



Artificial intelligence in glomerular diseases

Francesco P. Schena¹ · Riccardo Magistroni² · Fedelucio Narducci³ · Daniela I. Abbrescia⁴ · Vito W. Anelli³ · Tommaso Di Noia³

Received: 3 November 2021 / Revised: 19 December 2021 / Accepted: 21 December 2021 / Published online: 10 March 2022
© The Author(s), under exclusive licence to International Pediatric Nephrology Association 2022

Abstract

In this narrative review, we focus on the application of artificial intelligence in the clinical history of patients with glomerular disease, digital pathology in kidney biopsy, renal ultrasonography imaging, and prediction of chronic kidney disease (CKD). With the development of natural language processing, the clinical history of a patient can be used to identify a computable phenotype. In kidney pathology, digital imaging has adopted innovative deep learning algorithms (DLAs) that can improve the predictive capability of the examined lesions. However, at this time, these applications can only be used in research because there is no recognized validation to replace the conventional diagnostic applications. Kidney ultrasonography, used in the clinical examination of patients, provides information about the progression of kidney damage. Machine learning algorithms (MLAs) with promising results for the early detection of CKD have been proposed, but, still, they are not solid enough to be incorporated into the clinical practice. A few tools for glomerulonephritis, based on MLAs, are available in clinical practice. They can be downloaded on computers and cellular phones but can only be applied to unracial cohorts of patients. To improve their performance, it is necessary to organize large consortia with multiracial cohorts. Finally, in many studies MLA development has been carried out using retrospective cohorts. The performance of the models might differ in retrospective cohorts compared to real-world data. Therefore, the models should be validated in prospective external large cohorts.

Keywords Glomerulonephritis · Artificial Intelligence · Natural Language Processing · Machine Learning Algorithm · Deep Learning Algorithm · Clinical Outcome

Introduction

Glomerular diseases frequently cause chronic kidney disease (CKD) which is diagnosed on the basis of medical history, kidney biopsy, and ultrasonography. The clinical profile is characterized by the gradual deterioration of the kidney function progressing to kidney failure (KF).

Nephrologists' continuously raised awareness about CKD has generated a more accurate long-term follow-up of patients

and consequently a high number of electronic health records (EHRs). These big data contain administrative data, clinical information, digital pathology data of kidney biopsies, digital ultrasonography images, and longitudinal laboratory data of the clinical follow-up. They are a key source for understanding the disease's evolution and the response to treatment.

The process of big data requires an adequate and flexible approach in which artificial intelligence (AI) plays an important role [1]. AI is a branch of computer science that aims to create systems or methods to analyze big data. Figure 1 illustrates the subsets of AI. Fuzzy learning (FL), random forest (RF), decision trees (DT), and naïve Bayes (NB) are some AI algorithms. Machine learning (ML) is a subset of AI that accesses data and teaches computers to do natural brain functions through algorithms (MLA). MLAs can be divided into supervised and unsupervised learning algorithms. Some of the most commonly used MLAs are artificial neural network (ANN), support vector machine (SVM), k-nearest neighbors (KNN), logistic regression (LOG), and extreme

✉ Francesco P. Schena
paolo.schena@uniba.it

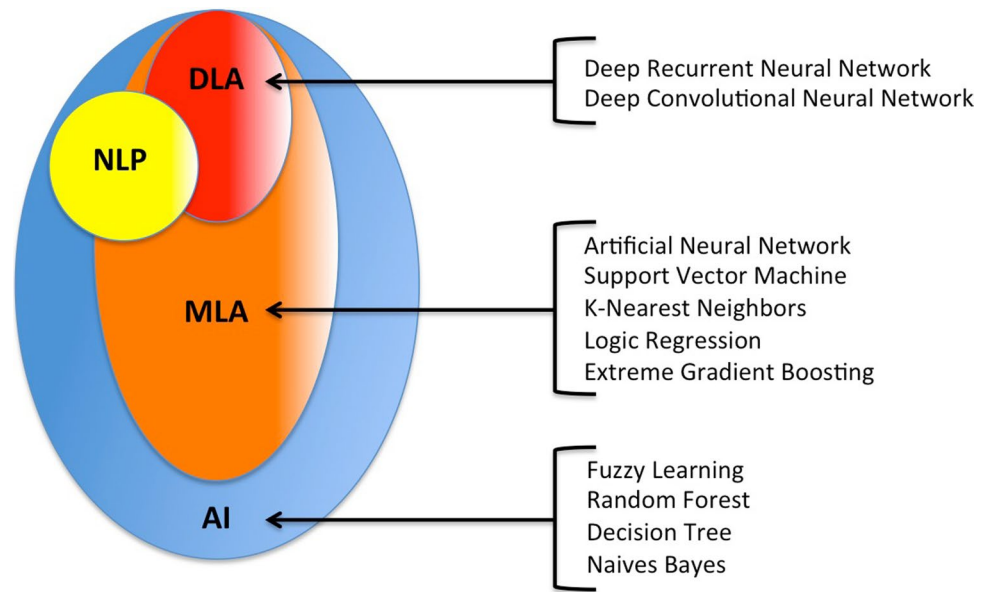
¹ Department of Emergency and Organ Transplantation, University of Bari, Bari, Italy

² Department of Nephrology, University of Modena, Modena, Italy

³ Department of Electrical and Information Engineering, Polytechnic of Bari, Bari, Italy

⁴ Schena Foundation, Bari, Italy

Fig. 1 List of algorithms used in the reviewed publications.
Abbreviations: AI=artificial intelligence; MLA= machine learning algorithms; DLA= deep learning algorithms; NLP= natural language processing



gradient boosting (xGB). A subset of ML is deep learning (DL) using algorithms (DLAs) to improve the performance of models. Deep recurrent neural network (DRNN) and deep convolutional neuronal network (DCNN) are currently used. Finally, natural language processing (NLP) extracts unstructured data and uses MLAs and DLAs to resolve tasks.

In this narrative review, we focus on the application of AI in the clinical history of patients with glomerular disease, using digital pathology data from kidney biopsies, digital renal ultrasonography images, and prediction of CKD/KF in observational clinical studies.

Natural language processing

EHRs are often formatted in a way that hinders direct processing by a machine. For this purpose, NLP plays a key role in healthcare [2]. It is based on a set of models, techniques, and algorithms able to a) understand and/or generate natural language, namely the human language used for communication, b) transform the text of the medical history into a format readable and understandable for a machine, c) identify disease risk factors, d) evaluate the efficiency of care and costs, and finally e) extract information from free-text clinical narratives within EHR databases.

The most exploited state-of-the-art tools for clinical information extraction are a) cTakes [3] that identifies the patient phenotype, detects medication discrepancies, and extracts adverse drug events, b) MetaMap [4] that maps biomedical texts to the Unified Medical Language System Metathesaurus and vice versa, and finally, and c) MedLEE that is widely used for pharmacoepidemiology and pharmacovigilance [5].

Another healthcare application of NLP is the development of Health Assistant Chatbots. In this case, NLP is used for both interacting with the user and processing text to identify relevant clinical information. Recently, Polignano et al. [6] have developed the HealthAssistantBot tool that can talk with patients and understand their symptoms, recommend doctors, and monitor treatments and health parameters.

NLP techniques are exploited for the dialog interface, the symptom checker, and the recommendation system. They can be used in structured and unstructured EHR-linked databases collecting real-world data from patients with glomerular diseases; these records belong to recently developed biobanks such as the kidney research network registry and database [7], and the Korean renal network system toward next-generation analysis [8]. Recently, the PCORnet®, a national patient-centered clinical research network, which incorporates data from millions of patients across nine clinical research networks and two health plan research networks including more than 300 hospitals in the US, has been used to validate a computational phenotype for nephrotic syndrome in children and adults [9]. This computable phenotype could be used by practitioners to improve the identification of patients with primary nephrotic syndrome. In conclusion, NLP can provide an opportunity to identify phenotype or unsuspected risk factors for glomerular diseases.

AI applied to kidney histology

The interpretation of kidney biopsy is the gold standard for the diagnosis of glomerular diseases. The advancements in digital pathology have opened up the introduction of innovative analysis tools in the clinic, among them the

application of DLAs, which is becoming dominant. These emerging models must be tested and validated rigorously to define their possible advantages over traditional manual reports, how they can support the traditional system, and in which limited areas they can eventually supersede it.

Manual reporting of kidney biopsy, while remaining irreplaceable, is limited. It is an intensive procedure, which requires the availability of an expert operator and a significant expenditure of technological and human resources. Even with the current cutting-edge digital and microscopic technologies and the most experienced human operators, the qualitative and semi-quantitative human assessment of the kidney biopsy is burdened by poor reproducibility both inter- and intra-operator [10, 11]. This issue is not specific to the pathology's characteristics, but it is rather common in medical imaging, in general, where the interpretation of data depends on human subjectivity. DLAs can significantly improve this process.

The approaches and purposes of DLAs in kidney pathology are articulated into three main processes: a) segmentation/cataloging, b) clinical classification, and c) other purposes. DLAs have been used for the automatic recognition of specific histological features applied in automatic *segmentation* (identification of glomeruli, tubules, arteries, veins, etc.) [12, 13] and *cataloging* (identification of the damaged structures, such as global and segmental sclerotic glomeruli, crescentic glomeruli, and atrophic tubules) [12–14]. Recently, Jayapandian et al. [15] have reported the automatic segmentation of six renal histologic structures with four staining techniques (hematoxylin and eosin, Masson trichrome, periodic acid–methenamine, and periodic acid–Schiff staining), using the digital kidney biopsies from the Nephrotic Syndrome Study Network dataset. The study was the first to apply ML segmentation of multiple pathology features simultaneously with a combination of staining techniques. Furthermore, because of the multicenter nature of the Consortium, this study highlights the effective capability of DLAs using heterogeneous samples with pre-analytic variations in tissue acquisition, processing, and slide preparation.

Clinical classification has a diagnostic and a prognostic purpose. Two classical examples are the IgAN Oxford Classification [16, 17] and the Diabetic Nephropathy Classification of the Renal Pathology Society [18]. The advantages of MLAs over traditional clinical classifications are traced back to important key elements: the algorithm evaluates the histological preparation in time, and, therefore, the cost is much lower than a human operator. The classifier demonstrates a performance comparable to that of the human counterpart, and in some cases, it is superior to human classifiers. Finally, one of the most appreciable characteristics of DLAs is the reproducibility of the results, one of the most problematic and well-documented weaknesses in pathology reporting.

The classifiers can be divided into three categories: unbiased classifiers, biased classifiers [11] and mixed-form classifiers (biased–unbiased) [19].

The classifiers most represented in the literature are biased: they use assumptions concerning the recognition of pathological features that are significant in the study of the disease, make use of their previous experience and literature or their past inclusion in validated manual classifications. A biased classification algorithm has the purpose of recognizing the constitutive features of a pre-existing manual classifier and substantially reproducing its logic. In this case, the algorithm replaces the human pathologist and the classifier operates on data derived from this automated recognition. An example of this approach is the work of Ginley et al. [11] who created a complex pipeline for the extraction of pathological features from the identified glomeruli. Finally, they analyzed the extracted data with a DRNN algorithm. The classifier showed predictive capabilities similar to those obtained by human scoring according to the pathologic classification of Diabetic Nephropathy of the Renal Pathology Society [18].

The unbiased classifier does not depend upon pre-conceived knowledge. In this case, the algorithm is not instructed to recognize specific pathology features but is free to create correlations between the imaging information and the diagnostic or prognostic outcome of the classifier. Kolachalama et al. [10] applied this approach in the classification of 171 biopsies, regardless of the clinical indication. They were performed and labeled based on clinical outcomes such as creatinine, eGFR, proteinuria, and 1–3- or 5-year survival. The clinical predictions were superior to those obtained from experienced pathologists.

An example of the mixed biased–unbiased approach is the work of Sheehan et al. [19], which had two phases. In the first phase, the assumption (biased approach) that the glomerular and non-glomerular compartments have two different information contents was made. The algorithm was instructed to identify the glomerular structures according to a standard segmentation protocol. In the second phase, they asked the algorithm to predict the genotype of the analyzed samples (wild type *versus* knockout mice) based on an unbiased analysis. This analysis was conducted separately for the glomerular and non-glomerular compartments. As for the glomerular compartment, the algorithm showed that it independently learned to distinguish the two genotypes based on different mesangial expansion, a phenotype previously characterized in this animal model. Notably, the application of the unbiased methodology on the non-glomerular compartment allowed the identification of unprecedented tubular phenotype, consisting of the presence of tubular vacuolation and alterations of the brush border. The tubular alterations were missed on a previous intensive traditional scanning by trained pathologists. Although the study was

conducted with an experimental mouse model, the results can certainly be extended to research with human samples. This work demonstrates that the application of an unbiased approach discloses the inherent analytical potential of DLA in identifying new histological characteristics that could be missed by a manual approach.

DLAs have also great versatility in other research applications. De Haan et al. [20] applied DLAs for the virtual staining by recoloring slides starting from a specific staining technique (H&E) to others (Masson's trichrome, periodic acid–Schiff, and Jones silver stain). The pathologists significantly improved the accuracy of the preliminary diagnoses performed with H&E alone, after accessing the virtual stainings. There was no significant difference in the accuracy of the diagnoses obtained with the virtual stainings compared to the histochemical stainings. Finally, the quality of the virtual stainings did not reveal appreciable differences with the histochemistry techniques as evaluated by the pathologists in a blind revision. This approach has several interesting practical applications. H&E staining can be prepared in a much faster time than special staining, and combining it with virtual staining allows for additional information obtained in advance when rapid reporting is needed (e.g., crescentic glomerulonephritis or transplant rejection). If the replacement of the histochemical staining with the virtual is validated, this would reduce the amount of the specimens to be processed, with increased safety and, laboratory efficiency alongside the cost reduction.

In another alternative application of DLAs, Gupta et al. [21] developed an approach to make ground truth annotation more efficient. The potential of DLAs in pattern recognition exceeds that of MLAs, but at the cost of a significant amount of annotated data for sufficient generalization (the ability of abstraction in recognizing the structure with respect to the examples provided). The authors showed that the segmentation of kidneys was repeatedly improved when pathologists interact with automatically generated annotations throughout the training process. The process involves an iterative closed loop in which the algorithm produces a segmentation proposal, and then it is reviewed by a human operator. The segmentation updated by the human curator retrains the algorithm that then produces an improved segmentation proposal. At this point, the loop repeats itself for a certain number of iterations. This strategy was defined by the authors as 'Human in the loop', and it significantly reduced the annotation effort required for the segmentation of large datasets of microscopy images.

Interestingly, Yang et al. [22] proposed a combined ML model based on a Mask Region-CNN and a long short-term memory RNN, to identify and classify glomerular lesions detected in kidney biopsy from 60 patients affected by different glomerulonephritides (MCD/FSGS, MN, IgAN, LN, and DN). The CNN algorithm detected and segmented

glomeruli stained with H&E, PAS, PAM, and TRI. Three essential steps were realized by the MLAs as glomerulus detection, disease classification, and lesion identification. The best performance of the AI system was observed in LN and DN. Specifically, the combined AI system accelerated the diagnostic workflow of the pathologists and reduced the poor interobserver agreement. However, the study had some important limitations: i) the small dataset of glomerulonephritides; ii) the identification of lesions in a single section; iii) the study focused only on the glomerular compartment; and finally iv) no data from the immunofluorescence and electron microscopy were considered. However, this study is the first step to reduce the workflow of the nephropathologists in the kidney biopsy report.

AI applied to kidney immunofluorescence

Direct immunofluorescence (IF) of the kidney biopsy sample is a central technique in kidney pathology. It has a high signal-to-noise ratio, which makes the analysis of these preparations effective. Traditionally, these preparations are observed under a fluorescence microscope, an intensive and time-consuming process. Like any other medical image interpreted by humans, the IF report is also affected by low reproducibility. Importantly, the direct IF has only limited viewing time due to the fluorescence bleaching. Therefore, the transition to a digital approach is certainly desirable. Nonetheless, commercial scanners suitable for digitizing whole IF have a lower frequency of use in pathology laboratories compared to histochemically stained scanners. Thus, the availability of large datasets of IF images in this field is low and publications poor. Zhao et al. [23] published data from a large direct IF whole slide image dataset and demonstrated the ability of their algorithm to automatically identify glomeruli. Ligabue et al. [24] applied DLAs to the classification of direct IF on kidney biopsy samples. Starting from a large dataset of glomeruli images, they trained an algorithm to identify the main classification categories traditionally used in IF reporting (appearance, distribution, and location). The DCNN algorithm achieved an accurate definition of the main pathology features. The concordance between the algorithm and the ground truth was not inferior to that for experienced pathologists. The classification was performed by the algorithm in 62.5 ms per image, which was 117.3 times faster than the average human evaluator.

AI applied to kidney ultrasonography

Ultrasonography (US) of the kidneys is routinely used in the clinical examination of patients with glomerulonephritis before the kidney biopsy. However, during the course of the

disease the kidney tissue deteriorates, a condition that may lead to CKD/KF. The progression of kidney damage is documented by serial kidney biopsies, but this invasive approach is declined by patients and physicians due to complications and contraindications. Kidney US is generally performed after the biopsy which may provide decisive information about the progression of kidney damage. The US evaluation shows hyperechogenic regions in the parenchymal cortex, making the differentiation between cortex and medulla. To overcome interobserver variability in the interpretation of the kidney, MLAs have been used for analytical standardization after preprocessing of US images, kidney segmentation, feature extraction, and image classification. Bommanna et al. [25] performed a computer-aided kidney diagnosis system based on three steps: general processing, pre-processing, and analysis. Two decision support systems were used: the multilayer back propagation network and the FLS. Results showed that this algorithm had high classification efficiency with minimum training and testing time. The system was validated by medical experts.

Kuo et al. [26] selected 4,505 high-quality kidney US images from 1,299 patients to train a DLA to predict the eGFR impairment from the US images using a DCNN. Next, they predicted an irreversible CKD status with an XGB algorithm. The values of the kidney function obtained with DCNN were used to identify the CKD status defined by the value of eGFR. The relationship between the AI model and the eGFR was expressed by a Pearson coefficient of 0.741. The accuracy value of the model was 85.6% higher than that obtained from experienced nephrologists. Thus, this model was able to transform kidney US imaging into an effective real-time screening tool for the prediction of kidney function.

A comprehensive approach to analyze and classify the CKD stages based on the kidney US images was conducted by Chen et al. [27]. They collected 798 kidney US images from 205 patients with different stages of CKD evaluated by eGFR. They pre-processed and analyzed the kidney parenchyma, central sinus, and the whole kidney. The SVM algorithm was chosen because of the limited amount of US images with accurate annotations used for the training and testing datasets. The test results confirmed the efficiency and effectiveness of the proposed model for the diagnosis of the CKD stage. This study demonstrated that the developed model could help nephrologists to understand the physiological changes during the progression of kidney damage. Moreover, this model could help to choose appropriate treatments. The system performed better on the right kidney than the left, due to more regular images.

Kim and Ye [28] classified different stages of CKD (normal, mild-moderate, and severe) extracting data from three regions of interest (ROI): the cortex of the kidney, the boundary between the cortex and medulla, and the medulla.

They used the gray-level co-occurrence matrix algorithm. A total of 58 parameters plus the size of the kidney were used. Then, an ANN, consisting of 64 inputs (first layer), 10 hidden layers, and 3 outputs, was applied to obtain the three stages of CKD. The classification accuracy was 95.4%. This algorithm can be implemented for the diagnosis of CKD with renal US.

Zhang et al. [29] applied radiomics for the first time to 623 US images of only the kidney parenchyma obtained before kidney biopsy from 68 adult patients with documented membranous nephropathy ($n=46$) and IgA nephropathy ($n=22$). To reduce bias, the images were divided into training and test datasets. After the extraction of 180-dimensional radiomics features based on the ROI and selection through the least absolute shrinkage and selection operator (LASSO) logistic regression model, the investigators applied four MLAs (LOG, SVM, RF, and KNN) to distinguish between the two types of glomerulonephritis. The RF algorithm had the best specificity (87.5%), while LOG had the highest sensitivity (88.9%) and accuracy (76.4%). The main limitation of this study was the lack of a validation test in an independent external cohort of patients. However, the kidney biopsy report included the diagnosis and documented the main active lesions that may benefit from therapy, and this aspect cannot be elucidated by US.

To this purpose, Athavale et al. [30] developed and validated a DLA, based on kidney US images, to quantify interstitial fibrosis and tubular atrophy (IFTA) in 352 patients who had different stages of IFTA. The images were processed using the UNet technique, and the classification was performed using the DCNN and XGB algorithms. Patients were assigned to four grades of IFTA according to renal histology (grade 1 = 0–24%; grade 2 = 25–49%; grade 3 = 50–74%; grade 4 = 75–100%). The sensitivity analysis showed an accuracy of over 84% that increased to 90% when classification performance was measured in 268 patients with long-term multiple US images. The highest prediction value was observed in patients with lupus nephritis (LN). The performance of the system did not improve the diagnosis when combining the obtained data with baseline laboratory findings. Because of the increased incidence of IFTA alterations during the course of the disease and repeated kidney biopsies, the use of DLA in kidney US may give additional information on the outcome of the disease. Moreover, this system reduces the cost of additional kidney biopsies and suggests therapeutic decisions such as no immunosuppressive therapy in the presence of a high percentage of IFTA. This study is proof of principle that DLA can be used to predict the progression of IFTA.

A further improvement of noninvasive kidney US techniques that predict IFTA was carried out by Zhu et al. [31]. They combined images of shear wave elastography (SWE), color Doppler flow US, and laboratory data in predicting the

progression of IFTA in 117 patients with different grades of IFTA. They did not find linear relationships between SWE and the IFTA grades among patients with different CKD stages. Therefore, they applied the RF algorithm to the combined data. MLA showed that SWE of the left kidney cortex and sinus were significant for determining the IFTA grades. In conclusion, the investigators using this algorithm improved the diagnostic performance of the traditional US for the different grades of IFTA.

AI applied to clinical outcomes

The use of landmark methods in large datasets can create prediction models and eliminate bias due to the changes of the predictor variables in time, thus, providing unbiased estimations of the conditional time-to-event probabilities [32]. However, longitudinal data are somewhat biased due to varied periods between the collection of data within and across patients. Also, patients are tracked for different periods. Data mining techniques such as classification, clustering, and combination rules are necessary to obtain unknown EHRs from databases.

MLAs replace the traditional statistical models recognizable to most clinicians [33]. They can manage a larger number of EHRs such as continuous and non-continuous variables that exceed the capacity of classical statistical methods. They constitute the latest progress in the field of bioinformatics and computational biology. To improve the accuracy of MLAs in the first phase of training, many attributes are used and then the number of variables is reduced to improve the accuracy. We report the application of MLAs in membranous nephropathy (MN), IgA nephropathy (IgAN), lupus nephritis (LN), and diabetic kidney disease (DKD).

Membranous nephropathy

Naumovic et al. [34] used ANN to evaluate the factors of poor prognosis and the efficacy of various therapies in 124 biopsy-proven idiopathic MN patients with a follow-up of at least 2 years (Table 1). Twenty-one demographic, clinical, laboratory, and histology data were used as input for the development of ANN, while five disease outcomes (death, initiation of dialysis, kidney function, remission, and relapse of the nephrotic syndrome) were the output parameters. Based on the received therapy for 6 months, patients were divided into three groups (corticosteroids or cyclophosphamide and small doses of corticosteroids or corticosteroids and chlorambucil). The ANN algorithm revealed a favorable outcome for 80 patients, while 44 had a poor outcome. A favorable outcome was observed in 62 patients and an unfavorable outcome in the other 62 subjects. The difference between the expected and observed positive therapeutic

Table 1 Application of machine learning algorithms in membranous nephropathy (MN) and immunoglobulin A nephropathy (IgAN)

Authors (year)	Disease	Population	Data set size	Validation set size	Learning model	Comparison with other MLA	Accuracy (%)	Input variables	Output variables	Limitations
Naumovic et al. (2010)	MN	C	124		ANN	SOM	100	21	Therapy efficacy	No external validation
Schena et al. (2021)	IgAN	C	948	167	ANN		86	7	KF	Tool developed only in Caucasians
Liu et al. (2018)	IgAN	A	262		RF		95	12	KF	No external validation
Chen et al. (2019)	IgAN	A	1,022	1,025	XGB	DT, RF, SVM, ANN, LOG	84	10	KF; 50% reduction of eGFR	Tool developed only in Asians
Zhang et al. (2020)	IgAN	A	1,167		XGB	RF, SVM, KNN, ELM, CB	78	16	KF	No external validation

C: Caucasians; A: Asians; ANN: artificial neural network; SOM: self-organizing maps; KF: kidney failure; RF: random forest; DT: decision tree; SVM: support vector machine; LOG: logic regression; KNN: K-nearest neighbor; ELM: extreme learning machine; CB: cut boost; extreme gradient boosting machine; DNN: deep neural network

effect was 14.52% that indicated the percentage of therapy inefficiency. Next, the investigators analyzed the accuracy of the algorithm in the three groups of patients receiving different therapies. The accuracy was very low in the first group, increased in the second group and reached the value of 82% in the third group. In conclusion, these results demonstrated the superiority of corticosteroids combined with alkylating agents in patients with idiopathic MN.

IgA nephropathy

In the last ten years, Schena et al. developed different ANN tools to predict KF in patients affected by biopsy-proven IgAN [35–37] (Table 1). Recently, they have developed a new ANN model [38] using a retrospective cohort of 948 patients with biopsy-proven IgAN, many of them from the VALIGA cohort [39]. The model was composed of seven variables, and the histological score of the kidney biopsy was carried out using the Oxford classification. The tool was based on two ANNs, the first of which predicted KF (yes or no) and the second the time it would take to develop KF. This updated model was more efficient than the previous developed. The classifier model had a performance of 0.82 in patients with a follow-up of five years and 0.89 in a follow-up of 10 years. Similar results were obtained over a 25-year follow-up period. Next, the model was validated by an external independent cohort of 167 IgAN patients. The DialCheck tool is available on the web site (www.igan.poliba.it), and it can be downloaded on Android cellular phones from the Play Store application and also on iOS cellular phones from the App Store.

A few AI studies on the prediction of KF in Chinese IgAN patients have been published (Table 1). Liu et al. [40] used an RF model to predict KF in 262 IgAN patients with a median follow-up of 4.66 years. Six variables (gender, age, hypertension, serum creatinine, 24-h proteinuria, and histological grading) were used to train the model, which had an AUC value of 92.57%. Then, additional variables with a calculated high rank (hemoglobin, albumin, eGFR, BUN, albuminuria, and C3 staining) were included and improved the AUC value to 95.45%. The study has some important weaknesses, such as the input of serum creatinine and eGFR that already includes the values of serum creatinine, or inclusion of albuminuria together with proteinuria that takes into consideration albuminuria. Finally, the model was not validated in an independent external cohort of patients.

Later, Chen et al. [41] developed a new tool using the database of the Nanjing Glomerulonephritis Registry. The XGB model was compared with other algorithms. The XGB had the highest values in the derivation and validation cohort (0.89 and 0.84, respectively). The 36 variables used in the training were reduced to 10 using the Shapley additive explanation method. Then, the XGB algorithm was

combined with the scoring scale model including three variables (tubular-atrophy/interstitial fibrosis, global sclerosis, and proteinuria) selected by stepwise Cox regression. Both algorithms were incorporated into the Nanjing IgAN risk stratification system tool, which is available online. The combined event of KF and 50% reduction in eGFR within 5 years after the kidney biopsy diagnosis were considered the outcomes. The limitations of this study were the use of a 50% reduction in eGFR that is used in randomized prospective clinical trials, and the validation carried out only in Chinese IgAN patients.

Zhang et al. [42] compared the XGB algorithm with other models to predict KF within 5 years in a cohort of 1,167 pediatric Chinese patients. First, they used 36 independent variables related to patient information, and then, the number was reduced to 16 attributes by the chi-square analysis. Performance analysis demonstrated that the XGB model had the best AUC (0.85) and accuracy (78.6%). A strength point of this study is the unique model developed in children with IgAN. However, it was not validated in an external independent cohort.

In conclusion, different MLAs have been used to predict KF in IgAN patients. However, they were validated in different racial populations. Further studies should be done to develop an international AI tool that predicts KF in IgAN patients of different races.

Lupus nephritis

Systemic lupus erythematosus (SLE) is a disease characterized by abnormal self-hyperactivity of B cells and differentiation of plasmablast and plasma cells secreting various autoantibodies. Circulating immune complexes are deposited in the kidneys developing LN. Moreover, hyperactive T cells contribute to the large production of type 1-interferon.

Kegerreis et al. [43] integrated transcriptomics data of three whole blood data sets to study the gene expression patterns of SLE in the active phase (SLE disease activity index of 6 or higher) and upregulation of DNAJC13 and IRF4, and downregulation of RPL22 with consistent fold changes (Table 2). Then, the authors used three MLAs (generalized linear model, KNN, and RF) to separate active from non-active SLE patients. The RF classifier with 25 variables showed the best performance (83% accuracy). Based on raw gene expression, the data indicated that the disease activity was sensitive to distinguish the disease phases and to provide personalized medication to SLE patients.

The systemic lupus collaborating clinics damage index (SDI) score is used to quantify the chronic damage occurring in 50% of SLE patients. Ceccarelli et al. [44] used the RNN algorithm to predict the development of chronic damage in a cohort of SLE patients with long-term follow-up divided into two groups: 38 patients with

Table 2 Application of machine learning algorithms in systemic lupus erythematosus (SLE) and lupus nephritis (LN)

Authors (year)	Disease	Population	Data set size	Validation set size	Learning model	Comparison with other MLA	Accuracy (%)	Input variables	Output variables	Limitations
Keggarreis et al. (2020)	SLE	C	413		RF	GLM, KNN	83	25	Active SLE	No external validation
Ceccarelli et al. (2017)	SLE	C	413		RNN	FFNN, LOG	77	15	Chronic damage	No external validation
Rajimehr et al. (2002)	SLE	A	242		ANN	LOG	68	11	LN development	No external validation
Chen et al. (2021)	LN	A	1,694		XGB		81	30	Renal flare	No external validation
Wolf et al. (2016)	LN	AA/AM	140		RF		79	30	Progressive kidney damage	No external validation
Helget et al. (2021)	LN	AA/AM	83		LOG, RF, SVMML, ANN, SVMR			8	Response to therapy	Low number of patients

C: Caucasians; A: Asians; AA: Afro-Americans; AM: Americans; ANN: artificial neural network; RF: random forest; XGB: extreme gradient boosting; SVM: support vector machine; LOG: logic regression; KNN: K-nearest neighbor; ELM: extreme learning machine; GLM: generalized linear model; RNN: recurrent neural network; FFNN: feed-forward neural network; SVMML: SVM linear; SVMR: SVM Gaussian

SDI = 0 at baseline, who developed chronic damage during the follow-up, and 94 patients without chronic damage at baseline who did not develop chronic damage (SDI = 0) (Table 2). The investigators used the same features of SDI (persistent and irreversible damage for at least 6 months assessed in 12 organ systems) as input data. The AUC of RNN to predict chronic damage was 0.77. RNN was compared with other algorithms that had an AUC greater than 0.50. The results indicated that RNN was suitable to identify patients at high risk of chronic organ damage when a sequence of input data was included in the algorithm during the follow-up. The main limitation of this study is the absence of the validation process of the tool in an independent cohort of SLE patients.

LN occurs in up to 50% of SLE patients at 5 years after diagnosis. Rajimehr et al. [45] used the ANN algorithm to predict LN in 242 SLE patients. They designed a list of 11 clinical and laboratory criteria to predict LN that were used to compare the ANN with the LOG model and with predictions from five experienced specialists (Table 2). The comparative analysis revealed that ANN had the best value of sensitivity (73.77%), specificity (62.96%), and overall accuracy (68.69%). Unfortunately, the investigators did not include serum creatinine, proteinuria, and hypertension in the list of the input variables of ANN.

Renal flare in LN is associated with poor kidney outcomes. Chen et al. [46] used the XGB algorithm to predict the renal flare at 5 years after remission in 1,694 Chinese LN patients (Table 2). The algorithm included 59 variables that were reduced to 30 by the stepwise Cox regression model. The algorithm had a good predictive performance with high values of C-index. Interestingly, partial remission was the most important predictive variable in the XGB model. This suggests that induction therapy should be prolonged to more than 6 months in patients with partial remission. Unfortunately, this study was not validated in an independent Caucasian cohort of LN patients.

The guidelines of the American College of Rheumatology recommend the change of therapy after 6 months if the patient does not respond. However, in these patients the kidney damage progresses during these 6 months. MLAs have been used by clinicians in defining such responses. Wolf et al. [47] used a panel of urinary parameters (cytokines, chemokines, growth factors) that could better reflect the kidney inflammatory process to predict the outcome of LN. Data from 140 LN patients were collected at the time of kidney biopsy before starting therapy (Table 2). Traditional clinical biomarkers and urinary parameters were used in the development of an RF algorithm that provided better outcome prediction (AUC = 79%) than traditional biomarkers. However, the study has important limitations because the urinary biomarkers were not normalized to creatininuria and the algorithm was not validated in an external independent

cohort of patients. Moreover, patients received different induction therapy protocols.

Recently, Helget et al. [48] have proposed a clinical proof-of-concept tool formed by five MLAs to predict the outcome of LN and to facilitate the discussion with patients concerning therapy (Table 2). They used data from a prospective longitudinal cohort database of 83 LN patients with a follow-up of at least 12 months after kidney biopsy. Forty-one responded to therapy, and 42 were non-responders. A combination of 20 variables was first considered, and then eight of them (activity and chronicity score, interstitial inflammation and fibrosis, urine protein-to-creatinine ratio, white blood cells, hemoglobin, and the cross-validated AUC value of each model) were selected for five algorithms. Next, a web-based tool was used where the operator included the variables above reported at the time of kidney biopsy. The principal limitation of this study was that the cohort of patients were mainly Afro-American, even though this disease is more frequent in this population.

In conclusion, different tools have been developed to predict SLE, LN and the outcome, but these tools should be validated in prospective clinical studies enrolling patients with different racial profiles.

Diabetic kidney disease

DKD is characterized by the presence of either microalbuminuria, proteinuria, impaired eGFR or all of the above. In 2011, the National Center for Chronic Disease Prevention and Health Promotion demonstrated that 30% of patients with type 1 or 2 diabetes mellitus develop clinical complications with DKD being one of the most frequent complications. It is necessary to prevent such disease complications at early stages. Most prediction models based on multivariable regression analysis or MLAs have been applied to predict the development of long-term (i.e., retinal, cardiovascular, and renal) and short-term (i.e., hypoglycemia) complications [49] (Table 3).

To facilitate the reuse of EMRs in research, the National Institutes of Health funded the Informatics for Integrating Biology and the Bedside National Center for Biomedical Computing to provide an open source of clinical data [50]. Song et al. [51] used a retrospective cohort of 15,645 diabetes mellitus patients of whom 37% were affected by DKD (Table 3). All patients had at least one value of eGFR or albuminuria. Data of 15 types of clinical observations were extracted from EMRs. Four algorithms were compared, and the XGB algorithm had the best performance (AUC 82%) in predicting DKD 6 months before the onset of renal complications. A set of 440 ensemble features was predictive of DKD development; some of which would merit further research.

Table 3 Application of machine learning algorithms in diabetes mellitus (DM) and in DM type 2

Authors (year)	Disease	Population	Data set size	Validation set size	Learning model	Comparison with other MLA	Accuracy (%)	Input variables	Output variables	Limitations
Song et al. (2019)	DM	AA/AM	15,645		GBM	DNN, ELAST-NET		440	DKD	No external validation
Leung et al. (2010)	DM	A	673		SVM, RF	PLSR, DT, NB, ANN	94	15	DKD	No external validation
Meza-Palacios et al. (2017)	DM2	C	70		FL			7	DKD	Low number of patients
Makino et al. (2019)	DM2	A	64,000		LOG			24	DKD	
Katsuki et al. (2018)	DM2	A	30,810		CA		66		DKD	No external validation

C: Caucasians; A: Asians; AA: Afro-Americans; AM: Americans; DM: diabetes mellitus; DKD: diabetic kidney disease; ANN: artificial neural network; RF: random forest; FL: fuzzy learning; PLSR: partial least square regression; NB: Naïve Bayes; CA: convolutional autoencoder; DT: decision tree; SVM: support vector machine; LOG: logistic regression; GBM: gradient boosting machine; ELAST NET: elastic net regression; DNN: deep neural network

The DKD prediction has also been evaluated by Leung et al. [52] who used a multistaged strategy based on MLAs and mathematical models in an 8-year prospective case–control cohort of 673 diabetic patients of whom 119 were affected by DKD (Table 3). Three categories of attributes were included into various ML programs: a) clinical and genetic variables, b) genetic variables only, and c) clinical variables only. In the training phase, seven MLAs were compared using the receiver operator curve analysis. In the second phase, the trained models selected a subset of variables with the best performance. The NB algorithm and partial least squares regression showed the best performance in the training phase. SVM and RF had the best performance for all subgroups of variables in the testing phase. Thus, 10 clinical variables and 5 genetic attributes used in the SVM and RF predicted DKD in subjects at high risk. The strengths of this study were: a) the comparison of different MLAs, b) the inclusion of 10 phenotypes and 5 genotypes in the models, and c) the development of a novel model based on two MLAs to predict DKD. However, the model was not validated in large cohorts of Chinese and non-Chinese patients.

To control the outcome of DKD, Meza-Palacios et al. [53] developed an FL algorithm in 60 type 2 diabetes patients attending hospital checks for DKD control (Table 3). The consultants' diagnoses were compared with the results obtained by FL with seven input variables (serum creatinine, eGFR, uric acid, dyslipidemia, glycemia, hypertension, and presence of type 2 diabetes). FL was successful in 93% of the cases. Thus, this tool may aid physicians in the control of DKD and the prescription of adequate therapy. The variables that had the highest impact on DKD control were serum creatinine, eGFR, and hypertension. Good control of the nephropathy was observed when GFR oscillated between 100 and 200 ml/min/1.73 m² and serum creatinine between 0.4 and 0.9 mg/dL. The investigators underlined that this supportive tool can be used by general practitioners in villages, towns, and sites where there are no specialist doctors after adequate training. This study has two important limitations: the absence of albuminuria as an input variable and the validation test of the tool in an independent external cohort of patients.

The progression of kidney damage was evaluated by Makino et al. [54] who developed an ANN model predictive of CKD in a cohort of more than 64,000 type 2 diabetes patients (Table 3). Interestingly, they first processed texts from medical examinations and nutrition consultations using latent Dirichlet allocation, which was used for topic extraction. Next, 24 variables were selected and collected longitudinally during the follow-up using a convolutional autoencoder system that was composed of 5 hidden layers. The prediction of CKD six months after the first visit was assigned as “aggravated (CKD stage 2–5)” or “stable (CKD stage 1)” kidney function. Among the chosen variables, the

increase of proteinuria strongly impaired the clinical course of DKD.

Finally, to predict DKD aggravation from stage 1 to 2, Katsuki et al. [55] used a stacked convolutional autoencoder model in a real-world EHR, based on long-term data obtained from 30,810 Japanese patients (Table 3). DKD aggravation was predicted 6 months after the last record in the EHR. The model was composed of five layers. The best predictive variables in patients who progressed from stage 1 to 2 were the increasing values of creatine phosphokinase, and body mass index, and the decreasing values of eGFR over time. However, the model showed a low accuracy (66%) and lacked validation.

In conclusion, different AI models have been used to predict CKD in patients with DM. Interestingly, some of them predicted the progression of the kidney damage. None of them was validated in large cohorts of DM patients. Therefore, it is necessary to organize prospective clinical studies in multicenter consortia before moving toward their use in clinical practice.

Future considerations

In this review, we analyzed AI applications on the medical history of patients, imaging from digital pathology, and ultrasonography, and in clinical outcomes of patients with primary and secondary glomerular diseases.

The main limitation in the application of MLAs in the histopathology of glomerulonephritis is the low number of cases included in the studies due to the low incidence of these diseases. Therefore, it is necessary to organize a large community of nephrologists and pathologists interested in digital pathology and outcome who can create consortia with large datasets of patients, renal structure images, and laboratory data using standardized and reproducible methods (e.g., serum creatinine and daily or spot proteinuria). Furthermore, the lack of standard image datasets including all possible variations can be overcome by increasing data availability of animal tissues with a given disease for the training of MLAs [56] or the use of autoencoders trained to identify specific stained renal structures between entities of glomerulonephritis [57]. Currently, MLAs in image analysis can be considered only for research purposes because there are no recognized certifications that would permit replacing the conventional diagnostic applications. MLAs are promising techniques of image analysis for institutions where conventional imaging procedures are controlled. Moreover, they may improve the work of skilled pathologists by integrating different variables such as tissue fixation, stain coloration, and image acquisition.

MLAs for the early detection of CKD and GFR estimation by kidney US are promising but still not solid enough

to be incorporated into clinical practice. To improve the performance of these models further multicenter studies should be conducted in large cohorts of patients, using standardized imaging protocols and homogeneous preparation for US imaging. Moreover, the models must be tested in large follow-up studies. Finally, an effort to apply automatic kidney segmentation before diagnosis using DLAs with high accuracy is necessary.

As for the clinical outcome studies, a few tools for specific glomerulonephritides are available in clinical practice. They can be downloaded on electronic devices but can only be applied to cohorts of one race. Also, many studies used to develop ML tools to predict the progression of glomerular disease and KF in small cohorts. Furthermore, the data of the training and testing sets were validated in external independent cohorts of patients. Therefore, it is necessary to organize large consortia with multiracial cohorts of patients, some of whom could be included in the validation study. Finally, most studies on the development of MLAs were carried out using retrospective cohorts. It has been demonstrated that the performance of the models might differ in retrospective cohorts when compared to real-world data. Therefore, the models must be validated in large prospective cohorts.

In conclusion, the development of AI-based clinical decision support systems in glomerular diseases needs, first, the strict collaboration between physicians, bioinformaticians, and AI researchers, and then, the validation of the developed MLAs in multicenter consortia with large cohorts of different racial profiles.

Acknowledgements The work has been financially supported by a grant of the Italian Ministry of University (PON RI ARS01_00876). The funder had no role in the design of the study.

Declarations

Conflict of interest All authors declare no competing interests.

References

- Beam AL, Kohane IS (2018) Big Data and Machine Learning in Health Care. *JAMA* 319:1317–1318. <https://doi.org/10.1001/jama.2017.18391>
- Koleck TA, Theresa A, Bourne PE, Bakken S (2019) Natural language processing of symptoms documented in free-text narratives of electronic health records: a systematic review. *J Am Med Inform Assoc* 26:364–379. <https://doi.org/10.1093/jamia/ocy173>
- Savova GK, Masanz JJ, Ogren PV, Zheng J et al (2010) Mayo clinical Text Analysis and Knowledge Extraction System (cTAKES): architecture, component evaluation and application. *J Am Med Inform Assoc* 17:507–513. <https://doi.org/10.1136/jamia.2009.001560>
- Aronson AR, Lang FM (2010) An overview of MetaMap: historical perspective and recent advances. *J Am Med Inform Assoc* 17:229–236. <https://doi.org/10.1136/jamia.2009.002733>
- Shang N (2014) Integrating domain knowledge to improve signal detection from electronic health records for pharmacovigilance. UT SBMI Dissertation (Open Access). 26
- Polignano M, Narducci F, Iovine A, Musto C et al (2020) HealthAssistantBot: a personal health assistant for the Italian language. *IEEE Access* 8:107479–107497. <https://doi.org/10.1109/ACCESS.2020.3000815>
- Eikstadt RN, Desmond HE, Lindner C, Chen LY et al (2021) The Development and Use of an EHR-Linked Database for Glomerular Disease Research and Quality Initiatives. *Glomerular Dis* 1:173–179. <https://doi.org/10.1159/000518187>
- Kang E, Kim Y, Kim YC, Kim E et al (2020) Biobanking for glomerular diseases: a study design and protocol for KOREA Renal biobank NETwoRK System TOWard NExt-generation analysis (KORNERSTONE). *BMC Nephrol* 21:367. <https://doi.org/10.1186/s12882-020-02016-z>
- Oliverio AL, Marchel D, Troost JP, Ayoub I et al (2021) Validating a computable phenotype for nephrotic syndrome in children and adults using PCORnet® data. *Kidney* 360. 2:1979–1986. <https://doi.org/10.34067/KID.0002892021>
- Kolachalama VB, Singh P, Lin CQ, Mun D et al (2018) Association of Pathological Fibrosis With Renal Survival Using Deep Neural Networks. *Kidney Int Rep* 3:464–475. <https://doi.org/10.1016/j.ekir.2017.11.002>
- Ginley B, Lutnick B, Jen KY, Fogo AB et al (2019) Computational Segmentation and Classification of Diabetic Glomerulosclerosis. *J Am Soc Nephrol* 30:1953–1967. <https://doi.org/10.1681/ASN.2018121259>
- Zeng C, Nan Y, Xu FT, Lei Q et al (2020) Identification of glomerular lesions and intrinsic glomerular cell types in kidney diseases via deep learning. *J Pathology* 252:53–64. <https://doi.org/10.1002/path.5491>
- Chagas P, Souza L, Araújo I, Aldeman N et al (2020) Classification of glomerular hypercellularity using convolutional features and support vector machine. *Artif Intell Med* 103:101808. <https://doi.org/10.1016/j.artmed.2020.101808>
- Bueno G, Fernandez-Carrobles MM, Gonzalez-Lopez L, Deniz O (2020) Glomerulosclerosis identification in whole slide images using semantic segmentation. *Comput Methods Programs Biomed* 184:105273. <https://doi.org/10.1016/j.cmpb.2019.105273>
- Jayapandian CP, Chen Y, Janowczyk AR, Palmer MB et al (2021) Development and evaluation of deep learning-based segmentation of histologic structures in the kidney cortex with multiple histologic stains. *Kidney Int* 99:86–101. <https://doi.org/10.1016/j.kint.2020.07.044>
- Roberts IS, Cook HT, Troyanov S, Alpers CE et al (2009) The Oxford classification of IgA nephropathy: pathology definitions, correlations, and reproducibility. *Kidney Int* 76:546–556. <https://doi.org/10.1038/ki.2009.168>
- Cattran DC, Coppo R, Cook HT, Fehally J et al (2009) The Oxford classification of IgA nephropathy: rationale, clinicopathological correlations, and classification. *Kidney Int* 76:534–545. <https://doi.org/10.1038/ki.2009.243>
- Tervaert TW, Mooyaart AL, Amann K, Cohen AH et al (2010) Pathologic classification of diabetic nephropathy. *J Am Soc Nephrol* 21:556–563. <https://doi.org/10.1681/ASN.2010.10010>
- Sheehan S, Mawe S, Cianciolo RE, Korstanje R, Mahoney JM (2019) Detection and Classification of Novel Renal Histologic Phenotypes Using Deep Neural Networks. *Am J Pathol* 189:1786–1796. <https://doi.org/10.1016/j.ajpath.2019.05.019>

20. de Haan K, Zhang Y, Zuckerman JE, Liu T et al (2021) Deep learning-based transformation of H&E stained tissues into special stains. *Nat Commun* 12:4884. <https://doi.org/10.1038/s41467-021-25221-2>
21. Gupta L, Klinkhammer BM, Boor P, Merhof D, Gadermayr M (2019) Iterative learning to make the most of unlabeled and quickly obtained labeled data in histology. *Proceedings of The 2nd International Conference on Medical Imaging with Deep Learning* 105: 215–224
22. Yang K, Leea C, Wangb HS, Shun-Chen Q et al (2021) Glomerular disease classification and lesion identification by machine learning. *Biomed J* S2319–4170(21):00111–00116. <https://doi.org/10.1016/j.bj.2021.08.011>
23. Zhao K, Tang YJJ, Zhang T, Carvajal J et al (2018) DGDI: A Dataset for Detecting Glomeruli on Renal Direct Immunofluorescence. Presented at 2018 Digital Image Computing: Techniques and Applications, Camberra, Australia, December 10–13, 2018
24. Ligabue G, Pollastri F, Fontana F, Leonelli M et al (2020) Evaluation of the Classification Accuracy of the Kidney Biopsy Direct Immunofluorescence through Convolutional Neural Networks. *Clin J Am Soc Nephrol* 15:1445–1454. <https://doi.org/10.2215/CJN.03210320>
25. Bommanna RK, Madheswaran M, Thyagarajah K (2008) A Hybrid Fuzzy-Neural System for Computer-Aided Diagnosis of Ultrasound Kidney Images Using Prominent Features. *J Med Syst* 32:65–83. <https://doi.org/10.1007/s10916-007-9109-5>
26. Kuo CC, Chang CM, Liu KT, Lin WK et al (2019) Automation of the kidney function prediction and classification through ultrasound-based kidney imaging using deep learning. *NPJ Digit Med* 2:29. <https://doi.org/10.1038/s41746-019-0104-2>
27. Chen CJ, Pai TW, Hsu HH, Lee CH et al (2019) Prediction of chronic kidney disease stages by renal ultrasound imaging. *Enterp Inf Syst* 14:1–18. <https://doi.org/10.1080/17517575.2019.1597386>
28. Kim D-H, Ye S-Y (2021) Classification of Chronic Kidney Disease in Sonography Using the GLCM and Artificial Neural Network. *Diagnostics* 11:864. <https://doi.org/10.3390/diagnostics11050864>
29. Zhang L, Chen Z, Feng L, Guo L et al (2021) Preliminary study on the application of renal ultrasonography radiomics in the classification of glomerulopathy. *BMC Med Imaging* 21:115. <https://doi.org/10.1186/s12880-021-00647-8>
30. Athavale AM, Hart PD, Itteera M, Cimbaluk D et al (2021) Development and Validation of a Deep Learning Model to Quantify Interstitial Fibrosis and Tubular Atrophy From Kidney Ultrasonography Images. *JAMA Netw Open* 4:e2111176. <https://doi.org/10.1001/jamanetworkopen.2021.11176>
31. Zhu M, Ma L, Yang W, Tang L et al (2021) Elastography ultrasound with machine learning improves the diagnostic performance of traditional ultrasound in predicting kidney fibrosis. *J Formosan Med Ass*. <https://doi.org/10.1001/10.1016/j.jfma.2021.08.011>
32. Wells BJ, Chagin KM, Li L, Hu B et al (2015) Using the landmark method for creating prediction models in large datasets derived from electronic health records. *Health Care Manag Sci* 18:86–92. <https://doi.org/10.1007/s10729-014-9281-3>
33. Obermeyer Z, Emanuel EJ (2016) Predicting the Future - Big Data, Machine Learning, and Clinical Medicine. *N Engl J Med* 375:1216–1219. <https://doi.org/10.1056/NEJMp1606181>
34. Naumovic R, Furuncic D, Jovanovic D, Stosovic et al (2010) Application of artificial neural networks in estimating predictive factors and therapeutic efficacy in idiopathic membranous nephropathy. *Biomed Pharmacother* 64:633–638. <https://doi.org/10.1016/j.biopha.2010.06.003>
35. Di Noia T, Ostuni VC, Pesce F, Binetti G et al (2013) An end stage kidney disease predictor based on an artificial neural networks ensemble. *Expert Syst Appl* 40:4438–4445. <https://doi.org/10.1016/j.eswa.2013.01.046>
36. Diciolla M, Binetti G, Di Noia T, Pesce F et al (2015) Patient classification and outcome prediction in IgA nephropathy. *Comput Biol Med* 66:278–286. <https://doi.org/10.1016/j.combiomed.2015.09.003>
37. Pesce F, Diciolla M, Binetti G, Naso D et al (2016) Clinical decision support system for end-stage kidney disease risk estimation in IgA nephropathy patients. *Nephrol Dial Transpl* 31:80–86. <https://doi.org/10.1093/ndt/gfv232>
38. Schena FP, Anelli VW, Trotta J, Di Noia T et al (2021) Development and testing of an artificial intelligence tool for predicting end-stage kidney disease in patients with immunoglobulin A nephropathy. *Kidney Int* 99:1179–1188. <https://doi.org/10.1016/j.kint.2020.07.046>
39. Coppo R, Troyanov S, Bellur S, Catran D et al (2014) Validation of the Oxford classification of IgA nephropathy in cohorts with different presentations and treatments. *Kidney Int* 86:828–836. <https://doi.org/10.1038/ki.2014.63>
40. Liu Y, Zhang Y, Liu D, Tan X et al (2018) Prediction of ESRD in IgA Nephropathy Patients from an Asian Cohort: A Random Forest Model. *Kidney Blood Press Res* 43:1852–1864. <https://doi.org/10.1159/000495818>
41. Chen T, Li X, Li Y, Xia E et al (2019) Prediction and Risk Stratification of Kidney Outcomes in IgA Nephropathy. *Am J Kidney Dis* 20:300–309. <https://doi.org/10.1053/j.ajkd.2019.02.016>
42. Zhang P, Wang R, Shi N (2020) IgA Nephropathy Prediction in Children with Machine Learning Algorithms. *Future Internet* 12:230. <https://doi.org/10.3390/fi12120230>
43. Kegerreis B, Catalina MD, Bachali P, Geraci NS et al (2018) Machine learning approaches to predict lupus disease activity from gene expression data. *Sci Rep* 9:9617. <https://doi.org/10.1038/s41598-019-45989-0>
44. Ceccarelli F, Sciandrone M, Perricone C, Galvan G et al (2017) Prediction of chronic damage in systemic lupus erythematosus by using machine-learning models. *PLoS One* 12:e0174200. <https://doi.org/10.1371/journal.pone.0174200>
45. Rajimehr R, Farsiu S, Montaser Kouhsari L, Bidari A et al (2002) Prediction of lupus nephritis in patients with systemic lupus erythematosus using artificial neural networks. *Lupus* 11:485–492. <https://doi.org/10.1191/0961203302lu226oa>
46. Chen Y, Huang S, Chen T, Liang D et al (2021) Machine Learning for Prediction and Risk Stratification of Lupus Nephritis Renal Flare. *Am J Nephrol* 52:152–160. <https://doi.org/10.1159/000513566>
47. Wolf BJ, Spainhour JC, Arthur JM, Janech MG, Petri M, Oates JC (2016) Development of Biomarker Models to Predict Outcomes in Lupus Nephritis. *Arthritis Rheumatol* 68:1955–1963. <https://doi.org/10.1002/art.39623>
48. Helget LN, Dillon DJ, Wolf B, Parks LP et al (2021) Development of a lupus nephritis suboptimal response prediction tool using renal histopathological and clinical laboratory variables at the time of diagnosis. *Lupus Sci Med* 8:e000489. <https://doi.org/10.1136/lupus-2021-000489>
49. Ellahham S (2020) Artificial Intelligence: The Future for Diabetes Care. *Am J Med* 133:895–900. <https://doi.org/10.1016/j.amjmed.2020.03.033>
50. Murphy SN, Weber G, Mendis M, Gainer V et al (2010) Serving the enterprise and beyond with informatics for integrating biology and the bedside (i2b2). *JAMA* 303:124–130. <https://doi.org/10.1136/jamia.2009.000893>
51. Song X, Waitman LR, Hu Y, Yu ASL et al (2019) Robust clinical marker identification for diabetic kidney disease with ensemble feature selection. *J Am Med Informatics Ass* 26:242–253. <https://doi.org/10.1093/jamia/ocy165>
52. Leung RKK, Wang Y, Ma RCW, Luk AOY et al (2013) Using a multi-staged strategy based on machine learning and mathematical modeling to predict genotype-phenotype risk patterns in diabetic kidney disease: a prospective case-control cohort analysis. *BMC Nephrol* 14:162. <https://doi.org/10.1186/1471-2369-14-162>

53. Meza-Palacios R, Aguilar-Lasserre AA, Ureña-Bogarín EL, Vázquez-Rodríguez CF et al (2017) Development of a fuzzy expert system for the nephropathy control assessment in patients with type 2 diabetes mellitus. *Expert Syst Appl* 72:335–343. <https://doi.org/10.1016/j.eswa.2016.10.053>
54. Makino M, Yoshimoto R, Ono M, Itoko T et al (2019) Artificial intelligence predicts the progression of diabetic kidney disease using big data machine learning. *Sci Rep* 9:11862. <https://doi.org/10.1038/s41598-019-48263-5>
55. Katsuki T, Ono M, Koseki A, Kudo M et al (2018) Risk Prediction of Diabetic Nephropathy via Interpretable Feature Extraction from EHR Using Convolutional Autoencoder. *Stud Health Technol Inform* 247:106–110. <https://doi.org/10.3233/978-1-61499-852-5-106>
56. Chang H, Han J, Zhong C, Snijders AM et al (2018) Unsupervised transfer learning via multiscale convolutional sparse coding for biomedical applications. *IEEE Trans Pattern Anal Mach Intell* 40:1182–1194. <https://doi.org/10.1109/TPAMI.2017.2656884>
57. Uzunova H, Schultz S, Handels H, Ehrhardt J (2019) Unsupervised pathology detection in medical images using conditional variational autoencoders. *Int J Comput Assist Radiol Surg* 14:451–461

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.