Machine learning in nephrology: scratching the surface

Qi Li¹, Qiu-Ling Fan², Qiu-Xia Han¹, Wen-Jia Geng³, Huan-Huan Zhao¹, Xiao-Nan Ding¹, Jing-Yao Yan¹, Han-Yu Zhu³

¹Department of Nephrology, Chinese People’s Liberation Army General Hospital, Chinese People’s Liberation Army Institute of Nephrology, State Key Laboratory of Kidney Diseases, National Clinical Research Center for Kidney Diseases, Beijing Key Laboratory of Kidney Diseases, Beijing 100853, China; ²Department of Nephrology, The First Affiliated Hospital of China Medical University, Shenyang, Liaoning 110001, China; ³Department of Nephrology, Guangdong Provincial Hospital of Chinese Medicine, Nephrology Institute of Guangdong Provincial Hospital of Chinese Medicine, The Second Affiliated Hospital of Guangzhou University of Chinese Medicine, Guangzhou, Guangdong 510120, China.

Abstract

Machine learning shows enormous potential in facilitating decision-making regarding kidney diseases. With the development of data preservation and processing, as well as the advancement of machine learning algorithms, machine learning is expected to make remarkable breakthroughs in nephrology. Machine learning models have yielded many preliminaries to moderate and several excellent achievements in the fields, including analysis of renal pathological images, diagnosis and prognosis of chronic kidney diseases and acute kidney injury, as well as management of dialysis treatments. However, it is just scratching the surface of the field; at the same time, machine learning and its applications in renal diseases are facing a number of challenges. In this review, we discuss the application status, challenges and future prospects of machine learning in nephrology to help people further understand and improve the capacity for prediction, detection, and care quality in kidney diseases.

Keywords: Machine learning; Nephrology; Kidney diseases

Introduction

Chronic kidney disease (CKD) is a major public health problem worldwide. Although estimates of CKD prevalence vary widely within and between countries, it is undeniable that CKD patients face the risk of serious consequences such as end-stage renal disease (ESRD) or cardiovascular disease, which is a growing global health burden. While significant progress has been made in prevention and treatment in recent decades, more efforts are needed to reverse the situation.

Machine learning (ML) is a kind of artificial intelligence (AI). The core of it is algorithmic methods, which enable the machine to solve problems without specific computer programming. The wide application of ML in the medical field helps to promote medical innovation, reduce medical costs, and improve medical quality. However, related research on solving clinical problems through ML in nephrology still needs to increase. Understanding the purpose and method of ML application and the current situation of its application in nephrology is a prerequisite for correctly addressing and overcoming these challenges.

Overview of ML

ML helps computers possess the same ability to learn, identify, and judge as human beings. It is mainly about the development and deployment of algorithms and usually uses statistical tools to determine behavior. ML technology can be broadly divided into supervised learning, unsupervised learning, and reinforcement learning according to different modeling needs.

Supervised learning

Supervised learning, such as logistic regression (LR), naive Bayesian classification, support vector machine (SVM), and random forest (RF), is the most common form of ML used in medical research. Each instance of supervised learning consists of an input object (usually a vector) and the desired output value (also known as a supervised signal). The application of supervised learning is very extensive. However, there are still limits in the application of this method in complex optimal control problems.

Quick Response Code: www.cmj.org

Access this article online

DOI: 10.1097/CM9.0000000000000694

© 2020 The Chinese Medical Association, produced by Wolters Kluwer, Inc. under the CC-BY-NC-ND license. This is an open access article distributed under the terms of the Creative Commons Attribution-Non Commercial-No Derivatives License 4.0 (CCBY-NC-ND), where it is permissible to download and share the work provided it is properly cited. The work cannot be changed in any way or used commercially without permission from the journal.

Chinese Medical Journal 2020;133(6)
Received: 28-10-2019 Edited by: Yuan-Yuan Ji

Correspondence to: Dr. Han-Yu Zhu, Department of Nephrology, Chinese People’s Liberation Army General Hospital, 28 Fuqiu Road, Haidian District, Beijing 100853, China
E-Mail: hanyuzzu301@126.com

10.1097/CM9.0000000000000694
Deep neural networks (DNNs) are the basis of DL. The signatures of high-dimensional DNNs capture thousands of subtle properties of histologic images that can be used to predict other endpoints, including disease status and physiological outcomes. Recent work has shown that DNNs have achieved human expert-level performance in natural and biomedical image classification tasks. The ability to generate assumptions, the adaptability to heterogeneous dataset analysis, and the rapid-diffusion open-source DL programs allow DL to play an important role in promoting medical development.

Convolutional neural networks

With the development of image processing, convolutional neural networks (CNNs) have gained traction in histology dataset classification. Construction of CNNs imitates the visual perception mechanism of organisms and can conduct supervised and unsupervised learning. These DL networks used in the diagnosis and treatment of a variety of diseases can also be retrained through population-specific datasets. CNNs have been proven to exceed human performance in visual target recognition.

Unsupervised learning

If the learning sample given does not contain category information, it is unsupervised learning. Similar to k-means clustering, unsupervised learning optimally divides samples into different categories based on the characteristics of training data without corresponding labels. In addition, unsupervised learning can capture intrinsic morphometric patterns in histology sections, which may play a key role in pathological diagnosis. In the future, unsupervised learning is likely to narrow the gap between human intelligence and AI.

Reinforcement learning

Reinforcement learning describes and solves the problem of agents maximizing returns or achieving specific goals through rewarding learning strategies in interaction with the environment. The Markov decision process, a common model of reinforcement learning, can capture the uncertainty associated with treatment outcomes and the randomness of the underlying processes, particularly well-suited for modeling sequential decision-making issues such as optimal dosing strategies for chronic diseases, to find the best drug dosage sequence.

What is deep learning?

In machine learning, deep learning (DL) stands out. DL is a specific type of ML method that is very suitable for processing large quantities of data as input without the need for a clear feature selection step. It can be trained to find complex patterns in big data with a high degree of precision.

ML in nephrology

Clinical big data are a valuable asset. With the continuous expansion of digital data in all aspects of health care and the development of AI technology, ML can be combined with clinical big data for disease diagnosis, prognosis, and other risk prediction. In the past few years, many different ML methods have been proposed and applied in medicine and computational biology. In addition to the diagnosis and treatment of clinical diseases by electronic health records (EHRs), they have also performed well in the field of medical image analysis and genomics.

Unlike the relatively mature application of big data and ML in the field of cardiovascular disease, the lack of evidence and the limitation of the scope of research in kidney disease have led to the fact that nephrology has not yet benefited significantly from the clinical application of big data and ML. In recent years, the use of precision medicine has made great progress in the field of nephrology. The Nephrotic Syndrome Study Network (NEPTUNE) is implementing the concept of precision medicine to develop new disease definitions through a comprehensive, multilevel analysis of disease processes in observational cohort studies. The combination of ML and big data will be an important factor in promoting precision medicine. Although it is only in its infancy, there are bright prospects and a future in the study of renal pathology and the risk prediction of kidney disease.

Renal pathology

The application of ML in biological image analysis is powerful and rapid and has been proven to be a reliable method for the analysis of malignant tumors such as cancer. In nephrology, biological image analysis by ML can be used in the diagnosis of renal pathology, which is the gold standard for the diagnosis of renal diseases. This diagnostic process, in turn, affects a series of treatment options and prognoses. Prevailing methods for glomerular assessment remain manual, labor-intensive, and non-standardized. Recently, to save manpower and time, and especially to improve the accuracy of diagnosis without bias, efforts have been made to automate the quantification of glomerular injury.

Segmentation of glomeruli and tubules

Segmentation is the basis of automatic pathological diagnosis that accurately identifies the structures of glomeruli and tubules from renal pathological images.
In the early days, an unsupervised semi-automated workflow was depicted by Sarder et al.\(^4\) and Ginley et al.\(^5\) for localization and segmentation of glomerular features. Although the localization accuracy reached 87%, the degree of glomerular injury was not reported, and the sample size was limited (15 fields with 148 glomeruli). Maree et al.\(^6\) used a supervised learning method for the identification of glomeruli and established a model with 95% precision and 81% recall. With the continuous development of CNNs in the field of image processing, Pedraza et al.\(^7\) first proposed the use of CNNs for transfer learning to identify and segment the glomeruli from the background. However, due to the lack of consideration of the location of glomeruli in large tissue areas where there are multiple glomeruli, their use in automated workflows was limited. To speed up and improve this process, Bukowy et al.\(^8\) proposed a new and robust method for glomerular localization by CNNs. Using color deconvoluted and normalized grayscale images, a glomerular localizer composed of two serially arranged ML classifiers for automatic identification of glomeruli within whole kidney sections was developed. The average precisions of used rat and human kidneys were 96.94% and 80.20%, and the average recall rates were 96.79% and 81.67%, respectively. The model laid the foundation for the automatic scoring of glomerular injury. Using this method, Kannan et al.\(^9\) demonstrated the ability of DL to assess complex tissue structures from digital human kidney biopsies. This CNN model can accurately discriminate non-glomerular images from normal or partially sclerosed and globally sclerosed images (sensitivity, 0.558; specificity, 0.999).

Furthermore, it was proven that the usage of the ML algorithm can not only clearly segment the glomeruli from the kidney image but also distinguish the renal tubule. Sheehan et al.\(^10\) used SVM classification to extract the features of renal tubules in mice (true positive rate, 92%; false-positive rate, 10%). Using 200 cores on the Vermont Advanced Compute Cluster, the glomerular segmentation pipeline can segment the full-sized mouse kidney section in approximately 40 min, allowing analysis of more glomeruli than manual completion. Additionally, the usage of SVM identifies changes in glomeruli and tubules in knock-out and wild-type genotypic mice, and scores pathological changes of mesangial matrix expansion and the degree of vacuolation of tubules. The ability of ML to identify and quantify the specific features of renal tissue structures, specifically, to facilitate segmentation of renal tissue and accurate pathological scoring and to help discover new histopathological features, has been demonstrated. Table 1 summarizes the specific research contents and ML methods of several recent articles mentioned above.

### Combination with clinical indicators

Kolachalama et al.\(^11\) first applied the analytic technique based on ML and renal pathological images in clinics. Patient-specific trichromatic images are trained and modeled by CNNs as input, and clinical indicators, including chronic kidney disease (CKD) stage, serum creatinine and nephrotic-range proteinuria at the time of biopsy, and 1-, 3-, and 5-year renal survival, are set as output in the study. Six CNN models were trained. Comparing the performance of CNN with that of an experienced renal pathologist to predict the CKD stage, the kappa-values were 0.519 and 0.051, respectively. The area under curve (AUC) was 0.912 (CNN) versus 0.840 (pathologist-estimated fibrinosis score [PEFS]) for creatinine models and 0.867 (CNN) versus 0.702 (PEFS) for proteinuria models. AUC values of the CNN models for 1-, 3-, and 5-year renal survival were 0.878, 0.875, and 0.904, respectively, whereas the AUC values for PEFS model were 0.811, 0.800, and 0.786, respectively. All the results demonstrate the effectiveness and clinical feasibility of using DN structures in renal pathology. These predictions may provide added value for biopsy results and, along with other clinical assessments, provide more accurate care management and follow-up strategies for biopsy patients. However, no prospective validation has been performed yet.

### Kidney diseases

With the introduction of the Economic and Clinical Health Information Technology act of 2009, the use of EHRs has increased dramatically.\(^12\) EHR data contain important information about the evolution of diseases. Processed by

---

### Table 1: Application of machine learning in renal pathology.

<table>
<thead>
<tr>
<th>Purpose</th>
<th>Author</th>
<th>Year</th>
<th>Kidney source</th>
<th>Algorithm</th>
<th>Size of training dataset</th>
<th>Size of testing dataset</th>
<th>Segmentation</th>
<th>Features extracted</th>
<th>Validation</th>
<th>Accuracy</th>
<th>FT score</th>
<th>Precision</th>
<th>Recall</th>
<th>Specificity</th>
<th>Sensitivity</th>
</tr>
</thead>
<tbody>
<tr>
<td>Identification and segmentation of glomeruli</td>
<td>Bukowy et al.(^13)</td>
<td>2018</td>
<td>Rat, Human</td>
<td>CNN, R-CNN</td>
<td>23,140 glomeruli (from 72 kidneys of rat)</td>
<td>1447 glomeruli (from additional 13 rat kidneys, six human samples)</td>
<td>ROIs; glomeruli</td>
<td>Trichrome image features</td>
<td>Cross and external validation</td>
<td>99.22%</td>
<td>96.94%</td>
<td>62.3%</td>
<td>95.06%</td>
<td>95.06%</td>
<td>62.3%</td>
</tr>
<tr>
<td>Identification and segmentation of NPS and GS glomeruli</td>
<td>Kannan et al.(^14)</td>
<td>2019</td>
<td>Human</td>
<td>CNN</td>
<td>120 patients</td>
<td>31 patients</td>
<td>ROIs; glomeruli</td>
<td>Trichrome image features</td>
<td>Cross validation</td>
<td>95.06%</td>
<td>62.3%</td>
<td>–</td>
<td>99.9%</td>
<td>55.4%</td>
<td>–</td>
</tr>
<tr>
<td>Three distinctions: 1. Glomerulus and tubule; 2. WT and KO glomeruli; 3. WT and KO tubuli</td>
<td>Sheehan et al.(^15)</td>
<td>2018</td>
<td>Rat</td>
<td>SVM</td>
<td>1. Eight images; 2. 20 glomeruli; 3. Tubule: 1000 image patches; 1. One image; 2. 78 glomeruli; 3. Tubules: 21,300 image patches</td>
<td>DNN-based abstract QIFs</td>
<td>ROIs; glomeruli</td>
<td></td>
<td>Cross validation</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
</tbody>
</table>

NPS: Normal or partially sclerosed; GS: Globally sclerosed; WT: Wild-type; KO: Knock-out; CNN: Convolutional neural network; SVM: Support vector machine; ROI: Region of interest; QIF: Quantitative image features; –: No relevant information mentioned in the article.
ML and widely used for modeling, EHRs can potentially help to diagnose and predict the progress of renal diseases and renal function damage. According to the data in EHRs, it is possible to achieve a more comprehensive understanding of the health status of patients and accurate prediction of the risk of obtaining specific diseases. In recent years, in the era of big data and technological innovation, using various ML algorithms to analyze standardized health data collected routinely and carrying out large-scale observational research is crucial and popular.[37] The prediction model established by EHR data is expected to promote individualized diagnosis and treatment of renal diseases and improve the quality of medical treatment.

**Diagnosis and prognosis of kidney diseases**

**CKD**

It has long been found that ML is more likely to predict eGFR accurately.[38] Table 2 lists the specific research contents and results of the following articles. Early studies found that taking advantage of AI is more likely to accurately predict the progress of various CKD. Tangri et al.[40] using routine laboratory data such as eGFR and albuminuria in EHRs, established a highly accurate proportional hazards model (COX) for CKD patients with renal failure and implemented an external validation (C-statistic, 0.921).

In addition, it has been proven that the use of temporal information in the model improves the predictive ability of renal deterioration.[40] The inclusion of temporal medical history information better predicts the loss of renal function and identifies high-risk patients with short-term renal failure.

With the continuous advancement of AI technology, various more complex and advanced ML algorithms have been explored and applied in modeling. Norouzi et al.[41] used 10-year clinical records of newly diagnosed CKD patients and the adaptive neurofuzzy inference system to predict the renal failure timeframe of CKD. The model accurately estimates eGFR changes for all sequential periods (normalized mean absolute error < 5%). In addition to using a variety of supervised learning algorithms for modeling, Perott et al.[42] used the unsupervised learning method latent Dirichlet allocation to process longitudinal clinical data and establish an accurate predictive model for the progression of CKD phase III to phase IV (C-statistic, 0.849).

Currently, numerous limitations remain. Although high performing CKD risk prediction models have been increasingly established, the actual effects of the models still need further exploration. After modeling, it is necessary to better calibrate and externally validate the results and verify the impacts on the outcome assessment before incorporating them into the guiding principles.[43] The research progress of two kinds of specific CKDs, diabetic kidney diseases (DKDs), and immunoglobulin A nephropathy (IgAN), are discussed as follows. The applications of ML in DKD and IgAN are listed in Table 3.

**DKD**

In the field of diabetes research, many studies have used various types of ML methods, such as artificial neural networks (ANNs) or decision trees (DTs), to predict the incidence of diabetes and determine the risk factors.[44-47] Diabetic complications have a considerable impact on quality of life and mortality, so many scientists are also working on the application of ML in the diagnosis and treatment of diabetic complications. Based on CNNs, Gulshan et al.[48] created a data-driven model for the detection of diabetic retinopathy using more than 100,000 clinical images. The high sensitivity and specificity of the model for the diagnosis of diabetic retinopathy make it a milestone of AI application in medicine.

DKD is also one of the most serious microvascular complications of diabetes and the most important cause of ESRD. The early non-invasive diagnosis of diabetic nephropathy (DN) is one of the current research hotspots. However, AI technology has not been widely used in this field.

**Prediction of the onset of DKD:** Scientists have modeled the prediction of the occurrence of DKD. Cho et al.[49] collected the medical data of 4321 diabetic patients and followed up for 10 years to develop a new visualization system by SVM classification. The model predicts the onset of DKD 2 to 3 months before the actual diagnosis with high prediction performance (AUC, 0.969). The research
was the first to use data mining technology and an advanced ML algorithm to predict DKD. Although the result was not accurate or complete enough, it provided scientists with a new idea for achieving effective and appropriate treatment strategies in the early stage of the disease. Rodriguez-Romero et al.[50] used the data of 10,251 patients screened from the Action to Control Cardiovascular Risk in Diabetes (ACCORD) trial and six kinds of ML algorithms to model DKD risk prediction. The study found that the RF (sensitivity >0.72, accuracy >0.73) and LR (sensitivity >0.76, accuracy >0.8) models showed the best predictive performance in both the training and testing databases, and among the included features, the decline in eGFR was the most important factor reflecting the development of DKD. However, a lack of external validation may cause the model to be only applicable to patients in the ACCORD trial. Recently, Ravizza et al.[51] used a LR-based Roche/IBM algorithm and real-world data (RWD) model for the onset of CKD in diabetic patients (AUC, 79.37%) and compared the performance with the results of multiple published large randomized controlled trials (RCTs). A total of 522,416 patients with diabetes from the IBM Explorys Database were screened for modeling and cross-validation, and 82,912 diabetic patients’ data from the Indiana Network for Patient Care database were screened for external validation. This demonstrates that the Roche/IBM algorithm based on RWD is consistent or even more accurate with research using RCT data. This evidence may change the current medical decision-making pattern that is mainly based on clinical RCTs.

In addition to using only EHR data, genotype was also a crucial feature of patients for modeling. Leung et al.[52] selected genotypes based on published studies and conducted structured clinical assessments of the included patients and incorporated the selected genotypes and phenotypic features into the ML program for modeling. The results reveal that the accuracy of the model based on a genotype-phenotype combination is higher than that of the model with genotype or phenotype only. Among the algorithms used, the SVM and RF algorithms show better performance (accuracy >0.9). With the advancement of technology and further research, the prediction of the onset of DKD in advance and even its application in clinical practice is no longer infeasible.

**Identification of DN and non-diabetic renal disease:** The Kidney Disease Outcomes Quality Initiative published in 2007, proposed the diagnostic criteria of DN and non-diabetic kidney disease (NDRD). However, further evaluation found that the diagnostic efficacy of the guidelines for Chinese patients was not satisfactory enough.[53] Therefore, 929 patients with type 2 diabetes who underwent renal biopsy from 2005 to 2017 were screened by Liu et al. [unpublished]. Based on the clinical test and renal pathological diagnosis results, differential diagnosis models of DN and NDRD based on SVM and RF were established. The sensitivity of the two models was 84.23% and 84.80%, respectively, and the specificity was 89.85% and 90.58%, respectively. Compared with other studies using the binary LR method,[54,55] the efficiency of the two models was higher.

---

### Table 3: Application of machine learning in DKD and IgAN.

<table>
<thead>
<tr>
<th>Purpose</th>
<th>Author Year</th>
<th>Journal</th>
<th>Study type</th>
<th>Sample size</th>
<th>Algorithms</th>
<th>Statistical analysis</th>
<th>Validation</th>
<th>AUROC (C-statistic)</th>
<th>Accuracy</th>
<th>Sensitivity</th>
<th>Specificity</th>
<th>F1 score</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>DKD</strong></td>
<td><strong>Prediction of the onset of DKD by EHRs</strong></td>
<td>Cho et al. 2008[56]</td>
<td>Ann Intern Med</td>
<td>Retrospective</td>
<td>252</td>
<td>SVM, LR</td>
<td>ROC, AUC, F1 score, t-test</td>
<td>Cross validation</td>
<td>96.9%</td>
<td>94%</td>
<td>93%</td>
<td></td>
</tr>
<tr>
<td><strong>Prediction of the onset of DKD by EHRs</strong></td>
<td>Rodriguez-Romero et al. 2019[57]</td>
<td>Clin Transl Sci</td>
<td>Retrospective</td>
<td>10,251</td>
<td>LR, DT, RF, SVM, SMO, NB</td>
<td>ROC</td>
<td>Cross validation</td>
<td>&gt;95%</td>
<td>&gt;95%</td>
<td>&gt;95%</td>
<td>–</td>
<td></td>
</tr>
<tr>
<td><strong>Prediction of the onset of DKD by EHRs</strong></td>
<td>Ravizza et al. 2019[58]</td>
<td>Nat Med</td>
<td>Retrospective</td>
<td>605,328</td>
<td>LR, RF</td>
<td>ROC, AUC</td>
<td>Cross validation</td>
<td>&gt;95%</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td></td>
</tr>
<tr>
<td><strong>Prediction of the onset of DKD by genotype-phenotype risk patterns</strong></td>
<td>Duan et al. 2013[59]</td>
<td>BMC Nephrol</td>
<td>Prospective</td>
<td>673</td>
<td>SVM, NB, RF, PLS, NNet, RF, DT</td>
<td>ROC, AUC</td>
<td>Cross validation</td>
<td>&gt;95%</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td></td>
</tr>
</tbody>
</table>

DKD: Diabetic kidney disease; IgAN: Immunoglobulin A nephropathy; EHRs: Electronic health records; ESRD: End stage renal disease; SVM: Support vector machine; LR: Logistic regression; RF: Random forest; SL: Simple logistic; SMO: Sequential minimal optimization; NB: Naïve Bayes; PLS: Partial least squares regression; NNet: Feed-forward neural networks; RPart: Recursive partitioning and regression trees; ANN: Artificial neural network; k-NN: k-Nearest neighbor; GBM: Gradient boosting machine; ROC: Receiver operating characteristic curve; AUC: Area under curve; H-L test: Hosmer-Lemeshow test; K-M analysis: Kaplan-Meier analysis; –: No relevant information mentioned in the article.
Currently, only single-center retrospective studies in the field have been carried out by the ML method. A multicenter prospective study with large mass data needs to be conducted and verified externally. Although it is only the beginning of relevant studies, a new idea for the diagnosis of DN by AI technology without renal biopsy is provided.

IgAN

IgAN is a major cause of renal failure. Early identification of patients with renal failure is useful for the prognosis and treatment of the disease. Multiple types of ML algorithms can effectively help predict the progression of ESRD in patients with IgAN.

Prediction of ESRD progression based on ANN: As early as 1998, a predictive model (sensitivity, 86.4%; specificity, 87.5%) was established based on ANN for the progression of ESRD in patients with IgAN. [56] This proves that modeling by ML algorithms more accurately predicts the progression of renal function damage in patients with IgAN than experienced renal specialists. However, the model is based on the information of only 54 patients and without validation. With the development of data preservation, processing and ML technology, a well-trained ANN model was developed into an online clinical decision support system (CDSS) used for quantitative risk assessment and time prediction of ESRD in IgAN patients in 2016 (www.IgAN.net). [57] The model was trained and tested based on the clinical data of 1040 IgAN patients confirmed by biopsy from different population cohorts. The performance of the model was satisfactory in different races (AUC > 90%). However, due to the inclusion of few Asians, further validation is necessary for its application.

Prediction of ESRD progression based on RF: Recently, Han et al. [58] predicted the progression of ESRD in IgAN patients by six ML algorithms (LR, RF, SVM, DT, ANNs, and k-nearest neighbor). Comparing the effectiveness of six models, the RF model can best predict the progression of ESRD (sensitivity, 80.6%; specificity, 95.29%). In the study of Liu et al. [59] the AUC of the RF model reached 97.29% after considering the results of C3 staining according to the contribution of predictors. The algorithm is used in a variety of progressive diseases to help clinicians stage and manage patients.

Prediction of ESRD progression based on gradient boosting machine: In addition to using parallel integration methods such as RF, the serial integration method is also used for models such as gradient boosting machines (GBMs). Chen et al. [60] constructed an accurate model for predicting the progression of ESRD in IgAN patients (C-statistic, 0.84) by XGBoost. Additionally, based on stepwise Cox regression, a scoring scale model (SSM) for risk stratification was constructed to identify specific patient groups with the same risk of progression. This XGBoost prediction model and the SSM were incorporated into the Nanjing IgAN Risk Stratification System, which is available online. ML provides a more favorable tool for strengthening the individualized treatment and management of IgAN patients.

However, current studies only focus on the prediction of ESRD in patients with IgAN, and there is no research modeling for the risk prediction of suffering IgAN, which may be the next direction of future efforts.

Acute kidney injury

Regardless of whether it is in high-income or low-income countries, the seven-day mortality rate for acute kidney injury (AKI) is 10% to 12%. [61] Efforts to improve the clinical outcomes of AKI have focused on early diagnosis and customized treatment. Early evaluation of AKI reduces mortality and improves renal prognosis. The model established by ML methods can not only realize early dynamic monitoring based on the real objective data of all patients but also save the time and energy of physicians. [62]

With the determination of AKI clinical practice guidelines and the significant growth of the application of EHRs in the field of big data, a large quantity of EHR data and ML algorithms have begun to play an important role in the clinical research of AKI. It is now an important tool for AKI diagnosis and prediction. [63] The establishment of a CDSS based on a self-learning predictive model may be used in in-hospital AKI monitoring in future clinical practice. [64]

The application of ML in AKI is listed in Table 4.

Early assessment of AKI

Early assessment of AKI can reduce mortality and improve renal prognosis. GBM is a common method for predicting the onset of AKI at present. [65-68] Lee et al. [69] modeled the prediction of AKI after liver transplantation and cardiac surgery by several ML algorithms. GBM showed the best performance in both studies, and the AUC of liver transplantation-related research was 90%, while the other was 78%. Additionally, Huang et al. [68] established a hazard prediction model for AKI after percutaneous coronary intervention (PCI) based on GBM. The study included a large quantity of data from 947,091 patients who underwent PCI surgery to establish a baseline model. In addition, temporal validation was conducted with data from 970,869 hospitalized patients. The AUC of the GBM model was 78.5% better than that of the baseline LR model (AUC, 75.3%). Therefore, advanced algorithms and massive data have the potential to provide accurate risk estimation.

With the cooperation of the University College London and DeepMind, [66] the era of early prediction of AKI and practical application in clinical diagnosis and treatment by the ML method is coming. Using data from 703,782 patients from multiple centers and stratified by a time window every six hours, a recurrent neural network-based AKI risk prediction model (AUC, 0.921) was established. AKI events were predicted within a 48-h window. However, the area under the precision-recall curve was just 29.7%, which represents a ratio of two false alerts for every true alert.

Prediction of AKI death risk

The prediction of the death risk in patients with AKI was also modeled. Using the medical information mart for intensive care III database, Lin et al. [70] developed a
mortality prediction model (AUC, 0.866, accuracy, 0.728) for AKI patients based on RF. However, the model slightly overestimates the mortality of patients with a low risk of death and underestimates the mortality of patients with high mortality.

With the further deepening of the research, ML-assisted monitoring may bring useful results to AKI and reduce the resulting morbidity and mortality. This inspiring prospect should be tested in further research.

**Dialytic treatments**

ML can be widely used in dialysis prescription and monitoring, complication management, death prediction, and so on. There is also great potential for application in child dialysis. The application of ML in dialysis is listed in Table 5. There have been several positive research results in dialytic treatments. Scientists organized the first scientific conference on the application status and future development of AI in dialysis at the Hospital Universitari Bellvitge (L’Hospitalet, Barcelona). In the conference, the experience of AI in dialysis was reviewed, and the obstacles, challenges, and future applications in this field were discussed. All the results indicate a wide application and good prospect for AI in renal dialysis. In the future, AI may significantly change clinical practice in terms of hemodialysis.

**Prescription management**

The adequacy of dialysis is difficult to determine, and its definition is constantly changing. Urea concentration in blood is a key factor in determining the dose for hemodialysis. Direct dialysate quantification (DDQ) is the most direct and accurate method for determining the removal rate of urea nitrogen (UN) in dialysis, but it is difficult to collect all the dialysate used for quantitative analysis. With the progress of AI technology, neural networks (NNs) can be used to predict changes in solute removal rate of urea nitrogen (UN) in dialysis, but it is difficult to collect all the dialysate used for quantitative analysis. There are several positive research results in dialytic treatments. Scientists organized the first scientific conference on the application status and future development of AI in dialysis at the Hospital Universitari Bellvitge (L’Hospitalet, Barcelona). In the conference, the experience of AI in dialysis was reviewed, and the obstacles, challenges, and future applications in this field were discussed. All the results indicate a wide application and good prospect for AI in renal dialysis. In the future, AI may significantly change clinical practice in terms of hemodialysis.

**Anemia management**

CKD anemia is one of the most common complications in dialyzed ESRD patients. The use of erythropoiesis-stimulating agents (ESAs) improves the treatment of CKD anemia, but owing to the risk of death and serious cardiovascular events, the treatment dose should vary from person to person. However, there are still many difficulties in choosing the best therapeutic dose. Patient status, pharmacokinetics, and pharmacodynamics of ESA are key factors for correct long-term dose-response
Table 5: Application of machine learning in dialysis.

<table>
<thead>
<tr>
<th>Purpose</th>
<th>Author</th>
<th>Year</th>
<th>Journal</th>
<th>Study type</th>
<th>Sample size</th>
<th>Assessing methods</th>
<th>Statistical analysis</th>
<th>Validation</th>
<th>AUC (C-statistic)</th>
<th>Error</th>
</tr>
</thead>
<tbody>
<tr>
<td>Prescription management</td>
<td>Ald et al[74]</td>
<td>2001</td>
<td>Am J Kidney Dis</td>
<td>Prospective</td>
<td>30</td>
<td>ANN, DDQ</td>
<td>ANOVA, χ², test, t-test</td>
<td>Cross validation</td>
<td>–</td>
<td>Kt/V: ER &lt; 10.9%</td>
</tr>
<tr>
<td>Prediction of the changes in solute concentration during and after HD</td>
<td>Fernandez et al[74]</td>
<td>2005</td>
<td>Clin Nephrol</td>
<td>Prospective</td>
<td>113</td>
<td>ANN, Std, mURR, 3mye</td>
<td>ROC, AUC</td>
<td>–</td>
<td>&gt;90.20%</td>
<td>Kt/V: ER &lt; 12.7%</td>
</tr>
<tr>
<td>Comparison of the estimators of HD dose and finding decision thresholds</td>
<td>Fernandez et al[74]</td>
<td>2001</td>
<td>Blood Purif</td>
<td>Prospective</td>
<td>133</td>
<td>ANN, Smye, Daugirdas</td>
<td>BAM, MSE, MDE</td>
<td>Cross validation</td>
<td>–</td>
<td>eqU: MDE = 0.22 ± 7.71 mg/ ml, eqKt/V: MDE = 0.01 ± 0.15 mg/ml</td>
</tr>
<tr>
<td>Anemia management</td>
<td>Barbieri et al[75]</td>
<td>2015</td>
<td>Comput Biol Med</td>
<td>Retrospective</td>
<td>4135</td>
<td>ANN</td>
<td>ME, MAE, RMSE</td>
<td>Cross validation</td>
<td>–</td>
<td>Hb: MAE &lt; 0.6 g/dl</td>
</tr>
<tr>
<td>Prediction of treatment response to anemia in patients with ESRD</td>
<td>Akbilgic et al[76]</td>
<td>2019</td>
<td>Kidney Int Rep</td>
<td>Retrospective</td>
<td>27,615</td>
<td>RF, ER, ANN, SVM, k-NN</td>
<td>ROC, AUC, C-statistic, Mann-Whitney U, t-test</td>
<td>Cross validation</td>
<td>&gt;71.8%</td>
<td>–</td>
</tr>
</tbody>
</table>

HD: Hemodialysis; ESRD: End stage renal disease; SCD: Sudden cardiac death; ANN: Artificial neural network; DDQ: Direct dialysate quantification; Std: Standard urea removal ratio; mURR: Modified urea removal ratio; RF: Random forest; SVM: Support vector machine; LR: Logistic regression; k-NN: k-Nearest neighbor; LASSO: Least absolute shrinkage and selection operator; CART: Classification and regression tree; ANOVA: Analysis of variance; ROC: Receiver operating characteristic curve; AUC: Area under curve; BAM: Bland-Altman agreement method; ME: Mean error; MSE: Mean-square error; MAE: Mean absolute error; RMSE: Root mean square error; MDE: Mean difference error; ER: Error; –: No relevant information mentioned in the article.

Death prediction

Hemodialysis patients may die as a result of many complications during or after hemodialysis. The mortality rate is high in the first year of dialysis.[76,81] Accurate estimation of post-dialysis mortality can help patients and clinicians make decisions as early as possible. Akbilgic et al[73] used the RF algorithm to construct four different mortality prediction models with 27,615 patients within 30, 90, 180, and 365 days after dialysis transition. The C-statistics of the models were 0.7185, 0.7446, 0.7504, and 0.7488, respectively, which shows good internal effectiveness and replication. The model accurately predicted the short-term post-dialysis mortality of patients after ESRD events, but there is still a lack of external verification.

Of the causes of death, sudden cardiac death is a serious cardiovascular complication in dialysis patients. The ML method has been revealed to be helpful in the prediction of sudden cardiac death.[82] Based on the information of patients before and after dialysis, the RF prediction model was constructed by Goldstein et al.[82] The C-statistic of the model is 0.799. Because the ability to predict sudden death decreases with the increase in the predicted time range, the data collected were more conducive to assessing short-term risks than long-term risks.

Challenges and future prospects

The remaining challenges

While the application of ML in health care and other areas is surprising, it is still in its infancy in the field of kidney disease. Challenges face both ML and its application in nephrology.
Challenges for ML

In the development of machine learning, it is still necessary to constantly improve ML algorithms to meet the challenges in conventional clinical practice. Due to the limitations of the development of ML, the inherent logic behind most ML models is similar to a black box, which is still difficult to explain to doctors.

Additionally, the ethics of ML must be considered. Although some guidelines have emerged, such as the Singapore Model Artificial Intelligence Governance Framework,\(^{[83]}\) to guide private sector organizations on how to use AI on an ethical basis, the development of clear clinical guidelines still lags far behind the progress of AI technology. It is necessary to develop clear guidelines as soon as possible to standardize its clinical application. Although ML models perform well with more training data, a balance needs to be found between privacy and regulatory requirements with the use of large and diverse datasets to improve the accuracy of ML models.\(^{[84,85]}\)

Challenges for nephrology

In addition to the limitation of the development of the ML algorithm, there are also many challenges in the application of ML in nephrology.

Challenges for clinical data collection and processing

In the process of data collection, it may be difficult to collect enough electronic medical record information, and different medical institutions lack unified standards and irregular, incomplete, or missing data, resulting in low data quality and difficulty in extracting meaningful information from the data. It is necessary to ensure the as much integrity and authenticity of the data as possible.\(^{[86]}\) The most commonly mentioned limitation in various studies is the lack of external verification. Only a few studies have carried out external verification of the model to prove the accuracy and practicability of the model. This is also related to the difficulty of data acquisition. At present, the data of each unit are basically non-shared, and the data are very scattered.

Medical data are large, confusing and complex. Data must be properly mapped and pre-processed before they are used for modeling. This step is the basis for the model because the accuracy is highly dependent on the reliability of the data in terms of clinical reflection.\(^{[87]}\) Regardless of how much effort is invested in improving the algorithm, any inaccuracy in the model will seriously limit the accuracy that the ML algorithms can achieve. Therefore, it is a very considerable challenge to obtain high-quality data and make it available.

Additionally, we have to consider the cost in the modeling process. It takes considerable manpower and material resources to collect data manually, to store data by computers, or to process data by large-scale computing power. Before modeling, the investment of funds is not a small challenge.

Challenge for renal pathological diagnosis

The published research is only at the stage of accurate identification of glomeruli. The sample size of the study is generally small, and the pictures used for modeling are basically from pathological sections of the kidneys of mice.\(^{[32,34]}\) Although AI has developed rapidly, due to the complexity of pathological manifestations of various renal diseases and their close relationship with clinical indicators, the automated pathological diagnosis of specific renal diseases based on images has not yet been published. It is not possible to completely replace pathologists for the renal pathological diagnosis of all kinds of kidney diseases through comprehensive patient information. This needs to be supported by a large quantity of data and confirmed by prospective studies. The DN framework and image processing operations across different clinical practices and image datasets need to be further validated to verify this technique in the distribution and spectrum of all lesions encountered in typical renal pathology services.

Future directions

Although most nephrologists are not familiar with the basic principles of medical AI now, in the future, through the collaboration of nephrologists and AI researchers, it is possible to use AI technology to build a big database for CKD research and establish a high-efficiency model that can be widely used in the diagnosis and treatment of CKD.

Objective and accurate AI techniques may be truly used in clinical renal pathological diagnosis to assist pathologists in discovering subtle differences between diseases that are indistinguishable to the naked eye. Even on this basis, an app may be developed to help patients identify renal biopsy results and predict renal prognosis.

Through the improvement of clinical data preservation, processing technology and the establishment of a resource sharing platform, kidney disease risk models based on multicenter massive data will be more reliable. When the performance of the model is high enough, it may one day be able to replace renal biopsy to achieve non-invasive diagnosis. The future is full of unknowns, but everything seems to be inevitable.

Conclusions

AI is playing an increasingly important role in many medical fields, helping doctors in most steps of patient management. In nephrology, AI has been used to predict the risk and progression of disease, as well as for hemodialysis prescriptions and follow-up. Although it is just scratching the surface now, with the progress of technology, the accumulation of data and the increase in investment, ML technology will make a major breakthrough in precision medicine in the field of nephrology.

Funding

This work was supported by grants from the National Natural Science Foundation of China (Nos. 61971441, 61671479, and 81804056), and the National Key R&D Program of China (No. 2016YFC1303500).

Conflicts of interest

None.
References


How to cite this article: Li Q, Fan QL, Han QX, Geng WJ, Zhao HH, Ding XN, Yan JY, Zhu HY. Machine learning in nephrology: scratching the surface. Chin Med J 2020;133:687–698. doi: 10.1097/CM9.0000000000000694