End-to-End Glomerulonephritis Diagnosis by Machine Learning on Periodic Acid-Schiff (PAS) Sections with the MILxFormer Architecture

Pietro A. Cicalese, MSc,[1] Huy Q. Vo,[1] Syed Rizvi, MSc,[1] Sándor Turkevi-Nagy, MD,[2] Jean-Baptiste Gibier, Prof, [3] Prof, Surya Seshan, Prof,[4] Prof. Savino Sciascia, Prof,[5] Samira Zare, MSc,[1] Gloria Bueno, Prof,[6] Anibal Pedraza,[6] Katharina Stolle,[7] Mahmoud Abbas, MD,[8] Bernd Schröppel, Prof,[9] Nicola Altini, MSc,[10] Vitoantonio Bevilacqua, Prof,[10] Paola Pontrelli, Prof,[11] Francesco Pesce, Prof,[11,12] , Loreto Gesualdo, Prof,[11] Avi Rosenberg, Prof,[13] , Michele Rossini, MD,[11] MD, Chandra Mohan*, Prof,[1] Prof., Hien V. Nguyen*,Prof,[1] Jan U. Becker*, MD[14]

[1] Department of Biomedical Engineering, University of Houston, Houston, USA;[2] Department of Pathology, Albert Szent-Györgyi Health Centre, University of Szeged, Szeged, Hungary; [3] Department of Pathology, Pathology Institute, Lille University Hospital (CHU), Lille, France; [4] Department of Pathology, Weill-Cornell Medical Center/New York Presbyterian Hospital, New York, NY, USA; [5] University Center of Excellence on Nephrologic, Rheumatologic and Rare Diseases with Nephrology and Dialysis Unit and Center of Immuno-Rheumatology and Rare Diseases (CMID), Coordinating Center of the Interregional Network for Rare Diseases of Piedmont and Aosta Valley, San Giovanni Bosco Hub Hospital, Turin, Italy; [6] VISILAB Research Group, University of Castilla–La Mancha, Ciudad Real, Spain
[7]Sektion Nephrologie, Klinik für Innerere Medizin 1, Universität zu Bonn, Germany; [8] Gerhard-Domagk Institute of Pathology, University Hospital of Münster, Germany; [9] Department of Nephrology, University Hospital of Ulm, Germany; [10]Department of Electrical and Information Engineering (DEI), Polytechnic University of Bari, 70126 Bari, BA, Italy; [11] Nephrology, Dialysis and Transplantation Unit, Department of Emergency and Organ Transplantation, University of Bari Aldo Moro, Bari, Italy; [12] Division of Renal Medicine, "Fatebenefratelli Isola Tiberina – Gemelli Isola", Rome, Italy; [13] Department of Pathology, Johns Hopkins University, Baltimore, USA; [14] Institute of Pathology, University Hospital of Cologne, Cologne, Germany
***authors contributed equally**

Background and Aims:

Machine learning (ML) holds great promise for improving diagnostics, prognostication and theranostics in nephropathology. So far, applications have not gone much further than segmentation of tissue compartments on whole slide images (WSIs) of paraffin sections. As a proof-of-concept study, we describe the development of a diagnostic classifier for glomerulephritis based on expert-annotated or automatically segmented glomerular transections from periodic-acid Schiff (PAS) paraffin sections only.

General Concept: This end-to-end pipeline extracts with the first module (segmentation) glomerular crops from a PAS WSI, generating a "bag of glomerular crops": the second module then uses the Bayesian MILxFormer architecture to predict the glomerulonephritis class.

Method:

A total of n=350 biopsies from 5 institutions with 12 classes of glomerulonephritis IgA nephropathy (IgAN), membranous nephropathy (Membranous), anti-glomerular basement membrane antibody GN (ABMGN), infection-associated GN (IAGN), ANCA-associated GN (ANCA-GN), idiopathic membranoproliferative GN (MPGN), SLE GN class IV (SLE-GN-IV), cryglobulinemic GN (CryoGN), C3 GN (C3-GN), dense deposit disease (DDD), fibrillary GN (FibrillaryGN) and proliferative GN with monoclonal immunoglobulin deposits (PGNMID) were included in the study with their respective PAS sections. Glomerular transections (crops) were expert-annotated by a nephropathologist and automatically segmented with our own transformer-based segmentation model trained on 100 biopsies with thrombotic microangiopathies and a range of vascular, vasculitic and glomerular diseases closely resembling/mimicking thrombotic microangiopathies.

For classification, we divided the cohort into 5 folds for internal cross-validation, performed sample size augmentation with various methods (including shifts in resolution/scale, AutoAugment and others) and trained our proprietary self-attention-based MILx architecture on an EfficientNet backbone with selection of glomerular crop batches by soft Markov chain Monte Carlo sampling in a semi-supervised fashion, with diagnostic class labels for each biopsy.

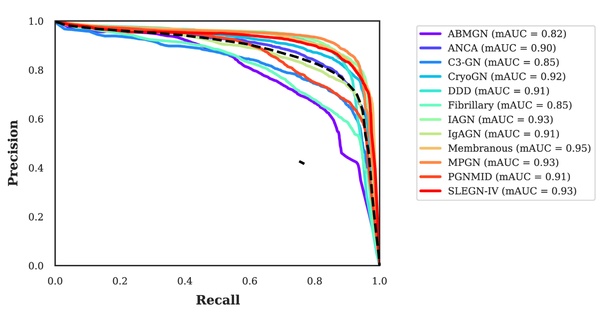
We compared the performance of our proprietary architecture on both expert-annotated and automatically segmented glomerular crops with a recently published benchmark architecture (CLAM) for multiple-instance learning in histopathology.

Results:

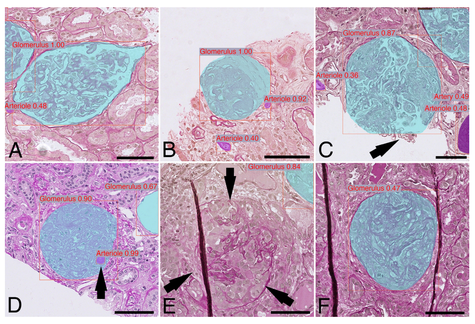
Automatic glomerular segmentation performance was excellent with mean AUC and sensitivity (mean average recall) over all classes at 0.904, with near perfect mean average specificity (0.994), as expected best for Membranous, worst for ABMGN.

Classification performance of MILx with expert-annotated glomerular crops as inputs had a mean balanced accuracy of 0.84, with AUC metrics in descending order of 0.97 for Membranous, 0.89 for ABMGN, 0.88 for IgAN, 0.86 for Fibrillary, 0.83 for MPGN, 0.80 for ANCA-GN, 0.79 for DDD, 0.78 for PGNMID, 0.75 for IAGN, 0.73 for SLE-GN-IV and CryoGN, 0.67 for C3-GN. Performance with MILx was similar for automatically segmented glomerular crops as input

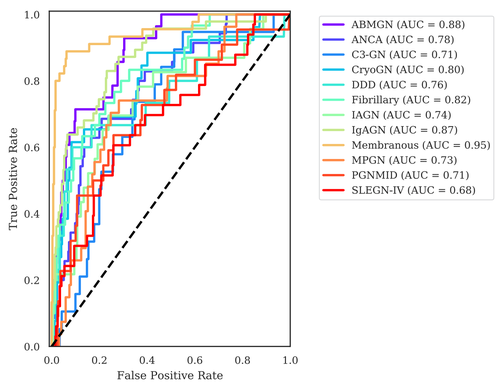
On this dataset, MILx outperformed CLAM with both entire WSIs as well as expert-annotated glomerular crops as inputs (mean balanced accuracy of 0.72) by a significant margin.



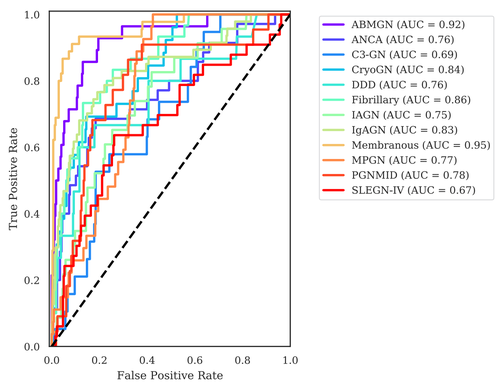
Performance of the glomerular segmentation module on the entire set of 350 PanGN biopsies as precision/recall curves.



Exemplary glomerular predictions (automatic segmentations) and mis-predictions with their certainty generated by the first module of our end-to-end diagnostic pipeline.



Mean receiver operating characteristics (ROC) for the MILxFormer classifier module when applied to expert segmented glomeruli.



Mean receiver operating characteristics (ROC) for the MILxFormer classifier module when applied to automatically segmented glomeruli.

GT Diagnostic Class	N	Pred	PPV	NPV	Sens	Spec	F1 Score	Accuracy	Balanced Accuracy
ABMGN	28	Top-1	46.43	95.34	46.43	95.34	46.43	46.43	70.89
		Top-2	71.43	97.52	71.43	97.52	71.43	71.43	84.47
		Top-3	82.14	98.45	82.14	98.45	82.14	82.14	90.3
ANCA	35	Top-1	30.56	92.36	31.43	92.06	30.99	31.43	61.75
		Top-2	56.76	95.53	60.0	94.92	58.33	60.0	77.46
		Top-3	66.67	96.5	68.57	96.19	67.61	68.57	82.38
C3-GN	20	Top-1	6.25	94.61	5.26	95.47	5.71	5.26	50.37
		Top-2	42.86	96.13	31.58	97.58	36.36	31.58	64.58
		Top-3	61.11	97.59	57.89	97.89	59.46	57.89	77.89
CryoGN	27	Top-1	36.0	94.77	34.62	95.06	35.29	34.62	64.84
		Top-2	51.72	96.57	57.69	95.68	54.55	57.69	76.69
		Top-3	62.07	97.51	69.23	96.6	65.45	69.23	82.92
DDD	15	Top-1	37.5	96.49	20.0	98.51	26.09	20.0	59.25
		Top-2	60.0	97.35	40.0	98.81	48.0	40.0	69.4
		Top-3	66.67	97.36	40.0	99.1	50.0	40.0	69.55
Fibrillary	30	Top-1	33.33	95.36	53.33	90.0	41.03	53.33	71.67
		Top-2	48.78	96.76	66.67	93.44	56.34	66.67	80.05
		Top-3	62.16	97.76	76.67	95.63	68.66	76.67	86.15
IAGN	23	Top-1	23.81	94.53	21.74	95.11	22.73	21.74	58.42
		Top-2	33.33	94.88	26.09	96.33	29.27	26.09	61.21
		Top-3	47.06	95.5	34.78	97.25	40.0	34.78	66.02
IgAGN	47	Top-1	61.54	92.6	51.06	95.05	55.81	51.06	73.06
		Top-2	75.0	95.42	70.21	96.37	72.53	70.21	83.29
		Top-3	79.07	95.77	72.34	97.03	75.56	72.34	84.69
Membranous	45	Top-1	73.58	97.98	86.67	95.41	79.59	86.67	91.04
		Top-2	82.0	98.67	91.11	97.05	86.32	91.11	94.08
		Top-3	82.0	98.67	91.11	97.05	86.32	91.11	94.08
MPGN	27	Top-1	17.65	93.35	22.22	91.33	19.67	22.22	56.78
		Top-2	37.5	95.28	44.44	93.81	40.68	44.44	69.13
		Top-3	51.61	96.55	59.26	95.36	55.17	59.26	77.31
PGNMID	22	Top-1	16.67	94.08	9.09	96.95	11.76	9.09	53.02
		Top-2	55.0	96.67	50.0	97.26	52.38	50.0	73.63
		Top-3	66.67	97.57	63.64	97.87	65.12	63.64	80.75
SLEGN-IV	33	Top-1	23.33	91.87	21.21	92.74	22.22	21.21	56.98
		Top-2	40.74	93.19	33.33	94.95	36.67	33.33	64.14
		Top-3	54.84	94.98	51.52	95.58	53.12	51.52	73.55
All Classes	350	Top-1	38.15	94.37	38.86	94.18	37.95	38.86	66.52
		Top-2	57.31	96.16	57.71	95.99	57.11	57.71	76.85
		Top-3	66.95	96.98	67.14	96.85	66.68	67.14	81.99

Classification performance for the MILxFormer classification module on expert annotated (segmented) glomeruli (all values presented as percentages).

		Prediction Diagnostic Class										
GT Diagnostic Class	Pred	ABMGN	ANCA	C3-GN	Cryo-GN	DDD	Fibrillary	IAGN	IgAGN	Mem-branous	MPGN	PGNMID	SLEGN-IV
ABMGN	Top-1	13	7	0	0	0	2	1	0	1	1	0	3
	Top-2	20	3	0	0	0	1	0	0	1	1	0	2
	Top-3	23	1	0	0	0	0	0	0	1	1	0	2
ANCA	Top-1	4	11	0	0	1	2	3	3	4	3	0	4
	Top-2	1	21	0	0	1	2	3	1	2	3	0	1
	Top-3	1	24	0	0	1	2	3	0	2	2	0	0
C3-GN	Top-1	1	1	1	2	0	4	1	0	1	6	0	2
	Top-2	1	1	6	2	0	2	1	0	1	4	0	1
	Top-3	1	1	11	2	0	0	0	0	1	3	0	0
CryoGN	Top-1	1	1	3	9	1	3	0	1	0	5	1	1
	Top-2	1	0	2	15	1	3	0	1	0	1	1	1
	Top-3	1	0	2	18	0	3	0	1	0	0	0	1
DDD	Top-1	2	0	0	2	3	2	1	1	0	3	0	1
	Top-2	0	0	0	2	6	1	1	1	0	3	0	1
	Top-3	0	0	0	2	6	1	1	1	0	3	0	1
Fibrillary	Top-1	3	1	0	0	0	16	1	1	5	1	1	1
	Top-2	2	1	0	0	0	20	1	1	2	1	1	1
	Top-3	0	1	0	0	0	23	1	1	2	0	1	1
IAGN	Top-1	1	1	1	2	0	2	5	3	0	3	2	3
	Top-2	1	1	1	2	0	2	6	2	0	3	2	3
	Top-3	0	1	1	2	0	2	8	2	0	2	2	3
IgAGN	Top-1	0	9	3	0	0	5	2	24	1	0	0	3
	Top-2	0	5	2	0	0	2	2	33	1	0	0	2
	Top-3	0	5	2	0	0	2	1	34	1	0	0	2
Mem-branous	Top-1	0	0	0	0	0	2	0	3	39	0	1	0
	Top-2	0	0	0	0	0	1	0	2	41	0	1	0
	Top-3	0	0	0	0	0	1	0	2	41	0	1	0
MPGN	Top-1	0	0	3	2	2	2	2	0	0	6	5	5
	Top-2	0	0	1	2	1	2	1	0	0	12	4	4
	Top-3	0	0	1	0	1	1	1	0	0	16	3	4
PGNMID	Top-1	0	1	2	5	0	7	1	0	1	3	2	0
	Top-2	0	1	0	3	0	4	0	0	1	2	11	0
	Top-3	0	1	0	3	0	1	0	0	1	2	14	0
SLEGN-IV	Top-1	3	4	3	3	1	1	4	3	1	3	0	7
	Top-2	2	4	2	3	1	1	3	3	1	2	0	11
	Top-3	2	2	1	2	1	1	2	2	1	2	0	17

Confusion matrix for the MILxFormer classification module when applied to expert segmented glomerular crops. Correct predictions (Pred) are highlighted in green, the most common mis-predictions in red.

Conclusion:

This proof-of-concept-study indicates that nephropathology-specific architectures like our MILx can be trained for complex tasks on relatively small biopsy cohorts. We should be able to deliver an end-to-end-pipeline for this diagnostic and other tasks based on training sets with case-labels provided by trusted institutions with only minimal expert labeling or annotation required.