

Automated Prediction of Kidney Failure in IgA Nephropathy with Deep Learning from Biopsy Images

Francesca Testa,¹ Francesco Fontana,¹ Federico Pollastri,² Johanna Chester,³ Marco Leonelli,¹ Francesco Giaroni,¹ Fabio Gualtieri,³ Federico Bolelli ², Elena Mancini,⁴ Maurizio Nordio,⁵ Paolo Sacco,⁶ Giulia Ligabue,³ Silvia Giovannella,³ Maria Ferri ³, Gaetano Alfano ¹, Loreto Gesualdo,⁷ Simonetta Cimino,⁸ Gabriele Donati,^{1,3} Costantino Grana,² and Riccardo Magistroni ^{1,3}

Abstract

Background and objectives Digital pathology and artificial intelligence offer new opportunities for automatic histologic scoring. We applied a deep learning approach to IgA nephropathy biopsy images to develop an automatic histologic prognostic score, assessed against ground truth (kidney failure) among patients with IgA nephropathy who were treated over 39 years. We assessed noninferiority in comparison with the histologic component of currently validated predictive tools. We correlated additional histologic features with our deep learning predictive score to identify potential additional predictive features.

Design, setting, participants, & measurements Training for deep learning was performed with randomly selected, digitalized, cortical Periodic acid–Schiff–stained sections images (363 kidney biopsy specimens) to develop our deep learning predictive score. We estimated noninferiority using the area under the receiver operating characteristic curve (AUC) in a randomly selected group (95 biopsy specimens) against the gold standard Oxford classification (MEST-C) scores used by the International IgA Nephropathy Prediction Tool and the clinical decision supporting system for estimating the risk of kidney failure in IgA nephropathy. We assessed additional potential predictive histologic features against a subset (20 kidney biopsy specimens) with the strongest and weakest deep learning predictive scores.

Results We enrolled 442 patients; the 10-year kidney survival was 78%, and the study median follow-up was 6.7 years. Manual MEST-C showed no prognostic relationship for the endocapillary parameter only. The deep learning predictive score was not inferior to MEST-C applied using the International IgA Nephropathy Prediction Tool and the clinical decision supporting system (AUC of 0.84 versus 0.77 and 0.74, respectively) and confirmed a good correlation with the tubulointerstitial score ($r=0.41$, $P<0.01$). We observed no correlations between the deep learning prognostic score and the mesangial, endocapillary, segmental sclerosis, and crescent parameters. Additional potential predictive histopathologic features incorporated by the deep learning predictive score included (1) inflammation within areas of interstitial fibrosis and tubular atrophy and (2) hyaline casts.

Conclusions The deep learning approach was noninferior to manual histopathologic reporting and considered prognostic features not currently included in MEST-C assessment.

CJASN 17: 1316–1324, 2022. doi: <https://doi.org/10.2215/CJN.01760222>

Due to the number of contributing authors, the affiliations are listed at the end of this article.

Correspondence:

Dr. Riccardo Magistroni, Division of Nephrology, Dialysis and Renal Transplantation, Azienda Ospedaliera Universitaria Policlinico di Modena, Via del Pozzo 71, 41124 Modena, Italy. Email: riccardo.magistroni@unimore.it

Introduction

The Oxford classification (MEST-C) (1–3) is the current gold standard for the histologic prognostic definition of IgA nephropathy. However, because the quality of current evidence is mostly on the basis of retrospective studies, most recent clinical guidelines do not suggest its application for clinical adoption in guiding treatment (4). Prospective, robust data on the Oxford classification in formal clinical trials are missing mostly due to its relatively new introduction (2009) with recent important revisions (2017) (3), high costs associated with complex histologic evaluations, and poor intra- and inter-reproducibility among pathologists (1,2,5–9).

Digital pathology has enabled innovative analyses using emerging technologies, including artificial intelligence (AI). AI offers a unique opportunity for efficient development of scoring systems. Advantages of AI may include low implementation costs and high reproducibility—two areas of current weakness in manual histologic scoring. However, before clinical adoption, AI algorithms must be developed, tested, and rigorously validated.

Automated pathologic analysis using deep learning, a branch of AI, enables the development of innovative applications in kidney pathology (10–12). Deep learning uses convolutional neural networks (CNNs),

artificial counterparts of the neural networks of the brain, enabling computer learning without explicit programming. The computer is exposed to examples of desired input-output behavior, from which it automatically extracts features that are useful for a particular task (13). Previously, we effectively used a deep learning approach to automatically extract classification features from kidney biopsy immunofluorescence images (14). Deep learning of histologic images can also be used for diagnostic purposes, but the purpose of this study was to develop a histologic deep learning predictive score (DLPS) for IgA nephropathy, assessed against ground truth (kidney failure) among a consecutive cohort of patients with IgA nephropathy collected over a 39-year period. Noninferiority was assessed in comparison with the Oxford classification adopted by the current gold standard International IgA Nephropathy Prediction Tool (IIPT) (15) and the currently available clinical decision support system for estimating the risk of kidney failure in IgA nephropathy (CDSS) (16). Additional histologic features will be correlated with our DLPS to

identify any potential predictive features not assessed by validated predictive tools.

Materials and Methods

This retrospective, longitudinal, single-center study considered biopsy specimens with primitive IgA nephropathy diagnoses registered between 1982 and 2021, extracted from our center's dedicated kidney biopsy registry. IgA nephropathy diagnoses secondary to Henoch-Schönlein purpura, lupus nephritis, chronic liver diseases, and other immunologic disorders were not considered. We excluded biopsy specimens from transplanted kidneys or those with unavailable histopathologic slides. Specimens were visually inspected to identify technically inadequate images for specimen discoloration or those damaged during archival storage; as required, new histologic sections were prepared and stained using Periodic acid-Schiff (PAS).

All selected biopsy specimens were codified according to patient and biopsy date (multiple biopsy specimens could

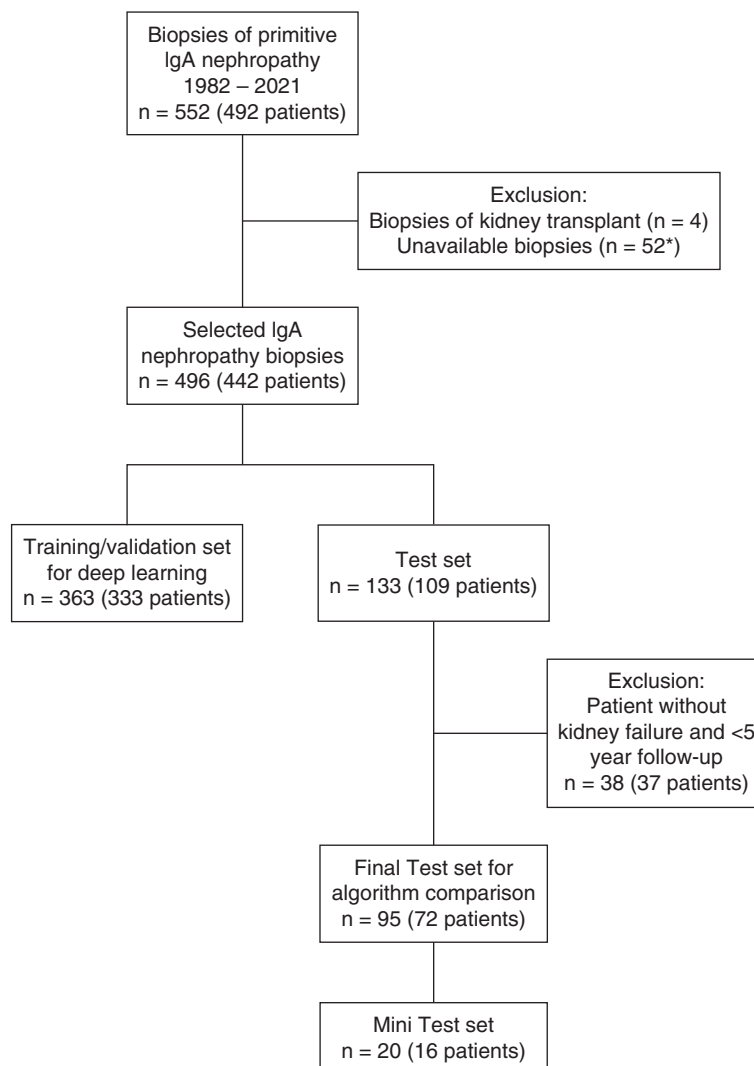


Figure 1. | Flow chart of kidney biopsy specimen selection, digitalization, and randomization in the training/validation set and test set.

*A total of 15 biopsy specimens were provided for the VALIGA study (25); 37 biopsy specimens are missing for other reasons.

have been selected for a single patient). Biopsy specimens were randomly assigned to training for deep learning (“training/validation set”) or a “test set.” Among those assigned to the test set, biopsy specimens from patients with either kidney failure or no event and with at least 5 years follow-up were further selected and included in the “final test set,” according to the IIPT’s maximum prognostic prediction time of 60 months. The final test set was used for comparison between our DLPS with the gold standard Oxford classification content of the IIPT and CDSS (see Figure 1).

Further analysis included the evaluation of additional histologic features (not included in the Oxford classification) on a “mini-test set” comprising 20 selected biopsy specimens (the ten best and ten worst DLPS scores) from the final test set. The analysis of biopsy slides and the acquisition of clinical data included in this study was approved by the local ethics committee (prot. 434/2019/OSS/AOUMO).

Patient Data

Demographic characteristics, clinical features, and patient outcomes were collected retrospectively. For each reported parameter, patients with missing data were not considered (patients included for each parameter analysis are specified in Table 1). GFR was estimated using the 2009 Chronic Kidney Disease Epidemiology Collaboration equation (17) and mean BP was calculated as (systolic BP $\times$ 2+diastolic BP)/3. CKD stage was defined according to the Kidney Disease Outcomes Quality Initiative guidelines (18). Patient outcome specified time to kidney failure, defined as the need for dialysis or kidney transplant. Patient clinical and outcome data were unknown to pathologists.

Patient follow-up data were considered up to the most recent consultation. To verify any loss to long-term follow-up, we investigated regional dialysis and transplantation registries for any additional patient outcomes. Data regarding the start date of replacement therapy and patient mortality were extracted. For follow-up of patients living in other Italian regions, local dialysis and transplantation registries were consulted. For any patients living in regions without these registries, follow-up data were censored up to the last available consultation at our center.

Kidney Biopsy: Preparation and Annotation

All specimens were obtained according to a specific kidney biopsy protocol. Briefly, during an ultrasound-guided percutaneous kidney biopsy procedure, two portions of kidney tissue were sampled from the cortical area (with a Bard Max-Core semiautomatic 16- or 18-gauge 3- to 16-cm needle; Bard Peripheral Vascular Inc., Tempe, AZ): one for immunofluorescence analysis and the other for light microscopy. After Bouin protocol fixation, the microscopy specimen was embedded in paraffin and 3- μ m-thick sections were sliced and stained with hematoxylin and eosin and other special stains (PAS, Masson Trichromic, Silver Jones, Congo red). Only PAS-stained specimens were considered in this study.

Digital slide microphotographs were acquired at high magnification (40 $\times$ /0.65 NA) by a whole slide scanner for digital pathology (D-Sight; Menarini Diagnostics, Florence, Italy). Native images (GXP format) were converted (TIFF

Table 1. Demographic and clinical characteristics of all patients included in the training/validation set and the test set

Characteristic	Value
Demographics	
Patients, <i>n</i>	442
Age, yr	37 (27–51)
Female sex, <i>n</i> (%)	137 (31)
Clinical characteristics	
Creatinine, mg/dl	1.2 (1.0–1.7)
eGFR, ml/min per 1.73 m ²	71 (44–97)
eGFR slope, ml/min per 1.73 m ² per yr	–1.2 (–3.8 to –0.8)
eGFR, ml/min per 1.73 m ² and CKD stage, %	
>90 (stage 1)	31
89–60 (stage 2)	29
59–45 (stage 3a)	15
44–30 (stage 3b)	10
29–15 (stage 4)	11
<15 (stage 5)	4
Urine protein-creatinine ratio ^a	1.0 (0.5–2.0)
Hb, g/dl ^b	13.7 (12–15)
IgA, mg/dl ^c	335 (262–424)
Systolic BP, mm Hg ^d	130 (120–150)
Diastolic BP, mm Hg ^d	80 (75–90)
Baseline therapy, %	
ACE-I or ARB therapy only	18
Immunosuppressive therapy only	15
ACE-I and immunosuppressive therapy	37
No therapy	30
Follow-up, yr	6.7 (2.4–14.6)
Kidney failure, <i>n</i> (%)	88 (20)
Ten-year kidney survival, %	78 (73–82)

Unless otherwise indicated, values are presented as median (25th–75th percentile). GFR was estimated according to the Chronic Kidney Disease Epidemiology Collaboration formula. Hb, hemoglobin; ACE-I, angiotensin-converting enzyme inhibitor; ARB, angiotensin II receptor blocker.

^aData available for 385 patients.

^bData available for 156 patients.

^cData available for 235 patients.

^dData available for 275 patients.

format at full resolution) for subsequent annotation (QuPath software; University of Edinburgh; <https://github.com/qupath/qupath/releases/tag/v0.3.0>) (19), carried out by a pathology technician (F. Gualtieri) and revised by a second, dedicated kidney pathologist (F.F.).

Each biopsy image was manually annotated to obtain compartmentalization of the section according to capsular, medullary, and cortical areas (Figure 2, Supplemental Table 1). All subsequent analyses were performed on the entire cortical area only.

Mesangial (M), endocapillary (E), segmental sclerosis (S), tubulointerstitial damage (T), and crescent (C) (MEST-C) parameters were annotated according to the current classification (1–3), quantified as zero or one for M, E, and S, or zero, one, or two for T and C. The MEST-C classification was manually scored by the pathologist (F.F.).

Digitalized Image Preparation and Datasets

Due to computational resource availability constraints (an entire biopsy section image is composed of about 6

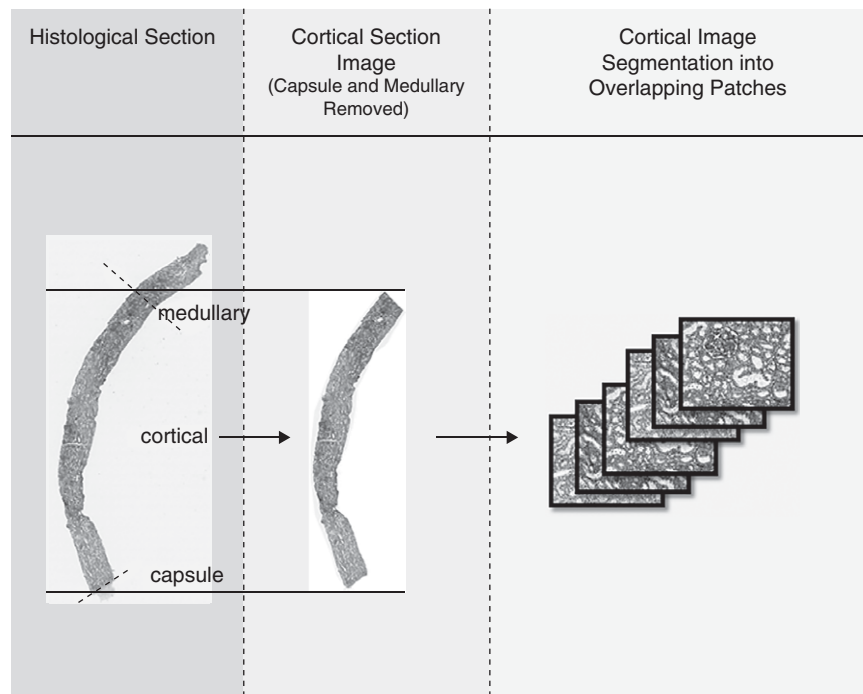


Figure 2. | Workflow of slide segmentation and production of overlapping patches.

billion pixels), training of a CNN was not possible for an elevated number of biopsy specimens. An important parameter for effective training is appropriate image size and resolution. A preprocessing pipeline, which employs several classic computer vision techniques to split whole-slide images into overlapping patches, assisted in achieving resizing (e.g., 512×512) and, therefore, input image dimensions were reduced. The dataset, composed of 496 biopsy specimens (1369 sections), yielded 7199 patches (Figure 2).

Deep Learning Training (Training/Validation Set)

Deep learning training was conducted on the training/validation set according to ground truth, following a soft-label strategy, including time periods of patient follow-up and time to eventual kidney failure (20). Coding of ground truth according to a scheme is outlined in Supplemental Table 2.

CNN Design

The ResNet-18 architecture (21), trained to evaluate multiple 256×512 patches simultaneously and yield biopsy specimen-level predictions, was used. Data augmentation was performed by randomly flipping and rotating each input image before feeding it to the network, because neither of these two transformations change the semantic content of the image. Moreover, input image contrast was randomly altered so the network could not distinguish between older and newer biopsy specimens.

Regularization techniques are usually excluded during inference when the goal is the best possible prediction for a single, novel image. In this study, inference was run multiple times and data augmentation techniques were used (22). During the training/validation procedure, all available

images were shown to the model and weights were updated accordingly. A total of 80 epochs (reiteration of training) were performed; at each epoch, four randomly selected patches derived from the cortical area of each section were chosen. Therefore, the probability of patches from each cortical area never being shown to the model was considered negligible.

Transfer learning was exploited by pretraining CNNs using the open-source ImageNet (23,24) and then finetuned to correctly classify histopathologic images. The learning rate was set during the fine-tuning process and a weighted crossentropy loss was used.

Prognostic Accuracy of DLPS versus Prediction Score Calculators

The DLPS output is a risk score ranging from zero to one. The score is proportional to the risk of kidney failure. The DLPS was applied to the independent final test set to generate a histologic risk prediction. The DLPS was compared with the IIPT and CDSS prediction score calculators, which are both based on MEST-C parameters (1–3). Because IIPT and CDSS prediction scores also require clinical data, we applied standardized data (median cohort clinical data) for each biopsy IIPT and CDSS predictive score generated. This was designed to eliminate the effect of clinical data (Supplemental Table 4).

DLPS Correlated with Additional Histologic Parameters (Mini-Test Set)

An additional manual histologic feature analysis, not considered by the MEST-C score, was performed on the mini-test set and was correlated to the DLPS.

Table 2. Distribution and multivariable analysis of the MEST-C parameters (gold standard) in the study's test set

Parameter	Descriptive Statistics, Median (25th–75th Percentile)	Multivariable Survival Analysis	
		Hazard Ratio (95% Confidence Interval)	P Value
M	0 (0–1)	1.92 (1.00 to 3.7)	0.05
E	1 (0–1)	0.72 (0.37 to 1.4)	0.34
S	1 (0–1)	2.1 (0.99 to 4.43)	0.05
T	0 (0–1)	2.06 (1.3 to 3.28)	0.002
C	0 (0–1)	1.91 (1.04 to 3.48)	0.03

n=133 biopsy specimens. M, mesangial; E, endocapillary; S, segmental sclerosis; T, tubulointerstitial; C, crescent.

Statistical Analysis

Kidney survival curves were generated using the Kaplan–Meier method. Uni- and multivariable analyses of MEST-C and clinical characteristics for time to kidney failure outcome were calculated by Cox regression. DLPS, IIPT, and CDSS were evaluated and compared according to the following metrics: accuracy, area under the receiver operating characteristic curve (AUC), F1 score, precision, recall, and specificity.

Correlations between our DLPS, IIPT and CDSS, the individual MEST-C parameters, and additional histologic features were assessed by Pearson correlation. For additional histologic correlation, the Bonferroni correction for repeated analyses was applied. We used Stata 11.2 software (Stata-Corp, College Station, TX).

Results

Patients

The participants' clinical characteristics are outlined in Table 1. Median (25th–75th percentile) patient follow-up was 6.7 (2.5–14.6) years with a 10-year kidney survival of 78% (95% confidence interval [95% CI], 73% to 82%). Uni- and multivariable analysis of clinical characteristics in

predicting kidney survival are outlined in Supplemental Table 3. Creatinine levels, urine protein-creatinine ratio, and mean BP were confirmed as significant clinical risk predictors for kidney failure ($P<0.001$).

Biopsy Specimens

The flow chart of kidney biopsy selection is depicted in Figure 1. From the 496 selected biopsy specimens meeting inclusion criteria, digitization yielded 1369 images (median of three slices per biopsy) and overlapping patches comprised 7199 images (Supplemental Table 1). The training/validation set included 363 biopsy specimens (73%), and the randomly generated test set (balanced for kidney failure outcome distribution) included 133 biopsy specimens (27%).

MEST-C Predictors of Kidney Survival in the Test Set

The distribution of MEST-C parameters of the test set are outlined in Table 2. The MEST parameters were included in a multivariable analysis for predictors of kidney survival. As previously proven (3,25), the T parameter is the most statistically predictive parameter for kidney failure (hazard ratio [HR], 2.06; 95% CI, 1.3 to 3.28; $P=0.002$). The C parameter shows a good predictive value (HR, 1.91; 95% CI, 1.04 to 3.48; $P=0.03$), and the M and S parameters show only marginal statistical significance (HR, 1.92; 95% CI, 1.00 to 3.7; $P=0.05$ and HR, 2.1; 95% CI, 0.99 to 4.43; $P=0.05$, respectively).

Prognostic Accuracy of DLPS versus Prediction Score Calculators

The prognostic accuracy of DLPS was comparable with both the prediction score calculators in terms of accuracy, AUC, F1 score, precision, recall, and specificity ($P=0.4$) (Table 3). According to receiver operating characteristic curve analysis, the AUC for DLPS was insignificantly higher (AUC=0.84) than those of the histologic components of the prediction score calculators (0.77 and 0.74, respectively; $P=0.3$) (see Figure 3).

Table 3. The deep learning predictive score, the International IgAN Prediction Tool, and the clinical decision support system for estimating the risk of end-stage kidney disease in IgA nephropathy prediction versus actual incidence of kidney failure in the final test set

Data Classification and Metrics	Deep Learning Predictive Score	International IgAN Prediction Tool	Clinical Decision Support System
Correctly predicted/actual kidney failure (biopsy specimens)	17/29	17/29	18/29
Correctly predicted/actual no kidney failure (biopsy specimens)	65/66	56/66	49/66
Accuracy	0.82	0.77	0.71
AUC	0.84	0.77	0.74
F1 score	0.67	0.61	0.56
Precision	0.77	0.63	0.51
Recall	0.59	0.59	0.62
Specificity	0.98	0.85	0.74

n=95 biopsy specimens. AUC, area under the receiver operating characteristic curve.

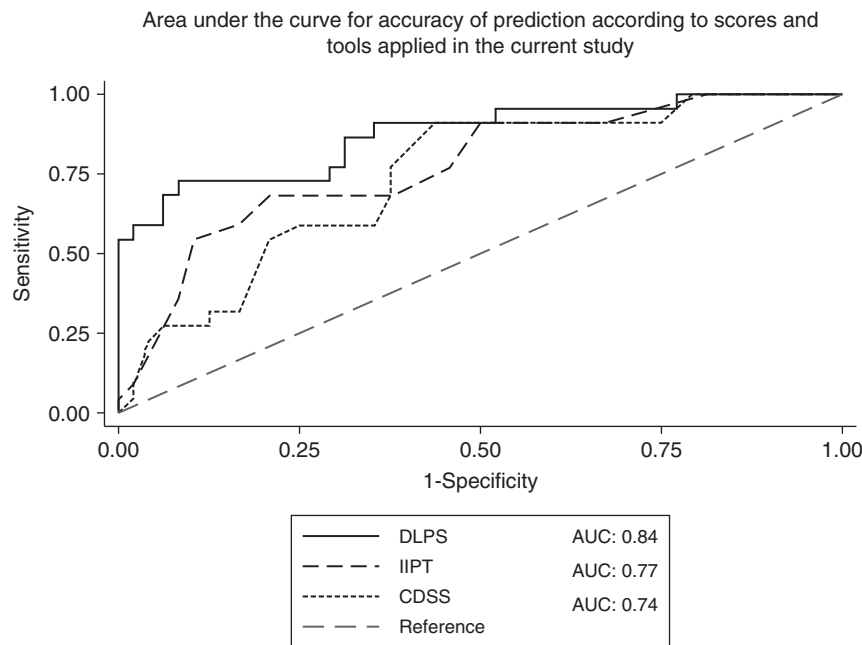


Figure 3. | Receiver operating characteristic curves of prediction accuracy of our deep learning predictive score (DLPS), the International IgA Nephropathy Prediction Tool (IIPT), and the clinical decision supporting system for estimating the risk of kidney failure in IgA nephropathy (CDSS). Analysis of the final test set ($n=95$ biopsy specimens). AUC, area under the receiver operating characteristic curve.

Correlations of MEST-C and Prediction Score Calculators Parameters with DLPS

Correlations between the MEST-C parameters, the histologic component of the prediction score calculators (standardized for clinical parameters), and our DLPS are shown in Table 4. A significant correlation between the prediction score calculators (IIPT and CDSS) with our DLPS is highlighted (R of 0.39 [$P<0.001$] and 0.33 [$P<0.001$], respectively). No specific correlations were observed between DLPS and the M, E, S, and C parameters. Additionally, the T parameter was highly correlated with the IIPT ($R=0.96$, $P<0.001$) and the CDSS ($R=0.79$, $P<0.001$), but only moderately with our DLPS ($R=0.41$, $P<0.001$). A significant negative correlation with the E parameter was observed with both prediction score calculators (R of -0.29 [$P<0.001$] and -0.47 [$P<0.001$], respectively) and was not replicated by our DLPS.

DLPS Correlated with Additional Histologic Parameters in Mini-Test Set

Because the MEST-C parameters were poorly correlated with our DLPS scores, additional histologic features were assessed in the mini-test set (the ten highest and lowest scoring DLPS biopsy specimens). Infiltration of interstitial fibrosis and tubular atrophy ($R^2=0.70$, $P=0.001$) and the number of hyaline casts ($R^2=0.70$, $P<0.001$) were the most statistically correlated additional histologic features with our DLPS score (see Table 5).

Discussion

The main objective of this study was to compare the prognostic efficacy of an automated deep learning histology score with the manual MEST-C approach. Because our

DLPS does not incorporate clinical data, it cannot in any way replace the current gold standard for prognostic definition of IgA nephropathy (the IIPT). Our results suggest that, compared with the histologic component of IIPT and CDSS predictive scores, our deep learning automated histologic scoring system was statistically noninferior. Further, results from this study confirm the importance of the T parameter, which was also highly correlated with our DLPS. The analysis of additional histopathologic parameters highlights that our DLPS correlates with multiple features not included in the MEST-C scoring evaluation, the strongest being inflammation within areas of interstitial fibrosis and tubular atrophy and hyaline casts. Despite these operational differences, the DLPS was accurate and this automated histologic approach may eliminate issues of human subjective bias and save time and costs compared with the “traditional” manual approach, although data do not prove superiority of DLPS to the MEST-C scores.

Post hoc identification of the features used by deep learning algorithms to carry out their predictions are inherently difficult to identify and, in the deep learning field, the “black box” is a well-known problem (26). The availability of MEST-C parameters in our dataset enabled a direct comparison for these features, thereby partially alleviating this limit. Analysis highlighted that the DLPS is highly correlated to the T score (Table 4). The prognostic significance of the T score is not new to literature; many authors have also reported its significance (1–3,25,27,28). This study highlights that our DLPS independently developed a sensitivity toward this parameter, albeit with less correlation than the IIPT and CDSS. Despite having a high prognostic capacity, our DLPS lacks any correlation with parameters related to glomerular structures (M, E, S, and C). We speculate the very low correlation with C may be due to the scarce

Table 4. Correlations between our deep learning predictive score, the International IgAN Prediction Tool, the clinical decision support system for estimating the risk of end-stage kidney disease in IgA nephropathy, and MEST-C parameters

Score	Deep Learning Predictive Score	International IgAN Prediction Tool	Clinical Decision Support System	Mesangial	Endocapillary	Segmental Sclerosis	Tubulointerstitial	Crescent
DLPS	1.00							
IIPT	0.39 ^a	1.00						
CDSS	0.33 ^a	0.85 ^a	1.00					
M	0.14	0.33 ^a	0.36 ^a	1.00				
E	−0.07	−0.29 ^a	−0.47 ^a	0.07	1.00			
S	0.09	0.21	0.37 ^a	0.21	0.11	1.00		
T	0.41 ^a	0.96 ^a	0.79 ^a	0.20	−0.20	0.15	1.00	
C	0.12	−0.05	−0.30 ^a	0.09	0.28 ^b	0.10	0.01	1.00

DLPS, deep learning predictive score; IIPT, International IgA Nephropathy Prediction Tool; CDSS, clinical decision support system for estimating the risk of kidney failure in IgA nephropathy; M, mesangial; E, endocapillary; S, segmental sclerosis; T, tubulointerstitial; C, crescent.

^a $P < 0.001$.

^b $P < 0.01$.

representation of this feature in our sample (see Table 2). However, it must be emphasized that our DLPS did show a very low correlation with the parameters relating to mesangial hypercellularity (M) and segmental glomerulosclerosis (S). M and S have both previously been confirmed in two large validation studies of the Oxford classification (25,28). Furthermore, the relatively low correlation of the T parameter in our DLPS outcome compared with the IIPT and CDSS suggests other features not included in the MEST-C score are considered by our DLPS in generating risk prediction. We attempted to identify these additional histologic features through correlation with the DLPS for biopsy specimens with the worst and best DLPS (extreme values). Interestingly, the most significant histologic parameters that emerged were

inflammation within areas of interstitial fibrosis and tubular atrophy and hyaline casts.

Inflammation within areas of interstitial fibrosis and tubular atrophy is associated with adverse outcomes in kidney transplantation and is related to T cell-mediated rejection. Although its biologic and functional relationship requires further investigation, T-cell infiltration appears to be a prerequisite for inflammation within areas of interstitial fibrosis and tubular atrophy development, as demonstrated in a large sequential kidney transplant biopsy study (29). Similarly, CD20-positive B cells form a prominent part of the interstitial infiltrating cells in IgA nephropathy (30). Tissue-infiltrated B cells secrete proinflammatory cytokines, chemokines, and Igs, which further exaggerate kidney inflammation by attracting more lymphocytes and provoking resident kidney cells, leading to kidney fibrosis and functional deterioration (31). Inflammation was initially evaluated by the researchers of the Oxford classification but, during the score definition, the analysis distinguishing between inflammation in scarred or unscarred tissue was abandoned due to very low reproducibility (2). Eventually, inflammation within areas of interstitial fibrosis and tubular atrophy could represent a more granular description of tubulointerstitial lesions in IgA nephropathy, but further analyses are requested to confirm this hypothesis.

Hyaline casts are associated with thyroidization-type tubular atrophy (32). The increased presence of these lesions in patients with high DLPS scores is likely associated to more advanced forms of tubular atrophy. Hyaline casts were not specifically scored in the development of the Oxford classification definition (2).

Our study suggests an important precedent in defining automated histology-based prognostic scores. Our data suggest that the deep learning approach has the potential to develop histologic prognostic models that are noninferior to those developed by a manual, human-based approach. Furthermore, our study suggests this innovative approach can achieve a prognostic histologic score relatively quickly, requiring fewer resources compared with the traditional approach.

Table 5. Additional histologic features (not included in MEST-C) scored for 20 selected images from the test set, including the ten biopsy specimens with the highest deep learning predictive score and ten lowest deep learning predictive score

Histologic Feature	<i>r</i>	<i>P</i> Value
IFTA	0.56	0.01
i-IFTA	0.70	0.001 ^a
i-non-IFTA	−0.41	0.07
IFV	0.62	0.008
Arteriolar hyalinosis	0.38	0.10
Glomeruli, <i>n</i>	0.24	0.31
Sclerotic glomeruli, <i>n</i>	0.56	0.01
Sclerotic glomeruli, %	0.57	0.009
Acute tubular injury	0.59	0.006
Tubulitis	−0.34	0.15
Peritubular capillaritis	−0.22	0.35
Hyaline casts	0.82	<0.001 ^a
RBC casts	0.20	0.40

r, Pearson correlation coefficient; IFTA, interstitial fibrosis and tubular atrophy; i-IFTA, inflammation in areas of IFTA; IFV, narrowing of arterial vessels due to intimal fibrosis; RBC, red blood cell.

^a $P < 0.003$ (statistically significant threshold according to Bonferroni correction).

This study has limitations related to the retrospective and monocentric nature of data collection. The algorithm was developed on the basis of histopathologic data only and additional factors, such as the heterogeneity between patients' medical treatments (steroid, immunosuppressive drug, the use of angiotensin-converting enzyme inhibitors and angiotensin receptor blockers) and other clinical and demographic data, were not considered in the assessment of predictive accuracy, but will be analyzed in future research. Larger, prospective studies, including individual clinical data, are necessary to validate our DLPS and to confirm its external validity.

In conclusion, this study suggests the deep learning approach is noninferior to manual histopathologic reporting and may consider prognostic features not currently included in the MEST-C assessment. The application of DLPS to a full algorithm (including clinical data) should be developed and validated.

Disclosures

G. Donati reports receiving honoraria from B. Braun Avitum and having consultancy agreements with Medtronic. L. Gesualdo reports receiving research funding from Abionyx and Sanofi; receiving honoraria from Astellas, AstraZeneca, Estor, Fresenius, Travençolo, and Werfen; having consultancy agreements with AstraZeneca, Baxter, Chinook, Estor, GlaxoSmithKline, Medtronic, Mundipharma, Novartis, Pharmadoc, Retrophin, Roche, Sandoz, Sanofi, and Travençolo; serving on the board of directors for the European Renal Association–European Dialysis and Transplant Association, Renal Pathology Society, and Società Italiana Nefrologia; serving in an advisory or leadership role for *Journal of Nephrology* and *Nephrology Dialysis Transplantation*; and receiving royalties from McGraw-Hill Education (Italy) Srl. R. Magistroni reports receiving research funding from Omeros Pharmaceutical and Reata Pharmaceutical and having consultancy agreements with Otsuka Pharmaceutical. F. Pollastri reports being employed by AstraZeneca Computational Pathology Munich. All remaining authors have nothing to disclose.

Funding

This work was supported by the Athenaeum Research Grant (FAR) 2021 of the Department of Surgical, Medical, Dental and Morphological Sciences related to Transplant, Oncology and Regenerative Medicine, University of Modena and Reggio Emilia, Italy and FSE 2014/2020 Obiettivo tematico 10, Emilia Romagna, Italy grant 2538.

Acknowledgments

We thank Prof. Rosanna Coppo and Prof. Paolo Schena for the many suggestions that have produced a significant improvement of this work and for the final revision of the draft.

Author Contributions

F. Bolelli, S. Cimino, M. Ferri, F. Fontana, L. Gesualdo, F. Giaroni, S. Giovannella, F. Gualtieri, M. Leonelli, G. Ligabue, R. Magistroni, E. Mancini, M. Nordio, F. Pollastri, P. Sacco, and F. Testa were responsible for data curation; F. Bolelli, C. Grana, R. Magistroni, and F. Pollastri were responsible for formal analysis and software; F. Fontana and R. Magistroni were responsible for investigation; R. Magistroni and F. Pollastri were responsible for methodology; C. Grana and R. Magistroni were responsible for project administration and resources; R. Magistroni conceptualized the study,

provided supervision, wrote the original draft, and was responsible for funding acquisition, validation, and visualization; and G. Alfano, J. Chester, G. Donati, F. Fontana, M. Leonelli, R. Magistroni, and F. Testa reviewed and edited the manuscript.

Supplemental Material

This article contains the following supplemental material online at <http://cjasn.asnjournals.org/lookup/suppl/doi:10.2215/CJN.01760222/-/DCSupplemental>.

Supplemental Table 1. Coding of soft labels for deep learning training.

Supplemental Table 2. Kidney failure risk, calculated by unadjusted and multivariable Cox regression analysis of clinical characteristics.

Supplemental Table 3. Manual annotation of biopsy images.

Supplemental Table 4. Deep Learning Predictive Score (DLPS), MEST-C score, IgA Nephropathy Prediction Tool (IIPT), Clinical Decision Support System for Estimating the Risk of End-Stage Kidney Disease in IgA Nephropathy (CDSS), and kidney failure during the follow-up according to Supplemental Table 1 coding (Ground Truth) for all patients in the test set.

References

- Cattran DC, Coppo R, Cook HT, Feehally J, Roberts IS, Troyanov S, Alpers CE, Amore A, Barratt J, Berthouix F, Bonsib S, Bruijn JA, D'Agati V, D'Amico G, Emancipator S, Emma F, Ferrario F, Fervenza FC, Florquin S, Fogo A, Geddes CC, Groene HJ, Haas M, Herzenberg AM, Hill PA, Hogg RJ, Hsu SI, Jennette JC, Joh K, Julian BA, Kawamura T, Lai FM, Leung CB, Li LS, Li PK, Liu ZH, Mackinnon B, Mezzano S, Schena FP, Tomino Y, Walker PD, Wang H, Weening JJ, Yoshikawa N, Zhang H; Working Group of the International IgA Nephropathy Network and the Renal Pathology Society: The Oxford classification of IgA nephropathy: Rationale, clinicopathological correlations, and classification. *Kidney Int* 76: 534–545, 2009
- Roberts IS, Cook HT, Troyanov S, Alpers CE, Amore A, Barratt J, Berthouix F, Bonsib S, Bruijn JA, Cattran DC, Coppo R, D'Agati V, D'Amico G, Emancipator S, Emma F, Feehally J, Ferrario F, Fervenza FC, Florquin S, Fogo A, Geddes CC, Groene HJ, Haas M, Herzenberg AM, Hill PA, Hogg RJ, Hsu SI, Jennette JC, Joh K, Julian BA, Kawamura T, Lai FM, Li LS, Li PK, Liu ZH, Mackinnon B, Mezzano S, Schena FP, Tomino Y, Walker PD, Wang H, Weening JJ, Yoshikawa N, Zhang H; Working Group of the International IgA Nephropathy Network and the Renal Pathology Society: The Oxford classification of IgA nephropathy: Pathology definitions, correlations, and reproducibility. *Kidney Int* 76: 546–556, 2009
- Trimarchi H, Barratt J, Cattran DC, Cook HT, Coppo R, Haas M, Liu ZH, Roberts IS, Yuzawa Y, Zhang H, Feehally J; IgAN Classification Working Group of the International IgA Nephropathy Network and the Renal Pathology Society; Conference Participants: Oxford Classification of IgA nephropathy 2016: An update from the IgA Nephropathy Classification Working Group. *Kidney Int* 91: 1014–1021, 2017
- Kidney Disease: Improving Global Outcomes Glomerular Diseases Work Group: KDIGO 2021 Clinical Practice Guideline for the Management of Glomerular Diseases. *Kidney Int* 100: S1–S276, 2021
- Furness PN, Taub N: Interobserver reproducibility and application of the ISN/RPS classification of lupus nephritis—a UK-wide study. *Am J Surg Pathol* 30: 1030–1035, 2006
- Gamba G, Reyes E, Angeles A, Quintanilla L, Calva J, Peña JC: Observer agreement in the scoring of the activity and chronicity indexes of lupus nephritis. *Nephron* 57: 75–77, 1991
- Grootscholten C, Bajema IM, Florquin S, Steenberg EJ, Peutz-Kootstra CJ, Goldschmeding R, Bijl M, Hagen EC, van Houwelingen HC, Derksen RH, Berden JH: Interobserver agreement of scoring of histopathological characteristics and classification of lupus nephritis. *Nephrol Dial Transplant* 23: 223–230, 2008

8. Oni L, Beresford MW, Witte D, Chatzitoliou A, Sebire N, Abulaban K, Shukla R, Ying J, Brunner HI: Inter-observer variability of the histological classification of lupus glomerulonephritis in children. *Lupus* 26: 1205–1211, 2017
9. Wernick RM, Smith DL, Houghton DC, Phillips DS, Booth JL, Runckel DN, Johnson DS, Brown KK, Gaboury CL: Reliability of histologic scoring for lupus nephritis: A community-based evaluation. *Ann Intern Med* 119: 805–811, 1993
10. Bouteldja N, Klinkhammer BM, Bülow RD, Droste P, Otten SW, Freifrau von Stillfried S, Moellmann J, Sheehan SM, Korstanje R, Menzel S, Bankhead P, Mietsch M, Drummer C, Lehrke M, Kramann R, Floege J, Boor P, Merhof D: Deep learning-based segmentation and quantification in experimental kidney histopathology. *J Am Soc Nephrol* 32: 52–68, 2021
11. Bülow RD, Kers J, Boor P: Multistain segmentation of renal histology: First steps toward artificial intelligence-augmented digital nephropathology. *Kidney Int* 99: 17–19, 2021
12. de Haan K, Zhang Y, Zuckerman JE, Liu T, Sisk AE, Diaz MFP, Jen KY, Nobori A, Liou S, Zhang S, Riahi R, Rivenson Y, Wallace WD, Ozcan A: Deep learning-based transformation of H&E stained tissues into special stains. *Nat Commun* 12: 4884, 2021
13. LeCun Y, Bengio Y, Hinton G: Deep learning. *Nature* 521: 436–444, 2015
14. Ligabue G, Pollastri F, Fontana F, Leonelli M, Furci L, Giovanella S, Alfano G, Cappelli G, Testa F, Bolelli F, Grana C, Magistroni R: Evaluation of the classification accuracy of the kidney biopsy direct immunofluorescence through convolutional neural networks. *Clin J Am Soc Nephrol* 15: 1445–1454, 2020
15. Barbour SJ, Coppo R, Zhang H, Liu ZH, Suzuki Y, Matsuzaki K, Katafuchi R, Er L, Espino-Hernandez G, Kim SJ, Reich HN, Feehally J, Cattran DC: International IgA Nephropathy Network: Evaluating a new international risk-prediction tool in IgA nephropathy. *JAMA Intern Med* 179: 942–952, 2019
16. Schena FP, Anelli VW, Trotta J, Di Noia T, Manno C, Tripepi G, D'Arrigo G, Chesnaye NC, Russo ML, Stangou M, Papagianni A, Zoccali C, Tesar V, Coppo R: members of the VALIGA study: Development and testing of an artificial intelligence tool for predicting end-stage kidney disease in patients with immunoglobulin A nephropathy. *Kidney Int* 99: 1179–1188, 2021
17. Levey AS, Stevens LA, Schmid CH, Zhang YL, Castro AF 3rd, Feldman HI, Kusek JW, Eggers P, Van Lente F, Greene T, Corresh J; CKD-EPI (Chronic Kidney Disease Epidemiology Collaboration): A new equation to estimate glomerular filtration rate. *Ann Intern Med* 150: 604–612, 2009
18. National Kidney Foundation: K/DOQI clinical practice guidelines for chronic kidney disease: Evaluation, classification, and stratification. *Am J Kidney Dis* 39: S1–S266, 2002
19. Bankhead P, Loughrey MB, Fernández JA, Dombrowski Y, McArt DG, Dunne PD, McQuaid S, Gray RT, Murray LJ, Coleman HG, James JA, Salto-Tellez M, Hamilton PW: QuPath: Open source software for digital pathology image analysis. *Sci Rep* 7: 16878, 2017
20. Diaz R, Marathe A: Soft labels for ordinal regression. Presented at the 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition, Long Beach, CA, June 15–20, 2019
21. He K, Zhang X, Ren S, Sun J: Deep residual learning for image recognition. Presented at the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, June 27–30, 2016
22. Pollastri F, Bolelli F, Paredes R, Grana C: Augmenting data with GANs to segment melanoma skin lesions. *Multimedia Tools Appl* 79: 15575–15592, 2020
23. Krizhevsky A, Sutskever I, Hinton GE: ImageNet classification with deep convolutional neural networks. Presented at the Twenty-sixth Conference on Neural Information Processing Systems, Lake Tahoe, NV, December 3–8, 2012
24. Deng J, Dong W, Socher R, Li L-J, Li K, Fei-Fei L: ImageNet: A large-scale hierarchical image database. Presented at the 2009 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Miami, FL, June 20–25, 2009
25. Coppo R, Troyanov S, Bellur S, Cattran D, Cook HT, Feehally J, Roberts IS, Morando L, Camilla R, Tesar V, Lunberg S, Gesualdo L, Emma F, Rollino C, Amore A, Praga M, Feriozzi S, Segoloni G, Pani A, Cancarini G, Durlak M, Moggia E, Mazzucco G, Giannakakis C, Honsova E, Sundelin BB, Di Palma AM, Ferrario F, Gutierrez E, Asunis AM, Barratt J, Tardanico R, Perkowska-Ptasinska A; VALIGA study of the ERA-EDTA Immunonephrology Working Group: Validation of the Oxford classification of IgA nephropathy in cohorts with different presentations and treatments. *Kidney Int* 86: 828–836, 2014
26. Castelvechi D: Can we open the black box of AI? *Nature* 538: 20–23, 2016
27. Markowitz G: Glomerular disease: Updated Oxford Classification of IgA nephropathy: A new MEST-C score. *Nat Rev Nephrol* 13: 385–386, 2017
28. Lv J, Shi S, Xu D, Zhang H, Troyanov S, Cattran DC, Wang H: Evaluation of the Oxford Classification of IgA nephropathy: A systematic review and meta-analysis. *Am J Kidney Dis* 62: 891–899, 2013
29. Nankivell BJ, Shingde M, Keung KL, Fung CLS, Borrows RJ, O'Connell PJ, Chapman JR: The causes, significance and consequences of inflammatory fibrosis in kidney transplantation: The Banff i-IFTA lesion. *Am J Transplant* 18: 364–376, 2018
30. Heller F, Lindenmeyer MT, Cohen CD, Brandt U, Draganovici D, Fischereder M, Kretzler M, Anders HJ, Sitter T, Mosberger I, Kerjaschki D, Regele H, Schlöndorff D, Segerer S: The contribution of B cells to renal interstitial inflammation. *Am J Pathol* 170: 457–468, 2007
31. Zheng N, Xie K, Ye H, Dong Y, Wang B, Luo N, Fan J, Tan J, Chen W, Yu X: TLR7 in B cells promotes renal inflammation and Gd-IgA1 synthesis in IgA nephropathy. *JCI Insight* 5: e136965, 2020
32. Dvanajscak Z, Cossey LN, Larsen CP: A practical approach to the pathology of renal intratubular casts. *Semin Diagn Pathol* 37: 127–134, 2020

Received: February 10, 2022 **Accepted:** June 27, 2022

F.T. and F.F. contributed equally to this work.

Published online ahead of print. Publication date available at www.cjasn.org.

AFFILIATIONS

¹Division of Nephrology, Dialysis and Renal Transplantation, Azienda Ospedaliera Universitaria Policlinico di Modena, Modena, Italy

²Department of Engineering “Enzo Ferrari,” University of Modena and Reggio Emilia, Modena, Italy

³Department of Surgery, Medicine, Dental Medicine and Morphological Sciences, University of Modena and Reggio Emilia, Modena, Italy

⁴U.O. Nefrologia, Dialisi, Ipertensione, Istituto di Ricovero e Cura a Carattere Scientifico (IRCCS) Azienda Ospedaliero-Universitaria di Bologna, Bologna, Italy

⁵Nephrology and Dialysis Unit, Unità Locale Socio Sanitaria 15 (ULSS 15), Camposampiero–Cittadella, Padua, Italy

⁶Nephrology and Dialysis Unit, Azienda Sanitaria Locale 3 (ASL 3), Genoa, Italy

⁷Department of Emergency and Organ Transplantation, University of Bari “Aldo Moro,” Bari, Italy

⁸Nephrology and Dialysis, Azienda Unità Sanitaria Locale (AUSL) Modena, Modena, Italy