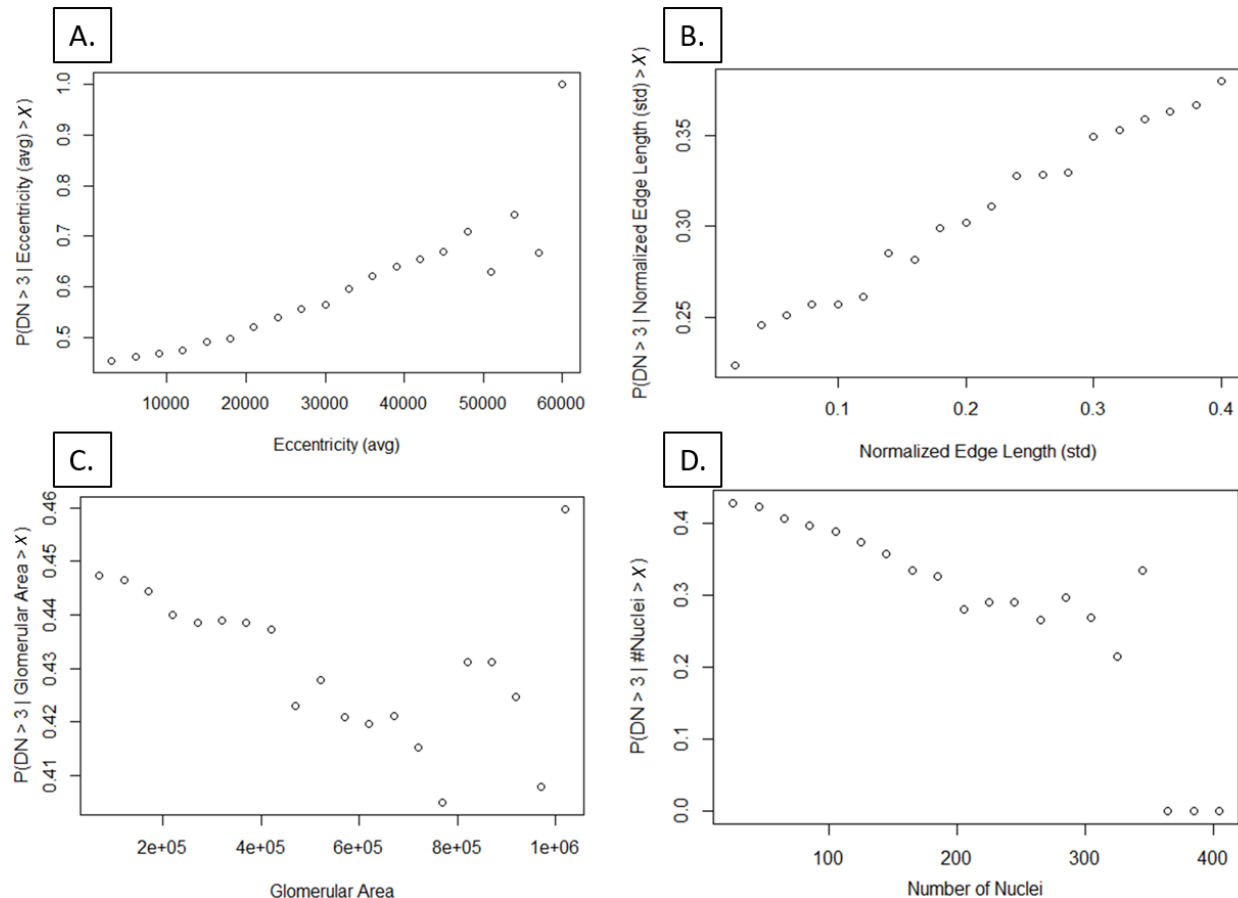


Figure 3: Conditional Probability Query Results. (A, B, C, D) Line graph showing change in probability of DN stage greater than or equal to three as feature value increases. Individual probabilities calculated using Monte Carlo particle filters of data distributions.

The differential influence of certain image features on probabilistic prediction of DN Stage was best observed through individual queries of the constructed model. Choosing a selection of features and varying the value by a set amount within the range of observed values resulted in the graphs above. These queries can also be altered to measure the probability that individual samples come from earlier stages of DN. For the later stages, it is known that the amount of mesangial expansion and hypercellularity decreases as the glomerulus loses significant function^[2]. These well-documented relationships are supported by the probability predictions for increasing numbers of nuclei and area measurements. MST features such as the standard deviation of normalized edge length and average eccentricity exhibit a larger range in probability prediction as the feature is increased. Feature values where the probability estimate range increases as the value range is also increased are more helpful in making accurate distinctions between different samples.

In order to test the robustness of our feature set and feature generation pipeline, we also separated the data based on the institution which they were obtained from. In this way we are hoping to make up for any potential minor stain

variabilities that the pipeline may have encountered and ensure that the relationships between features are preserved. Ideally, all of the BNs would have the same nodes and edge weights because all of the features were calculated the same way and from patients with the same disease.

Institution	Number of Glomeruli
Universidade Federal da Bahia (UFBA)	58
Kidney Translational Research Center (KTRC)	783
Vanderbilt University Medical Center (VUMC)	286

Table 1: Different sources of data. Table shows the three different institutions from which the glomeruli images were derived as well as the number of samples from each.

An independent BN was constructed based on the separated data sets using the same method as with the combined data set. When these three different networks were compared it was discovered that most of the fundamental relationships were preserved. The large variation in sample size could account for the varying edge weights between the different models but overall they were very similar. To quantify the differences between the four BNs, we used the Structural Hamming Distance metric (SHD). The SHD can be thought of as the total edge permutations necessary to transform one graph skeleton into the other.

Institution	Combined	KTRC	VUMC	UFBA
Combined	0			
KTRC	38	0		
VUMC	44	47	0	
UFBA	57	50	35	0

Table 2: Structural Hamming Distances between network pairs: Table containing the different SHD values for each pair of networks. The largest value (57) corresponds to the comparison between UFBA and the Combined datasets.

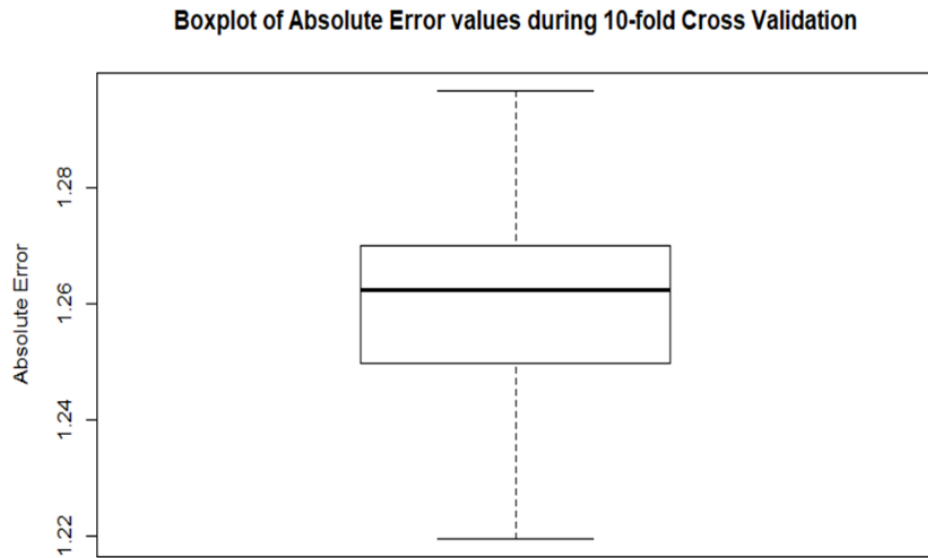


Figure 4: Boxplot of Mean Absolute Error values during 10-fold Cross Validation: Plot depicting the absolute error in predicted DN stage for one test set for full network trained on all samples.

Initial holdout predictions were based off of the network constructed using 70% of the total dataset. At each round of cross-validations, a random sampling of the full dataset was selected and a new BN was fit according to that dataset. The *bnlearn* in-built ‘predict’ function was used to combine the test sample’s feature values and the network’s learned parameters. The results of 10-fold cross-validation are shown in Fig. 8. The average absolute error across all folds ended up being 1.27.

It is not always necessary to input a value for each of the nodes in the network. This is advantageous if we are including a feature that is difficult to measure. Another reason this could be a boon is if we wanted to prune the network to a few select nodes without negatively impacting performance. The first step in determining which combination of nodes provides the most efficient performance is to list all of the possible networks that can be constructed. The total number of different combinations depends on the number of nodes one is working with.

$$\# \text{ com} = \sum_{n=1}^N C_n^N = C_1^N + C_2^N + C_3^N + \cdots + C_N^N$$

Equation 1: Number of possible networks from known number of nodes. This equation takes the number of nodes (N) and determines the number of networks needed to be constructed and tested.

From the equation above, it is easy to see how the possible number of combinations can shoot upwards if there are more than 10 nodes. For that reason, we selected a subset of eleven nodes to construct the sub-networks from based on their proximity to the Markov blanket of the DN stage node. The Markov blanket is a term used to describe the parents of a node, the children of a node, and the other parents of those children. Because DN stage is a terminal node (without children) then this subset contains only nodes that are parents of DN stage or strongly associated with those parents. To determine which combination is the best, we have to iterate through each of the possible

combinations and test that network on the combined dataset. We used 10-fold cross-validation again for each of the sub-networks.

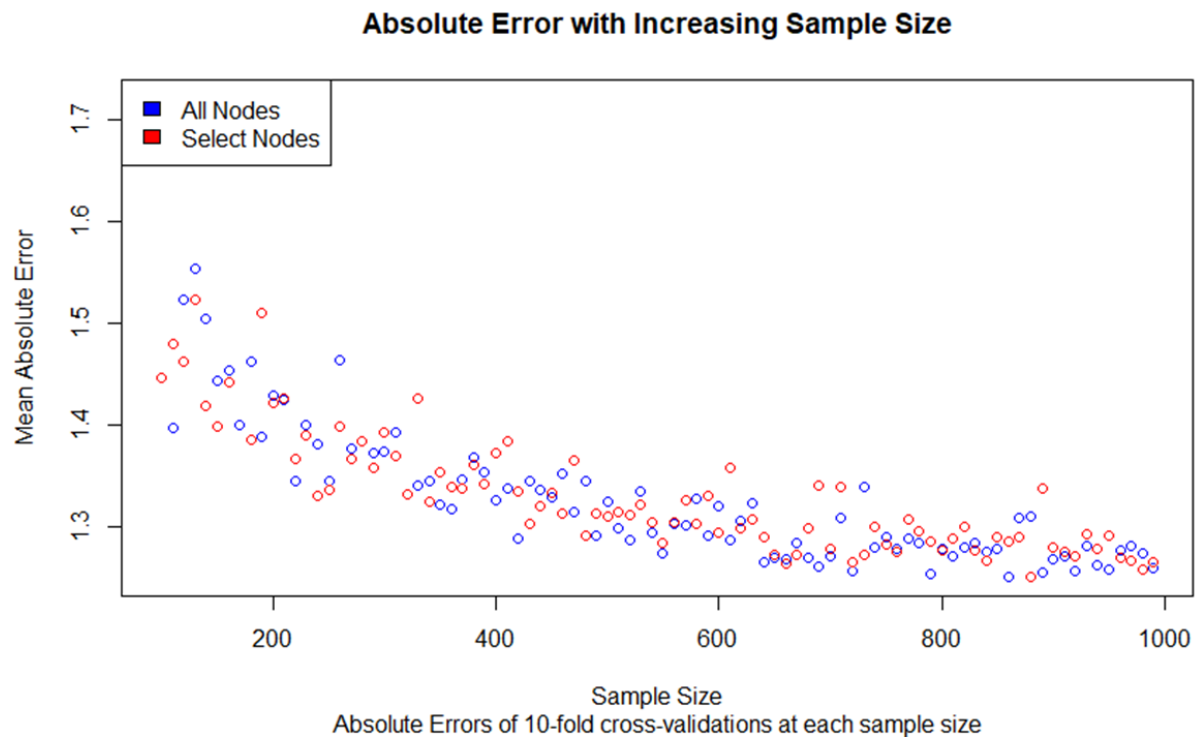


Figure 5: MAE with Increasing Sample Size: This graph shows the impact of increasing sample size on the classification accuracy of both the full network as well as the minimum MAE sub-network. Using only a fraction of the nodes, the sub-network had similar performance with similar response to increasing sample size.

The sub-network with the minimum measured MSE included the nodes: Leaf Fraction, Degrees/node, Edge Length average and standard deviation, Normalized Edge Length, and Eccentricity. All of these features are derived from the glomerular MSTs. The network including only these nodes also exhibited very similar behavior to the full network over a wide range of sample sizes.

4. DISCUSSION

Deep learning methods have an astronomical potential in medical image analysis, however, when it comes to understanding underlying mechanisms of disease progression they fall short in their interpretability. When designing computational tools with the goal of aiding clinicians, our first goal should always be to build off of the knowledge that decades of previous research has led to and Bayesian networks (BN) are advantageous in this regard. In this way we can pursue paths of research with validated methods and give better direction for treatment options. Another benefit of BN models is that they use simple probabilistic relationships with language that is very familiar to medical professionals.

5. CONCLUSION AND FUTURE WORK

By employing the MST for quantifying hidden features of glomerulus images, we can be sure that the same classifying criterion are applied to every sample. In this way, we can greatly reduce the amount of variability that appears in DN stage classification. Through the use of BN, we can determine the most important features for classification and focus on those features for developing queries. The resulting probabilistic queries can give clinicians an initial estimate for the likelihood of stage membership. Bayesian Networks are a very accessible form of “deep learning” methods due to their inclusion of user-defined features which are more readily translatable to a biological setting than latent space features of other methods. In regards to future work on this subject, we plan include clinical data into another model from which we have kidney biopsy images. With this model it will be possible to see the changing influence of variables such as eGFR, proteinuria, and albuminuria and how these compare to individual and grouped glomerular features. Being able to merge traditional measures for severity of DN progression with image features is an exciting field of research. Incorporating quantitative characteristics of patients’ glomeruli will give clinicians a new perspective of formerly imprecise diagnostic criteria.

6. ACKNOWLEDGEMENT

The project 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 Diabetic Complications Consortium grant DK076169; NIDDK grant R01 DK114485 & DK114485 02S1; NIDDK CKD Biomarker Consortium grant U01 DK103225; and NIDDK Kidney Precision Medicine Project grant U2C DK114886.

REFERENCES

- [1]. Corey Magee, D.J.G., Chris J. Watson, Derek P. Brazil, *Diabetic Nephropathy: a Tangled Web to Unweave*. Cardiovascular Drugs and Therapy, 2017. **31**(5-6): p. 13.
- [2]. Tervaert, T.W.C., et al., *Pathologic Classification of Diabetic Nephropathy*. Journal of the American Society of Nephrology, 2010. **21**: p. 556-563.
- [3]. Brandon Ginley, B.L., Kuang-Yu Jen, Agnes Fogo, Sanjay Jain, Avi Rosenberg, Vignesh Walavalkar, Gregory Wilding, John E. Tomaszewski, Rabi Yacoub, Giovanni Maria Rossi, Pinaki Sarder, *Computational Segmentation and Classification of Diabetic Glomerulosclerosis*. Journal of the American Society of Nephrology, 2019: p. 17.
- [4]. Gower, J.C. and G.J.S. Ross, *Minimum Spanning Trees and Single Linkage Cluster Analysis*. Journal of the Royal Statistical Society. Series C (Applied Statistics), 1969. **18**(1): p. 54-64.
- [5]. Jessica Lovelace Rainbolt, M.S., *The Use of Minimal Spanning Trees in Particle Physics*. Journal of Instrumentation, 2017. **12**(2): p. 29.
- [6]. Ginley, B., et al., *Computational segmentation and classification of diabetic glomerulosclerosis*. Journal of the American Society of Nephrology : JASN, To appear.
- [7]. Lutnick, B., 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.
- [8]. Arnout C. Ruifrok, D.A.J., *Quantification of histological staining by color deconvolution*. Analytical and Quantitative Cytology and Histology, 2001. **23**: p. 8.
- [9]. Simon, O., et al., *Examining structural changes in diabetic nephropathy using inter-nuclear distances in glomeruli: a comparison of variously automated methods*. Proceedings of SPIE (SPIE Medical Imaging 2018: Digital Pathology), 2018. **10581**: p. 105810B: 1-10.
- [10]. Freeman, L.C., *A Set of Measures of Centrality Based on Betweenness*. Sociometry, 1977. **40**(1): p. 6.
- [11]. P. Twarie, E.v.D., A. Hillebrand, C.J. Stam, *The minimum spanning tree: An unbiased method for brain network analysis*. NeuroImage, 2015. **104**: p. 11.
- [12]. Timo J. T. Koski, J.M.N., *A Review of Bayesian Networks and Structure Learning*. Mathematica Applicanda, 2012. **40**(1): p. 50.

- [13]. Scutari, M., *Learning Bayesian Networks with the bnlearn R Package*. Journal of Statistical Software, 2010. **35**(3): p. 21.
- [14]. Moral, P.D., *Nonlinear filtering: Interacting particle resolution*. Comptes Rendus de l'Académie des Sciences, 1997. **325**(6): p. 5.