

UC Davis

UC Davis Previously Published Works

Title

Examining structural patterns and causality in diabetic nephropathy using inter-glomerular distance and Bayesian graphical models

Permalink

<https://escholarship.org/uc/item/0rz8t5vv>

Authors

Majumdar, Aurijoy
Jen, Kuang-Yu
Jain, Sanjay
et al.

Publication Date

2019-02-01

DOI

10.1117/12.2513598

Peer reviewed



Published in final edited form as:

Proc SPIE Int Soc Opt Eng. 2019 February ; 10956: . doi:10.1117/12.2513598.

Examining Structural Patterns and Causality in Diabetic Nephropathy using inter-Glomerular Distance and Bayesian Graphical Models

Aurijoy Majumdar¹, Kuang-Yu Jen², Sanjay Jain³, John E. Tomaszewski¹, and Pinaki Sarder¹

¹Departments of Pathology and Anatomical Sciences, University at Buffalo

²Departments of Pathology, University at California at Davis

³Department of Medicine, Washington University School of Medicine in St. Louis

Abstract

In diabetic nephropathy (DN), hyperglycemia drives a progressive thickening of glomerular filtration surfaces, increased cell proliferation as well as mesangial expansion and a constriction of capillary lumens. This leads to progressive structural changes inside the Glomeruli. In this work, we make a study of structural glomerular changes in DN from a graph-theoretic standpoint, using features extracted from Minimal Spanning Trees (MSTs) constructed over intercellular distances in order to classify the “packing signatures” of different DN stages. We further investigate the significance of the competing effects of Volume change measured here in 2Dimensional Pixel span area (Area) on one hand and increased cell proliferation on the other in determining the packing patterns. Towards that we formulate the problem as Dynamic Bayesian Network (DBN). From our preliminary results we do postulate that volume expansion caused by internal pressure as capillary lumens constriction has perhaps has a greater effect in the early stages.

Keywords

Diabetic nephropathy; whole slide image analysis; Minimum Spanning Tree; Support Vector Machine; Graphical Models; Dynamic Bayesian Network; Medical Image processing

I. INTRODUCTION

Graph theoretic analysis is suitable for discerning some of the structural changes in Glomeruli that ensue as a result of DN progression. We look into Minimum Spanning Trees (MSTs) a graph-theoretical construct whereby the points of a network are joined using the shortest total connection distance, which have already proven useful for characterization of subtle structural changes in diverse fields^[8,10] as well as disease classification tasks^[1,11,12,14]. In particular, MSTs have the advantage of being independent of anisotropic image characteristics^[13]. We posit that MSTs or rather the change in specific characteristics

*Address all correspondence to: Pinaki Sarder; Tel: 716-829-2265, pinakisa@buffalo.edu.

(depth, avg. width, clustering tendency etc.) of MST structures might be synonymous with packing characteristics of the Glomeruli thereby signaling altered intra-Glomeruli dynamics and consequently DN Progression. We use Bayesian Graphical Models namely Dynamic Bayesian Networks (DBN) to represent the phenomenon of DN progression parametrically. We try to infer the correlation/causation at every DN stage between spatial parameters, such as Volume change (expressed as pixel span area) and increased cellularity as observed through the lens of MST.

II. METHODS & RESULTS

Imaging and data preparation.

Data consisted of 49 samples of Glomeruli with a DN stage of 1 (normal), with 47 glomeruli with a stage of 2 (mild progression) and 42 with stage3 (significant progression).

MST Feature Extraction.

After having constructed the MST by connecting centroids of nuclei inside the Glomeruli in the images, we extract features over the MST which signify by and large packing density patterns in the Glomeruli such as depth of tree, avg. nuclei per level of tree, no. of leaves per level, clustering features such as k-means clusters, Single-link clusters etc. in addition to averaged features such as avg. MST edge distance, avg. inter-nuclei distance for all nuclei in the Glomerulus and global features such as nuclei count and pixel span area. The MST for a Glomerulus would be unique and correspond to some extent, to the intra-Glomerulus cell proliferation dynamics. We postulate that the certain structural packing features as examined through the Glomeruli MST, would be characteristic of the stage of DN in progression.

Verification using Support Vector Machine.

In an effort to verify the veracity of our feature set we use it to train and test three Support Vector Machines (SVM) with Linear, Gaussian and RBF Kernel over stages 1 vs 2, 2 vs 3 and 1 vs 3 . With limited parameter tuning the results were 77%, 72% and 78% respectively. In each case we rank the support vectors and find that the area or rather change in area is the most important support vector in cases 1 vs 2 and 1 vs3, followed by cellularity. In case of 2 vs 3 we find k-means, area and cellularity to be close in rank. The distribution of area in the first two stages as well as Precision and Recall results are shown.

Formulating and Evaluating Dynamic Bayesian Graph.

After observing that the features extracted from the MST do have some statistical significance in predicting DN stage, we proceed to have a first attempt at solving a fundamental morphological question regarding the structural alterations observed in the Glomeruli. It is speculated that in progressing through the stages, glomerular expansion occurs from increased internal pressure as capillary lumens constrict and sclerose, which should increase overall inter-nuclear distance slightly whereas there is a competing phenomenon of increased number of nuclei which should decrease the overall inter nuclei distance. It is postulated that going from Stage 1 to 2 the volume expansion takes precedence over the cell proliferation effect in deciding the packing signature. In contrast in the later

stage 3, the increased cellularity in the expanded mesangial space has a more significance in determining the packing signature of the Glomeruli and consequently the MST structure.

In an attempt to order the predicted causes of altered cell structure of glomeruli in the order of influence, we encode the problem as a Dynamic Bayesian Network in order to capture the order of precedence in causes separately in the three stages. We proceed to model the two competing variables Volume expansion (modeled in 2D as pixel area span) which should correlate to an overall increase in additive quantities such as total MST edge length etc. and cellular proliferation which should act in the opposing direction to reduce the aforementioned quantities. The problem is then further divided into two stages Structure Learning and Parameter learning.

Structure Learning.

For learning the structure learning we use the paradigm of Maximum Likelihood Estimation over Maximum A Posteriori estimation. We also include a Bayesian Information Criterion as a regularization criterion for avoiding too complex formulations of the structure of the Bayesian Network. Thereon, we use Hill-Climbing Search over graph construction operators.

Parameter Learning.

In order to learn the parameters of the network structure we also use Maximum Likelihood Estimation over Maximum A Posteriori Estimation. This is due to ease of implementation.

Preliminary results analyzed from features obtained from three sets of Disease stages (1, 2 & 3) show a Dynamic Bayesian Network Structure which was learnt purely from the data which show that the effect of Pixel Area expansion (Volume) influences the packing pattern in the Glomeruli more so than the Cell proliferation initially in transitioning from Stage 1 to 2. Thereafter, the both of them have similar influence in stage 3. In the DBN Conditional Probability Distribution we find the edge weights of Area connecting other features having more weight than the Cellularity in stage 1 and 2, whereas in stage 3 the edge weights are comparable

III. DISCUSSION

The work represents a first study of useful structural features corresponding to the packing signature of each stage in DN. We successfully design and extract important clustering features indicative of the packing patterns. In this regard we speculate that competing effects of Glomerular expansion caused due to pressure from capillary constriction and intercellular distance reduction caused due to cellular proliferation do have different signatures in the cell packing process.

In addition we try to speculate which cause has greater influence over the cell distribution and packing in each stage of the disease and whether the influence changes over the progression timeline. Here we employ probabilistic graphical modeling in the form of Dynamic Bayesian networks in an attempt to probe into the answer. From preliminary observations we posit that in the initial stages the glomerular Volume expansion does parry

over cell proliferation in determining the packing pattern. Only in stage 3 does both of them reach equilibrium in their influence over the packing.

IV. CONCLUSION AND FUTURE WORKS

In conclusion this study tries a first pass in employing a novel technique using combinatorial features extracted from MSTs drawn over Glomeruli nuclei in order to classify DN stages (1,2 & 3). We essentially propose a pipeline in order to semi-automatically derive image features and construct meaningful combinatorial structures in the form of MSTs which pertain to packing signatures of intra-Glomerular dynamics which in turn are mainly governed by the stage of progression of DN. We show MSTs exhibit key features which can be used to classify DN stage. We wish to extract more intricate features such as degree of divergence, cluster coefficients, tree hierarchy overlap in order to gain further insights into the change in packing dynamics inside the Glomeruli.

In addition, we aim to solve the problem of causality statistically which is basically to find whether volume expansion or cellular proliferation is the primary cause of perturbation in Glomerular structure. We would like to improve upon the results and construct better graphical explanations of the data by improving features and most importantly encoding priors in the form of Multinomial and Gamma Distribution Priors and use marginal Likelihood Estimation which is much better suited for limited samples like our case.

ACKNOWLEDGEMENT

The project was supported by the faculty startup funds from the Jacobs School of Medicine and Biomedical Sciences, University at Buffalo, the University at Buffalo IMPACT award, NIDDK Diabetic Complications Consortium grant DK076169, and NIDDK grant R01 DK114485.

REFERENCES

- [1]. Magee C, Grieve DJ, Watson CJ et al., “Diabetic Nephropathy: a Tangled Web to Unweave,” *Cardiovasc Drugs Ther*, 31(5-6), 579–592 (2017). [PubMed: 28956186]
- [2]. Liu ZZ, Bullen A, Li Y et al., “Renal Oxygenation in the Pathophysiology of Chronic Kidney Disease,” *Front Physiol*, 8, 385 (2017). [PubMed: 28701959]
- [3]. Zeni L, Norden AGW, Cancarini G et al., “A more tubulocentric view of diabetic kidney disease,” *J Nephrol*, 30(6), 701–717 (2017). [PubMed: 28840540]
- [4]. National Chronic Kidney Disease Fact Sheet, 2014 Available from: https://www.cdc.gov/diabetes/pubs/pdf/kidney_factsheet.pdf
- [5]. Papadopoulou-Marketou N, Chrousos GP and Kanaka-Gantenbein C, “Diabetic nephropathy in type I diabetes: a review of early natural history, pathogenesis, and diagnosis,” *Diabetes/ Metabolism Research and Reviews*, 33(2), e2841–5 (2017).
- [6]. Tervaert TW, Mooyaart AL, Amann K. et al., “Pathologic classification of diabetic nephropathy,” *J Am Soc Nephrol*, 21(4), 556–63 (2010). [PubMed: 20167701]
- [7]. Kitada M, Ogura Y and Koya D, “Rodent models of diabetic nephropathy: their utility and limitations,” *Int J Nephrol Renovasc Dis*, 9, 279–290 (2016). [PubMed: 27881924]
- [8]. Dussert C, Rasigni G, Rasigni M et al., “Minimal spanning tree: A new approach for studying order and disorder,” *Physical Review B*, 34(5), 3528–3531 (1986).
- [9]. Rainbolt JL and Schmitt M, “The use of minimal spanning trees in particle physics,” *Journal of Instrumentation*, 12(02), P02009 (2017).
- [10]. Tewarie P, van Dellen E, Hillebrand A et al., “The minimum spanning tree: An unbiased method for brain network analysis,” *Neuroimage*, 104, 177–188 (2015). [PubMed: 25451472]

- [11]. Wan T, Cao J, Chen J et al., "Automated grading of breast cancer histopathology using cascaded ensemble with combination of multi-level image features," *Neurocomput*, 229(C), 34–44 (2017).
- [12]. Doyle S, Agner S, Madabhushi A et al., "Automated grading of breast cancer histopathology using spectral clustering with textural and architectural image features." 496–499.
- [13]. Coleman K, van Diest PJ, Baak JP et al., "Syntactic structure analysis in uveal melanomas," *Br J Ophthalmol*, 78(11), 871–4 (1994). [PubMed: 7848986]
- [14]. Choi H-K, Bengtsson E, Jarkrans T et al., "Minimum spanning trees (MST) as a tool for describing tissue architecture when grading bladder carcinoma." 615–620.
- [15]. Tesch GH and Allen TJ, "Rodent models of streptozotocin-induced diabetic nephropathy," *Nephrology (Carlton)*, 12(3), 261–6 (2007). [PubMed: 17498121]
- [16]. Ginley B, Tomaszewski JE, Yacoub R et al., "Unsupervised labeling of glomerular boundaries using Gabor filters and statistical testing in renal histology," *Journal of Medical Imaging*, 4(2), 021102: 1–12 (2017).
- [17]. Ruifrok AC and JOHnston DA, "Quantification of histochemical staining of color deconvolution," *Anal Quant. Cytol. Histol*, 23(4), 291–299 (2001). [PubMed: 11531144]
- [18]. Koller Daphne and Friedman Nir "Probabilistic Graphical Models" 2009
- [19]. Hastie, Tibshirani and Friedman "Elements of Statistical Learning" 2009

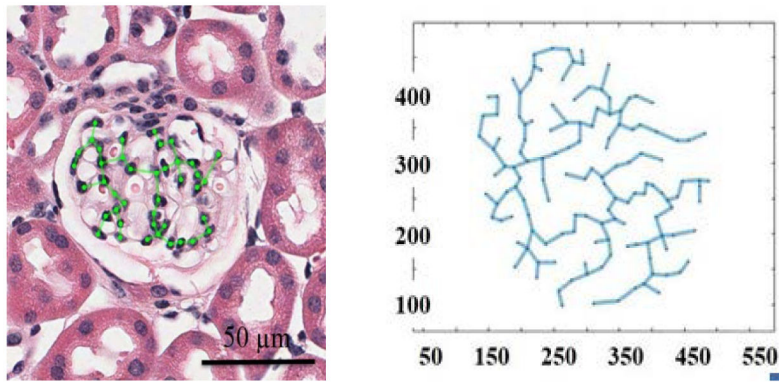


Fig 1a Minimum Spanning Tree constructed from Glomeruli. Fig 1b MST connecting the centroids of the nuclei.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

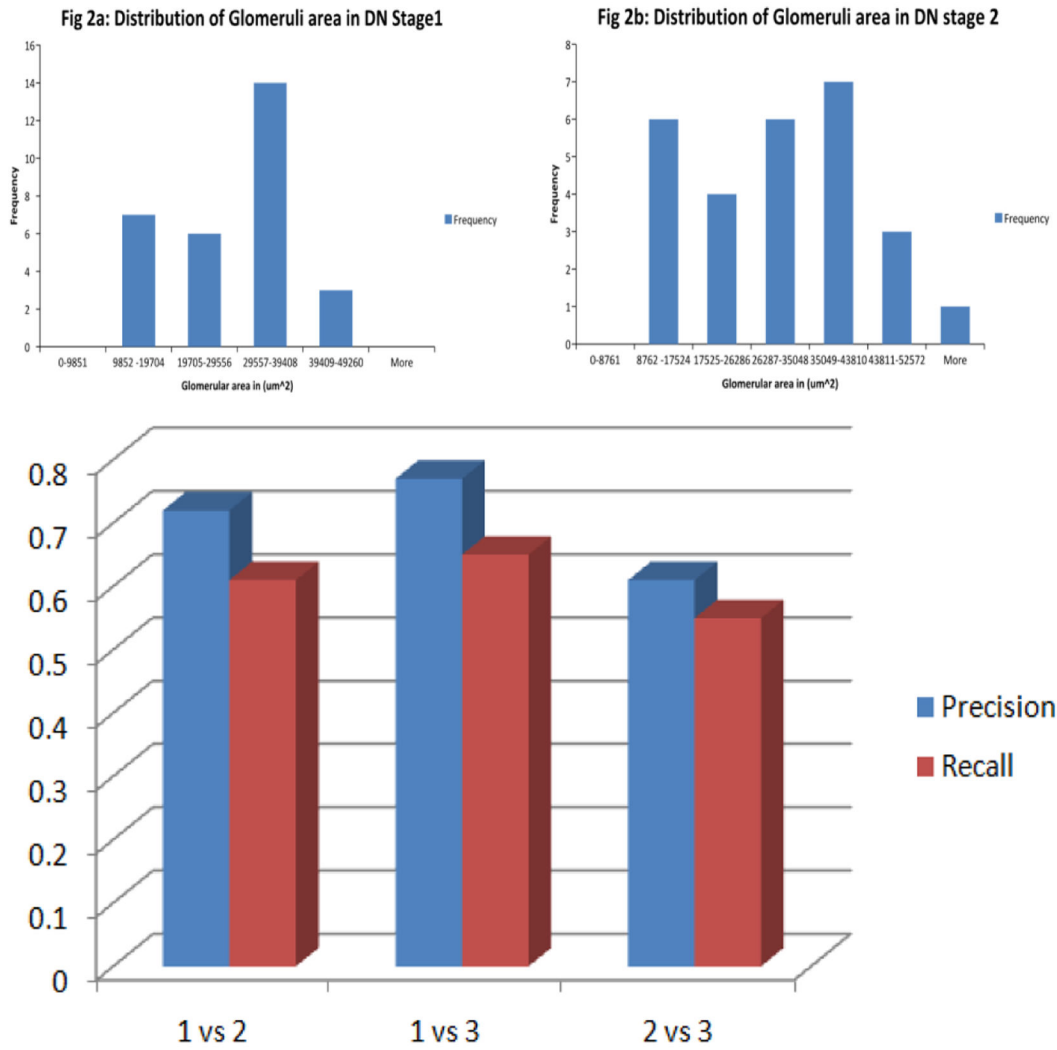


Fig 2a and Fig 2b show the Distribution of Glom Area which is the most relevant support vector in 1 vs 2 case and 1 vs 3 case. Fig 2c represents precision vs recall value for all the SVM experiments. Preliminary results show that area, cellularity and k-clusters are the most important support vectors.

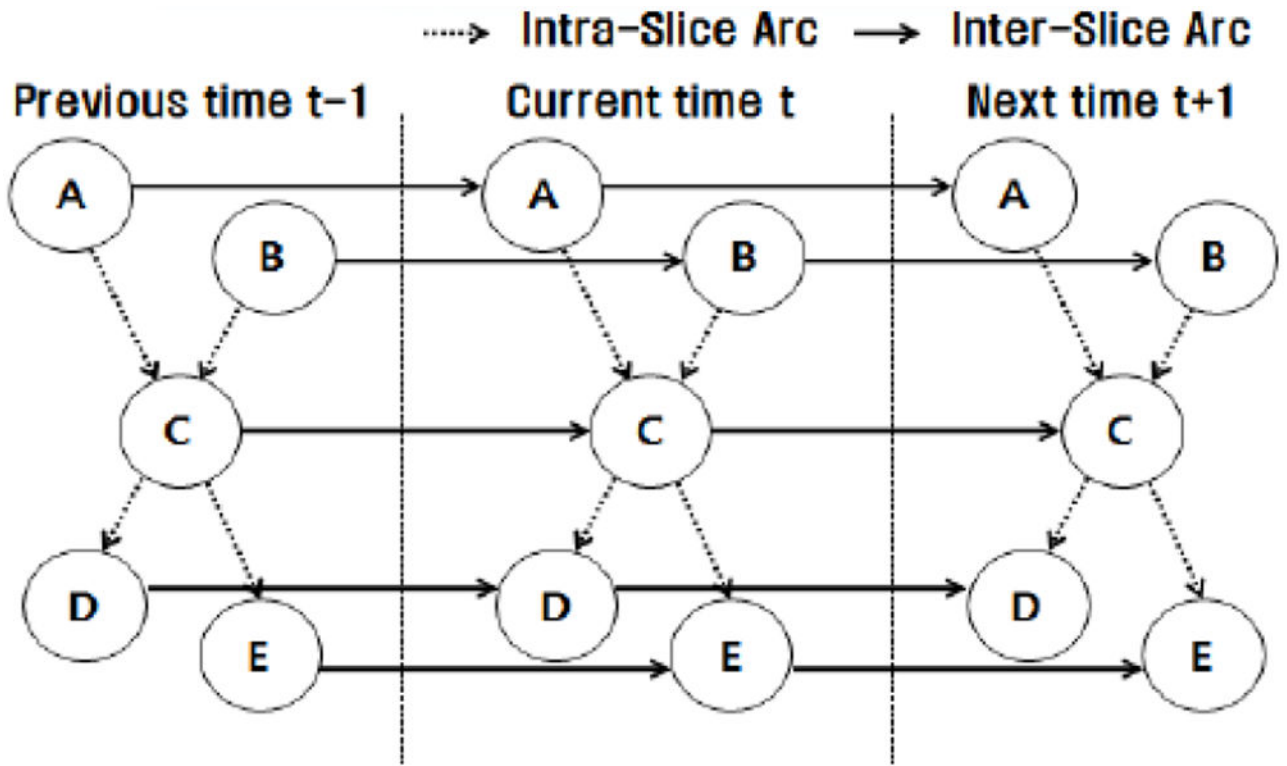
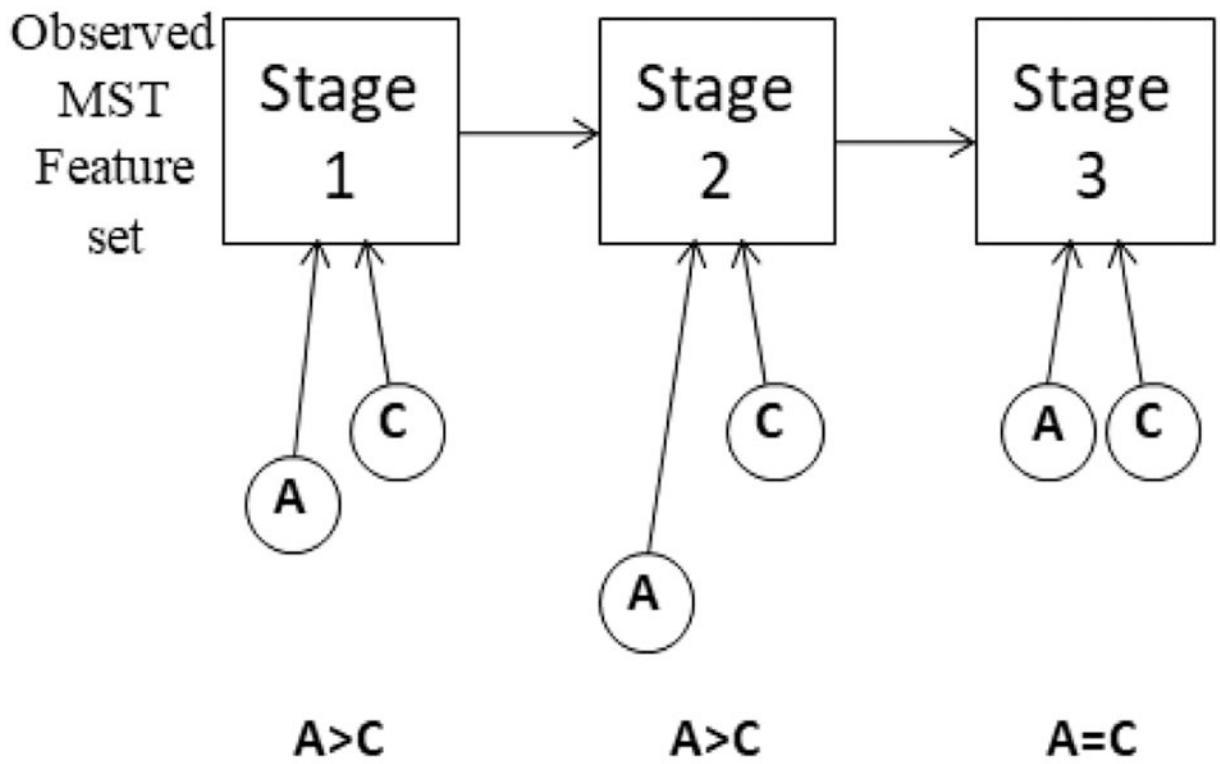


Fig3 represents an example of a candidate Dynamic Bayesian network where the letter and corresponding parameters are respectively: A: Area & Θ_A : Distribution in Area within that time slice. B: Cellularity & Θ_B : Distribution of Cellularity within time slice. C: Packing Distribution & Θ_C : Distribution of Packing Distribution within that time slice. D: Clusters and Θ_D : Distribution of cluster coefficient within time slice. E: MST edge length and Θ_E : Distribution of MST edge length within a time slice. The time $t-1$, t , $t+1$ slices themselves represent Disease stages 1, 2 and 3. The inter time slice connections are transition weights which will remain stationary for simplicity.



A: Volume (area span)

C: Nuclei Count

Fig 4 describes the relative edge weight in the three stages.