

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13

# Title: Evaluation of the Classification Accuracy of The Kidney Biopsy Direct Immunofluorescence through Convolutional Neural Networks

14  
15  
16  
17  
18

Authors: Ligabue Giulia<sup>1</sup>, Pollastri Federico<sup>2</sup>, Fontana Francesco<sup>3</sup>, Leonelli Marco<sup>3</sup>, Furci Luciana<sup>3</sup>, Giovanella Silvia<sup>1</sup>, Alfano Gaetano<sup>3</sup>, Cappelli Gianni<sup>1,3</sup>, Testa Francesca<sup>3</sup>, Bolelli Federico<sup>2</sup>, Grana Costantino<sup>2</sup> & Magistrone Riccardo<sup>1,3</sup>

19  
20

Running title: DeepIF: classification of immunofluorescence by CNNs.

- 21  
22  
23  
24  
25  
26  
27
- 1) Department of Surgery, Medicine, Dental Medicine and Morphological Sciences, University of Modena and Reggio Emilia, Modena, Italy
  - 2) Department of Engineering "Enzo Ferrari", University of Modena and Reggio Emilia, Modena, Italy
  - 3) Division of Nephrology, Dialysis and Renal Transplantation, AOU Policlinico di Modena, Italy

28  
29

Corresponding author: Magistrone Riccardo

30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

email: [riccardo.magistrone@unimore.it](mailto:riccardo.magistrone@unimore.it)

### Significance statement (max 120words, summary of work)

The classification of immunological deposits is a pivotal procedure of Kidney Pathology. It requires trained personnel and the images interpretation is a complex and slow procedure. Deep Learning has the potential to automate the image classification procedure and increase the accuracy and reproducibility. This approach can allow to extract new features important for the reclassification of kidney diseases in a diagnostic and prognostic perspective. 12259 kidney immunofluorescence images were used to train Convolutional Neural Networks to recognize the classic patterns reported in the immunofluorescence analysis. The algorithm developed the ability to recognize microscopic patterns and showed substantial agreement with three different pathologists. The algorithm can become a support tool in the routine kidney immunofluorescence analysis with time saving and reproducibility advantages.

For Peer Review

## Abstract

**Background.** Immunohistopathology is an essential technique in the diagnostic workflow of a kidney biopsy. Deep Learning is an effective tool in the elaboration of medical imaging. We wanted to evaluate the role of a Convolutional Neural Network as a support tool for kidney immunofluorescence reporting.

**Methods.** High magnification (400X) immunofluorescence images of kidney biopsies performed from the year 2001 to 2018 were collected. The report, adopted at the Division of Nephrology of the AOU Policlinico di Modena, describes the specimen in terms of '*Appearance*', '*Distribution*', '*Location*' and '*Intensity*' of the glomerular deposits identified with fluorescent antibodies against IgG, IgA, IgM immunoglobulins, C1q and C3 complement fractions, Fibrinogen, kappa and lambda light chains. The report was used as ground truth for the training of the Convolutional Neural Networks.

**Results.** 12259 immunofluorescence images of 2542 subjects undergoing kidney biopsy were collected. The test set analysis showed accuracy values between 0.79 ('Irregular Capillary wall' feature) and 0.94 ('Fine Granular' feature). The agreement test of the results obtained by the Convolutional Neural Networks with respect to the ground truth showed similar values to three pathologists of our center. Convolutional Neural Networks were 117.3 times faster than human evaluators in analyzing 180 image tests. A web platform, where it is possible to upload digitized images of immunofluorescence specimens, is available to evaluate the potential of our approach.

**Conclusion.** The data showed that the accuracy of Convolutional Neural Networks are comparable to that of pathologists experienced in the field.

## Introduction

The kidney histopathological diagnosis involves a combined and complementary approach of different microscopic techniques: essentially, they consist of light microscopy, immunohistopathology and electron microscopy. The kidney biopsy is frequently necessary to distinguish among diseases with similar clinical presentation. The evaluation of glomerular disease by light microscopy alone rarely allows a definitive diagnosis, for which the information from immunohistopathology and electron microscopy analysis are required. All these elements must finally be integrated with the patient's clinical history and laboratory findings to obtain the conclusive diagnosis. Immunofluorescence technique (or other immunohistopathologic approaches) is required to demonstrate deposits of immunoglobulins and complement components. Some specific kidney diseases such as, for example, IgA Nephropathy, anti-glomerular basal membrane glomerulonephritis and C3 glomerulopathy can only be diagnosed by the result of immunofluorescence. The analysis of the immunofluorescence pattern of deposition in glomerular diseases is a time-consuming activity, it relies on the availability of specific resources and requires an experienced operator for interpretation. In pathology, as in many other clinical disciplines, the ability of artificial intelligence algorithms to interpret medical images is taking on a dominant role. Some examples of the application of these technologies are emerging also in kidney pathology <sup>1, 2</sup>. An artificial neural network is an interconnected ensemble of simple processing elements, with the ability of learning how to perform different tasks directly from data. Each element, or *artificial neuron*, receives input values from several preceding nodes, processes them, and computes the output in order to forward it to the next set of neurons. When this hierarchic architecture is formed by several layers, the strategy takes the name of Deep Learning. Each node has a set of *weights* used to multiply its inputs. These weights are the processing ability of a neural network. Convolutional Neural Networks present weights in the form of small filters, which learn to recognize specific characteristics regardless of their position within the input image. They are randomly initialized, and then updated during the training process in order to optimize the final output. The learning strategy causes neural networks to adapt their filters in order to gain the ability to extract meaningful features directly from the set of images used during the training process. This is the reason why, unlike any other machine learning technique, deep learning algorithms require no hand-crafted feature. Given the aforementioned hierarchical architecture, automatically extracted features grow more complex as the depth of the network increases. However, very deep neural networks tend to learn to classify the training set with extremely high accuracy while failing to generalize their capabilities to samples that were not used during training. This phenomenon takes the name of *overfitting*<sup>3</sup>, and causes networks to learn to recognize each training image instead of the semantic content within. Therefore, many efforts have been devoted to finding valid regularization techniques <sup>4-6</sup>. Convolutional Neural Networks have been widely proved to outperform other strategies in countless computer vision tasks such as semantic segmentation, object detection, object classification and others <sup>7-9</sup>. This kind of architecture is able to process images in an extremely effective and computationally efficient manner. In fact, Convolutional Neural Networks have been extensively employed in many medical imaging fields such as skin lesion analysis <sup>10, 11</sup> and retinopathy or pneumonia<sup>12, 13</sup>, growing into an extremely powerful tool to support specialists in the clinical decision making. In this work, we aimed to exploit artificial neural networks to build an automated tool supporting the diagnostic process for the immunofluorescence classification of kidney biopsy.

## Methods

### Kidney biopsy and Immunofluorescence staining and image acquisition

The analysis of the immunofluorescence images used in this study was approved by the local Ethical Committee ('Comitato Etico dell'Area Vasta Emilia Nord', Protocol #434/2019/OSS/AOUMO). The immunofluorescence images were collected in our center during the years 2001 – 2018; they were obtained from native or kidney allograft biopsies, independently of the clinical indication for the biopsy

1  
2  
3 collection. The only inclusion criterion for the study was the availability of images and of the pathology  
4 report. In the kidney biopsy procedure, two cores of kidney tissue were obtained in the cortical kidney area  
5 by a semiautomatic needle (16 or 18 gauge x 16cm, Bard Max-Core). The core assigned to light microscopy  
6 was Bouin-fixed and routinely processed according to conventional procedures. The second core, prepared  
7 for direct immunofluorescence, was quickly frozen in liquid nitrogen sliced in 3 $\mu$ m cryo-section (Cryostat  
8 CM1950, Leica Microsystem) and stained with FITC-conjugated rabbit antisera directed against human  
9 immunoglobulin (IgG, IgA, IgM and their fractions: Lambda and Kappa chains), Fibrinogen (F) and  
10 complement components C1q and C3 (all antisera were from DakoCytomation). The slides were incubated  
11 with the primary antibodies at 20°C in the dark for 30min, then washed 2 times, 15min each, in PBS (IgG  
12 and C3 1:80; IgA, IgM, C1q; F, K, Lambda 1:40; each dilution with PBS1X). Images were captured at 400X  
13 apparent magnification with a fluorescence microscope (BX41 with U-RFL-T, Olympus) by a digital camera  
14 (XC30, Firmware version 4.0.2, Olympus) controlled by a dedicated software (CellB software, Olympus), and  
15 stored as gray-scale uncompressed TIFF images with 16 bits per pixel. All the images were captured with a  
16 standardized protocol and in particular at a fixed time of exposure. The quality and sharpness of the  
17 available images can be checked on line (<https://nephronn.ing.unimore.it/public/>) on a sample of 180  
18 images (Test set).  
19  
20  
21  
22  
23

## 24 25 Ground Truth Training and Test Sets

26  
27 The immunofluorescence report (Supplemental Material Table 1 adopted at our center describes the  
28 '*Appearance*', '*Distribution*' and '*Location*' of immune deposits. Furthermore, the '*Intensity*' of staining is  
29 described by a semi-quantitative rank with discrete values from 0 to 3 (by 0.5 intervals).  
30

31 A separate description is detailed for each anatomical compartment (glomeruli, tubules, interstitium,  
32 vessels), but in the present work only the glomerular compartment was considered. The standard report  
33 includes the evaluation of immunoglobulins, fibrinogen and complement deposition. The deposit  
34 '*Appearance*' is categorized in the following predefined classes: granular (coarse granular or fine granular),  
35 linear or pseudo-linear (Supplemental Material Figure 1). The '*Distribution*' of deposits is described as focal  
36 or diffuse, and segmental or global (Supplemental Material Figure 2). With regard to '*Location*', the  
37 following predefined classes were annotated: mesangial or capillary wall (Supplemental Material Figure 3).  
38 When possible, capillary wall sub-location has been defined: continuous regular capillary wall (sub-  
39 epithelial), discontinuous regular capillary wall, irregular capillary wall. All annotations underwent a quality  
40 control check and when needed they were corrected by an experienced kidney pathologist. Ground Truth  
41 was provided by F.L. and L.M (respectively Pathologist 3 and 2). All the pathologists reviewed the test set:  
42 F.F. (Pathologist 1) in particular did not contribute to the Ground Truth definition but participated to the  
43 test set evaluation. A detailed description of pathologist contribution is reported in the '*Author*  
44 '*Contribution*' section.  
45  
46  
47  
48  
49  
50

## 51 Convolutional Neural Networks Design

52 For all the features with the exception of '*Intensity*', the algorithm is required to assign an image to one of  $k$   
53 categories. In order to do so, a neural network processes the input image and outputs  $k$  values, each of which  
54 is the likelihood that the input sample belongs to the  $k$ -th class. One image can only belong to one class;  
55 therefore, the final prediction is the class with the highest score. A different strategy is employed to predict  
56 the '*Intensity*' of the deposits as a value ranging from 0 to 3 (with intervals of 0.5). For this task, a neural  
57 network is asked to output a probability for each of the possible intensity values (7 different classes). The  
58 final intensity prediction is obtained as:  
59  
60

$$\sum_k^n p_k * v_k$$

where  $n$  is the number of classes ( $n=7$ ),  $p$  is the probability that the network assigns to  $k$ th class and  $v$  is the value represented by  $k$ th class.

Convolutional neural networks designed for classification progressively reduce the image dimensions, while increasing the number of feature maps obtained after each layer. ResNet-101<sup>14</sup> exploits 101 convolutional layers to progressively transform input images composed of 512x512 pixels and the 3 channels (R, G, and B) in images with 16x16 pixels and 2048 feature maps. From each 16x16 features map, a single value is obtained as the average of the 256 pixels and, finally, a non-convolutional fully connected layer is trained to process the 2048 mean values of the channels of features and obtain the two output values, representing the final predicted probability for each class. Input images are given a square shape by adding black bands close to the borders when needed, and are then resized to 512x512.

In order to help Convolutional Neural Networks to generalize their processing abilities, and therefore avoid overfitting, two regularization techniques are applied during the training process: data augmentation<sup>4</sup> and dropout<sup>5</sup>. To perform data augmentation, each image is randomly flipped and rotated before feeding it to the network, since neither of these two transformations should change the content of the image, and therefore the deposit characteristics. To put dropout into use, feature maps are randomly set to zero before the fully connected layer, in order to boost the resilience of the model. Regularization techniques are excluded during inference, when the goal is to obtain the best possible prediction for a single, novel image.

In this work transfer learning is exploited by pre-training Convolutional Neural Networks using ImageNet<sup>15, 16</sup> and then fine-tuning them to correctly classify immunofluorescence images. The learning rate set to  $1^{-5}$  during the fine-tuning process, and a weighted cross-entropy loss is employed.

In order to improve the interpretability of the algorithm, the Grad-Cam method<sup>17</sup> is used to draw attention heatmaps. This is done by first automatically analyzing how each convolutional filter affects the predictive score of each class, and assigning a *weight* to each filter-class couple. Then, when a sample is fed to the network the *weights* are used to highlight which pixels of the input image had a bigger impact on the prediction of the final class.

#### *Assessment of Convolutional Neural Networks Performance*

The Convolutional Neural Networks classification performance in feature recognition of 'Appearance', 'Distribution' and 'Location' is assessed using four different metrics:

$$Accuracy = \frac{True\ Positives + True\ Negatives}{True\ Positives + True\ Negatives + False\ Positives + False\ Negatives};$$

'Accuracy' is the ratio between the correct predictions and the total predictions'

$$Recall = \frac{True\ Positives}{True\ Positives + False\ Negatives} \quad ; \quad \text{'Recall' has the same statistical meaning as 'Sensitivity'}$$

$$Precision = \frac{True\ Positives}{True\ Positives + False\ Positives} \quad ; \quad \text{'Precision' has the same statistical meaning as 'Positive Predictive Value'}$$

$$F1\ Score = 2 * \frac{Precision * Recall}{Precision + Recall} \quad ; \quad \text{'F1 Score' has the same statistical meaning as 'Weighted Average of Precision and Recall'}$$

Moreover, Receiver Operating Characteristic (ROC) curves are presented, to display the relation between True Positive Rate and False Positive Rate and to compute the Area Under the Curve (AUC).

True Positive Rate is equal to the Recall, while False Positive Rate is calculated according to the following equation:

$False\ Positive\ Rate = \frac{False\ Positives}{True\ Negatives + False\ Positives}$  ; False Positive Rate has the same statistical meaning as 'Fall-Out'

For the Convolutional Neural Networks performance of the 'Intensity' we represented the classification result in a confusion matrix. The confusion matrix is a specific type of data representation<sup>18</sup> that reports the classes of data provided by the algorithm on each row, while it represents the classes of observed data on each column (or vice versa). The name derives from the fact that it allows a simple visualization of mistakes that the system tends to make in the classification of objects.

Furthermore, we calculated Mean Absolute Error (MAE) and Mean Squared Error (MSE). Where  $f \in X$  represent predicted values and  $y \in Y$  represent ground truth labels, MAE and MSE are defined as:

$$MAE = \frac{\sum_{k=0}^n |y_k - x_k|}{n}$$

$$MSE = \frac{\sum_{k=0}^n (y_k - x_k)^2}{n}$$

Agreement between evaluators, Convolutional Neural Networks and ground truth was calculated by the Cohen's K<sup>19</sup>. The average between features K presented in Table 4 is reported as mean  $\pm$  standard deviation. All the statistical Analysis were performed by the software package STATA/IC 11.2 for Windows (STATA Corp 4905 Lakeway Drive, College Station TX 77845, USA).

## Results

### The dataset

The database of immunofluorescence analysis collects the reports of 2542 consecutive subjects that underwent kidney biopsy between October 2001 and December 2018 in the Division of Nephrology and Dialysis of the AOU Policlinico di Modena, Italy. For each specimen at least one image of a glomerulus for each antibody is available, but frequently the images of more than one glomerulus are stored for the same antibody. The main clinical characteristics of the patients are reported in Table 1. The biopsy samples were obtained from native or transplant kidneys, independently of the indication for the biopsy procedure. The only criterion for inclusion was the availability of images and of the pathology report. 2225 are reports of native kidneys while 317 reports describe biopsies of transplanted kidneys. The first ten more frequent final diagnoses for the native kidneys are reported in Supplemental Table 1. Each report describes the result of the immunofluorescence analysis of the kidney specimen of the patient for the following deposits: IgG, IgA, IgM, C1q, C3, Fibrinogen, kappa and lambda light chains. The definition of the features collected in this study are described in Figure 1-3 of supplemental materials. The distribution of the immunofluorescence features is reported in Supplemental Table 3 and 4. According to the prevalent histological diagnosis of IgA Nephropathy, the most frequent '*Location*' feature is mesangial, the most frequent '*Appearance*' feature is coarse granular and the most frequent '*Distribution*' feature is global / diffuse. The image dataset collects 12259 images. The partitioning of the images into a training set (11059 images), validation set (200 images) and test set (1000 images) is shown in Figure 1. For each feature, a different training, validation and test set was prepared. Given the fact that negative samples are overrepresented for every feature, it is crucial that the three subsets have similar positive/negative samples ratio. In particular, a test set containing no positive samples would be futile. The presence (or absence) of the investigated feature is thus taken into account when splitting the dataset, in order to force similar data distribution into every subset. Every other patient metadata is ignored during the partitioning and their distribution in the subsets is ruled by random distribution.

### Performance of the Convolutional Neural Networks

We trained the neural networks to recognize the four different '*Locations*': mesangial, continuous regular capillary wall, discontinuous regular capillary wall, irregular capillary wall. A general class named capillary wall obtained by merging all the different kinds of capillary wall locations was added to the analysis. Regarding '*Location*', the pattern recognition tasks are completely independent classification problems, since each image could present neither, none, or every combination of the investigated location characteristics. For each one of the tasks, a specific Convolutional Neural Network is trained to classify a single image as either presenting or not presenting the pattern of interest. The same approach is taken to tackle the deposit '*Appearance*' (linear/pseudolinear, coarse granular and fine granular) and the deposit '*Distribution*' (segmental and global). Since our dataset is composed of high-magnification images (400X) that do not allow a panoramic evaluation of the slice, the '*Location*' features 'diffuse' or 'focal' have been left out from the analysis. The discontinuous regular capillary wall and the linear/pseudolinear features were not further assessed because of the low performance of the Convolutional Neural Network on these parameters: the linear / pseudolinear and discontinuous regular capillary wall F1 scores were 0.26 and 0.09 respectively, much lower values than obtained by the other features (Table 2). We suspect that the bad performance of these features is mainly attributable to the poor representation in our dataset (linear / pseudolinear and discontinuous regular capillary respectively 88, 335 and 232 features - Supplemental Table 3). However, only the evaluation of a larger dataset with a greater representation of these features will allow to verify this hypothesis.



1  
2  
3 Table 2 represents the performance of the neural network to predict the immunofluorescence features. All  
4 the features of 'Location', 'Appearance' and 'Distribution' obtained levels of accuracy around 0.8 and above.  
5 The highest performance was obtained by the 'Appearance' feature 'fine granular' (0.94) followed by the  
6 'Location' feature 'continuous regular capillary wall' (0.91). The ROC curve of each feature is presented in  
7 Figure 3.  
8  
9

10 Eight representative examples of the ground truth and automatic feature classification used in the test set  
11 are depicted in Figure 2. Furthermore, the figures show the results obtained through the Grad-Cam method.  
12 The heatmaps indicate that neural networks correctly identify the glomerulus within an image, and search  
13 for the pattern in the appropriate section of the image. The final decision to classify the images is notably  
14 influenced by the deposits on the glomerular structure.  
15

16 For what concerns the 'Intensity', we were able to predict the intensity level of images with a good level of  
17 approximation as confirmed by the Mean Absolute Error (MAE) of 0.398, and a Mean Squared Error (MSE) of  
18 0.455, which are two of the most relevant metrics for regression tasks such as intensity prediction<sup>20</sup>. Table 3  
19 presents the confusion matrix of the seven intensity classes, where the predicted class for each sample is the  
20 closest one to the real output value (which is a continuous number between 0 and 3).  
21  
22

23 We have preliminary data suggesting that the algorithm is robust enough to correctly recognize the  
24 glomerular pattern even in the presence of unusual tubular changes. We refer to some peculiar conditions  
25 with the presence of deposits along the tubular basement membrane as, for example, can occur in lupus  
26 nephritis or in light chain deposition disease. We have included in the supplementary materials the figures,  
27 the Ground Truth and the interpretation of the Convolutional Neural Network of two cases (Supplemental  
28 Figure 4 and 5, Supplemental Table 7 and 8). Nonetheless, given the small number of these cases, this  
29 statement will have to be confirmed in the light of a broader series.  
30  
31  
32

### 33 Agreement between pathologists, ground truth and Convolutional Neural Network

34 We compared the classifications of a set of 20 images for each feature (180 test images overall) between  
35 three different pathologists and Convolutional Neural Networks with respect to ground truth (Supplemental  
36 Table 5). The human evaluators as well as the Convolutional Neural Network were blinded about clinical  
37 information related to the images. The data suggest a fair to moderate agreement between pathologists and  
38 ground truth; Convolutional Neural Networks had a performance with respect to ground truth comparable  
39 to human evaluators. We also calculated for each possible evaluator, Convolutional Neural Network and  
40 ground truth pair, an average between the Cohen's K of each feature (Table 4). Also in this case the data  
41 suggest a moderate agreement between the different evaluators and, in particular, Convolutional Neural  
42 Networks show comparable performance to human evaluators. The inter-agreement Cohen's K of every  
43 feature is reported in supplemental materials table 6.  
44  
45  
46

### 47 Time analysis

48 The three pathologists spent respectively 25, 23 and 18 minutes in the analysis of the 180 test images. This  
49 corresponds to an average time of evaluation per image of  $7.3 \pm 1.2$  seconds. The execution time of the  
50 Convolutional Neural Networks has been calculated on an Intel(R) Xeon(R) E5-2650 v2@2.60GHz CPU with  
51 32 cores and 126GB of RAM, running Ubuntu 18.04 LTS and using a Tesla K80 GPU with 12GB of memory.  
52 The time for the complete analysis of the 180 test images by the Convolutional Neural Networks was 11.2  
53 seconds. This corresponds to 62.5 milliseconds per image which is 117.3 times faster than the average human  
54 evaluator.  
55  
56  
57  
58  
59  
60

## Discussion

In this study we presented a Deep Learning approach for the automatic reporting of immunofluorescence specimens of kidney biopsy. The main finding of our work is that Convolutional Neural Networks achieve an accurate definition of the main pathology features usually collected from this type of specimen: 'Appearance', 'Distribution', 'Location' and 'Intensity' of the deposits. The comparison of agreement among three pathologists shows a moderate agreement and, in this respect, Convolutional Neural Networks show a similar performance to human evaluators. The interest in Deep Learning applied to nephropathology has grown significantly in recent years. The researchers adopted Deep Learning technology to achieve different goals: some focused on the automated identification of the main microscopic structures (glomeruli, tubules, vessels etc.)<sup>21-25</sup>, some authors worked on the segmentation of the identified structures<sup>2</sup>, and others have used Convolutional Neural Networks to obtain automated clinical classifications<sup>1,2</sup>. We can consider our approach as part of the strategies for developing a tool for assisting in automated clinical classification. In this work we did not deal with the problem of the preliminary extraction of glomerular structures, that was already addressed in many previous papers<sup>21-25</sup>. Since our dataset was already made up of high-magnification images of individual glomeruli historically collected in the last 18 years of manual reporting of the kidney specimens, we focused our attention on the pure classification of these images. Immunohistopathologic analysis (in particular the immunofluorescence approach) has never been previously addressed with Deep Learning techniques. We therefore focused on the ability to extract from these images the features normally collected by our pathologists during the examination. The immunofluorescence interpretation, like many other clinical imaging techniques, is a field in which the Deep Learning approach can offer a significant contribution. Indeed, any medical imaging technique involving human reporting is conditioned by a significant interpretative subjectivity. This poor reproducibility has been repeatedly documented in the old kidney histopathological literature<sup>26-29</sup> and has been recently confirmed in studies concerning the application of Deep Learning in kidney pathology<sup>1,2</sup>. The indices of agreement between human pathologists vary according to the metrics used, the preparation of the specimen, the histological parameters assessed, and the kidney disease considered; overall, reported agreement rate between human kidney pathologists is fair to moderate with agreement ratio ranging between 0.3 and 0.6. The agreement among the pathologists of our center settles within the moderate level, according to Cohen's K methodology<sup>30</sup>. In this scenario, the Convolutional Neural Network algorithms have ground truth agreement performances similar to the human operators. Our data suggest a high accuracy in identifying the analyzed features, which varies from 0.79 to 0.94. The feature in which the system performs best is the identification of the 'Fine Granular Appearance', which together with the 'Continuous Regular Capillary wall Location' (the second-best identified feature), are often associated in the definition of the subepithelial deposits of Membranous Glomerulonephritis. With regard to these two patterns, the levels of agreements expressed between pathologists and ground truth has proved to be completely analogous to that between Convolutional Neural Networks and ground truth. However, even if less outstanding, also the agreement relating to the remaining features shows a Convolutional Neural Networks performance comparable to the human evaluators. It can be assumed that the imperfect concordance between ground truth with respect to Convolutional Neural Network and pathologists may be attributable to a different analytical setting (direct microscopic visualization of the specimen compared to captured images). However, we believe that this is a partial and not a substantial explanation of the phenomenon. In fact, our data suggest that under the same analytical conditions, the pathologists show a rather modest agreement between them (table 4), a well-known phenomenon against which an automated approach can bring improvements. Even for Intensity, a critical parameter for the reporting, a substantial subjectivity is documented (data not shown). To the best of our knowledge, there are no literature data suggesting that a simple determination of the fluorescence intensity level could replace a subjective (or Convolutional Neural Network) assessment. The classification of intensity still remains a semi-quantitative evaluation as several pre-analytical variables are not completely controllable and do not allow a purely

1  
2  
3 quantitative evaluation. These include the variability of fluorescence emission between commercial  
4 antibody batches, the decay of UV lamp fluorescence during its lifetime and many other confounders that  
5 are not usually under control (lab temperature, solution batches, manual skill of the technician etc.).  
6

7 The reproducibility of the evaluation, which in itself already has enormous value, is not the only advantage  
8 that Deep Learning can introduce in this field; the speed of interpretation, the amount of data that can be  
9 analyzed and the relative economic savings all represent elements of great interest in the adoption of this  
10 technology. About this aspect the speed of execution of the analyses is crucial: the pathologists spent on  
11 average 22 minutes in the assessment of 180 images. Convolutional Neural Networks processed the same  
12 images in 11.2 seconds with comparable accuracy. This in itself represents 117.3 times greater efficiency than  
13 the human evaluator. Nevertheless, compared to the evaluation of a specimen in a real scenario, it must be  
14 taken into account that our simulation does not consider the times of identification of the glomeruli and the  
15 image acquisition from the microscope, since these were not goals of our study.  
16  
17  
18

19 Our group has developed a web platform at <https://nephronn.ing.unimore.it/> where it is possible to upload  
20 digitized images of glomeruli from immunofluorescence preparations at high magnification (400X). Once  
21 the image is loaded, it is analyzed and the prediction for each one of the 9 features currently available in  
22 our algorithm is presented. Due to the current hardware constraints, the web application relies on  
23 shallower neural networks than the ones described in this paper, causing a classification accuracy drop of  
24 about 5%. This platform will allow the kidney pathology community to directly experience the potential of  
25 our approach. In particular, this service will allow the development of possible collaborations with other  
26 kidney pathology centers for sharing images and contributing to the development of a larger common  
27 immunofluorescence database. At the moment, in fact, this experience is based on a relatively  
28 homogeneous data set; indeed, in this study, all the preparations were obtained from a single center that  
29 processed and digitized the images with a single protocol. It cannot be excluded that our algorithm may  
30 have lower performances when applied to images derived from other centers. Another limitation of our  
31 approach consists in the impossibility of recognizing with a significant accuracy the features that were less  
32 represented in our data set (discontinuous regular capillary wall and the linear / pseudoliner) or which are  
33 not captured by our standardized report (for example the amorphous / globular - IgM / Complement  
34 staining in segmental glomerulosclerosis). With collaborations already formalized with other laboratories,  
35 we plan to increase our data set and to incorporate these additional features into our algorithm in the near  
36 future. Another possible limitation of our approach could be identified in the double role of pathologist 2  
37 and pathologist 3 in defining the ground truth and in the analysis of the Testing set. However, we do not  
38 believe that this negatively affects the result of our work. Indeed, the dual role of pathologists 2 and 3  
39 would have possibly inflated their K value of concordance with the ground truth to the detriment of  
40 pathologist 1 and the Convolutional Neural Network. This phenomenon, which however does not emerge  
41 from the data (table 4), would eventually make the estimate of concordance K of the Convolutional Neural  
42 Network more conservative.  
43  
44  
45  
46  
47  
48

49 We can speculate that our algorithm can be part of a larger computer-aided pathology diagnostic platform.  
50 This platform, in the context of a network of nephrological centers engaged in kidney biopsy diagnostics,  
51 could provide for local scanning of immunofluorescence preparations (Whole Slide Imaging). The scanned  
52 image could be sent to the central processing center which will apply a glomeruli identification algorithm<sup>25</sup>  
53 and the subsequent analysis of the individual glomeruli for the automatic definition of the characteristics  
54 of 'Appearance', 'Distribution', 'Location' and 'Intensity'. An averaged summary of the results obtained by  
55 each glomerulus could therefore be organized in a report which would be sent to the local center for  
56 validation and implementation in the flowchart of the biopsy diagnostic process.  
57  
58

59 In conclusion, we presented a Deep Learning approach capable of recognizing with significant accuracy the  
60 main features normally collected in the reporting of kidney immunofluorescence. The data showed an

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

accuracy comparable to what is normally expressed by pathologists who are experts in the field. In the near future, the collaboration with other centers with experience in kidney pathology will allow us to complete our algorithm with features that are currently under-represented and will guarantee external validity of our analysis. We believe that the algorithm developed by our group can become a useful tool to support the reporting of kidney immunofluorescence.

For Peer Review

## Author contribution

Ligabue Giulia, acquisition of immunofluorescence images, draft and revision of the paper

Pollastri Federico, development of the Convolutional Neural Networks, draft and revision of the paper

Fontana Francesco, pathologic evaluation of the test set (Pathologist 1).

Leonelli Marco, pathologic definition of the Ground Truth and evaluation of the test set (Pathologist 2).

Furci Luciana, pathologic definition of the Ground Truth and evaluation of the test set (Pathologist 3).

Giovanella Silvia, acquisition of immunofluorescence images, revision of the paper

Alfano Gaetano, revision of the paper

Cappelli Gianni, revision of the paper

Testa Francesca, revision of the paper

Federico Bolelli, development of the web page, revision of the paper

Grana Costantino: design of the study, revision of the paper

Magistrone Riccardo: conception and design of the study, data analysis, draft and revision of the paper

All authors approve the final version of the manuscript.

## Acknowledgements

This work was supported by personal academic funds of Prof. Riccardo Magistrone.

## Disclosure

None of the authors of the work has any financial conflict of interest to submit this manuscript.

## Figures

**Figure 1.** The image dataset collects 12259 images. The partitioning of the images into a training set (11059 images), validation set (200 images) and test set (1000 images) is shown.

**Figure 2.** Each panel of the figure represents the result of the feature specific Convolutional Neural Networks elaboration of a test image. Inside each panel the picture on the left shows the original test image, while the picture on the right shows its heatmap. In the heatmap the red area shows the sections of the image most involved in the classification process by the Convolutional Neural Network. All the presented images were correctly classified by the specific Convolutional Neural Network. A) Parietal; B) Mesangial; C) Continuous Regular Capillary Wall; D) Irregular Capillary Wall; E) Coarse Granular; F) Fine Granular; G) Segmental; H) Global

**Figure 3.** ROC curves of the Convolutional Neural Networks performance of prediction of each feature.

## Table of Contents of Supplemental material

- **Supplemental Table 1. Immunofluorescence report.**
- **Supplemental Table 2. Final diagnoses of the kidney biopsies.**
- **Supplemental Table 3. Numeric distribution of the immunofluorescence features.**
- **Supplemental Table 4. Percent distribution of the immunofluorescence features.**
- **Supplemental Table 5. Ground truth agreement between pathologists and Convolutional Neural Networks.**
- **Supplemental Table 6. Inter-agreement Cohen’s K of each feature.**
- **Supplemental Table 7. Comparison of the report of a case of Lupus Nephritis (Supplemental Figure 4) between Ground Truth and Convolutional Neural Networks (CNN).**
- **Supplemental Table 8. Ground truth agreement between pathologists and Convolutional Neural Networks.**
  
- Supplemental Figure 1. Classification of the ‘*Appearance*’ of the deposits of the Immunofluorescence specimen.
- Supplemental Figure 2. Classification of the ‘*Distribution*’ of the deposits of the Immunofluorescence specimen.
- Supplemental Figure 3. Classification of the ‘*Location*’ of the deposits of the Immunofluorescence specimen.
- Supplemental figure 4. **IgG direct Immunofluorescence of a Lupus Nephritis case.**
- Supplemental figure 5. **IgG direct Immunofluorescence of a Light Chain deposition disease.**

## References

1. Ginley, B, Lutnick, B, Jen, KY, Fogo, AB, Jain, S, Rosenberg, A, Walavalkar, V, Wilding, G, Tomaszewski, JE, Yacoub, R, Rossi, GM, Sarder, P: Computational Segmentation and Classification of Diabetic Glomerulosclerosis. *J Am Soc Nephrol*, 30: 1953-1967, 2019.
2. Hermsen, M, de Bel, T, den Boer, M, Steenbergen, EJ, Kers, J, Florquin, S, Roelofs, J, Stegall, MD, Alexander, MP, Smith, BH, Smeets, B, Hilbrands, LB, van der Laak, J: Deep Learning-Based Histopathologic Assessment of Kidney Tissue. *J Am Soc Nephrol*, 30: 1968-1979, 2019.
3. Hawkins, DM: The problem of overfitting. *J Chem Inf Comput Sci*, 44: 1-12, 2004.
4. Pollastri, F, Bolelli, F, Paredes, R, Grana, C: Augmenting data with GANs to segment melanoma skin lesions. *Multimedia Tools and Applications*, 2019.
5. Srivastava, N, Hinton, G, Krizhevsky, A, Sutskever, I, Salakhutdinov, R: Dropout: A Simple Way to Prevent Neural Networks from Overfitting. *J Mach Learn Res*, 15: 1929-1958, 2014.
6. Ioffe, S, Szegedy, C: Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift. *ArXiv*, abs/1502.03167, 2015.
7. Litjens, G, Kooi, T, Bejnordi, BE, Setio, AAA, Ciompi, F, Ghafoorian, M, van der Laak, J, van Ginneken, B, Sanchez, CI: A survey on deep learning in medical image analysis. *Med Image Anal*, 42: 60-88, 2017.
8. Canalini, L, Pollastri, F, Bolelli, F, Cancilla, M, Allegretti, S, Grana, C: Skin Lesion Segmentation Ensemble with Diverse Training Strategies. Cham, Springer International Publishing, 2019 pp 89-101.
9. Bolelli, F, Baraldi, L, Pollastri, F, Grana, C: A Hierarchical Quasi-Recurrent approach to Video Captioning. *2018 IEEE International Conference on Image Processing, Applications and Systems (IPAS)*. 2018 pp 162-167.
10. Pollastri, F, Bolelli, F, Palacios, RP, Grana, C: Improving Skin Lesion Segmentation with Generative Adversarial Networks. *2018 IEEE 31st International Symposium on Computer-Based Medical Systems (CBMS)*. 2018 pp 442-443.
11. Esteva, A, Kuprel, B, Novoa, RA, Ko, J, Swetter, SM, Blau, HM, Thrun, S: Dermatologist-level classification of skin cancer with deep neural networks. *Nature*, 542: 115-118, 2017.
12. Gulshan, V, Peng, L, Coram, M, Stumpe, MC, Wu, D, Narayanaswamy, A, Venugopalan, S, Widner, K, Madams, T, Cuadros, J, Kim, R, Raman, R, Nelson, PC, Mega, JL, Webster, DR: Development and Validation of a Deep Learning Algorithm for Detection of Diabetic Retinopathy in Retinal Fundus Photographs. *JAMA*, 316: 2402-2410, 2016.
13. Kermany, DS, Goldbaum, M, Cai, W, Valentim, CCS, Liang, H, Baxter, SL, McKeown, A, Yang, G, Wu, X, Yan, F, Dong, J, Prasadha, MK, Pei, J, Ting, MYL, Zhu, J, Li, C, Hewett, S, Dong, J, Ziyar, I, Shi, A, Zhang, R, Zheng, L, Hou, R, Shi, W, Fu, X, Duan, Y, Huu, VAN, Wen, C, Zhang, ED, Zhang, CL, Li, O, Wang, X, Singer, MA, Sun, X, Xu, J, Tafreshi, A, Lewis, MA, Xia, H, Zhang, K: Identifying Medical Diagnoses and Treatable Diseases by Image-Based Deep Learning. *Cell*, 172: 1122-1131 e1129, 2018.
14. He, K, Zhang, X, Ren, S, Sun, J: Deep Residual Learning for Image Recognition. *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*: 770-778, 2015.
15. Alex, K, Sutskever, I, Hinton, GE: ImageNet Classification with Deep Convolutional Neural Networks. 1097--1105, 2012.
16. Deng, J, Socher, R, Fei-Fei, L, Dong, W, Li, K, Li, L-J: ImageNet: A large-scale hierarchical image database. *2009 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. 2009 pp 248-255.
17. Selvaraju, RR, Cogswell, M, Das, A, Vedantam, R, Parikh, D, Batra, D: Grad-CAM: Visual Explanations from Deep Networks via Gradient-Based Localization. *2017 IEEE International Conference on Computer Vision (ICCV)*: 618-626, 2016.
18. Tharwat, A: Classification assessment methods. *Applied Computing and Informatics*, 2018.
19. Cohen, J: A Coefficient of Agreement for Nominal Scales. *Educational and Psychological Measurement*, 20: 37-46, 1960.
20. Xu, Y, Pei, J, Lai, L: Deep Learning Based Regression and Multiclass Models for Acute Oral Toxicity Prediction with Automatic Chemical Feature Extraction. *J Chem Inf Model*, 57: 2672-2685, 2017.



21. Bukowy, JD, Dayton, A, Cloutier, D, Manis, AD, Staruschenko, A, Lombard, JH, Solberg Woods, LC, Beard, DA, Cowley, AW, Jr.: Region-Based Convolutional Neural Nets for Localization of Glomeruli in Trichrome-Stained Whole Kidney Sections. *J Am Soc Nephrol*, 29: 2081-2088, 2018.
22. Gadermayr, M, Dombrowski, AK, Klinkhammer, BM, Boor, P, Merhof, D: CNN cascades for segmenting sparse objects in gigapixel whole slide images. *Comput Med Imaging Graph*, 71: 40-48, 2019.
23. Gallego, J, Pedraza, A, Lopez, S, Steiner, G, Gonzalez, L, Laurinavicius, A, Bueno, G: Glomerulus Classification and Detection Based on Convolutional Neural Networks. *Journal of Imaging*, 4: 20, 2018.
24. Marsh, JN, Matlock, MK, Kudose, S, Liu, T-C, Stappenbeck, TS, Gaut, JP, Swamidass, SJ: Deep Learning Global Glomerulosclerosis in Transplant Kidney Frozen Sections. *bioRxiv*, 2018.
25. Zhao, K, Tang, YJJ, Zhang, T, Carvajal, J, Smith, DF, Wiliem, A, Hobson, P, Jennings, A, Lovell, BC: DGD: A Dataset for Detecting Glomeruli on Renal Direct Immunofluorescence. *2018 Digital Image Computing: Techniques and Applications (DICTA)*. 2018 pp 1-7.
26. Gamba, G, Reyes, E, Angeles, A, Quintanilla, L, Calva, J, Pena, JC: Observer agreement in the scoring of the activity and chronicity indexes of lupus nephritis. *Nephron*, 57: 75-77, 1991.
27. Grootcholten, C, Bajema, IM, Florquin, S, Steenbergen, 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.
28. 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.
29. 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.
30. Landis, JR, Koch, GG: The measurement of observer agreement for categorical data. *Biometrics*, 33: 159-174, 1977.

**Table 1. Main clinical characteristics of the patients.**

|                                   |                    |
|-----------------------------------|--------------------|
| Subjects                          | 2542               |
| M : F (%)                         | 62% : 38%          |
| Age (years)                       | 52 (38 - 66)       |
| Creatinine (mg/dl)                | 1.6 (1.0 - 2.7)    |
| eGFR (ml/min/1.73m <sup>2</sup> ) | 45 (22 - 78)       |
| CKD class:                        |                    |
| 1                                 | 18%                |
| 2                                 | 18%                |
| 3a                                | 13%                |
| 3b                                | 16%                |
| 4                                 | 18%                |
| 5                                 | 17%                |
| Hb (g/dl)                         | 12.0 (10.1 - 13.7) |
| Urine Protein/Creatinine          | 1.8 (0.5 - 4.3)    |
| Hematuria                         | 56%                |
| systolic BP (mmHg)                | 130 (120 - 142)    |
| diastolic BP (mmHg)               | 80 (70 - 84)       |
| Diabetes                          | 11%                |
| Hypertension                      | 37%                |
| Clinical presentation:            |                    |
| None                              | 11%                |
| Urinary Abnormalities             | 44%                |
| Nephrotic Syndrome                | 32%                |
| Nephritic Syndrome                | 9%                 |
| Macrohematuria                    | 4%                 |

M:F = male : female sex ratio; eGFR= estimated glomerular filtration rate; CKD = chronic kidney disease; Hb = hemoglobin; SBP = systolic blood pressure; DBP = diastolic blood pressure. With the exception of the male:female ratio, all the data are expressed as median (25° - 75° percentile) and percent.

**Table 2. Neural network performance in feature prediction.**

|                     | <b>Feature</b>                    | <b>Accuracy</b> | <b>Recall</b> | <b>Precision</b> | <b>F1 Score</b> | <b>AUC</b> |
|---------------------|-----------------------------------|-----------------|---------------|------------------|-----------------|------------|
| <b>Appearance</b>   | coarse granular                   | 0.84            | 0.61          | 0.34             | 0.44            | 0.85       |
|                     | fine granular                     | 0.94            | 0.76          | 0.43             | 0.47            | 0.83       |
| <b>Distribution</b> | segmental                         | 0.81            | 0.50          | 0.36             | 0.42            | 0.81       |
|                     | global                            | 0.82            | 0.74          | 0.87             | 0.79            | 0.89       |
| <b>Location</b>     | mesangial                         | 0.84            | 0.78          | 0.71             | 0.74            | 0.89       |
|                     | capillary wall                    | 0.81            | 0.77          | 0.66             | 0.71            | 0.87       |
|                     | continuous regular capillary wall | 0.91            | 0.82          | 0.75             | 0.78            | 0.94       |
|                     | irregular capillary wall          | 0.79            | 0.67          | 0.49             | 0.57            | 0.84       |

For Peer Review

**Table 3: Confusion Matrix of the intensity classes.**

Confusion matrix provides insight how prediction is correctly distributed over the seven different classes of intensity. The ground truth labels are given vertically and the predicted labels by the Convolutional Neural Network are written on the horizontal axis.

|     | 0    | 0.5  | 1    | 1.5  | 2    | 2.5  | 3    |
|-----|------|------|------|------|------|------|------|
| 0   | 0.82 | 0.11 | 0.04 | 0.00 | 0.02 | 0.00 | 0.01 |
| 0.5 | 0.31 | 0.29 | 0.22 | 0.05 | 0.07 | 0.02 | 0.05 |
| 1   | 0.16 | 0.16 | 0.28 | 0.06 | 0.24 | 0.00 | 0.10 |
| 1.5 | 0.16 | 0.10 | 0.20 | 0.04 | 0.20 | 0.02 | 0.28 |
| 2   | 0.03 | 0.03 | 0.10 | 0.10 | 0.34 | 0.00 | 0.39 |
| 2.5 | 0.06 | 0.03 | 0.07 | 0.01 | 0.25 | 0.03 | 0.55 |
| 3   | 0.00 | 0.00 | 0.02 | 0.00 | 0.06 | 0.00 | 0.92 |

**Table 4. Average agreement between evaluators.**

The average K of each feature  $\pm$  standard deviations of K of each feature is reported (180 test images overall). CNNs = Convolutional Neural Networks

| <b>Average</b> | Ground Truth    | CNNs            | Pathologist 1   | Pathologist 2   |
|----------------|-----------------|-----------------|-----------------|-----------------|
| Pathologist 3  | 0.39 $\pm$ 0.18 | 0.43 $\pm$ 0.22 | 0.56 $\pm$ 0.15 | 0.49 $\pm$ 0.17 |
| Pathologist 2  | 0.42 $\pm$ 0.07 | 0.54 $\pm$ 0.21 | 0.56 $\pm$ 0.16 |                 |
| Pathologist 1  | 0.45 $\pm$ 0.21 | 0.52 $\pm$ 0.15 |                 |                 |
| CNNs           | 0.49 $\pm$ 0.20 |                 |                 |                 |

For Peer Review

## Supplemental material

### Table of Contents of Supplemental material

- **Supplemental Table 1. Immunofluorescence report.**
- **Supplemental Table 2. Final diagnoses of the kidney biopsies.**
- **Supplemental Table 3. Numeric distribution of the immunofluorescence features.**
- **Supplemental Table 4. Percent distribution of the immunofluorescence features.**
- **Supplemental Table 5. Ground truth agreement between pathologists and Convolutional Neural Networks.**
- **Supplemental Table 6. Inter-agreement Cohen's K of each feature.**
- **Supplemental Table 7. Comparison of the report of a case of Lupus Nephritis (Supplemental Figure 4) between Ground Truth and Convolutional Neural Network.**
- **Supplemental Table 8. Ground truth agreement between pathologists and Convolutional Neural Networks.**
  
- Supplemental Figure 1. Classification of the '*Appearance*' of the deposits of the Immunofluorescence specimen.
- Supplemental Figure 2. Classification of the '*Distribution*' of the deposits of the Immunofluorescence specimen.
- Supplemental Figure 3. Classification of the '*Location*' of the deposits of the Immunofluorescence specimen.
- Supplemental figure 4. **IgG direct Immunofluorescence of a Lupus Nephritis case.**
- Supplemental figure 5. **IgG direct Immunofluorescence of a Light Chain deposition disease.**

**Supplemental Table 1. Immunofluorescence report.**

|           |              |   | IgG | IgA | IgM | C1q | C3 | F | Kappa | Lambda |
|-----------|--------------|---|-----|-----|-----|-----|----|---|-------|--------|
| GLOMERULI | appearance   | linear  |     |     |     |     |    |   |       |        |
|           |              | pseudolinear                                      |     |     |     |     |    |   |       |        |
|           |              | coarse granular                                   |     |     |     |     |    |   |       |        |
|           |              | fine granular                                     |     |     |     |     |    |   |       |        |
|           | distribution | diffuse/segmental                                 |     |     |     |     |    |   |       |        |
|           |              | diffuse/global                                    |     |     |     |     |    |   |       |        |
|           |              | focal/segmental                                   |     |     |     |     |    |   |       |        |
|           |              | focal/global                                      |     |     |     |     |    |   |       |        |
|           | location     | mesangial   |     |     |     |     |    |   |       |        |
|           |              | capillary wall                                    |     |     |     |     |    |   |       |        |
|           |              | continuous regular capillary wall (subepithelial) |     |     |     |     |    |   |       |        |
|           |              | capillary wall regular discontinuous              |     |     |     |     |    |   |       |        |
|           |              | irregular capillary wall (subendothelial)         |     |     |     |     |    |   |       |        |
|           |              | INTENSITY   |     |     |     |     |    |   |       |        |

The immunofluorescence report adopted at our center describes the 'Appearance', 'Distribution' and 'Location' of immune deposits. Furthermore, the 'Intensity' of staining is described by a semi-quantitative rank with discrete values from 0 to 3 (by 0.5 intervals). The report includes the evaluation of IgG, IgA, IgM immunoglobulins, complement fractions C1Q, C3, Fibrinogen, Kappa and Lambda light chains.

**Supplemental Table 2. Final diagnoses of the kidney biopsies.**

| <b>Diagnosis</b>   | <b>Freq</b> | <b>Percent</b> |
|--|-------------|----------------|
| IgA Nephropathy  | 401         | 16%            |
| Membranous Glomerulonephritis                                      | 232         | 9%             |
| Mesangial Proliferative Glomerulonephritis without immune deposits | 150         | 6%             |
| Diabetic Nephropathy   | 146         | 6%             |
| Mesangial Proliferative Glomerulonephritis with IgM Deposits       | 134         | 5%             |
| Chronic Interstitial Nephritis                                     | 111         | 4%             |
| Unclassified Nephropathy   | 98          | 4%             |
| Membranoproliferative Glomerulonephritis                           | 78          | 3%             |
| Benign Nephroangiosclerosis  | 65          | 3%             |
| Lupus Nephritis  | 64          | 3%             |
| Other Diagnoses  | 1063        | 41%            |
| <b>Total</b>   | <b>2225</b> | <b>100%</b>    |



**Supplemental Table 3. Numeric distribution of the immunofluorescence features.**

|                  |                     |   | <b>IgG</b> | <b>IgA</b> | <b>IgM</b> | <b>C1q</b> | <b>C3</b> | <b>F</b> | <b>Kappa</b> | <b>Lambda</b> | <b>Overall</b> |
|------------------|---------------------|---|------------|------------|------------|------------|-----------|----------|--------------|---------------|----------------|
| <b>GLOMERULI</b> | <b>appearance</b>   | linear  | 28         | 11         | 4          | 1          | 6         | 10       | 9            | 19            | <b>88</b>      |
|                  |                     | pseudolinear                                      | 139        | 75         | 18         | 2          | 14        | 22       | 45           | 20            | <b>335</b>     |
|                  |                     | coarse granular                                   | 55         | 78         | 277        | 87         | 260       | 28       | 74           | 127           | <b>986</b>     |
|                  |                     | fine granular                                     | 171        | 126        | 109        | 43         | 182       | 41       | 135          | 165           | <b>972</b>     |
|                  | <b>distribution</b> | diffuse/segmental                                 | 50         | 91         | 265        | 67         | 138       | 52       | 57           | 96            | <b>816</b>     |
|                  |                     | diffuse/global                                    | 565        | 734        | 656        | 279        | 987       | 370      | 618          | 937           | <b>5146</b>    |
|                  |                     | focal/segmental                                   | 32         | 61         | 209        | 59         | 134       | 71       | 37           | 51            | <b>654</b>     |
|                  |                     | focal/global                                      | 15         | 22         | 59         | 14         | 43        | 11       | 9            | 23            | <b>196</b>     |
|                  | <b>location</b>     | mesangial   | 154        | 682        | 556        | 183        | 661       | 186      | 292          | 613           | <b>3327</b>    |
|                  |                     | capillary wall                                    | 158        | 301        | 318        | 95         | 305       | 264      | 158          | 300           | <b>1899</b>    |
|                  |                     | continuous regular capillary wall (subepithelial) | 365        | 135        | 116        | 87         | 295       | 67       | 291          | 341           | <b>1697</b>    |
|                  |                     | discontinuous regular capillary wall              | 17         | 17         | 35         | 14         | 97        | 13       | 14           | 25            | <b>232</b>     |
|                  |                     | irregular capillary wall (subendothelial)         | 93         | 135        | 514        | 183        | 378       | 83       | 131          | 207           | <b>1724</b>    |

**Supplemental Table 4. Percent distribution of the immunofluorescence features.**

|           |              |   | IgG   | IgA   | IgM   | C1q   | C3    | F     | Kappa | Lambda |
|-----------|--------------|---|-------|-------|-------|-------|-------|-------|-------|--------|
| GLOMERULI | appearance   | linear  | 0.96% | 0.38% | 0.14% | 0.03% | 0.20% | 0.34% | 0.31% | 0.65%  |
|           |              | pseudolinear                                      | 5%    | 3%    | 0.61% | 0.07% | 0.48% | 0.75% | 1%    | 0.68%  |
|           |              | coarse granular                                   | 2%    | 3%    | 9%    | 3%    | 9%    | 0.96% | 3%    | 4%     |
|           |              | fine granular                                     | 6%    | 4%    | 4%    | 1%    | 6%    | 1%    | 5%    | 6%     |
|           | distribution | diffuse/segmental                                 | 2%    | 3%    | 9%    | 2%    | 5%    | 2%    | 2%    | 3%     |
|           |              | diffuse/global                                    | 19%   | 25%   | 22%   | 10%   | 34%   | 13%   | 21%   | 33%    |
|           |              | focal/segmental                                   | 1%    | 2%    | 7%    | 2%    | 5%    | 2%    | 1%    | 2%     |
|           |              | focal/global                                      | 1%    | 1%    | 2%    | 0.48% | 1%    | 0.38% | 0.31% | 0.78%  |
|           | location     | mesangial   | 5%    | 23%   | 19%   | 6%    | 23%   | 6%    | 10%   | 21%    |
|           |              | capillary wall                                    | 5%    | 10%   | 11%   | 3%    | 10%   | 9%    | 5%    | 10%    |
|           |              | continuous regular capillary wall (subepithelial) | 12%   | 5%    | 4%    | 3%    | 10%   | 2%    | 10%   | 12%    |
|           |              | discontinuous regular capillary wall              | 0.58% | 0.58% | 1%    | 0.48% | 3%    | 0.44% | 0.48% | 0.85%  |
|           |              | irregular capillary wall (subendothelial)         | 3%    | 5%    | 18%   | 6%    | 13%   | 3%    | 4%    | 7%     |

**Supplemental Table 5. Ground truth agreement between pathologists and Convolutional Neural Network.**

Agreement between human evaluators (three different pathologists) and Convolutional Neural Networks was calculated by the Cohen's Kappa (Kappa  $\pm$  standard deviation of 20 test images per feature).

|              | Feature                              | Pathologist 1<br>vs<br>Ground Truth | Pathologist 2<br>vs<br>Ground Truth | Pathologist 3<br>vs<br>Ground Truth | CNN<br>vs<br>Ground Truth |
|--------------|--------------------------------------|-------------------------------------|-------------------------------------|-------------------------------------|---------------------------|
| Appearance   | coarse granular                      | 0.40 $\pm$ 0.20                     | 0.50 $\pm$ 0.21                     | 0.60 $\pm$ 0.20                     | 0.40 $\pm$ 0.24           |
|              | fine granular                        | 0.30 $\pm$ 0.16                     | 0.30 $\pm$ 0.22                     | 0.20 $\pm$ 0.22                     | 0.30 $\pm$ 0.25           |
| Distribution | segmental                            | 0.41 $\pm$ 0.20                     | 0.50 $\pm$ 0.21                     | 0.10 $\pm$ 0.22                     | 0.20 $\pm$ 0.25           |
|              | global                               | 0.67 $\pm$ 0.21                     | 0.40 $\pm$ 0.18                     | 0.60 $\pm$ 0.22                     | 0.60 $\pm$ 0.21           |
| Location     | mesangial                            | 0.80 $\pm$ 0.22                     | 0.50 $\pm$ 0.21                     | 0.50 $\pm$ 0.21                     | 0.90 $\pm$ 0.11           |
|              | capillary wall                       | 0.60 $\pm$ 0.22                     | 0.40 $\pm$ 0.22                     | 0.40 $\pm$ 0.22                     | 0.50 $\pm$ 0.22           |
|              | continuous regular<br>capillary wall | 0.60 $\pm$ 0.20                     | 0.40 $\pm$ 0.18                     | 0.50 $\pm$ 0.22                     | 0.60 $\pm$ 0.21           |
|              | irregular capillary wall             | 0.30 $\pm$ 0.22                     | 0.40 $\pm$ 0.20                     | 0.20 $\pm$ 0.20                     | 0.40 $\pm$ 0.24           |

**Supplemental Table 6. Inter-agreement Cohen's K of each feature.**

Pathologist 1, 2, 3 (Pat.1, Pat. 2, Pat. 3); Ground Truth (GT); Convolutional Neural Network (CNNs).

| <b>Mesangial</b>                         | GT          | CNNs        | Pat.1       | Pat.2       |
|--|-------------|-------------|-------------|-------------|
| Pat.3                                    | 0.50 ± 0.22 | 0.59 ± 0.21 | 0.70 ± 0.18 | 0.34 ± 0.26 |
| Pat.2                                    | 0.50 ± 0.22 | 0.59 ± 0.21 | 0.50 ± 0.22 |             |
| Pat.1                                    | 0.80 ± 0.15 | 0.70 ± 0.18 |             |             |
| CNNs                                     | 0.90 ± 0.11 |             |             |             |
| <b>Capillary wall</b>                    | GT          | CNNs        | Pat.1       | Pat.2       |
| Pat.3                                    | 0.40 ± 0.24 | 0.70 ± 0.18 | 0.60 ± 0.21 | 0.60 ± 0.21 |
| Pat.2                                    | 0.40 ± 0.24 | 0.90 ± 0.11 | 0.42 ± 0.23 |             |
| Pat.1                                    | 0.60 ± 0.21 | 0.51 ± 0.22 |             |             |
| CNNs                                     | 0.50 ± 0.22 |             |             |             |
| <b>Continuous Regular Capillary wall</b> | GT          | CNNs        | Pat.1       | Pat.2       |
| Pat.3                                    | 0.50 ± 0.22 | 0.31 ± 0.24 | 0.52 ± 0.21 | 0.34 ± 0.23 |
| Pat.2                                    | 0.40 ± 0.24 | 0.32 ± 0.27 | 0.74 ± 0.20 |             |
| Pat.1                                    | 0.60 ± 0.21 | 0.57 ± 0.22 |             |             |
| CNNs                                     | 0.60 ± 0.21 |             |             |             |
| <b>Irregular Capillary wall</b>          | GT          | CNNs        | Pat.1       | Pat.2       |
| Pat.3                                    | 0.20 ± 0.25 | 0.40 ± 0.24 | 0.48 ± 0.23 | 0.52 ± 0.24 |
| Pat.2                                    | 0.40 ± 0.24 | 0.40 ± 0.24 | 0.48 ± 0.23 |             |
| Pat.1                                    | 0.30 ± 0.25 | 0.30 ± 0.25 |             |             |
| CNNs                                     | 0.40 ± 0.24 |             |             |             |
| <b>Coarse Granular</b>                   | GT          | CNNs        | Pat.1       | Pat.2       |
| Pat.3                                    | 0.60 ± 0.21 | 0.76 ± 0.18 | 0.76 ± 0.18 | 0.66 ± 0.21 |
| Pat.2                                    | 0.50 ± 0.22 | 0.89 ± 0.13 | 0.66 ± 0.21 |             |
| Pat.1                                    | 0.40 ± 0.24 | 0.76 ± 0.18 |             |             |
| CNNs                                     | 0.40 ± 0.24 |             |             |             |
| <b>Fine Granular</b>                     | GT          | CNNs        | Pat.1       | Pat.2       |
| Pat.3                                    | 0.20 ± 0.25 | 0.30 ± 0.25 | 0.30 ± 0.25 | 0.70 ± 0.18 |
| Pat.2                                    | 0.30 ± 0.25 | 0.35 ± 0.25 | 0.35 ± 0.25 |             |
| Pat.1                                    | 0.10 ± 0.26 | 0.61 ± 0.30 |             |             |
| CNNs                                     | 0.30 ± 0.25 |             |             |             |
| <b>Segmental</b>                         | GT          | CNNs        | Pat.1       | Pat.2       |
| Pat.3                                    | 0.10 ± 0.26 | 0.06 ± 0.27 | 0.42 ± 0.23 | 0.18 ± 0.26 |

|       |                 |                 |                 |
|-------|-----------------|-----------------|-----------------|
| Pat.2 | $0.50 \pm 0.22$ | $0.43 \pm 0.25$ | $0.45 \pm 0.22$ |
| Pat.1 | $0.30 \pm 0.25$ | $0.38 \pm 0.22$ |                 |
| CNNs  | $0.20 \pm 0.25$ |                 |                 |

| <b>Global</b> | GT              | CNNs            | Pat.1           | Pat.2           |
|---------------|-----------------|-----------------|-----------------|-----------------|
| Pat.3         | $0.60 \pm 0.21$ | $0.35 \pm 0.26$ | $0.67 \pm 0.20$ | $0.55 \pm 0.23$ |
| Pat.2         | $0.40 \pm 0.24$ | $0.47 \pm 0.27$ | $0.86 \pm 0.16$ |                 |
| Pat.1         | $0.50 \pm 0.22$ | $0.38 \pm 0.28$ |                 |                 |
| CNNs          | $0.60 \pm 0.21$ |                 |                 |                 |

For Peer Review

**Supplemental Table 7. Comparison of the report of a case of Lupus Nephritis (Supplemental Figure 4) between Ground Truth and Convolutional Neural Network (CNN).**

|           |              |   | IgG Ground Truth | IgG CNN |
|-----------|--------------|---|------------------|---------|
| GLOMERULI | appearance   | linear  |                  |         |
|           |              | pseudolinear                                      |                  |         |
|           |              | coarse granular                                   |                  |         |
|           |              | fine granular                                     | yes              |         |
|           | distribution | segmental   |                  |         |
|           |              | global  | yes              | yes     |
|           | location     | mesangial   |                  |         |
|           |              | capillary wall                                    |                  |         |
|           |              | continuous regular capillary wall (subepithelial) | yes              | yes     |
|           |              | discontinuous regular capillary wall              |                  |         |
|           |              | irregular capillary wall (subendothelial)         |                  |         |
|           | Intensity    | 3   | 3                |         |

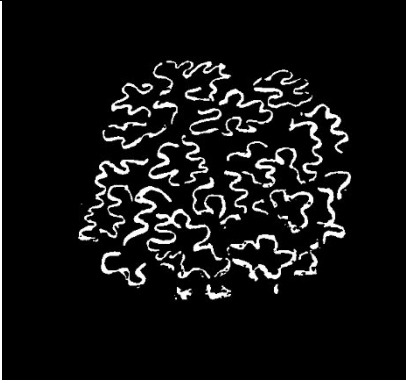
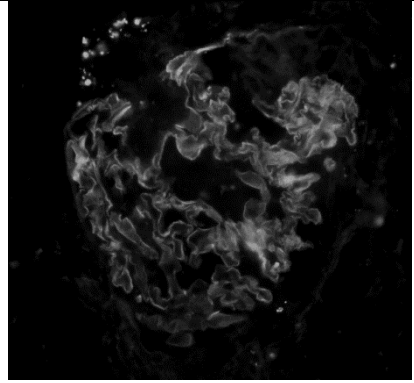
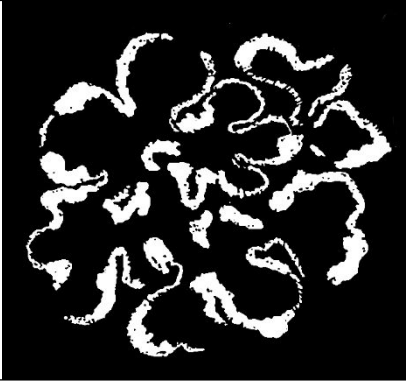
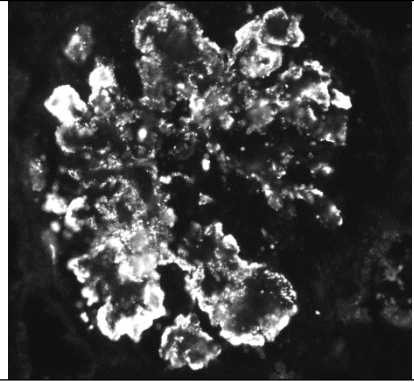
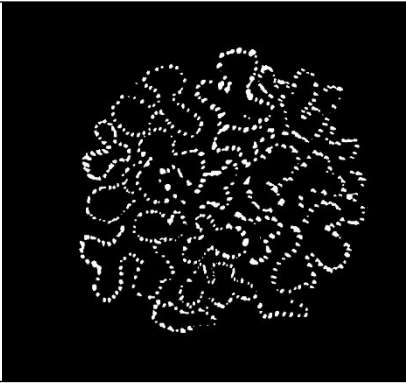
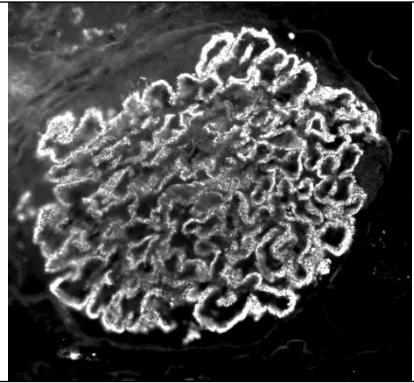
The Convolutional Neural Network could correctly recognize the deposits of the tuft (except for the fine granular appearance) despite of the presence of the extraglomerular deposits.

**Supplemental Table 8. Ground truth agreement between pathologists and Convolutional Neural Networks.**

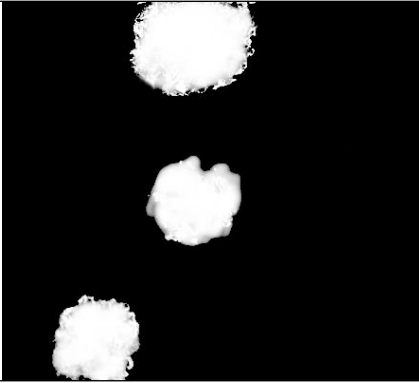
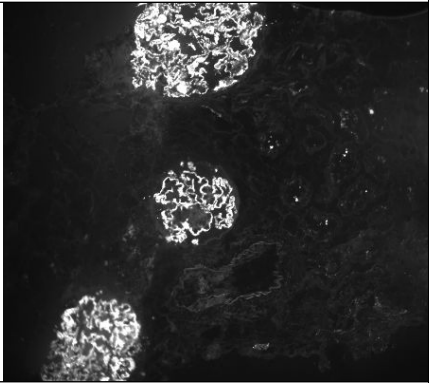

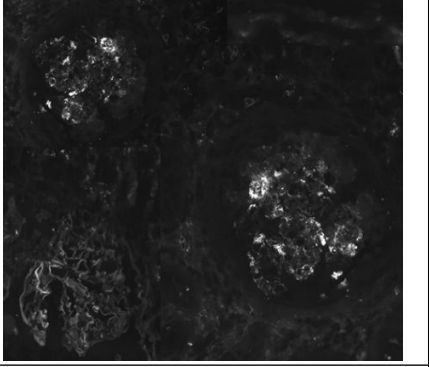

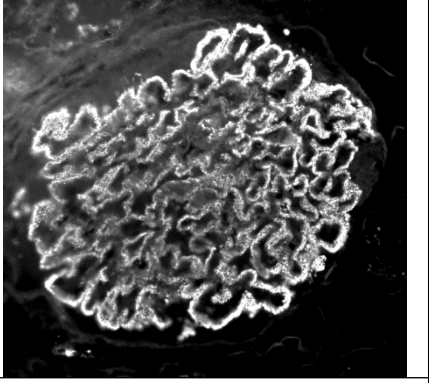

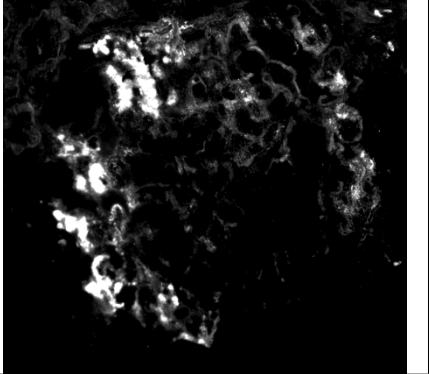
|           |              |   | Lambda Light Chain Ground Truth | Lambda Light Chain CNN |
|-----------|--------------|---|---------------------------------|------------------------|
| GLOMERULI | appearance   | linear  | yes                             |                        |
|           |              | pseudolinear                                      |                                 |                        |
|           |              | coarse granular                                   |                                 |                        |
|           |              | fine granular                                     |                                 |                        |
|           | distribution | segmental   |                                 |                        |
|           |              | global  | yes                             | yes                    |
|           | location     | mesangial   |                                 |                        |
|           |              | capillary wall                                    | yes                             | yes                    |
|           |              | continuous regular capillary wall (subepithelial) |                                 |                        |
|           |              | discontinuous regular capillary wall              |                                 |                        |
|           |              | irregular capillary wall (subendothelial)         |                                 |                        |
|           | Intensity    | 3   | 3                               |                        |

The Convolutional Neural Network (CNN) could correctly recognize the deposits of the tuft despite of the presence of the extraglomerular deposits. The Convolutional Neural Network could not recognize the linear appearance because it was not trained for this task.

**Supplemental Figure 1. Classification of the 'Appearance' of the deposits of the Immunofluorescence specimen<sup>1-3</sup>.**


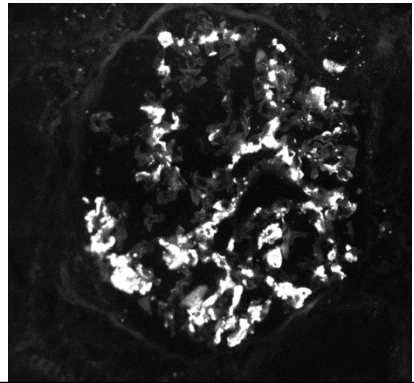
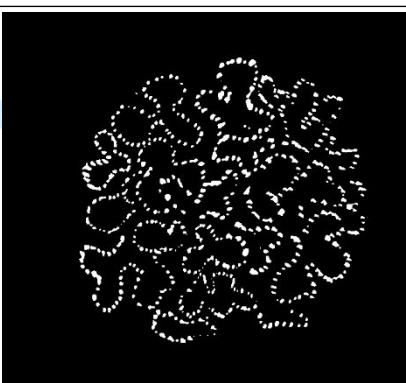
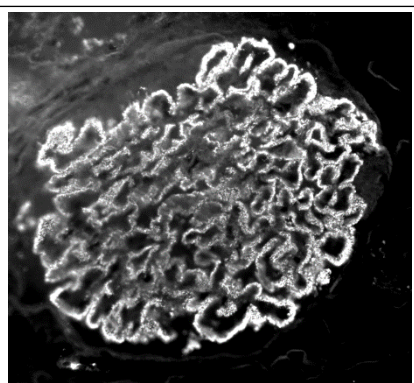
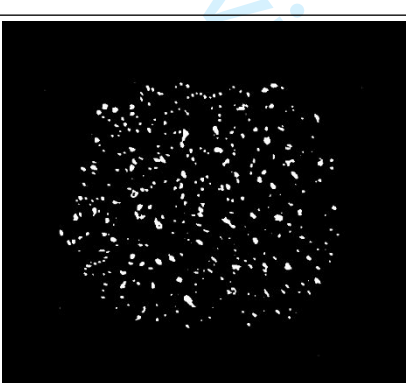
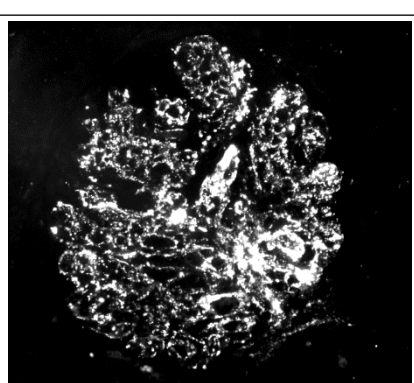

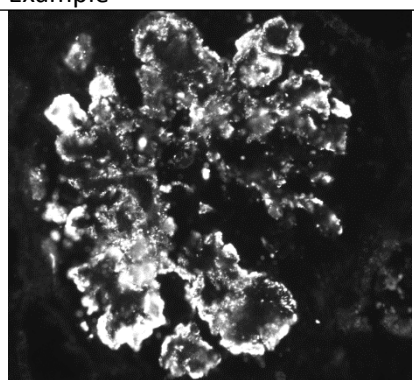
| <b>APPEARANCE</b>   |  |   |
|---|--|---|
| <b>LINEAR and PSEUDO-LINEAR</b>   | Scheme   | Example   |
| <p>The presence of the linear pattern is characteristic of the deposition of antibody against components of the glomerular basement membrane (GBM). The presence of circulant IgG against GBM is typical of the Goodpasture's syndrome<sup>4</sup>. Linear staining is also reported in diabetic nephropathy (pseudolinear) along the glomerular capillary walls with immunoglobulin G (IgG). In membranous glomerulonephritis the deposits are usually discrete and uniform, but in some cases, especially at the initial stage of the disease, they are small and confluent. In these instances, they could have a pseudolinear appearance.</p> |    |    |
| <b>COARSE GRANULAR</b>  | Scheme   | Example   |
| <p>Coarse granularity can be identified in different conditions. For example, this pattern can be found in the deposits of membranoproliferative lesions and in the sub-epithelial deposits of postinfectious glomerulonephritis.</p>   |  |  |
| <b>FINE GRANULAR</b>  | Scheme   | Example   |
| <p>Fine granularity as opposed to coarse granularity has smaller diameter deposits. This condition can, for example, be observed in the deposits of the membranous glomerulonephritis at the initial stages.</p>  |  |  |

**Supplemental Figure 2. Classification of the 'Distribution' of the deposits of the Immunofluorescence specimen<sup>1-3</sup>.**

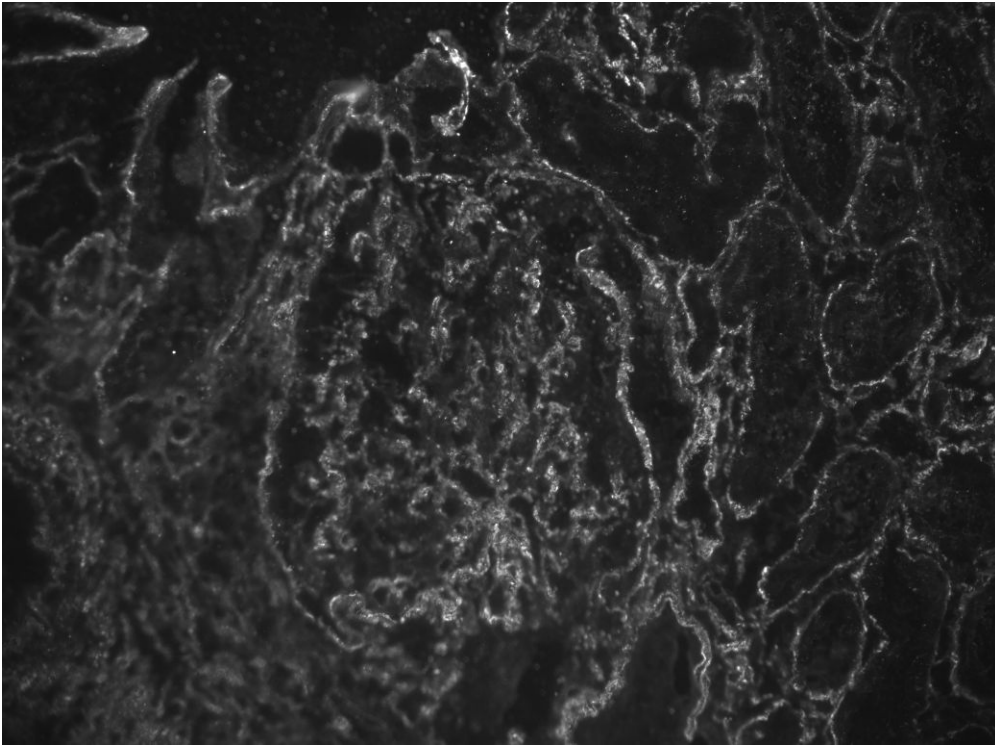
| <b>DISTRIBUTION</b>   |  |   |
|---|--|---|
| <b>DIFFUSE</b>  | Scheme   | Example   |
| The term 'diffuse' describes a lesion that is involving more than 50% of glomeruli. |    |    |
| <b>FOCAL</b>  | Scheme   | Example   |
| Focal describes a lesion that is involving less than 50% of glomeruli.              |   |   |
| <b>GLOBAL</b>   | Scheme   | Example   |
| The term 'global' describes a lesion that involves all of a glomerular tuft.        |  |  |
| <b>SEGMENTAL</b>  | Scheme   | Example   |
| The term 'segmental' describes a lesion that involves part of a glomerular tuft.    |  |  |



**Supplemental Figure 3. Classification of the 'Location' of the deposits of the Immunofluorescence specimen<sup>1-3</sup>.**

| <b>LOCATION</b>   |  |   |
|---|--|---|
| <b>MESANGIAL</b>  | Scheme   | Example   |
| Mesangial deposition of immunoglobulin is the typical pattern of the IgA Nephropathy. However, mesangial deposition can be founded in other conditions such as membranoproliferative glomerulonephritis or postinfectious glomerulonephritis.   |    |    |
| <b>CONTINUOUS REGULAR CAPILLARY WALL</b>  | Scheme   | Example   |
| The regular continuous capillary wall pattern corresponds to the identification of subepithelial deposits possible through ultramicroscopic analysis. This is the typical immunofluorescence finding of Membranous glomerulonephritis: a diffuse global granular deposits of immune reactants that follow the contour of the GBM. |   |   |
| <b>DISCONTINUOUS REGULAR CAPILLARY WALL</b>   | Scheme   | Example   |
| This pattern is also described as "bumps and humps" or "lumpy-dumpy", traditionally associated with postinfectious glomerulonephritis.  |  |  |
| <b>IRREGULAR CAPILLARY WALL</b>   | Scheme   | Example   |
| The irregular capillary wall pattern corresponds to the identification of subendothelial deposits, ribbon-like deposits, negative in the mesangium, giving the appearance of lobes. This pattern is frequently associated with membranoproliferative glomerulonephritis.  |  |  |

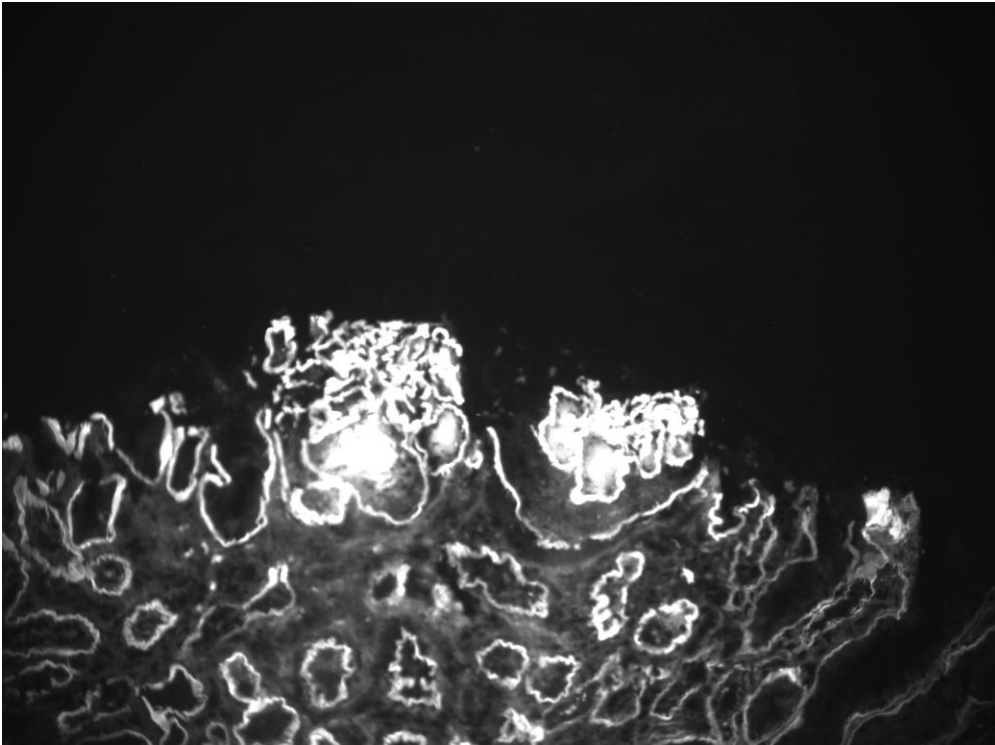
1  
2  
3 **Supplemental Figure 4. IgG direct Immunofluorescence of a Lupus Nephritis case (400X).** The image shows  
4 the presence of regular continuous capillary wall deposits of the glomerulus and basal membrane deposits  
5 of the tubules. The Convolutional Neural Network could correctly recognize the deposits of the tuft despite  
6 of the presence of the extraglomerular deposits (see Supplemental Table 7).  
7



32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

Review

1  
2  
3 **Supplemental Figure 5. IgG direct Immunofluorescence of a Light Chain deposition disease (400X).** The  
4 image shows the presence of deposits of the capillary wall of the glomerulus and basal membrane deposits  
5 of the tubules. The Convolutional Neural Network could correctly recognize glomerular deposits despite of  
6 the presence of the extraglomerular deposits (see Supplemental Table 8).  
7



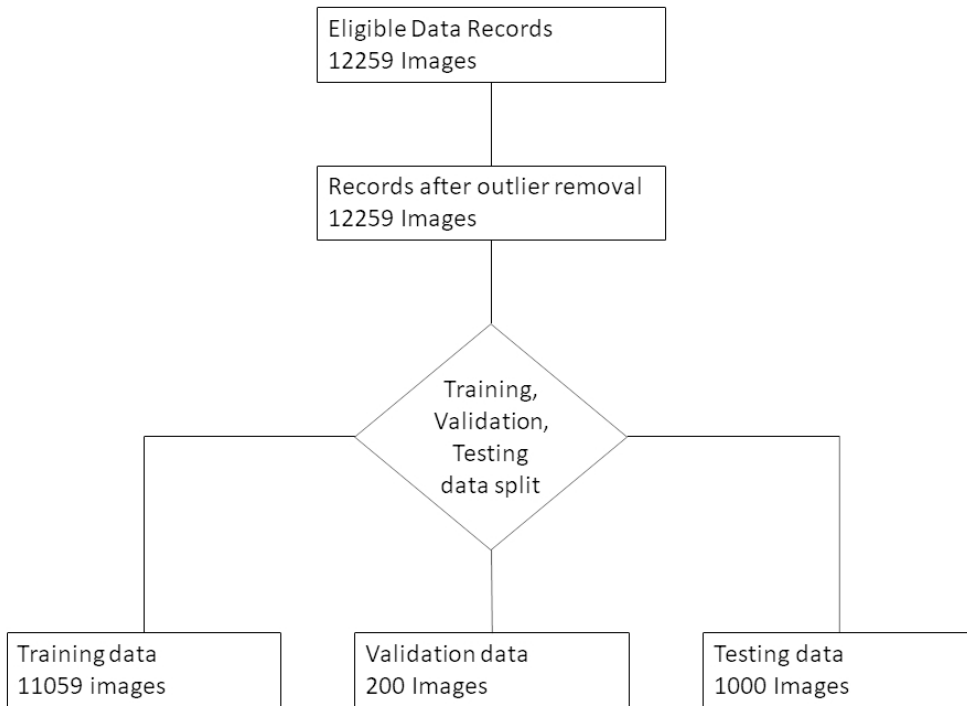
1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

## References

1. Wilson, CB, Dixon, FJ: Diagnosis of immunopathologic renal disease. *Kidney International*, 5: 389-401, 1974.
2. Morel-Maroger, L, Leatham, A, Richet, G: Glomerular abnormalities in nonsystemic diseases. Relationship between findings by light microscopy and immunofluorescence in 433 renal biopsy specimens. *Am J Med*, 53: 170-184, 1972.
3. D'Amico, G, Colasanti, G, Bestetti-Bosisio, M: *Le Glomerulonefriti Primitive* La Medicina Internazionale, 1977.
4. Lerner, RA, Glassock, RJ, Dixon, FJ: The role of anti-glomerular basement membrane antibody in the pathogenesis of human glomerulonephritis. *J Exp Med*, 126: 989-1004, 1967.

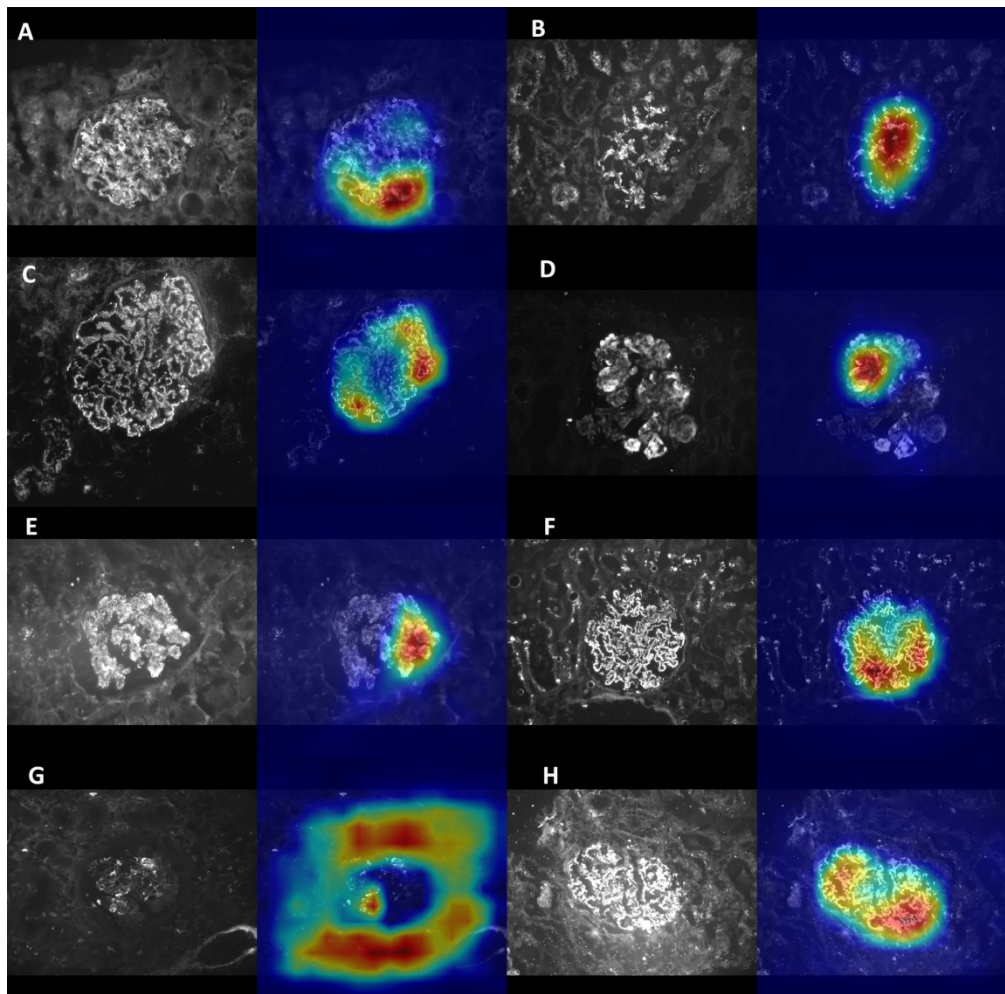
For Peer Review

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60



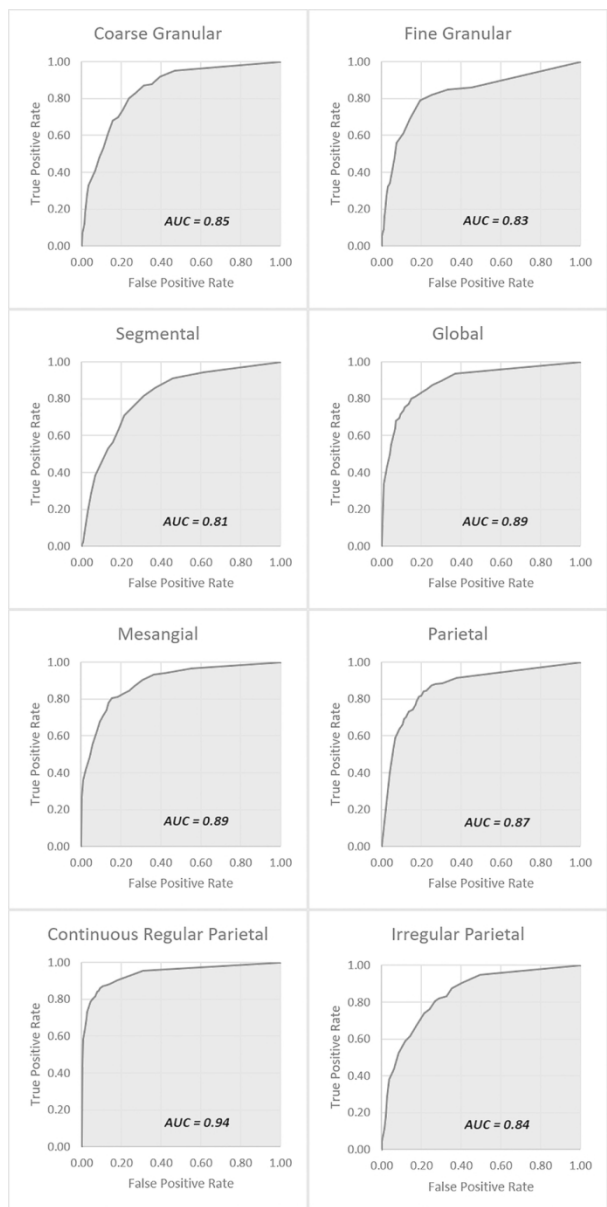
319x230mm (72 x 72 DPI)

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60



160x157mm (300 x 300 DPI)

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60



80x160mm (300 x 300 DPI)