ORIGINAL ARTICLE

Identification of glomerulosclerosis using IBM Watson and shallow neural networks

Francesco Pesce[1] · Federica Albanese[1] · Davide Mallardi[1] · Michele Rossini[1] · Giuseppe Pasculli[1,2] · Paola Suavo-Bulzis[1] · Antonio Granata[3] · Antonio Brunetti[4] · Giacomo Donato Cascarano[4] · Vitoantonio Bevilacqua[4] · Loreto Gesualdo[1]

Abstract

Background Advanced stages of different renal diseases feature glomerular sclerosis at a histological level which is observed by light microscopy on tissue samples obtained by performing a kidney biopsy. Computer-aided diagnosis (CAD) systems leverage the potential of artificial intelligence (AI) in healthcare to support physicians in the diagnostic process.

Methods We propose a novel CAD system that processes histological images and discriminates between sclerotic and non-sclerotic glomeruli. To this goal, we designed, tested, and compared two artificial neural network (ANN) classifiers. The former implements a shallow ANN classifying hand-crafted features extracted from Regions of Interest (ROIs) by means of image-processing procedures. The latter, instead, employs the IBM Watson Visual Recognition System, which uses a deep artificial neural network making decisions taking the images as input, without the need to design any procedure for describing images with features. The input dataset consisted of 428 sclerotic glomeruli and 2344 non-sclerotic glomeruli derived from images of kidney biopsies scanned by the Aperio ScanScope System.

Results Both AI approaches allowed to very accurately distinguish (mean MCC 0.95 and mean Accuracy 0.99) between sclerotic and non-sclerotic glomeruli. Although the systems may seem interchangeable, the approach based on feature extraction and classification would allow clinicians to gain information on the most discriminating features. In fact, further procedures could explain the classifier's decision by analysing which subset of features impacted the most on the final decision.

Conclusions We developed a customizable support system that can facilitate the work of renal pathologists both in clinical and research settings.

✉ Francesco Pesce
francesco.pesce@uniba.it

✉ Loreto Gesualdo
loreto.gesualdo@uniba.it

1 Nephrology, Dialysis, and Transplantation Unit, Department of Emergency and Organ Transplantation, University of Bari Aldo Moro, Bari, Italy

2 Department of Computer, Control and Management Engineering Antonio Ruberti (DIAG), La Sapienza University, Rome, Italy

3 Nephrology and Dialysis Unit, "Cannizzaro" Hospital, 95123 Catania, Italy

4 Department of Electrical and Information Engineering (DEI), Polytechnic University of Bari, Italy, Via Edoardo Orabona, 4, 70125 Bari, Italy

Graphical abstract

Identification of glomerulosclerosis using IBM Watson and shallow neural networks
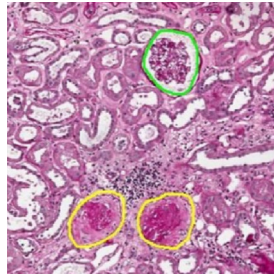
Methodology We compared two artificial neural network (ANN) classifiers. The former (A) implements a shallow ANN classifying hand-crafted features extracted from Regions of Interest (ROIs) by means of image-processing procedures. The latter (B), instead, employs the IBM Watson Visual Recognition System, which uses a deep ANN making decisions taking the images as input, without the need to design any procedure for describing images with features.
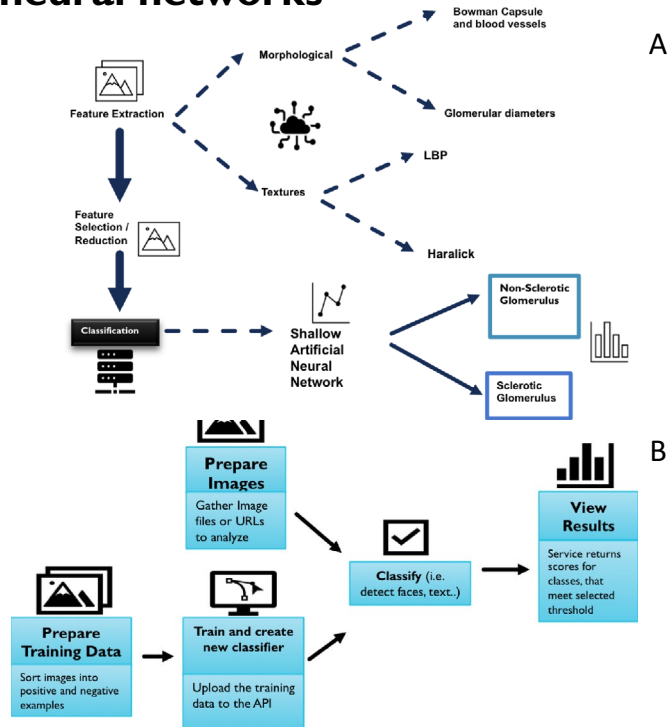
Conclusions Both AI approaches allowed to very accurately distinguish (mean MCC 0.95 and mean Accuracy 0.99) between sclerotic and non-sclerotic glomeruli. Although the systems may seem interchangeable, the approach based on feature extraction and classification would allow clinicians to gain information on the most discriminating features. In fact, further procedures could explain the classifier's decision by analysing which subset of features impacted the most on the final decision.

Keywords Renal biopsy · Glomerulosclerosis · IBM Watson · Artificial intelligence

Introduction

The primary histologic indicators of irreparable renal injury include interstitial fibrosis, tubular atrophy (IFTA) and glomerulosclerosis, which is the final pathological alteration of chronic kidney diseases [1]. It is characterised by the deposition of scar tissue, which replaces the renal parenchyma, and is quantified by renal pathologists to indicate the presence and extent of renal damage. However, such assessment can be variable among pathologists [2–5] with results often indicating decisions based on grading systems that may be applied differently in different institutions.

Several previous studies have applied various morphometric methods to improve the reproducibility and accuracy of IFTA assessment [6–9], and Machine Learning (ML) algorithms have already been successfully applied to glomerular segmentation by different research groups [7,

10–13] comprehending a whole-slide classifier to directly replicate a pathologist's assessment of IFTA and glomerulosclerosis on renal biopsy specimens [14].

In this work, we aimed to design a computer-aided diagnosis (CAD) system based on Artificial Intelligence (AI) to detect glomerulosclerosis automatically. Specifically, we designed, developed, tested, and compared two types of classifiers, both based on artificial neural networks (ANNs), namely ML algorithms capable of learning tasks, based on examples of input data and the desired output [15].

Precisely, the first approach (feature-based) implements a pipeline that, starting from the input images, extracts features describing the input data and makes the classifications based on these features. Instead, the second approach employs the IBM Watson Visual Recognition (WVR) framework, which is capable of making decisions based on the input images without the need to design procedures

for extracting hand-crafted features. IBM WVR uses Deep Learning (DL) algorithms to analyse and classify images [16, 17].

Materials and methods

Altogether, 26 kidney biopsies performed between 07/2011 and 02/2015 at the Department of Emergency and Organ Transplantations (DETO) of the Bari University Hospital were used. All kidney biopsies were stained with Periodic Acid-Schiff PAS [18] after formalin fixation and inclusion in paraffin. For each biopsy several 2–3 μm thick sections were cut from different levels of the tissue (at least 3 levels with 3 sections for each level). All biopsies were processed at the same institution. Slides were stained at different times (i.e. when such organ donation occurred). Each slide was scanned using an Aperio ScanScope at 20× with a resolution of 0.50 μm/pixel. For each slide, glomeruli were identified and manually annotated using the Aperio Image Scope tool by two independent renal pathologists. Glomeruli were labelled as "sclerotic" or "non-sclerotic" (Fig. 1a). After the manual labelling, we developed a MATLAB script to extract the Regions of Interest (ROI) employed for the subsequent stages.

The final dataset included 2772 glomeruli, 428 sclerotic and 2344 non-sclerotic ones, with a ratio between the two classes of 1/5.5.

For clarity's sake, we considered sclerotic glomeruli as belonging to the Positive class, whereas the non-sclerotic samples constituted the Negative class.

We split the dataset into two parts to perform the subsequent analyses: a training set (about 80% of the entire dataset) to train the classifiers, and a test set (about 20%) to evaluate the classification performance.

Regarding the WVR System by IBM, from the 80% of glomeruli belonging to the train set, the 10% of the samples was randomly withdrawn to create a validation set, used to choose the best models to test. Supplementary Tables 1 and 2 show the number of samples constituting the dataset processed by the classification systems for both the considered approaches. Only the test set was used to assess the final performance; the same test set was used for both approaches thus allowing for a better comparison of the two models.

Feature-based classification approach

The feature-based classification approach is based on the extraction of features from the input images through image processing techniques; the classification is then performed
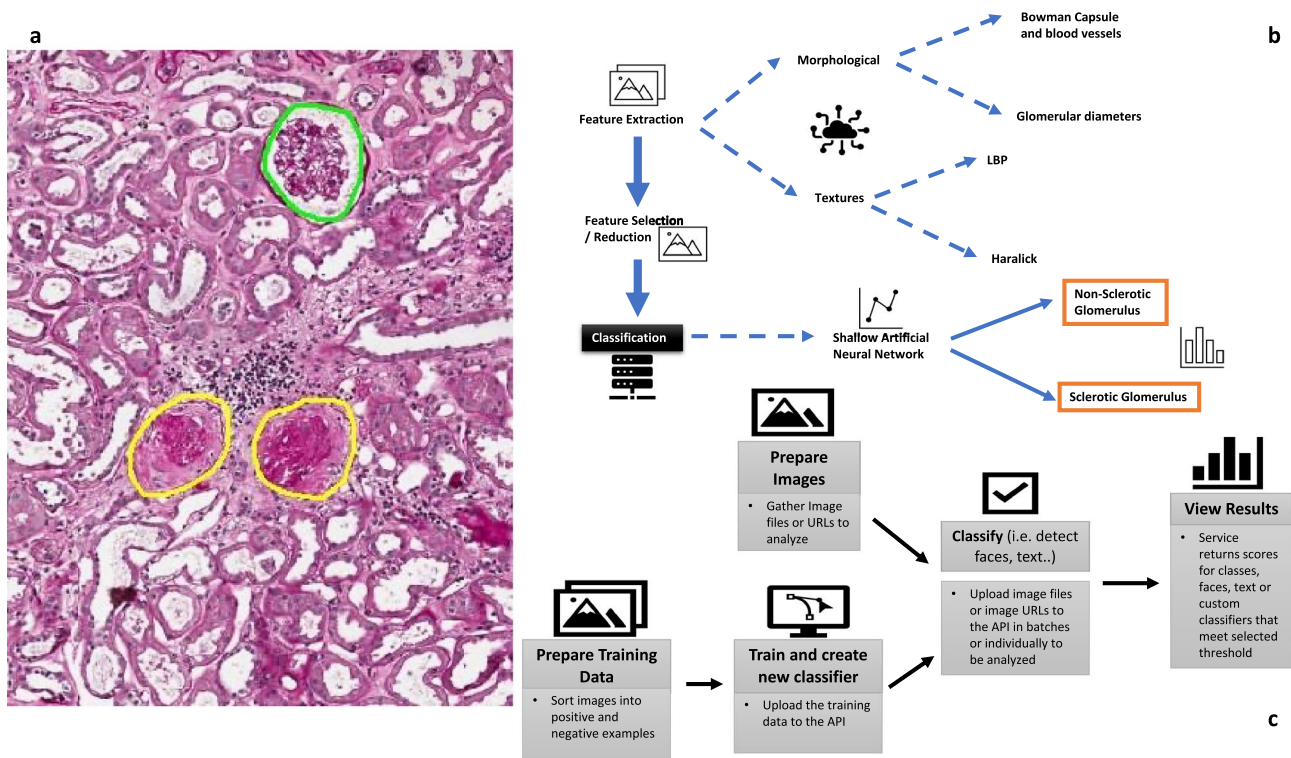
Fig. 1 **a** Glomeruli annotation. In the pre-processing stage, glomeruli were manually annotated by two renal pathologists. Non-sclerotic glomeruli were marked in green and sclerotic glomeruli in yellow. **b** (upper right quadrant). Feature based classification approach. **c** (lower right quadrant). IBM Watson Visual Recognition Workflow

with a supervised ML algorithm, namely a shallow ANN, allowing to characterise and distinguish between sclerotic and non-sclerotic glomeruli.

We designed and developed this model following three steps: (i) feature extraction, (ii) feature reduction and (iii) glomeruli classification. The workflow is described in Fig. 1b.

Regarding the feature extraction procedure, two morphological characteristics related to Bowman's capsule and Bowman's space were extracted after image processing procedures that were necessary due to the PAS staining of the images. Also, 148 textural features based on the well-known multi-radial colour LBP (mrcLBP) and Haralick algorithms were obtained. After extracting the features, a procedure to reduce the feature space was needed due to the high dimensionality of such data. The feature reduction process reduces the number of features considering the most useful ones, namely, those better contributing to the discrimination process while removing the irrelevant or redundant ones.

In order to do this, Principal Component Analysis (PCA) was performed, allowing to reduce the data to be considered in the subsequent phase to 95 components contributing to 99.9% of the variance of the input data.

After image processing and feature extraction procedures, we built a shallow ANN with one hidden layer. In order to select a suitable number of neurons for the hidden layer, we trained and cross-validated (tenfold cross-validation) the ANN, changing the number of neurons iteratively, from 1 to 95 (the number of input features). We then chose the configuration reporting the highest average Matthews Correlation Coefficient (MCC), i.e. the ANN configuration with 27 neurons in the hidden layer.

Concerning the training hyperparameters, all the configurations had the following:

1. Weight and bias initialization with the Nguyen-Widrow initialization algorithm;
2. Weight and bias update with the Scaled Conjugate Gradient algorithm;
3. Cross-Entropy loss function;
4. Hyperbolic tangent sigmoid transfer function for the hidden layer, whereas the softmax function was employed for the output layer;
5. Network training stopped if any of the following criteria were met:
    a. 100 training epochs reached;
    b. 6 consecutive validation fails;
    c. loss value reached 0;
    d. gradient performance reached 10^{-6};

Matlab Deep Learning Toolbox™ was employed to design, train and validate the ANN architectures.

In order to assess the robustness of the implemented workflow, we performed tenfold cross-validation and a final hard-voting procedure for making decisions of the test set. Furthermore, we performed ten runs of the classification pipeline in order to evaluate the performance variations with respect to the data contained in the folds. Precisely, the training dataset was split into tenfold; in turn, ninefold were used to train the network, whereas the last fold was used to validate it. Classification of the samples belonging to the test set was performed by considering a majority voting by the ten classifiers: the most supported class was then assigned to the specific sample.

In this feature-based approach, we faced the issue of the imbalance of the dataset by implementing two complementary strategies. Firstly, the MCC was evaluated as a general performance comparison among the folds. In fact, the MCC value is a measure of the quality of binary (two-class) classifications which considers the number of false positives and false negatives; thus, it is generally regarded as a balanced measure that can be used even if the classes are of very different sizes [19]. The second strategy, instead, considers the Receiving Operating Characteristic (ROC) in order to choose the correct classification threshold value. ROC curves plot the True Positive Rate (TPR) variations against the False Positive Rate (FPR), varying the threshold used for making the decision by the classifier. Selecting the most suitable threshold, such as the one providing us to obtain the higher Area Under the Curve (AUC), allowed us to reduce the classifier polarization due to the most represented class.

Matlab source code is available at the following Github repository: https://github.com/LabInfInd/glomerulosc"lerosis_identification_watson_ann.git.

IBM Watson Visual Recognition

IBM WVR, differently from the previous method, uses DL algorithms to analyse and classify images [16, 17]. DL is a branch of ML focused on algorithms based on models showing deep architectures characterised by multiple layers capable of extracting features that describe the input data at higher abstraction levels, i.e. Convolutional Neural Networks or Deep Neural Networks [20].

Five steps are needed to train and use a classification model on the IBM WVR system:

Prepare training data: sort images into positive or negative images. A set of images related to the classification task have to be collected. In order to optimize the training phase, the images should have similar size, resolution, and colour palette. With these images, two training sets must be created: a set with the positive images (containing the features the classifier should recognize) and

another with negative image examples (without features). The two training sets should not overlap;

Train and create new models: upload examples as training data. These two sets are uploaded to the WVR service that is available on IBM Cloud. The service automatically trains its neural network based on positive and negative image examples. At the end of this stage, a custom model has been created and will be available for usage in the Recognition service;

Prepare images: gather images to analyse. After training the model, any set of images can be uploaded to the Recognition service in order to be classified;

Analyse images: use the built-in capabilities or a custom model. The trained custom model classifies each image of the uploaded set;

View results: review the insights into your visual content. For each of the analysed images, the system returns the image associated class, a set of information that characterises the imputed image and the features recognized in the latter.

The workflow is depicted in Fig. 1c.

Thanks to the extreme versatility of IBM's proprietary algorithm for WVR, which allows users to train Watson AI on almost any visual content in order to create custom analysis models, we designed several models by combining the following variables:

- colour of the image (PAS staining or grayscale-converted images, to better understand how well the classifier discriminates analysing colours);
- size of images (original or resized to 224×224 pixel files according to Watson's guidelines);
- binary (two types of images provided: sclerotic glomeruli and non-sclerotic glomeruli images) or multi-class technique (two classes: sclerotic and non-sclerotic glomeruli);
- the number of images (to balance the number of samples per class, as suggested in WVR guidelines).

We employed different image augmentation procedures to balance the positive and negative image samples concerning the last point. Since the IBM WVR guidelines suggest creating models with an approximately equal number of positive and negative cases in the training set, data augmentation of the sclerotic glomeruli images was carried out.

Data augmentation is a method that increases the number of samples in a dataset creating synthetic images, in this case, by applying image transformations to the available samples. By doing this, we were able to balance the number of sclerotic (~ 300) and non-sclerotic images (~ 1600) in the training set.

Finally, two training data sets were created:

(i) the first one was obtained by subsampling the most represented class of samples, randomly selecting 313 non-sclerotic glomeruli (negative) from the negatives, and 307 sclerotic glomeruli (positive) images (model 300, no data augmentation needed);

(ii) the second dataset was generated with the data augmentation; thus, it contained 1667 negative samples and 1607 positive images (model 1600).

Results

Different metrics were considered to evaluate the performance of both classifiers. Specifically, we evaluated Accuracy, Precision, Recall, Specificity, F1 score and, as already mentioned, the MCC, considering True Positive, True Negative, False Positive and False Negative according to the Confusion Matrix.

Feature-based classification approach

Table 1 reports the average and the best performance obtained by the Feature-Based classifier whose confusion matrices are reported in Supplementary Tables 3a and 3b (respectively for the best and worst case scenarios). The results show that the feature-based Artificial Neural Network was able to discriminate sclerotic and non-sclerotic glomeruli with high performance (mean MCC = 0.95 and mean Accuracy = 0.99) and low variability (MCC std = 0.01 and Accuracy std < 0.00).

Average Precision and Recall were equal to 0.98 (± 0.01) and 0.93 (± 0.02), respectively, showing better performance in the identification of non-sclerotic glomeruli (all the non-sclerotic glomeruli were detected in the best case).

IBM Watson Visual Recognition

We created, for each dataset, eight balanced classifiers, considering the different images obtained through the processing described in the Methods section, thus obtaining 16 models altogether. Specifically, for each dataset, there were:

Table 1 Performance of the feature-based approach

	Mean + std	Best
Accuracy	0.9874 ± 0.0018	0.9914
Precision	0.9844 ± 0.0111	1.0000
Recall	0.9310 ± 0.0153	0.9425
MCC	0.9501 ± 0.0074	0.9659
Specificity	0.9974 ± 0.0019	1.000
F1-score	0.9568 ± 0.0065	0.9659

1. multi-class, PAS staining and original size model;
2. multi-class in grayscale and original size model;
3. multi-class resized and PAS staining model;
4. multi-class resized in grayscale model;
5. binary, PAS staining and original size model;
6. binary, PAS staining and resized model;
7. binary in grayscale and original size model;
8. binary resized in grayscale model.

A validation test was carried out on all 16 models to choose the most performing one. The results are shown in the following tables in terms of recall and specificity, considering a classification threshold set at 0.5.

The classifiers in the analysis provide a score between 0 and 1. This number indicates Watson's confidence in classifying an image as belonging to a certain class. The validation test is the same for each model. Every test had a cut-off of 0.5: if > 0.5, the glomerulus was considered as belonging to the tested class.

Based on the models' performances on the validation set (Table 2), we focused the analysis on the test set, considering only the model performing at best, i.e. the "binary resized in grayscale model". The test is the same for each model. Every test had a cut-off of 0.5: if the test resulted in a value > 0.5, the glomerulus was considered belonging to the tested class.

The obtained performances on the test set are reported in terms of Accuracy, Precision, Recall, MCC, Specificity and F1-score using intermediate augmented datasets (Supplementary Table 4).

Comparison between IBM Watson Visual Recognition and feature-based model

The test set used to compare the performance of the two classification approaches was the same for both models and consisted of 492 non-sclerotic glomeruli and 87 sclerotic glomeruli.

The results of the comparison between IBM WVR and the feature-based model are reported in Table 3, in terms of average performance (± standard deviation).

Table 3 Comparison between IBM WVR and the feature-based model

	IBM Visual Recognition	Feature-based
Precision	0.9647	0.9844 ± 0.0111
Recall	0.9425	0.9310 ± 0.0153
Specificity	0.9939	0.9974 ± 0.0019
Accuracy	0.9862	0.9874 ± 0.0018
MCC	0.9455	0.9501 ± 0.0074
F1	0.9535	0.9568 ± 0.0065

Basic model metrics with 300 images are reported for the classifier based on IBM WVR to avoid any interference due to the data augmentation

Evaluation metrics were good and comparable between the two systems and both the classification approaches reached high levels of performance. IBM WVR showed a higher recall, whereas precision was higher with the feature-based model. Both models, however, performed better in the identification of non-sclerotic glomeruli.

Focusing on the misclassifications of both the classifiers, most of the errors were due to low-quality images caused by technical artefacts, which even renal pathologists misinterpret and commonly discard in clinical practice (Supplementary Fig. 1).

Discussion

Since its advent, AI has always been recognized as a valid tool to assist the processing of virtually every data modality and ultimately enhance the human capability of handling and making sense of such data. ML, and, more specifically, DL, have long been part of our daily routines: computer vision tasks [21] (such as object detection, face recognition, action, and activity recognition), voice recognition of smartphones, autopilot of vehicles [22].

Healthcare, too, has acknowledged the potential support of AI in performing the most diverse tasks (such as diagnosis, therapeutic strategies, patient management) in a short

Table 2 Performance of IBM Watson on the validation dataset

	MODEL 300		MODEL 1600	
	Specificity (%)	Recall (%)	Specificity (%)	Recall (%)
Multiclass, PAS staining and original size	97.14	94.12	100	97.06
Binary, PAS staining and original size	97.14	97.06	100	97.06
Multiclass in grayscale and original size	97.14	97.06	100	100
Multiclass resized and PAS staining	97.14	91.18	100	97.06
Multiclass resized in grayscale	97.14	97.06	100	94.12
Binary, PAS staining and resized	97.14	97.06	99.46	100
Binary in grayscale and original size	97.14	97.06	100	100
Binary resized in grayscale	97.14	97.06	100	100

time and with the advantage of being cost-effective. AI can potentially be applied to every medical speciality; imaging has definitely been one of most prolific fields [23], especially in oncology (i.e. thoracic imaging, breast lesions [24], colonoscopy, brain tumours). IBM WVR has been successfully "trained" to detect abnormalities and extract textural features of the altered lung parenchyma that could be related to specific signatures of the Covid-19 virus [25].

The implementation and development of digital pathology, too, have been driven by the progress in ML and DL [26], and several AI systems have already been developed to assist physicians [27, 28].Through the development of computational image analysis tools for tissue interrogation, AI and ML have brought pathology to the forefront in this process of re-defining nephrology [29, 30]. In fact, AI applied to image processing can offer many advantages in terms of accuracy and workload management for renal pathologists, also potentially helping with the discovery of novel biomarkers in research settings [31–35]. Furthermore, the recent gathering of the Banff Digital Pathology Working Group demonstrates the strong interest in AI and will help to advance the use of such techniques in specific renal pathology fields [36] (e.g., renal transplantation).

In the automated analysis of kidney images, we propose a system which focuses on glomerulosclerosis. To do this, we tested two different ANN approaches, and both classifiers showed good performance in recognising glomerulosclerosis and discriminating between normal and sclerotic glomeruli. Although recent literature demonstrated that DL methodologies are able to perform better than traditional ML approaches [37–39], our results show that, in this case, performance of the feature-based approach remains comparable to the IBM WVR system and may offer some advantages. In particular, the description of regions with discriminative features and the implemented pipeline for designing the classifier made the feature-based model quite robust. Shallow ANNs, furthermore, seem to be more precise and more flexible since it is possible to customize the algorithm according to the number and quality of desired features. These advantages make the feature-based model potentially suitable for each field of application and at any level of complexity.

Another interesting note that emerges from this study is that even a general-purpose visual analytics tool like IBM WVR has led to very accurate results, though particular care was devoted to preparing the training set. Despite the user-friendly interface provided by IBM WVR, choosing the model that would perform better was not straightforward, and up to 16 models derived from different combinations of key input parameters were prepared to be tested. Namely, we worked on the colour (the original histological staining or greyscale), the size (original or resized as suggested by IBM WVR guidelines), the class definition ("binary": when only one class is defined e.g., the sclerotic glomerulus versus anything else, or "multi class" e.g., both sclerotic and non-sclerotic glomeruli are used as separate classes to be recognized), and the number of images given that having roughly 50/50 positive and negatives is recommended). This latter parameter was particularly challenging. In order to balance the dataset, we used data augmentation, but such technique did not result in a linear improvement of the performance as shown by the profile of the MCC across the different models (Supplementary Table 4).

To the best of our knowledge, this is the first study exploiting and comparing the IBM WVR system for this particular task. Additional features to implement the final system will include the recognition of intermediate sclerotic lesions, other renal compartments and the automatic annotation of the glomeruli as different ANNs have already been created for this purpose [40, 41].

Author contributions The specific contribution of each author is as follows. FP and LG conceptualized the study. FP supervised the whole analysis and wrote the manuscript. AG provided the biopsy samples. MR annotated the renal biopsies. FA and DM prepared the datasets and ran IBM Watson. AB, GDC and VB trained the feature-based model. GP and PSB edited and reviewed the manuscript. All authors have seen and approved the submitted version of the manuscript and agreed to be accountable for all aspects of the work.

Availability of data and material Data are available from the Authors upon reasonable request and with permission of "Azienda Ospedaliera Policlinico di Bari".

Declarations

Conflict of interest The author declares that they have no conflict of interest.

Ethical statement Ethical approval for this study was provided by the Independent Ethical Committee of Azienda UniversitariaOspedaliera Consorziale - Policlinico, Bari, Italy.

References

1. Djudjaj S, Boor P (2019) Cellular and molecular mechanisms of kidney fibrosis. Mol Aspects Med 65:16–36

2. Furness PN, Taub N, Assmann KJ, Banfi G, Cosyns JP, Dorman AM et al (2003) International variation in histologic grading is large, and persistent feedback does not improve reproducibility. Am J Surg Pathol 27:805–810

3. Gough J, Rush D, Jeffery J, Nickerson P, McKenna R, Solez K et al (2002) Reproducibility of the Banff schema in reporting protocol biopsies of stable renal allografts. Nephrol Dial Transplant 17:1081–1084

4. Grootscholten C, Bajema IM, Florquin S, Steenbergen EJ, Peutz-Kootstra CJ, Goldschmeding R et al (2008) Interobserver agreement of scoring of histopathological characteristics and classification of lupus nephritis. Nephrol Dial Transplant 23:223–230

5. Snoeijs MG, Boonstra LA, Buurman WA, Goldschmeding R, van Suylen RJ, van Heurn LW et al (2010) Histological assessment of pre-transplant kidney biopsies is reproducible and representative. Histopathology 56:198–202

6. Kolachalama VB, Singh P, Lin CQ, Mun D, Belghasem ME, Henderson JM et al (2018) Association of pathological fibrosis with renal survival using deep neural networks. Kidney Int Rep 3:464–475

7. Hermsen M, de Bel T, den Boer M, Steenbergen EJ, Kers J, Florquin S et al (2019) Deep learning-based histopathologic assessment of kidney tissue. J Am Soc Nephrol 30:1968–1979

8. Servais A, Meas-Yedid V, Noël LH, Martinez F, Panterne C, Kreis H et al (2011) Interstitial fibrosis evolution on early sequential screening renal allograft biopsies using quantitative image analysis. Am J Transplant 11:1456–1463

9. Farris AB, Adams CD, Brousaides N, Della Pelle PA, Collins AB, Moradi E et al (2011) Morphometric and visual evaluation of fibrosis in renal biopsies. J Am Soc Nephrol 22:176–186

10. Ginley B, Lutnick B, Jen KY, Fogo AB, Jain S, Rosenberg A et al (2019) Computational segmentation and classification of diabetic glomerulosclerosis. J Am Soc Nephrol 30:1953–1967

11. Lutnick B, Ginley B, Govind D, McGarry SD, LaViolette PS, Yacoub R et al (2019) An integrated iterative annotation technique for easing neural network training in medical image analysis. Nat Mach Intell 1:112–119

12. Pedraza A, Gallego J, Lopez S, Gonzalez L, Laurinavicius A, Bueno G (2017) Glomerulus classification with convolutional neural networks. In: Medical image understanding and analysis: 21st annual conference, MIUA, Edinburgh, United Kingdom

13. Bukowy JD, Dayton A, Cloutier D, Manis AD, Staruschenko A, Lombard JH et al (2018) Region-based convolutional neural nets for localization of glomeruli in trichrome-stained whole kidney sections. J Am Soc Nephrol 29:2081–2088

14. Ginley B, Jen KY, Han SS, Rodrigues L, Jain S, Fogo AB, Zuckerman J, Walavalkar V, Miecznikowski JC, Wen Y, Yen F, Yun D, Moon KC, Rosenberg A, Parikh C, Sarder P (2021) Automated computational detection of interstitial fibrosis, tubular atrophy, and glomerulosclerosis. J Am Soc Nephrol. https://doi.org/10.1681/ASN.2020050652

15. He K, Zhang X, Ren S et al (2016) Deep residual learning for image recognition. In: Proceedings of the IEEE computer society conference on computer vision and pattern recognition, pp 770–778

16. Bhattacharjee B, Hill ML, Wu H, Chandakkar PS, Smith JR, Wegman MN (2017) Distributed learning of deep feature embeddings for visual recognition tasks. IBM J Res Dev 61(4/5):4:1-4:8

17. Ahmed I, House R, Deilma N, Deilma N, Luo L (2019) (IBM Canada, Markham, Ontario, Canada) "Custom visual recognition model with Watson studio". In: CASCON '19: Proceedings of the 29th annual international conference on computer science and software engineering, November, pp 376–377

18. Young M, Leslie SW (2020) Renal biopsy. StatPearls. StatPearls Publishing, Treasure Island

19. Boughorbel SB (2017) Optimal classifier for imbalanced data using Matthews Correlation Coefficient metric. PLoS ONE 12(6):e0177678

20. Yamashita R, Nishio M, Do RKG et al (2018) Convolutional neural networks: an overview and application in radiology. Insights Imaging 9:611–629

21. Voulodimos A, Doulamis N, Doulamis A, Protopapadakis E (2018) Deep learning for computer vision: a brief review. Comput Intell Neurosci 1:7068349

22. Grigorescu S, Trasnea B, Cocias T et al (2020) A survey of deep learning techniques for autonomous driving. J Field Robotics 37:362–386

23. Hosny A, Parmar C, Quackenbush J, Schwartz LH, Aerts HJWL (2018) Artificial intelligence in radiology. Nat Rev Cancer 18(8):500–510

24. Bevilacqua V, Brunetti A, Guerriero A et al (2019) A performance comparison between shallow and deeper neural networks supervised classification of tomosynthesis breast lesions images. Cogn Syst Res 53:3–19

25. https://developer.ibm.com/recipes/tutorials/ai-can-support-covid19-diagnosis-from-medical-imaging/

26. Niazi MKK, Parwani AV, Gurcan MN (2019) Digital pathology and artificial intelligence. Lancet Oncol 20:e253–e261

27. Venugopalan J, Tong L, Hassanzadeh HR et al (2021) Multimodal deep learning models for early detection of Alzheimer's disease stage. Sci Rep 11:3254

28. Doi K (2007) Computer-aided diagnosis in medical imaging: historical review, current status and future potential. Comput Med Imaging Graph 31(4–5):198–211

29. Barisoni L, Lafata KJ, Hewitt SM, Madabhushi A, Balis UGJ (2020) Digital pathology and computational image analysis in nephropathology. Nat Rev Nephrol. https://doi.org/10.1038/s41581-020-0321-6

30. Becker JU, Mayerich D, Padmanabhan M, Barratt J, Ernst A, Boor P, Cicalese PA, Mohan C, Nguyen HV, Roysam B (2020) Artificial intelligence and machine learning in nephropathology. Kidney Int 98(1):65–75

31. Zhang Z, Xie Y, Xing F et al (2017) MDNet: a semantically and visually interpretable medical image diagnosis network. In: IEEE computer society conference on computer vision and pattern recognition, pp 6428–6436

32. Holzinger A, Malle B, Kieseberg P, et al. (2017) Towards the augmented pathologist: challenges of explainable-AI in digital pathology. arXiv preprint arXiv:1712.06657

33. Huo Y, Deng R, Liu Q, Fogo AB, Yang H (2021) AI applications in renal pathology. Kidney Int 99:1309–1320

34. Thongprayoon C, Kaewput W, Kovvuru K et al (2020) Promises of big data and artificial intelligence in nephrology and transplantation. J Clin Med 9:1107

35. Hermsen M, Smeets B, Hilbrands L, van der Laak J (2020) Artificial intelligence: is there a potential role in nephropathology? Nephrol Dial Transplant. https://doi.org/10.1093/ndt/gfaa181

36. Farris AB, Moghe I, Wu S et al (2020) Banff Digital Pathology Working Group: going digital in transplant pathology. Am J Transplant 20:2392–2399

37. Kim GB, Jung KH, Lee Y, Kim HJ, Kim N, Jun S, Seo JB, Lynch DA (2018) Comparison of shallow and deep learning methods on classifying the regional pattern of diffuse lung disease. J Digit Imaging 31(4):415–424

38. Ososkov G, Goncharov P (2017) Shallow and deep learning for image classification. Opt Mem Neural Netw 26:221–248

39. Chauhan S, Vig L, De Filippo De Grazia M, Corbetta M, Ahmad S, Zorzi M (2019) A comparison of shallow and deep learning methods for predicting cognitive performance of stroke patients from MRI lesion images. Front Neuroinform 13:53

40. Bevilacqua V et al (2017) An innovative neural network framework to classify blood vessels and tubules based on Haralick features evaluated in histological images of kidney biopsy. Neurocomputing 228:143–153

41. Kannan S et al (2019) Segmentation of glomeruli within trichrome images using deep learning. Kidney Int Rep 4: 955–962